

Article

A Genome-First Approach to Estimate Prevalence of Germline Pathogenic Variants and Risk of Pancreatic Cancer in Select Cancer Susceptibility Genes

Esteban Astiazaran-Symonds ^{1,*}, Jung Kim ¹ , Jeremy S. Haley ² , Sun Young Kim ^{1,3} , H. Shanker Rao ², Regeneron Genetics Center ⁴, David J. Carey ², Douglas R. Stewart ¹ and Alisa M. Goldstein ^{1,*} 

¹ Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, 9609 Medical Center Drive, Rockville, MD 20850, USA; jung.kim2@nih.gov (J.K.); sunyoung.kim@nih.gov (S.Y.K.); drstewart@mail.nih.gov (D.R.S.)

² Geisinger Clinic, Geisinger Health System, 100 N Academy Ave, Danville, CA 17821, USA; jshaley@geisinger.edu (J.S.H.); hsr Rao@geisinger.edu (H.S.R.); djcarey@geisinger.edu (D.J.C.)

³ Department of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N Broadway, Baltimore, MD 21205, USA

⁴ Regeneron Genetics Center, 777 Old Saw Mill River Rd, Tarrytown, NY 10591, USA; support@regeneron.com

* Correspondence: esteban.astiazaransymonds@nih.gov (E.A.-S.); goldstea@mail.nih.gov (A.M.G.); Tel.: +1-240-276-7233 (A.M.G.)

Simple Summary: Prevalence and cancer risk estimates derived from the evaluation of affected individuals in the clinic are subject to ascertainment bias. This limitation can be mitigated using a genome-first approach in which genotypic data is analyzed before knowing a patient's phenotype. Our study aimed to analyze two large cohorts with available exome and phenotype data unselected for a specific diagnosis from a genome-first perspective focusing on six pancreatic cancer predisposition genes. We provide estimates of (1) prevalence of heterozygotes for the general population and for individuals with pancreatic cancer and (2) cancer risk for pancreatic cancer for each gene evaluated. For mutation carriers, we found an elevated risk of pancreatic cancer for most genes evaluated, with variation among genes. This work expands our knowledge of the complex genetics of this cancer and will help identify patients at the highest risk who could benefit from future screening or therapeutic strategies.

Abstract: Patients with germline pathogenic variants (GPV) in cancer predisposition genes are at increased risk of pancreatic ductal adenocarcinoma (PDAC), the most common type of pancreatic cancer. The genes most frequently found to harbor GPV in unselected PDAC cases are *ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *CHEK2*, and *PALB2*. However, GPV prevalence and gene-specific associations have not been extensively studied in the general population. To further explore these associations, we analyzed genomic and phenotypic data obtained from the UK Biobank (UKB) and Geisinger MyCode Community Health Initiative (GHS) cohorts comprising 200,600 and 175,449 participants, respectively. We estimated the frequency and calculated relative risks (RRs) of heterozygotes in both cohorts and a subset of individuals with PDAC. The combined frequency of heterozygous carriers of GPV in the general population ranged from 1.22% for *CHEK2* to 0.05% for *CDKN2A*. The frequency of GPV in PDAC cases varied from 2.38% (*ATM*) to 0.19% (*BRCA1* and *CDKN2A*). The RRs of PDAC were elevated for all genes except for *BRCA1* and varied widely by gene from high (*ATM*) to low (*CHEK2*, *BRCA2*). This work expands our understanding of the frequencies of GPV heterozygous carriers and associations between PDAC and GPV in several important PDAC susceptibility genes.

Keywords: cancer risk; pancreatic cancer; cancer genetics; pancreatic ductal adenocarcinoma; inherited cancer



Citation: Astiazaran-Symonds, E.; Kim, J.; Haley, J.S.; Kim, S.Y.; Rao, H.S.; Genetics Center, R.; Carey, D.J.; Stewart, D.R.; Goldstein, A.M. A Genome-First Approach to Estimate Prevalence of Germline Pathogenic Variants and Risk of Pancreatic Cancer in Select Cancer Susceptibility Genes. *Cancers* **2022**, *14*, 3257. <https://doi.org/10.3390/cancers14133257>

Received: 7 June 2022

Accepted: 1 July 2022

Published: 2 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pancreatic cancer is one of the deadliest cancers to date, with high mortality despite the improvement in the treatments of cancers in general. This is in part due to the late discovery of pancreatic cancer, which makes treatment and management challenging [1]. Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer and is associated with the worst prognosis [2].

It has been estimated that 5–10% of PDAC cases have a genetic basis [3–6]. Many germline pathogenic variants (GPV) in cancer predisposition genes that are part of several inherited cancer syndromes, such as hereditary breast and ovarian cancer syndrome and Lynch syndrome, have been identified in patients with PDAC [3].

A recent systematic review identified *ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *CHEK2*, and *PALB2* as the most frequent genes harboring GPV in individuals with PDAC not meeting criteria for any cancer predisposition syndrome with varying degrees of family history [4]. However, the risk of PDAC related to GPV, the frequency of GPV carriers in PDAC patients, and the prevalence of PDAC in GPV carriers in these genes is not completely understood.

Furthermore, much of what is known about the genetics of PDAC has occurred when individuals are evaluated in the clinic due to relevant personal or family history of PDAC and undergo molecular diagnostic testing. While identifying GPV in individuals with PDAC provides a helpful perspective regarding the association of these variants and the risk of developing this cancer, this approach may overestimate associations because of the potential for ascertainment bias.

One way to mitigate this concern is using a “genome-first” approach. Notably, there is a growing number of biobanks of population-based and health system cohorts with available germline exome data. These projects allow for innovative ways to explore PDAC by the analysis of population genomics data and identification of genetic variation, followed by subsequent abstraction of relevant clinical phenotypes that have been observed. A genome-first approach allows for a more accurate measurement of the prevalence of GPV and risk of developing PDAC in patients with this diagnosis independent of previous personal or family history with reduced ascertainment bias, especially when more than one biobank is analyzed.

In this study, we analyzed exome data from participants from two databases of population genomics data [UK Biobank (UKB) and the Geisinger MyCode Community Health Initiative (GHS) cohorts] [7,8] to identify GPV, calculate the prevalence of heterozygous carriers, and estimate the risk of developing PDAC for individuals harboring GPV in six important pancreatic cancer susceptibility genes.

2. Materials and Methods

We focused on the six genes with the highest frequencies of GPV in individuals with PDAC unselected for family history: *ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *CHEK2*, and *PALB2*. We analyzed GPV in these six genes in participants of two large exome databases (UKB and GHS) and evaluated their relationship with pancreatic cancer (Table S1).

2.1. Exome Databases, Sequencing, and Variant Annotation

UKB is a large population-based prospective study in the United Kingdom with adult participants at recruitment, with extensive matching phenotypic and genomic data. The approach used at the Regeneron Genetics Center to perform exome sequencing (ES) in DNA samples from the UKB study, as well as quality metrics and variant filtering, has been previously described in detail [9–11]. The present study focused on the 200,600 participants with exome data available in late 2020. The cohort is mostly of European descent (87.7%) and the average age is 69 years.

Geisinger is an integrated health system in south-central and northeastern Pennsylvania. All Geisinger patients are eligible to participate in the MyCode Community Health Initiative, a system-wide biorepository of blood and DNA samples for broad future research. Participants agree to allow their sample data to be linked to information in their

Geisinger health records. MyCode enrollment is open to all patients regardless of the medical conditions they have. MyCode DNA samples were exome sequenced by the Regeneron Genetics Center using IDT exon capture probes as previously described [8]. Coverage depth was sufficient to provide more than 90% coverage of the targeted bases for 99% of samples. Alignments and variant calling were based on the GRCh38 human genome reference sequence. The cohort is mostly of European descent (96.5%) and the average age is 57 years. The Geisinger study population consisted of the first 175,449 MyCode participants to have available exome sequence data.

Exome and phenotype (PDAC) data from a total of 376,049 participants were included in the study. The exome data were annotated with snpEFF [12] and ANNOVAR [13]. Subsequently, we predicted pathogenicity by ClinVar [14] (data obtained 8 June 2021) and InterVar (version 2.1.3) [15]. Variants determined to be pathogenic or likely pathogenic by these tools were then curated based on the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) [16]. Briefly, variants were evaluated based on allele frequency using The Genome Aggregation Database (gnomAD v2.1.1) (data obtained 10 November 2021), a resource of summarized data from large-scale sequencing projects (<https://gnomad.broadinstitute.org/>), effect on protein (truncating vs. non-truncating), aggregate in silico predictors from Franklin Genoox (<https://franklin.genoox.com/clinical-db/home>) (data obtained 15 December 2021), and evidence obtained from functional studies and case-control studies from the literature.

2.2. Electronic Health Records (EHR) Review

We used data from the EHR to identify participants with PDAC in the UKB and GHS cohorts. Specifically, clinical phenotypes of PDAC were determined using validated phenotype algorithms that use International Classification Diseases (ICD) diagnosis codes (C25) for PDAC. The code for tumors of the endocrine pancreas (C25.4) was excluded. In addition to longitudinal electronic health record data, we also queried the Geisinger cancer registry, which contains information on all patients diagnosed with cancer at a Geisinger facility; the cancer registry data is also contributed to the National Cancer Database [17].

2.3. Statistical Analysis

Prevalences of GPV for each of the six genes were tabulated for the case patients and noncarrier controls in the UKB and GHS, and 95% confidence intervals (CI) were estimated. 2×2 contingency tables were then used to test for association of variants and a diagnosis of PDAC by relative risk (RR) and 95% CI. The significance of the association was evaluated using Fisher's exact test values ($p < 0.05$).

3. Results

3.1. Frequency of GPV in Six PDAC Susceptibility Genes in All Participants and in Individuals with PDAC

First, as the frequency of heterozygous carriers of these GPV in the general population is not well understood, we analyzed GPV in these six genes in the 376,049 participants from both cohorts. Overall, 1.22% of participants carried a GPV in *CHEK2* followed by *ATM* (0.46%), *BRCA2* (0.38%), *PALB2* (0.16%), *BRCA1* (0.15%), and *CDKN2A* (0.05%) (Table 1). No individuals were found to be homozygous for variants in any of the genes evaluated.

Next, we focused on the individuals in both cohorts who were found to have a diagnosis of PDAC. Table 2 shows the frequency of GPV in these PDAC participants for each of the six genes by cohort, combined and compared with reports from the literature. Overall, 6.94% (5.99% from UKB and 7.60% from GHS) of PDAC patients had a GPV in one of the six genes investigated. Individually, in UKB with 417 reported PDAC patients, 2.64% ($n = 11$) had *ATM* GPV, followed by *CHEK2* (1.44%, $n = 6$) and *BRCA2* (1.20%, $n = 5$). In the GHS cohort with 592 patients with PDAC, 2.70% of PDAC patients had a *CHEK2* GPV followed by *ATM* (2.19%) and *BRCA2* (1.52%).

Table 1. Number and prevalence of heterozygotes for GPV for each cohort by gene.

| | Number and Prevalence (%, in Parentheses) of Heterozygotes for GPV in UKB (n = 200,619) | Number and Prevalence (%, in Parentheses) of Heterozygotes for GPV in GHS (n = 175,449) | Number and Prevalence (%, in Parentheses) of Heterozygotes for GPV in Both Cohorts (n = 376,068) |
|---------------|--|--|--|
| <i>ATM</i> | 839 (0.42%) | 885 (0.50%) | 1724 (0.46%) |
| <i>BRCA1</i> | 223 (0.11%) | 356 (0.20%) | 579 (0.15%) |
| <i>BRCA2</i> | 710 (0.35%) | 717 (0.40%) | 1427 (0.38%) |
| <i>CDKN2A</i> | 137 (0.07%) | 47 (0.027%) | 184 (0.05%) |
| <i>CHEK2</i> | 1619 (0.80%) | 2987 (1.69%) | 4606 (1.22%) |
| <i>PALB2</i> | 397 (0.20%) | 205 (0.12%) | 6022 (0.16%) |

GPV: germline pathogenic variant; UKB: UK Biobank; GHS: Geisinger MyCode Health Initiative.

Table 2. Frequency of germline pathogenic variants in participants with pancreatic cancer.

| | UKB (n = 417) | | GHS (n = 592) | | Both Cohorts (n = 1009) | Astiazaran- Symonds et al., 2021 |
|---------------|--|---|--|---|-------------------------------|--|
| | Number of PDAC Participants with GPVs | Frequency of GPVs in PDAC Participants | Number of PDAC Participants with GPVs | Frequency of GPVs in PDAC Participants | Cumulative Frequency | Cumulative Frequency |
| ALL P/LP | 25 * | 5.99% | 45 | 7.60% | 6.94% | - |
| <i>ATM</i> | 11 | 2.64% | 13 | 2.19% | 2.38% | 2.52% |
| <i>BRCA1</i> | 0 | 0.0% | 2 | 0.34% | 0.19% | 0.99% |
| <i>BRCA2</i> | 5 | 1.20% | 9 | 1.52% | 1.38% | 2.90% |
| <i>CDKN2A</i> | 2 | 0.48% | 0 | 0.0% | 0.19% | 0.98% |
| <i>CHEK2</i> | 6 | 1.44% | 16 | 2.70% | 2.18% | 1.15% |
| <i>PALB2</i> | 2 | 0.48% | 5 | 0.84% | 0.69% | 0.65% |

PDAC: Pancreatic ductal adenocarcinoma; GPV: germline pathogenic variant; UKB: UK Biobank; GHS: Geisinger MyCode Health Initiative. * Addition of all individuals (26) and total (25) are not equal due to one individual carrying two variants.

3.2. Prevalence of PDAC in All Participants and in Individuals Harboring GPV in the Six Genes

The prevalence of individuals with PDAC in UKB was 417 (0.23%) and 592 (0.34%) in the GHS cohort, for a total of 1009 PDAC patients in both cohorts combined (0.27%). Next, to understand the number of patients who developed PDAC and who also carried GPV in the six evaluated genes, we calculated the prevalence of PDAC in this subset of patients (Table 3). Although based on small numbers, the highest prevalence of PDAC in UKB was observed for *CDKN2A* (1.45%) and in the GHS cohort for *PALB2* (2.43%). Across both cohorts combined, 24 individuals with an *ATM* GPV (1.39%), 22 subjects with a *CHEK2* GPV (1.16%), and 14 participants with a *BRCA2* GPV (0.98%) were found to have PDAC. At a variant level, *ATM* (Figure 1) and *BRCA2* (Figure 2) showed the widest spectrum of allelic variation in PDAC participants, with 22 and 14 unique variants, respectively. In contrast, the *CHEK2* variants c.470T>C (p.Ile157Thr) and p.Glu457fs (c.1100delC) were observed in 19 of the 22 GPV heterozygous carriers. The p.Arg170fs variant in *PALB2* was the only *PALB2* GPV detected more than once in PDAC patients, while variants in *BRCA1* and *CDKN2A* were all unique. None of these individuals were related. Supplementary Table S1 presents all the GPV in these six genes identified in patients with PDAC. The median age at cancer diagnosis for individuals with GPV in any gene evaluated was 66.9 for UKB (vs. 65.4 in those without GPV) and 65 for GHS (vs. 66.9 in those without GPV). None of the differences reached statistical significance ($p < 0.05$).

Table 3. Prevalence of PDAC in participants harboring a GPV (i.e., heterozygous) by gene.

| | UKB Participants Harboring GPVs Who Developed PDAC | GHS Participants Harboring GPVs Who Developed PDAC | Total Number of Heterozygotes for GPVs Who Developed PDAC |
|---------------|--|--|---|
| <i>ATM</i> | 11/839 (1.31%) | 13/885 (1.47%) | 24/1724 (1.39%) |
| <i>BRCA1</i> | 0/223 (0%) | 2/356 (0.56%) | 2/579 (0.34%) |
| <i>BRCA2</i> | 5/710 (0.70%) | 9/717 (1.25%) | 14/1427 (0.98%) |
| <i>CDKN2A</i> | 2/137 (1.45%) | 0/47 (0%) | 2/184 (1.09%) |
| <i>CHEK2</i> | 6/1619 (0.37%) | 16/2987 (0.53%) | 22/4606 (0.48%) |
| <i>PALB2</i> | 2/397 (0.50%) | 5/205 (2.44%) | 7/602 (1.16%) |

PDAC: Pancreatic ductal adenocarcinoma; GPV: germline pathogenic variant; UKB: UK Biobank; GHS: Geisinger MyCode Health Initiative.

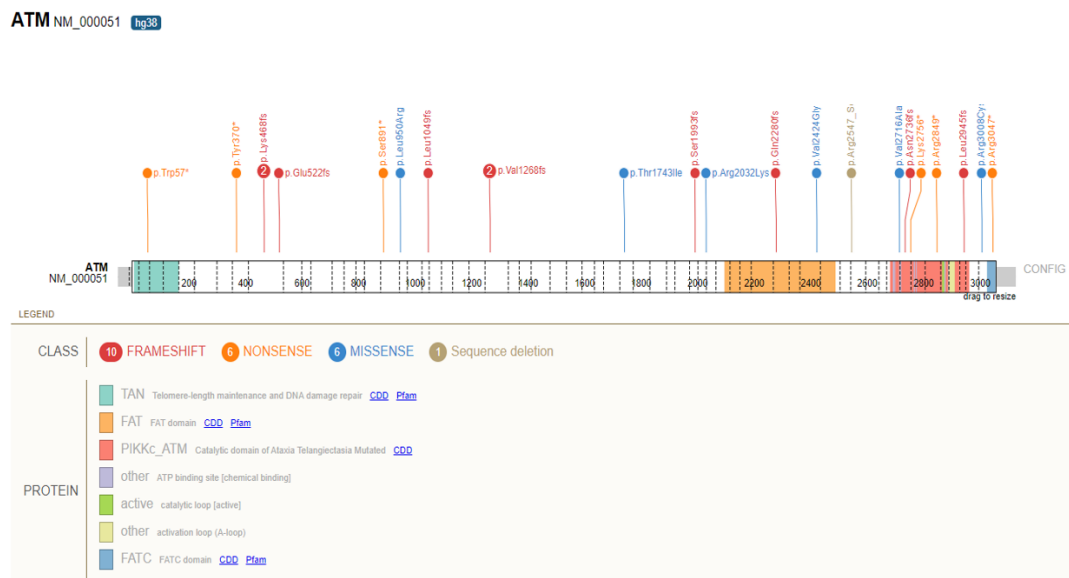


Figure 1. Distribution of germline pathogenic variants in *ATM* identified in participants with pancreatic ductal Adenocarcinoma from the UK Biobank and Geisinger MyCode Health Initiative cohorts. Splice site variant (c.8786 + 1G > A) not shown.

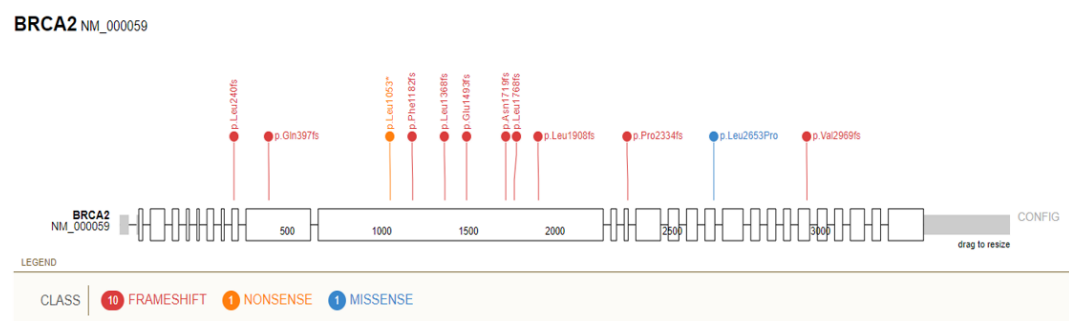


Figure 2. Distribution of germline pathogenic variants in *BRCA2* identified in participants with pancreatic ductal adenocarcinoma from the UK Biobank and Geisinger MyCode Health Initiative cohorts. Splice site variants (c.7977-1G > C and c.8487 + 1G > C) not shown.

3.3. Risk of PDAC in Participants Harboring GPV in *ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *CHEK2*, and *PALB2*

Next, we estimated the RR of PDAC in individuals harboring GPV for each of the six genes in the two cohorts (Table 4 and Figure 3). The highest RR was observed for *CDKN2A* (RR = 7, 95% CI 1.9–25), followed by *ATM* (RR = 6.4, 95% CI 3.6–11.6), and *BRCA2* (RR = 3.4, 95% CI 1.4–7.9) in UKB and *PALB2* (RR = 7.4, 95% CI 3.1–17.9), followed by *ATM* (RR = 4.4,

95% CI 2.6–7.6), *BRCA2* (RR = 3.8, 95% CI 2.0–7.2), and *CHEK2* (RR = 1.6, 95% CI 1.0–2.6) in GHS. These RRs were significantly increased ($p < 0.05$) for *ATM* and *BRCA2* in both UKB and GHS, whereas *CDKN2A*, *CHEK2*, and *PALB2* were each significant in only a single cohort.

Table 4. The relative risk of pancreatic cancer in participants harboring a germline pathogenic variant for each gene.

| Gene | UKB | | GHS | |
|---------------|----------------|-----------------|----------------|-----------------|
| | RR (95% CI) | <i>p</i> -Value | RR | <i>p</i> -Value |
| <i>ATM</i> | 6.4 (3.6–11.6) | <0.0001 * | 4.4 (2.6–7.6) | <0.0001 * |
| <i>BRCA1</i> | N/A | >0.9999 | 1.7 (0.4–6.7) | 0.3176 |
| <i>BRCA2</i> | 3.4 (1.4–7.9) | 0.0151 * | 3.8 (2–7.2) | 0.0007 * |
| <i>CDKN2A</i> | 7 (1.9–25) | 0.0312 * | N/A | >0.9999 |
| <i>CHEK2</i> | 1.8 (0.8–3.9) | 0.1482 | 1.6 (1–2.6) | 0.0411 * |
| <i>PALB2</i> | 2.4 (0.7–8.8) | 0.1888 | 7.4 (3.1–17.9) | 0.0006 * |

PDAC: Pancreatic ductal adenocarcinoma; GPV: germline pathogenic variant; UKB: UK Biobank; GHS: Geisinger MyCode Health Initiative; N/A: not calculable; RR: Relative risk. * *p*-value < 0.05 (statistically significant).

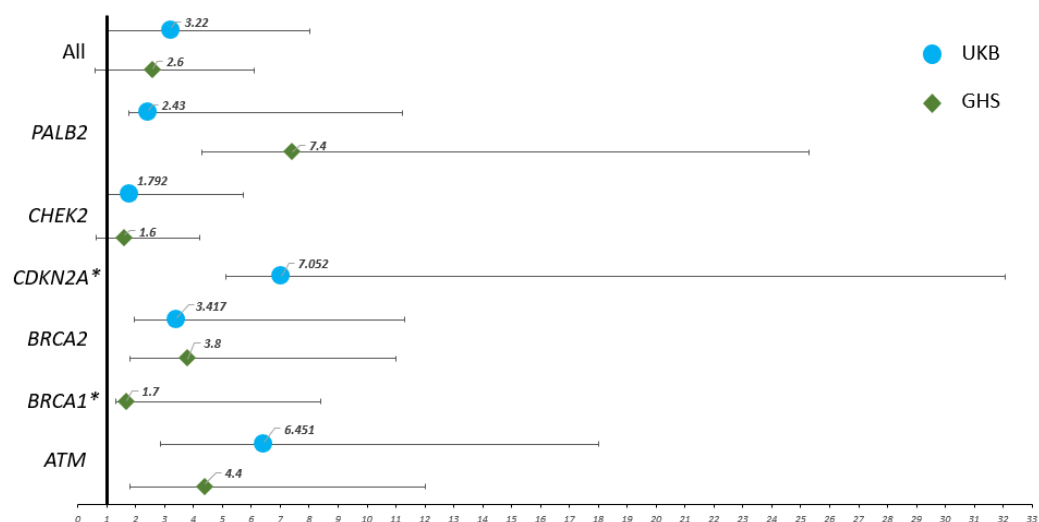


Figure 3. The relative risk of pancreatic ductal adenocarcinoma in participants harboring germline pathogenic variants for each gene. PDAC: pancreatic ductal adenocarcinoma, GPV: germline pathogenic variant. UKB: UK Biobank; GHS: Geisinger MyCode Health Initiative. * Variants identified only in one cohort.

4. Discussion

In this study, we used a genome-first approach to analyze genomic and EHR-linked data from 2 large cohorts to identify individuals who harbor GPV in a select group of genes that are frequently detected in PDAC patients (*ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *CHEK2*, and *PALB2*). We found the highest prevalence of heterozygotes for GPV in *CHEK2*, followed by *ATM* and *BRCA2*. The prevalence of PDAC among GPV carriers showed some variation across the two cohorts, but overall, the prevalence ranged from 0.3% in *BRCA1* to 1.4% in *ATM*. Finally, all genes had a significantly increased RR for either or both cohorts, except for *BRCA1*.

The prevalence of heterozygous carriers for GPV in the genes evaluated has been estimated in patients with cancer [18–21], but their prevalences in the general population are less clear. Moreover, some estimates have been variant-specific (founder variants in *BRCA1*, *BRCA2*, and *CHEK2*) [22,23]. Previous studies analyzing data from gnomAD controls v2.1, have reported a prevalence for heterozygous carriers of known GPV in *CHEK2* (0.67%), *BRCA2* (0.35%), *PALB2* (0.1%), and *BRCA1* (0.38%) [24], while the carrier frequency for *ATM* has been calculated at 0.5–1% [25]. The genome-first approach utilized

here allowed us to estimate prevalences in a population unselected for a specific diagnosis and, therefore, may more closely resemble the general population. While prevalences for *ATM*, *BRCA1*, *BRCA2*, and *PALB2* in this study are consistent with prevalences observed in the literature, we calculated that as many as 1.22% of individuals in both cohorts unselected for personal or family history of cancer harbored GPV in *CHEK2*. This frequency was mostly driven by the *CHEK2* founder variants c.470T>C (p.Ile157Thr) and p.Glu457fs (c.1100delC), which could explain the higher frequency in the UKB and GHS cohorts. Indeed, higher prevalences for these two variants have been observed in different Northern European populations [26], disproportionally represented in these two cohorts. Lastly, we estimated a prevalence of *CDKN2A* GPV carriers of 0.03–0.07% for the two cohorts.

Furthermore, although our sample of PDAC patients was relatively small, our estimates of the prevalence of GPV in these participants were mostly consistent with what has been reported in the literature [4]. Of note, we saw some differences in the prevalence of heterozygotes for these GPV in UKB and GHS, which could be in part due to age differences between the two cohorts. Studies have previously discussed the complex genetics of PDAC characterized by a great heterogeneity of genes involved and an overall lower risk when compared to GPV in the same genes for other cancers [4]. Generally, genes and variants have been divided into high-risk genes with a $RR \geq 5.0$, moderate-risk genes with a $RR 1.5–5.0$, and low-risk genes or genetic modifiers with a RR of 1.01–1.5 [27]. Findings in this study suggest that while these six genes increase the risk of developing PDAC, there is wide variability in the degree of risk across genes. Overall, we saw a high risk of PDAC in *ATM* (both cohorts) and in *CDKN2A* (UKB) and *PALB2* (GHS). A moderate risk was seen for *BRCA2* (both cohorts) and *PALB2* (UKB). Low risk was seen for *CHEK2* (both cohorts). The high risk of PDAC in *ATM* heterozygotes has been described before in a multicenter cohort [28]. Interestingly, this finding contrasts with the more moderate risk that has been estimated for breast cancer in this gene [29]. Li et al. recently reported a moderate risk of PDAC for *BRCA1* heterozygotes (2.36 $RR = 2.36$; 95% CI, 1.51 to 3.68) and for *BRCA2* heterozygotes ($RR = 3.34$; 95% CI, 2.21 to 5.06) [30]. Of note, although we observed an RR of 1.7 for PDAC in *BRCA1* heterozygotes and the risk in other studies had been calculated to be 2 to 4-fold [31], our results were not statistically significant. For *CHEK2*, a two-fold association has been reported for patients with familial pancreatic cancer in a recent study from Poland, but the association was not seen in unselected PDAC patients or individuals in familial pancreatic cancer families who did not have cancer [32]. In the case of moderate-risk and low-risk genes, there are even more challenges in deriving management recommendations when compared to more well-studied genes (e.g., *BRCA1* and *BRCA2*). For this reason, GPV carriers are generally recommended to undergo individualized management for PDAC in which other genetic (genetic background evidenced by positive family history, variants in other genes, etc.) and non-genetic factors likely play a role [33].

With the increased availability of clinical-grade sequencing, recommendations regarding which individuals should be genetically tested are expanding. The National Cancer Council Network (NCCN) now recommends germline genetic testing for all individuals with PDAC regardless of age (NCCN Guidelines: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic. Version 1.2020)

Another way to evaluate the associations between PDAC and these genes, which is made possible by the genome-first approach, is by examining the prevalence of PDAC in the subset of unselected individuals harboring GPV. Indeed, this approach allowed for the identification of individuals who carried GPV in cancer predisposition genes who did not have a reported diagnosis of PDAC but who could be at increased risk for it. Although other cancers were not evaluated, it is possible that some individuals who carried GPV in a cancer predisposition gene would not have been ascertained through a traditional phenotype-first approach in the clinic due to a lack of personal or family history of cancer. Furthermore, individuals with GPV, especially in high-risk genes, could possibly benefit from innovative technologies in early PDAC screening, including circulating tumor DNA,

exosomes in blood, artificial intelligence, metabolomics, and ion mobility spectrometry, which have shown promising prospects [34].

From the genes evaluated here, *BRCA1*, *BRCA2*, and *PALB2* are part of the ACMG Secondary Findings v3.0 list [35] and laboratories are now recommended to report GPV in these genes even in individuals without a personal or family history of cancer when next generation sequencing is performed [36]. Genome-first approaches may provide some preliminary data to assist with decision-making related to secondary findings.

Recent advances in targeted therapies for a growing list of inherited cancers have revolutionized treatment for these patients. Indeed, the PARP inhibitor rucaparib (Olaparib) is now approved as maintenance therapy for advanced PDAC with GPV in *BRCA1* and *BRCA2*. Furthermore, a recent investigator-initiated, single-arm phase II study assessed its role on advanced PDAC with GPV in *PALB2* and demonstrated that it is a safe and effective therapy for platinum-sensitive cancers, further expanding the role of this drug in patients with inherited cancer [37].

This study had several limitations, one of which was the small sample of individuals with PDAC. For this reason, the examination of risk for some of the genes may have been underpowered. Related, the 95% CI for risk estimates for some genes (*PALB2*, *CDKN2A*) were particularly wide and the results must be interpreted carefully. In addition, although other genetic and nongenetic factors may modify risk, we could not incorporate these factors into the statistical analyses. Further, the use of ES data in our study precluded us from evaluating gross deletions and insertions, which account for up to 11–13% of *BRCA1* and 2–3% of *BRCA2* [38,39], and ~20% of *CDKN2A* [40] GPV. Moreover, although our cohorts might be more similar to the general population than most case-control studies, the GHS cohort is a health-system-based cohort and thus may include a higher proportion of cancer cases than what is seen in the general population. Finally, both cohorts include participants of mostly European ancestry, which might limit the utility of these findings in more diverse populations. Finally, the study relied on available reported phenotypic data for participants in both cohorts. For this reason, it is possible that individuals with PDAC were not included in the study if the specific ICD10 code was not reported.

5. Conclusions

In this study, we analyzed EHR-linked exome data from two large cohorts and focused on a group of genes where GPV are frequently identified in individuals with PDAC to explore the complex associations between germline variation and the development of this cancer. We estimated and confirmed the prevalences of heterozygous carriers for these genes. Overall, we found varying degrees of PDAC risk in the genes assessed, from high (*ATM*) to low (*CHEK2*). This work expands our understanding of the frequency of variation in these genes in the general population and the associations between the development of PDAC in individuals harboring GPV in these genes, especially for *ATM*, *CHEK2*, and *PALB2*, where associations are not as well studied as for other genes. The availability of large exome databases is undoubtedly an invaluable new tool that provides another perspective to our understanding of the intricate relations between genetic factors and the development of pancreatic cancer.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14133257/s1>. Table S1. Germline pathogenic variants in participants with pancreatic cancer by gene and cohort.

Author Contributions: Conceptualization, E.A.-S. and A.M.G.; methodology, E.A.-S., A.M.G., J.K. and J.S.H.; formal analysis, J.K. and J.S.H.; investigation, E.A.-S. and A.M.G.; resources, H.S.R., R.G.C., D.J.C., A.M.G. and D.R.S.; data curation, E.A.-S., J.K., S.Y.K. and A.M.G.; writing—original draft preparation, E.A.-S. and A.M.G.; writing—review and editing, A.M.G., J.K., D.J.C., D.R.S., S.Y.K. and E.A.-S.; supervision, A.M.G.; project administration, A.M.G.; funding acquisition, A.M.G. and D.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute, NIH, Bethesda, MD. The work of SK was in part supported by an NIH Institutional National Research Award training grant (T32GM007471). Exome sequencing of Geisinger DNA samples was supported by a research contract with the Regeneron Genetics Center.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting Tables 1–4, Figures 1–3, and Supplementary material are available upon request from the corresponding author.

Acknowledgments: This research has been conducted using the UK Biobank Resource under application 54389. This work utilized the computational resources of the NIH High-Performance Computing Biowulf cluster.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **2021**, *71*, 7–33. [[CrossRef](#)]
2. Rahib, L.; Smith, B.D.; Aizenberg, R.; Rosenzweig, A.B.; Fleshman, J.M.; Matrisian, L.M. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **2014**, *74*, 2913–2921. [[CrossRef](#)] [[PubMed](#)]
3. Petersen, G.M. Familial pancreatic cancer. *Semin. Oncol.* **2016**, *43*, 548–553. [[CrossRef](#)] [[PubMed](#)]
4. Astiazaran-Symonds, E.; Goldstein, A.M. A systematic review of the prevalence of germline pathogenic variants in patients with pancreatic cancer. *J. Gastroenterol.* **2021**, *56*, 713–721. [[CrossRef](#)]
5. McWilliams, R.R.; Rabe, K.G.; Olswold, C.; De Andrade, M.; Petersen, G.M. Risk of malignancy in first-degree relatives of patients with pancreatic carcinoma. *Cancer* **2005**, *104*, 388–394. [[CrossRef](#)] [[PubMed](#)]
6. Brentnall, T.A. Cancer surveillance of patients from familial pancreatic cancer kindreds. *Med. Clin. N. Am.* **2000**, *84*, 707–718. [[CrossRef](#)]
7. Sudlow, C.; Gallacher, J.; Allen, N.; Beral, V.; Burton, P.; Danesh, J.; Downey, P.; Elliott, P.; Green, J.; Landray, M.; et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **2015**, *12*, e1001779. [[CrossRef](#)]
8. Carey, D.J.; Fetterolf, S.N.; Davis, F.D.; Faucett, W.A.; Kirchner, H.L.; Mirshahi, U.; Murray, M.F.; Smelser, D.T.; Gerhard, G.S.; Ledbetter, D.H. The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research. *Genet. Med.* **2016**, *18*, 906–913. [[CrossRef](#)]
9. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O’Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [[CrossRef](#)]
10. Van Hout, C.V.; Tachmazidou, I.; Backman, J.D.; Hoffman, J.D.; Liu, D.; Pandey, A.K.; Gonzaga-Jauregui, C.; Khalid, S.; Ye, B.; Banerjee, N.; et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **2020**, *586*, 749–756. [[CrossRef](#)]
11. Backman, J.D.; Li, A.H.; Marcketta, A.; Sun, D.; Mbatchou, J.; Kessler, M.D.; Benner, C.; Liu, D.; Locke, A.E.; Balasubramanian, S.; et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **2021**, *599*, 628–634. [[CrossRef](#)] [[PubMed](#)]
12. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)] [[PubMed](#)]
13. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)] [[PubMed](#)]
14. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **2018**, *46*, D1062–D1067. [[CrossRef](#)] [[PubMed](#)]
15. Li, Q.; Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* **2017**, *100*, 267–280. [[CrossRef](#)] [[PubMed](#)]
16. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **2015**, *17*, 405–424. [[CrossRef](#)] [[PubMed](#)]
17. Shivakumar, M.; Miller, J.E.; Dasari, V.R.; Gogoi, R.; Kim, D. Exome-Wide Rare Variant Analysis from the DiscovEHR Study Identifies Novel Candidate Predisposition Genes for Endometrial Cancer. *Front. Oncol.* **2019**, *9*, 574. [[CrossRef](#)] [[PubMed](#)]

18. Borecka, M.; Zemankova, P.; Vocka, M.; Soucek, P.; Soukupova, J.; Kleiblova, P.; Sevcik, J.; Kleibl, Z.; Janatova, M. Mutation analysis of the PALB2 gene in unselected pancreatic cancer patients in the Czech Republic. *Cancer Genet.* **2016**, *209*, 199–204. [[CrossRef](#)] [[PubMed](#)]
19. Cremin, C.; Lee, M.K.; Hong, Q.; Hoeschen, C.; MacKenzie, A.; Dixon, K.; McCullum, M.; Nuk, J.; Kalloger, S.; Karasinska, J.; et al. Burden of hereditary cancer susceptibility in unselected patients with pancreatic ductal adenocarcinoma referred for germline screening. *Cancer Med.* **2020**, *9*, 4004–4013. [[CrossRef](#)] [[PubMed](#)]
20. Ghiorzo, P.; Fornarini, G.; Sciallero, S.; Battistuzzi, L.; Belli, F.; Bernard, L.; Bonelli, L.; Borgonovo, G.; Bruno, W.; De Cian, F.; et al. CDKN2A is the main susceptibility gene in Italian pancreatic cancer families. *J. Med. Genet.* **2012**, *49*, 164–170. [[CrossRef](#)]
21. Grant, R.C.; Selander, I.; Connor, A.A.; Selvarajah, S.; Borgida, A.; Briollais, L.; Petersen, G.M.; Lerner-Ellis, J.; Holter, S.; Gallinger, S. Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology* **2015**, *148*, 556–564. [[CrossRef](#)] [[PubMed](#)]
22. Meijers-Heijboer, H.; Ouweland, A.V.D.; Klijn, J.; Wasielewski, M.; De Snoo, A.; Oldenburg, R.; Hollestelle, A.; Houben, M.; Crepin, E.; Van Veghel-Plandsoen, M.; et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat. Genet.* **2002**, *31*, 55–59. [[PubMed](#)]
23. Hartge, P.; Struewing, J.P.; Wacholder, S.; Brody, L.C.; Tucker, M.A. The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am. J. Hum. Genet.* **1999**, *64*, 963–970. [[CrossRef](#)] [[PubMed](#)]
24. Petridis, C.; Arora, I.; Shah, V.; Moss, C.L.; Mera, A.; Clifford, A.; Gillett, C.; Pinder, S.E.; Tomlinson, I.; Roylance, R.; et al. Frequency of Pathogenic Germline Variants in CDH1, BRCA2, CHEK2, PALB2, BRCA1, and TP53 in Sporadic Lobular Breast Cancer. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 1162–1168. [[CrossRef](#)] [[PubMed](#)]
25. Cavaciuti, E.; Lauge, A.; Janin, N.; Ossian, K.; Hall, J.; Stoppa-Lyonnet, D.; Andrieu, N. Cancer risk according to type and location of ATM mutation in ataxia-telangiectasia families. *Genes Chromosomes Cancer* **2005**, *42*, 1–9. [[CrossRef](#)] [[PubMed](#)]
26. Apostolou, P.; Fostira, F.; Papamentzelopoulou, M.; Michelli, M.; Panopoulos, C.; Fountzilias, G.; Konstantopoulou, I.; Voutsinas, G.E.; Yannoukakos, D. CHEK2 c.1100delC allele is rarely identified in Greek breast cancer cases. *Cancer Genet.* **2015**, *208*, 129–134. [[CrossRef](#)]
27. Foulkes, W.D. Inherited susceptibility to common cancers. *N. Engl. J. Med.* **2008**, *359*, 2143–2153. [[CrossRef](#)]
28. Hsu, F.C.; Roberts, N.J.; Childs, E.; Porter, N.; Rabe, K.G.; Borgida, A.; Ukaegbu, C.; Goggins, M.G.; Hruban, R.H.; Zogopoulos, G.; et al. Risk of Pancreatic Cancer Among Individuals With Pathogenic Variants in the ATM Gene. *JAMA Oncol.* **2021**, *7*, 1664–1668. [[CrossRef](#)]
29. Thompson, D.; Duedal, S.; Kirner, J.; McGuffog, L.; Last, J.; Reiman, A.; Byrd, P.; Taylor, M.; Easton, D.F. Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl. Cancer Inst.* **2005**, *97*, 813–822. [[CrossRef](#)]
30. Li, S.; Silvestri, V.; Leslie, G.; Rebbeck, T.R.; Neuhausen, S.L.; Hopper, J.L.; Nielsen, H.R.; Lee, A.; Yang, X.; McGuffog, L.; et al. Cancer Risks Associated With BRCA1 and BRCA2 Pathogenic Variants. *J. Clin. Oncol.* **2022**, *40*, 1529–1541. [[CrossRef](#)]
31. Pilarski, R. The Role of BRCA Testing in Hereditary Pancreatic and Prostate Cancer Families. *Am. Soc. Clin. Oncol. Educ. Book* **2019**, *39*, 79–86. [[CrossRef](#)] [[PubMed](#)]
32. Lener, M.R.; Kashyap, A.; Kluźniak, W.; Cybulski, C.; Soluch, A.; Pietrzak, S.; Huzarski, T.; Gronwald, J.; Lubiński, J. The Prevalence of Founder Mutations among Individuals from Families with Familial Pancreatic Cancer Syndrome. *Cancer Res. Treat.* **2017**, *49*, 430–436. [[CrossRef](#)] [[PubMed](#)]
33. Tischkowitz, M.; Balmaña, J.; Foulkes, W.D.; James, P.; Ngeow, J.; Schmutzler, R.; Voian, N.; Wick, M.J.; Stewart, D.R.; Pal, T. Management of individuals with germline variants in PALB2: A clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, *23*, 1416–1423. [[CrossRef](#)] [[PubMed](#)]
34. Yang, J.; Xu, R.; Wang, C.; Qiu, J.; Ren, B.; You, L. Early screening and diagnosis strategies of pancreatic cancer: A comprehensive review. *Cancer Commun.* **2021**, *41*, 1257–1274. [[CrossRef](#)] [[PubMed](#)]
35. Miller, D.T.; Lee, K.; Chung, W.K.; Gordon, A.S.; Herman, G.E.; Klein, T.E.; Stewart, D.R.; Amendola, L.M.; Adelman, K.; Bale, S.J.; et al. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, *23*, 1381–1390. [[CrossRef](#)]
36. Miller, D.T.; Lee, K.; Gordon, A.S.; Amendola, L.M.; Adelman, K.; Bale, S.J.; Chung, W.K.; Gollob, M.H.; Harrison, S.M.; Herman, G.E.; et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2021**, *23*, 1391–1398. [[CrossRef](#)]
37. Reiss, K.A.; Mick, R.; O’Hara, M.H.; Teitelbaum, U.; Karasic, T.B.; Schneider, C.; Cowden, S.; Southwell, T.; Romeo, J.; Izzur, N.; et al. Phase II Study of Maintenance Rucaparib in Patients with Platinum-Sensitive Advanced Pancreatic Cancer and a Pathogenic Germline or Somatic Variant in BRCA1, BRCA2, or PALB2. *J. Clin. Oncol.* **2021**, *39*, 2497–2505. [[CrossRef](#)]
38. Truty, R.; Paul, J.; Kennemer, M.; Lincoln, S.E.; Olivares, E.; Nussbaum, R.L.; Aradhya, S. Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. *Genet. Med.* **2019**, *21*, 114–123. [[CrossRef](#)]
39. LaDuca, H.; Polley, E.C.; Yussuf, A.; Hoang, L.; Bs, S.G.; Hart, S.N.; Yadav, S.; Hu, C.; Na Ms, J.; Goldgar, D.E.; et al. A clinical guide to hereditary cancer panel testing: Evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genet. Med.* **2020**, *22*, 407–415. [[CrossRef](#)]
40. Goldstein, A.M.; Chan, M.; Harland, M.; Gillanders, E.M.; Hayward, N.K.; Avril, M.-F.; Azizi, E.; Bianchi-Scarra, G.; Bishop, D.T.; Pailleters, B.B.-D.; et al. High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. *Cancer Res.* **2006**, *66*, 9818–9828. [[CrossRef](#)]