Title: Functional characterization of all *CDKN2A* missense variants and comparison to in silico models of pathogenicity

Authors:

Hirokazu Kimura[1], Kamel Lahouel[2], Cristian Tomasetti[3], and Nicholas J. Roberts[1,4,]*.

Affiliations:

[1]Department of Pathology, the Johns Hopkins University School of Medicine; Baltimore, 21287, USA.

[2]Division of Integrated Genomics, Translational Genomics Research Institute; Phoenix, 85004, USA.

[3]Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope; Duarte, 91010, USA.

[4]Department of Oncology, the Johns Hopkins University School of Medicine; Baltimore, 21287, USA.

*Corresponding author. Email: nrobert8@jhmi.edu

Abstract

Interpretation of variants identified during genetic testing is a significant clinical challenge. In this study, we developed a high-throughput CDKN2A functional assay and characterized all possible *CDKN2A* missense variants. We found that 40% of all missense variants were functionally deleterious. We also used our functional classification to assess the performance of in silico models that predict the effect of variants, including recently reported models based on machine learning. Notably, we found that all in silico models similarly when compared to our functional classifications with accuracies of 54.6 – 70.9%. Furthermore, while we found that functionally deleterious variants were enriched within ankyrin repeats, rarely were all missense variants at a single residue functionally deleterious. Our functional classifications are a resource to aid the interpretation of *CDKN2A* variants and have important implications for the application of variant interpretation guidelines, particularly the use of in silico models for clinical variant interpretation.

Introduction

Genetic testing of patients with cancer to identify variants associated with an increased cancer risk and sensitivity to targeted therapies is becoming more common as broad testing criteria are integrated into clinical care guidelines (Goggins et al., 2020; Stoffel et al., 2019). American College of Medical Genetics (ACMG) provides a framework to integrate multiple types of evidence, including variant characteristics, disease epidemiology, clinical information, and functional classifications, to interpret variants in any gene (Richards et al., 2015). In silico variant effect predictors are also integrated into AMCG variant interpretation guidelines as supporting evidence to aid classification of variants. While numerous models have been developed, varied accuracy, poor agreement between models, and inflated performance on publicly available data have been reported (Cubuk et al., 2021; Jaffe et al., 2011; Wilcox et al., 2022). Recently developed variant effect predictors aim to overcome these limitations by incorporating deep-learning based protein structure predictions and by not training on human annotated datasets (Brandes et al., 2023; Cheng et al., 2023; Gao et al., 2023). However, post-development assessment of machine learning based variant effect predictors to determine accuracy on novel experimental datasets and suitability for clinical use are limited.

Variants that cannot be classified as either pathogenic or benign are categorized as variants of uncertain significance (VUSs). However, while pathogenic and benign variants identified during genetic testing are clinically actionable, VUSs are the cause of deep uncertainty for patients and their health care providers as an unknown fraction are functionally deleterious and therefore, likely pathogenic. For example, individuals with germline VUSs in a pancreatic cancer susceptibility gene would not be eligible for clinical surveillance programs that are associated with improved patient outcomes, unless they otherwise meet family history criteria (Goggins et al., 2020; Stoffel et al., 2019). Similarly, patients with breast or pancreatic cancer and a germline *BRCA2* VUS would not be eligible for treatment with olaparib, a poly (ADP-ribose) polymerase inhibitor (Golan et al., 2019; Tutt et al., 2021). Reclassification of VUSs into pathogenic or benign strata has real-world, life-or-death consequences that necessitate a high degree of accuracy.

Germline VUSs in hereditary cancer genes are a frequent finding in patients with cancer and frequently can be reclassified as pathogenic on the basis of in vitro functional characterization (Kimura et al., 2022). In patients with pancreatic ductal adenocarcinoma (PDAC), germline

CDKN2A VUSs affecting p16^{INK4a}, most often rare missense variants, are found in up to 4.3% of patients (Chaffee et al., 2018; Kimura et al., 2021; McWilliams et al., 2018; Roberts et al., 2016; Shindo et al., 2017; Zhen et al., 2015). As functional data from well-validated in-vitro assays are incorporated into ACMG variant interpretation guidelines, we recently determined the functional consequence of 29 *CDKN2A* VUSs identified in patients with PDAC using an in vitro cell proliferation assay (Kimura et al., 2022; Richards et al., 2015). We found that over 40% of VUSs assayed were functionally deleterious and could reclassified as likely pathogenic.

Functional characterization is time-consuming, expensive, and requires technical and scientific expertise. These limitations hinder assessment of in silico variant effect predictors and patient access to functional data that may allow reclassification of VUSs into clinically actionable strata. As *CDKN2A* VUSs will continue to be identified in patients with cancer undergoing genetic testing, we developed a multiplexed functional assay to provide a broad interpretation framework for *CDKN2A* variants. We characterized all possible *CDKN2A* missense variants and compared our functional classifications to including recently developed in silico models based on machine learning to determine the accuracy of variant effect predictions.

Results

Functional characterization of *CDKN2A* missense variants

We utilized a codon optimized *CDKN2A* sequence for our multiplexed functional assay (Appendix 1-table 1). Expression of codon optimized CDKN2A or three synonymous CDKN2A variants, p.L32L, p.G101G, and p.V126V, in PANC-1, a PDAC cell line with a homozygous deletion of *CDKN2A*, resulted in significant reduction is cell proliferation (P value < 0.0001; Figure 1-figure supplement 1A). Conversely, expression of three pathogenic variants, p.L32P, p.G101W, and p.V126D, in PANC-1 cells did not result in any significant changes in cell proliferation. To determine if there were unappreciated selective effects during in vitro culture, we generated a CellTag library based on the pLJM1 plasmid that contained twenty non-functional 9 base pair barcodes of equal representation. We then transduced PANC-1 cells stably expressing codon optimized CDKN2A with the CellTag library and determined representation of each barcode in the cell pool, before and after in vitro culture. We found no statistically significant changes in barcode representation, indicating that representation of a pool of functionally neutral variants is stable over a period of in vitro culture representing our assay time course (Figure 1-figure supplement 1B).

We next determined whether we could identify functionally deleterious *CDKN2A* variants at a single residue when all amino acid variants were assayed simultaneously. We generated lentiviral expression libraries for two CDKN2A amino acid residues, p.V126 and p.R144, that include pathogenic and benign variants, respectively. Each lentiviral expression library contained all amino acid variants (19 missense and 1 synonymous variant) at a single residue. We then transduced PANC-1 cells with each of the lentiviral expression libraries and determined the representation of each variant in the resulting cell pool before and after a period of in vitro culture (Figure 1A, B). Synonymous variants, p.V126V and p.R1441R, as well as a previously reported benign variant, p.R144C, either decreased or maintained their representation in the cell pool during in vitro culture. Representation of a previously reported pathogenic variant, p.V126D, increased in the cell pool. Notably, several other variants including p.V126R, p.V126W, p.V126K, and p.V126Y, increased in representation, suggesting that additional amino acid changes at this residue are functionally deleterious (Figure 1A).

Next, to functionally characterize 2,964 *CDKN2A* missense variants, we generated 156 CDKN2A lentiviral expression libraries where each library contained all possible amino acid substitutions at a single residue. PANC-1 cells were then transduced with each of the lentiviral expression libraries and representation of each *CDKN2A* variant in the resulting cell pool determined before and after a period of in vitro culture. Variant read counts were then analyzed using a gamma generalized linear model and variants with statistically significant P values were classified as functionally deleterious. Variants with P values that did not reach statistical significance were classified as functionally deleterious.

We found that 1,182 of 2,964 missense variants (39.9%) characterized in our assay were functionally deleterious and that 1,782 variants (60.1%) were functionally neutral (Figure 2A, Appendix 1-table 2). In general, our results were consistent with previously reported classifications. Thirty-two variants identified in patients with cancer and previously reported to be functionally deleterious in published literature and/or reported in ClinVar as pathogenic or likely pathogenic were characterized as functionally deleterious in our assay (Figure 2B, Appendix 1-table 3) (Chaffee et al., 2018; Chang et al., 2016; Horn et al., 2021; Hu et al., 2018; Kimura et al., 2022; McWilliams et al., 2018; Roberts et al., 2016; Zhen et al., 2015). Of 162 benign variants, including 156 synonymous variants and six missense variants previously reported to be functionally neutral in published literature and/or reported in ClinVar as benign or likely benign, all were characterized as functionally neutral in our assay (Figure 2B, Appendix 1-table 3) (Kimura et al., 2022; McWilliams et al., 2018; Roberts et al., 2016). Similarly, of 50 missense variants classified as VUSs in ClinVar that also have previously reported functional data, only two variants, p.G35Q and p.G101R, had divergent functional classifications. Both variants were previously reported to have functionally neutral effects but were characterized as functionally deleterious in our assay (Figure 2C, Appendix 1-table 2, 3).

<u>Comparison to in silico prediction algorithms</u>

As in silico predictions of variant effect are integrated into ACMG variant interpretation guidelines as supporting evidence, we compared the ability of different algorithms, including recently described algorithms that incorporate deep-learning models of protein structure to predict the functional consequence of *CDKN2A* missense variants. Using our functional classifications for all *CDKN2A* missense variants as truth, we compared our functional classifications to predictions from CADD, PolyPhen-2, SIFT, VEST, AlphaMissense, ESM1b, and PrimateAI-3D. Predictions for all missense variants (1,182 functionally deleterious and 1,782 functionally neutral) were available for comparison for all algorithms, except CADD and PrimateAI-3D, where 910 (349 functionally deleterious and 561 functionally neutral) and 904 (349 functionally deleterious and 555 functionally neutral) missense variants had predictions available respectively (Appendix 1-table 4). In silico variant effect predictors performed similarly across a broad range of performance characteristics (Appendix 1-table 5). Accuracy of in silico model predictions were 54.6 – 70.9% (CADD – 56.8%; PolyPhen-2 – 54.6%; SIFT – 61.4%; VEST – 70.8%; AlphaMissense – 70.9%; ESM1b – 64.9%; and PrimateAI-3D; 66.7%) (Figure 3). We also assessed sensitivity, specificity, positive predictive value, and negative predictive value for each model. We found that sensitivity was 0.15 – 0.90 (CADD – 0.86; PolyPhen-2 – 0.9; SIFT – 0.64; VEST – 0.67; AlphaMissense – 0.69; ESM1b – 0.77; and PrimateAI-3D – 0.15), specificity was 0.31 – 0.99 (CADD – 0.39; PolyPhen-2 – 0.32; SIFT – 0.60; VEST – 0.74; AlphaMissense – 0.72; ESM1b – 0.57; and PrimateAI-3D – 0.99), positive predictive value was 0.46 – 0.93 (CADD – 0.47; PolyPhen-2 – 0.46; SIFT – 0.51; VEST – 0.63;

AlphaMissense – 0.62; ESM1b – 0.54; and PrimateAI-3D – 0.93), and negative predictive value was 0.65 – 0.82 (CADD – 0.81; PolyPhen-2 – 0.82; SIFT – 0.71; VEST – 0.77; AlphaMissense – 0.78; ESM1b – 0.79; and PrimateAI-3D – 0.65).

Distribution of functionally deleterious variants

Functionally deleterious missense variants were not distributed evenly across CDKN2A. CDKN2A contains four ankyrin repeats that mediate protein-protein interactions, ankyrin repeat 1 at codon 11-40, ankyrin repeat 2 at codon 44-72, ankyrin repeat 3 at codon 77-106, and ankyrin repeat 4 at codon 110-139 (Goldstein, 2004; Ruas and Peters, 1998; Sun et al., 2010) (Figure 2-figure supplement 1A). Functionally deleterious variants were enriched in ankyrin repeat 1 (44.8%, adjusted P value = 8.5×10^{-4}), ankyrin repeat 2 (47.4%, adjusted P value = 1.4×10^{-6}), and ankyrin repeat 3 (55.0%, adjusted P value = 6.1×10^{-22}), and depleted in ankyrin repeat 4 (27.2%, adjusted P value = 1.6×10^{-8}), non-ankyrin repeat residue 1-10 (22.0%, adjusted P value = 1.5×10^{-5}), and residues 140-156 (9.1%, adjusted P value = 4.5×10^{-30}) (Figure 2-figure supplement 1B). Moreover, functionally deleterious variants were further enriched within 10 residue subregions of ankyrin repeats 1-3, with 59.5% of variants in residues 14 – 23 of ankyrin repeat 1, 63.5% of variants in residues 46-55 of ankyrin repeat 2, and 81.5% of variants in residues 80-89 of ankyrin repeat 3 being classified as functionally deleterious (Figure 2D, Appendix 1-table 2). Furthermore, analysis of functionally deleterious variants may highlight critical and non-critical resides for CDKN2A function. Across all single residues, the mean percent of functionally deleterious missense variants was 39.9% (95% confidence interval: 34.7% – 45.0%) (Figure 2-figure supplement 1C). At six amino acid residues, p.G55, p.P81, p.H83, p.D84, p.G89, and p.P114, all missense variants were functionally deleterious. These residues are conserved between human and murine p16 (Byeon et al., 1998). Furthermore, p.H83 has been reported to stabilized peptide loops connecting the helix-turn-helix structure exhibited by four ankyrin repeats (Byeon et al., 1998), whereas p.D84 and p.G89 are located in a 20-residue region reported to interact with CDK4 and CDK6 (Fåhraeus et al., 1996). Conversely, at 18 residues, no missense variant was characterized as functionally deleterious (Appendix 1-table 2).

Functional effect of *CDKN2A* somatic mutations

Somatic alterations in *CDKN2A* are a frequent finding in many types of cancer. However, not all somatic alterations are unequivocally deleterious to protein function. Missense somatic mutations, for example, are particularly challenging to functionally interpret and the presence of a functionally neutral somatic mutation may impact patient care (Tung et al., 2020). To understand the functional effect of missense somatic mutations in *CDKN2A*, we functionally classified mutations reported in the Catalogue Of Somatic Mutations In Cancer (COSMIC) (Forbes et al., 2009), The Cancer Genome Atlas (TCGA) (Muddabhaktuni and Koyyala, 2021), patients with cancer undergoing sequencing at The Johns Hopkins University School of Medicine (JHU), Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets Clinical Sequencing Cohort (MSK-IMPACT) (Cheng et al., 2015). Overall, 355 unique missense somatic mutations were reported, of which 193 (54.4%) were functionally deleterious (Appendix 1-table 6). The percent of missense somatic mutations that were classified as functionally deleterious was greater than the percent of all possible *CDKN2A* missense variants that were classified as functionally deleterious, suggesting enrichment of functionally deleterious

missense changes among somatic mutations (Figure 2A, Appendix 1-table 6). The prevalence of functionally deleterious missense somatic mutations was similar in COSMIC, TCGA, JHU, and MSK-IMPACT with 55.9%, 68.0%, 65.7%, and 66.7% of mutations being classified as functionally deleterious, respectively (Figure 4A, Appendix 1-table 6). Similar to all functionally deleterious variants, functionally deleterious missense somatic mutations were not distributed evenly across *CDKN2A*. Functionally deleterious somatic mutations were enriched within the ankyrin repeat 3 (Figure 4B, Appendix 1-table 6), We found that 29.0%, 37.1%, 50.8%, and 42.9% of all functionally deleterious missense somatic mutations occurred within ankyrin repeat 3 in COSMIC, TCGA, JHU, and MSK-IMPACT, respectively, and within this domain, 62.7%, 73.1%, 72.7%, and 70.8% of functionally deleterious mutations were in residues 80-89 in COSMIC, TCGA, JHU, and MSK-IMPACT, respectively (Figure 4B).

We were also able to assess the contribution of functionally deleterious *CDKN2A* missense somatic mutations in COSMIC, TCGA, JHU, and MSK-IMAPCT by cancer type, and found that 44.4 – 95.7% of reported *CDKN2A* missense somatic mutations were functionally deleterious when stratified by cancer type (Figure 4-figure supplement 1A-D). When considering missense somatic mutation reported in all databases, there was enrichment of functionally deleterious mutations in melanoma (84.8%; adjusted P value – 0.019) and depletion of functionally deleterious mutations in colorectal adenocarcinoma (49.0%; adjusted P value = 2.6×10^{-4}) (Figure 4-figure supplement 2). As the proportion of missense somatic mutations that were functionally deleterious was less in colorectal carcinoma compared to other types of cancer, we assessed whether somatic mutations in mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) were associated with the functional status of *CDKN2A* missense somatic mutations. Thirty-six samples in COSMIC had a *CDKN2A* missense somatic mutation, of which 12 samples (33.3%) had a somatic mutation in a mismatch repair gene. We found that 3 of 12 samples (25%) with a somatic mutation in a mismatch repair gene had a functionally deleterious *CDKN2A* missense somatic mutation compared to 12 of 24 samples (50%) without a somatic mutation in a mismatch repair gene (Fisher's exact test; P = 0.2821).

Discussion

VUSs in hereditary cancer susceptibility genes, predominantly rare missense variants, are a frequent finding in patients undergoing genetic testing and the cause of significant uncertainty. ACMG variant interpretation guidelines incorporate functional data, as well as other evidence such as in silico predictions of variant effect, to aid classification of variants as either pathogenic or benign. *CDKN2A* VUSs are a frequent finding in patients with PDAC. We previously found that over 40% of *CDKN2A* VUSs were functionally deleterious and therefore could be reclassified as likely pathogenic using ACMG variant interpretation guidelines. In this study, we developed a validated high-throughput in vitro assay and functionally characterized 2,964 *CDKN2A* missense variants, representing all possible single amino acid variants. We found that 1,182 missense variants (39.9%) were functionally deleterious. These pre-defined functional characterizations are resource for the scientific community and can be integrated into variant interpretation schema, such as those from the ACMG, to aid classification of *CDKN2A* germline variants and somatic mutations.

Our comprehensive characterization of all possible *CDKN2A* missense variants allowed us to assess the ability of in silico algorithms – including recently published predictors based on machine learning AlphaMissense, ESM1b, and PrimateAI-3D – to predict the pathogenicity or

functional effect of *CDKN2A* missense variants. We found that all in silico variant effect predictors assessed performed similarly. Overall, PolyPhen-2 had highest sensitivity and negative predictive value at 0.90 and 0.82 respectively. PrimateAI-3D had the highest specificity and positive predictive values at 0.99 and 0.93, respectively. However, while specificity and positive predictive values were high, Primate AI-3D lowest sensitivity and negative predictive values at 0.15 and 0.65 respectively. Highest accuracy was observed with AlphaMissense at 70.9%, closely followed by VEST at 70.8%. Given that reclassification of VUSs in hereditary cancer genes into inappropriate strata has significant implications for patients, integration of in silico models, including those utilizing machine learning, may be premature. Ultimately, our data support current ACMG guidelines that include in silico predictions of variant effect as supporting evidence of pathogenicity or benign impact.

Our study also provides other insight for the implementation of variant interpretation guidelines. ACMG guidelines include presence of a missense variants at a residue with a previously reported pathogenic variant as moderate evidence of pathogenicity. We found that the mean percent of functionally deleterious missense variants per residue was 39.9% and that only six amino acid residues were all missense variants functionally deleterious. These data suggest, at least for *CDKN2A*, that the presence of a pathogenic missense variant at a residue should be used with caution when classifying other missense variants at the same residue.

Our high-throughput functional assay characterized variants based upon a broad cellular phenotype, cell proliferation, in a single PDAC cell line. However, there appear to be limited cell-specific and assay-specific differences in functional classifications of *CDKN2A* variants. In our previous study, we characterized 29 *CDKN2A* VUSs in three PDAC cell lines, using cell proliferation and cell cycle assays, and found agreement between all functional classifications (Kimura et al., 2022). Nevertheless, our assay may not encompass all cellular functions of *CDKN2A* and underestimate the number of functionally deleterious missense variants. Moreover, in this study, we found that functionally deleterious effects were enriched among somatic missense mutations compared to all *CDKN2A* missense variants, further supporting the veracity of our functional classifications. Furthermore, in silico models to predict variant effect may characterize function (CADD, PolyPhen-2, and SIFT) or pathogenicity (VEST, AlphaMissense, ESM1b, and PrimateAI-3D). As our assay does not directly characterize pathogenicity, to compare in silico assays, we assumed that predicted pathogenic variants were functionally deleterious.

In this study, we determined functional classifications for all possible *CDKN2A* missense variants to aid variant classifications. Furthermore, comparison of our functional classifications to in silico variant effect predictors, including recently described algorithms based on machine learning, provides performance benchmarks and supports current recommendations integrating data computational data into variant interpretation guidelines.

Methods

<u>Cell lines</u>

PANC-1 (American Type Culture Collection, Manassas, VA; catalog no. CRL-1469), a human PDAC cell line with a homozygous deletion of *CDKN2A* (Caldasl et al., 1994) and 293T (American Type Culture Collection; catalog no. CRL-3216), a human embryonic kidney cell line, were maintained in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific Inc.,

Waltham, MA; catalog no.11995-065) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific Inc.; catalog no. 26140-079). Cell line authentication and mycoplasma testing were performed using the GenePrint 10 System (Promega Corporation, Madison, WI; catalog no. B9510) and the PCR-based MycoDtect kit (Greiner Bio-One, Monroe, NC; catalog no. 463 060) (Genetics Resource Core Facility, The Johns Hopkins University, Baltimore, MD).

CDKN2A somatic mutation data

CDKN2A ($p16^{INK4}$; NP_000068.1) missense somatic mutation data was obtained from the Catalogue Of Somatic Mutations In Cancer (Forbes et al., 2009), The Cancer Genome Atlas (Muddabhaktuni and Koyyala, 2021), patients with cancer undergoing sequencing at The Johns Hopkins University School of Medicine (Baltimore, MD), Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets Clinical Sequencing Cohort (Cheng et al., 2015).

Plasmids

pHAGE-CDKN2A (Addgene, Watertown, MA; plasmid no. 116726) was created by Gordon Mills & Kenneth Scott (Ng et al., 2018). pLJM1 (Addgene; plasmid no. 91980) was created by Joshua Mendell (Golden et al., 2017). pLentiV_Blast (Addgene, plasmid no. 111887) was created by Christopher Vakoc (Tarumoto et al., 2020). psPAX2 (Addgene, plasmid no. 12260) was created by Didier Trono), and pCMV-VSV-G (Addgene, plasmid no. 8454) was created by Bob Weinberg (Stewart et al., 2003).

CDKN2A expression plasmid libraries

CDKN2A cDNA from pHAGE-CDKN2A was subcloned into the pLJM1 plasmid as previously described (Kimura et al., 2022). Codon-optimized *CDKN2A* cDNA using $p16^{INK4A}$ amino acid sequence (NP_000068.1), was designed (Appendix 1-table 1) and pLJM1 containing codon optimized CDKN2A (pLJM1-CDKN2A) generated (Twist Bioscience, South San Francisco, CA). Then, 156 plasmid libraries were synthesized using pLJM1-CDKN2A such that each library contained all possible 20 amino acids variants (19 missense and 1 synonymous) at a given position.

Single variant CDKN2A expression plasmids

Individual pLJM1-CDKN2A expression constructs for *CDKN2A* missense variants, p.L32L, p.L32P, p.G101G, p.G101W, p.V126D, and p.V126V were generated using the Q5 Site-Directed Mutagenesis kit (New England Biolabs, Ipswich, MA; catalog no. E0552). Primers used for site-directed mutagenesis are given in Appendix 1-table 7. Integration of each *CDKN2A* variant was confirmed using Sanger sequencing (Genewiz, Plainsfield, NJ) using the CMV Forward sequencing primer (CGCAAATGGGCGGTAGGCGTG). The manufacturer's protocol was followed unless otherwise specified.

CellTag plasmid library

Twenty nonfunctional 9 base pair barcodes "CellTags" were subcloned into pLentiV_Blast using Q5® Site-Directed Mutagenesis kit (New England Biolabs; catalog no. E0552) (Biddy et al., 2018). Primers used to generate each CellTag plasmid are given in Appendix 1-table 7. Integration of each CellTag was confirmed using Sanger sequencing (Genewiz) (sequencing primer: AACTGGGAAAGTGATGTCGTG). The manufacturer's protocol was followed unless otherwise specified. CellTag plasmids were then pooled to form a CellTag plasmid library with equal representation of each CellTag plasmid.

Lentivirus production

Lentivirus production was performed as previously described with the following modifications (Kimura et al., 2022). pLJM1 lentiviral expression vectors (plasmid libraries and single variant expression plasmids) and lentiviral packaging vectors (psPAX2 and pCMV-VSV-G) were transfected into 293T cells using Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific, Waltham, MA; catalog no. L3000008). Media was collected at 24 hours and 48 hours, pooled, and lentiviral particles concentrated using Lenti-X Concentrator (Clontech, Mountain View, CA; catalog no. 631231).

Lentiviral transduction

PANC-1 cells were used for CDKN2A plasmid library and single variant CDKN2A expression plasmid transductions. PANC-1 cells previously transduced with pLJM1-CDKN2A (PANC-1^{CDKN2A}) were used for CellTag library transductions. Briefly, 1 x 10^5 cells were cultured in media supplemented with 10 ug/ml polybrene and transduced with 4 x 10^7 transducing units per mL of lentivirus particles. Cells were then centrifuged at 1,200 x g for 1 hour. After 48 hours of culture at 37^oC and 5% CO_2, transduced cells were selected using 3 µg/ml puromycin (CDKN2A plasmid libraries and single variant CDKN2A expression plasmids) or 5 µg/ml blasticidin (CellTag plasmid library) for 7 days. After selection, cells were trypsinized for passage into T150 flasks and DNA collection. T150 flasks were cultured for 2-5 weeks until confluent and then DNA collected. DNA was extracted from PANC-1 cells using the PureLink™ Genomic DNA Mini Kit (Invitrogen, Carlsbad, CA; catalog no. K1820-01).

Generation of sequence libraries

Library preparation and sequencing was performed as previously described with the following modifications (Kinde et al., 2011). For the 1st stage PCR, 3 target specific primers were designed to amplify *CDKN2A* amino acid positions 1 to 53, 54 to 110, and 111 to 156. Forward and reverse 1st stage primers contained 5' sequence, M13F (GTAAAACGACGGCCAGC) and M13R (CAGGAAACAGCTATGAC) respectively, to enable second stage of amplification and ligation of Illumina adapter sequences (Appendix 1-table 7). DNA was amplified with Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs; catalog no. M0494S). For the 1st stage PCR, each DNA sample was amplified in three reactions each containing 66 ng of DNA.1st stage PCR products for each sample were then pooled and purified using the Agencourt AMPure XP system (Beckman Coulter, Inc, Brea, CA; catalog no. A63881) into 50 µL of elution buffer. Purified PCR product was amplified in a 2nd stage PCR to add Illumina adaptor sequences and indexes (Appendix 1-table 7). PCR Amplification was performed with KAPA HiFi HotStart PCR Kit (Kapa Biosystems, Wilmington, MA; catalog no. KK2501) in 25 µL reactions

containing 5X KAPA HiFi Buffer - 5 µL, 10 mM KAPA dNTP Mix - 0.75 µL, 10 µM forward primer - 0.75 µL, 10 µM reverse primer - 0.75 µL. For the 1st stage PCR, 66 ng of template DNA and 12.5 µL, Q5 Hot Start High-Fidelity 2X Master Mix was used with the following cycling conditions: 98 ^{o}C for 30 seconds; 25 cycles of 98 ^{o}C for 10 seconds, 72 ^{o}C for 30 seconds, 72 ^{o}C for 25 seconds; 72 ^{o}C for 2 minutes. For the 2nd stage PCR, 0.25 µL of of 1st stage PCR product and 0.5 µL of 1 U/µL KAPA HiFi HotStart DNA Polymerase was used with the following cycling conditions: 95 ^{o}C for 3 minutes; 25 cycles of 98 ^{o}C for 20 seconds, 62 ^{o}C for 15 seconds, 72 ^{o}C for 1 minute. 2nd stage PCR products were purified with the Agencourt AMPure XP system (Beckman Coulter, Inc.; catalog no. A63881) into 30 µL of elution buffer. Samples were quantified by Qubit using dsDNA HS assay kit (Invitrogen; catalog no. Q33230).

Sequencing and analysis

Sequence libraries were pooled into groups of 16 samples and sequenced on the Illumina MiSeq System (Illumina, San Diego, CA) with the MiSeq Reagent Kit v2 (300 cycles) (Illumina catalog no. MS-102-2002) to generate 150 base pair paired-end reads. Samples were demultiplexed and FASTQ sequenced read files were generated with MiSeq control software 2.5.0.5 (Illumina). Paired sequence reads were then combined into a single contiguous sequence using Paired-End Read Merger (Zhang et al., 2014). Reads supporting each variant at a given amino acid position were counted using perl.

Functional characterization of *CDKN2A* variants using gamma generalized linear model

We determined if a variant has a fitness advantage by assessing the significance of the observed ratio $r_{v,cf}$ at confluence between the number of cells with a missense variant v and the number of cells with a synonymous variant at a given amino acid position. Under the assumption that the missense variant is neutral (null model), we assumed that the distribution of $r_{v,cf}$ can be explained by two key covariates: $r_{v,init}$, which represent the missense variant-to-synonymous variant ratio at day 9, and $p_{v,init}$, the proportion of the missense variant cells among other variants, including the synonymous variant, at the studied position. More specifically, given the variables $r_{v,init}$ and $p_{v,init}$, the ratio at confluence follows a distribution:

$$r_{v,cf} \sim \Gamma(\alpha, \beta_v)$$

where the mean u_v of the Gamma distribution is such that:

$$u_v = \frac{\alpha}{\beta_v} = r_{v,init}^{a}\ p_{v,init}^{b}.$$

Here, the parameters of the null model to estimate are $\alpha, a, and\ b$, where α, is the shape parameter of the Gamma distribution and is assumed to be the same for all variants. This model is a Gamma Generalized Linear Model (GLM) over the response variable $r_{v,cf}$ with a log-link function and covariates $log(r_{v,init})$ and $log(p_{v,init})$. Estimating the parameters will provide a null distribution of $r_{v,cf}$, generating a p-value for every observed $r_{v,cf}$ for any variant at a given position.

To estimate the parameters α, a, and b, we utilized three control experiments where the CellTag plasmid library was transduced into PANC-1^{CDKN2Aco} cells so that each CellTag represented a neutral variant. For a single experiment, every variant can be considered as wild-type, and we

test the other 19 variants against it, knowing that they are neutral and therefore follow the null distribution. This provides us with 19 x 20 triplets $\left(\mathrm{r_{v,cf}}, \mathrm{p_v^{init}}, \mathrm{r_v^{init}}\right)$, for every experiment, yielding 1140 datapoints when considering all three experiments together. To estimate the parameters using these 1140 data points, we fit the GLM corresponding GLM model using the sklearn.linear_model module.

After the estimation of parameters α, a, and b, every observation for a tested variant v at a given position of the triplet $\left(\mathrm{r_{v,cf}}, \mathrm{p_v^{init}}, \mathrm{r_v^{init}}\right)$ yields a p-value, defined as the probability of observing a ratio at confluence that is at least $r_{v,cf}$ given $p_{v,init}, r_{v,init}$ under the null Gamma model. As some variants were tested in repeated experiments, we combined their associated p-values into a single p-value using Fisher's method. Finally, to determine if a variant presents a fitness advantage, we apply a Benjamini-Hochberg estimator on all the tested variants p-values, fixing the False Discovery Rate at a level of 0.05.

Data visualization

Heat map of individual variant p-values by amino position was generated using R with the heatmaply package (Galili et al., 2018).

Cell proliferation assay

Cell proliferation assay were performed as previously described with the following modifications (Kimura et al., 2022). 1×10^5 cells were seeded into in vitro culture on day 0. Cell were counted on day 14 using a TC20 Automated Cell Counter (Bio-Rad Laboratories, Herclues, CA; catalog no. 1450102). Relative cell proliferation value was calculated as cell number normalized to empty vector control. Assays were repeated in triplicate. Mean cell proliferation value and standard deviation (s.d.) were calculated.

Variant effect predictions

Publicly available algorithms were used to predict the consequence of *CDKN2A* missense variants. Prediction algorithms used included: Combined Annotation Dependent Depletion (CADD) (Kircher et al., 2014), Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei et al., 2010), Sorting Intolerant From Tolerant (SIFT) (Kumar et al., 2009), Variant Effect Scoring Tool score (VEST) (Carter et al., 2013), AlphaMissense (*18*), ESM1b (Brandes et al., 2023), and PrimateAI-3D (Gao et al., 2023) (Appendix 1-table 4). PolyPhen-2, SIFT, VEST, AlphaMissense, and ESM1b prediction were available for all missense variants. CADD scores were available for 910 missense variants and where multiple CADD scores were possible, mean values were used. PrimateAI-3D prediction scores were available for 904 assayed missense variants.

Statistical analyses

Statistical analyses were performed using JMP v.11 (SAS, Cary, NC) and Python statsmodel package (version 0.14.0). Cell proliferation value means were compared with the Student's t-test. Proportions of functionally deleterious missense variants and somatic mutations were compared

with the Z-test and multiple test correction performed with the Bonferroni method. P values < 0.05 were considered statistically significant.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**:248–249. doi:10.1038/nmeth0410-248

Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, Morris SA. 2018. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**:219–224. doi:10.1038/s41586-018-0744-4

Brandes N, Goldman G, Wang CH, Ye CJ, Ntranos V. 2023. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat Genet* **55**:1512–1522. doi:10.1038/s41588-023-01465-0

Byeon IJL, Li J, Ericson K, Selby TL, Tevelev A, Kim HJ, O'Maille P, Tsai MD. 1998. Tumor suppressor p16INK4A: Determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol Cell* **1**:421–431. doi:10.1016/S1097-2765(00)80042-8

Caldasl C, Hahn SA, Luis T, Marks C, Schutte M, Seymour AB, Weinstein CL, Hruban RH, Yeo CJ, Kern SE. 1994. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat Genet* **8**:27–32.

Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. 2013. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14 Suppl 3**:S3. doi:10.1186/1471-2164-14-s3-s3

Chaffee KG, Oberg AL, McWilliams RR, Majithia N, Allen BA, Kidd J, Singh N, Hartman A-R, Wenstrup RJ, Petersen GM. 2018. Prevalence of Germline Mutations in Cancer Genes. *Genet Med* **20**:119–127. doi:10.1038/gim.2017.85.PREVALENCE

Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, Gao JJ, Socci ND, Solit DB, Olshen AB, Schultz N, Taylor BS. 2016. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* **34**:155–163. doi:10.1038/nbt.3391

Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Rose Brannon A, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF. 2015. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagnostics* **17**:251–264. doi:10.1016/j.jmoldx.2014.12.006

Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, Schneider RG, Senior AW, Jumper J, Hassabis D, Kohli P, Avsec Ž. 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science (80-)* **7492**. doi:10.1126/science.adg7492

Cubuk C, Garrett A, Choi S, King L, Loveday C, Torr B, Burghel GJ, Durkie M, Callaway A, Robinson R, Drummond J, Berry I, Wallace A, Eccles D, Tischkowitz M, Whiffin N, Ware

JS, Hanson H, Turnbull C, CanVIG-UK. 2021. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet Med* **23**:2096–2104. doi:10.1038/s41436-021-01265-z

Fåhraeus R, Paramio JM, Ball KL, Lain S, Lane DP. 1996. Inhibition of pRb phosphorylation and cell-cycle progression by a 20-residue peptide derived from P16CDKN2/INK4A. *Curr Biol*. doi:10.1016/s0960-9822(02)00425-6

Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA. 2009. COSMIC (the Catalogue of Somatic Mutations In Cancer): A resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* **38**:652–657. doi:10.1093/nar/gkp995

Galili T, O'Callaghan A, Sidi J, Sievert C. 2018. Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**:1600–1602. doi:10.1093/bioinformatics/btx657

Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, Yang Y, Dietrich ASD, Fiziev PP, Kuderna LFK, Sundaram L, Wu Y, Adhikari A, Field Y, Chen C, Batzoglou S, Aguet F, Lemire G, Reimers R, Balick D, Janiak MC, Kuhlwilm M, Orkin JD, Manu S, Valenzuela A, Bergman J, Rousselle M, Silva FE, Agueda L, Blanc J, Gut M, de Vries D, Goodhead I, Harris RA, Raveendran M, Jensen A, Chuma IS, Horvath JE, Hvilsom C, Juan D, Frandsen P, de Melo FR, Bertuol F, Byrne H, Sampaio I, Farias I, do Amaral JV, Messias M, da Silva MNF, Trivedi M, Rossi R, Hrbek T, Andriaholinirina N, Rabarivola CJ, Zaramody A, Jolly CJ, Phillips-Conroy J, Wilkerson G, Abee C, Simmons JH, Fernandez-Duque E, Kanthaswamy S, Shiferaw F, Wu D, Zhou L, Shao Y, Zhang G, Keyyu JD, Knauf S, Le MD, Lizano E, Merker S, Navarro A, Bataillon T, Nadler T, Khor CC, Lee J, Tan P, Lim WK, Kitchener AC, Zinner D, Gut I, Melin A, Guschanski K, Schierup MH, Beck RMD, Umapathy G, Roos C, Boubli JP, Lek M, Sunyaev S, O'Donnell-Luria A, Rehm HL, Xu J, Rogers J, Marques-Bonet T, Farh KKH. 2023. The landscape of tolerated genetic variation in humans and primates. *Science (80-)* **380**. doi:10.1126/science.abn8197

Goggins M, Overbeek KA, Brand R, Syngal S, Del Chiaro M, Bartsch DK, Bassi C, Carrato A, Farrell J, Fishman EK, Fockens P, Gress TM, Van Hooft JE, Hruban RH, Kastrinos F, Klein A, Lennon AM, Lucas A, Park W, Rustgi A, Simeone D, Stoffel E, Vasen HFA, Cahen DL, Canto MI, Bruno M. 2020. Management of patients with increased risk for familial pancreatic cancer: updated recommendations from the International Cancer of the Pancreas Screening (CAPS) Consortium. *Gut* **69**:7–17. doi:10.1136/gutjnl-2019-319352

Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, Park J-O, Hochhauser D, Arnold D, Oh D-Y, Reinacher-Schick A, Tortora G, Algül H, O'Reilly EM, McGuinness D, Cui KY, Schlienger K, Locker GY, Kindler HL. 2019. Maintenance Olaparib for Germline BRCA -Mutated Metastatic Pancreatic Cancer . *N Engl J Med* **381**:317–327. doi:10.1056/nejmoa1903387

Golden RJ, Chen B, Li T, Braun J, Manjunath H, Chen X, Wu J, Schmid V, Chang TC, Kopp F, Ramirez-Martinez A, Tagliabracci VS, Chen ZJ, Xie Y, Mendell JT. 2017. An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* **542**:197–202. doi:10.1038/nature21025

Goldstein AM. 2004. Familial melanoma, pancreatic cancer and germline CDKN2A mutations. *Hum Mutat* **23**:630. doi:10.1002/humu.9247

Horn IP, Marks DL, Koenig AN, Hogenson TL, Almada LL, Goldstein LE, Romecin Duran PA, Vera R, Vrabel AM, Cui G, Rabe KG, Bamlet WR, Mer G, Sicotte H, Zhang C, Li H, Petersen GM, Fernandez-Zapico ME. 2021. A rare germline CDKN2A variant (47T>G; p16-L16R) predisposes carriers to pancreatic cancer by reducing cell cycle inhibition. *J Biol Chem* **296**:1–11. doi:10.1016/J.JBC.2021.100634

Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, Lilyquist J, Na J, Moore R, Antwi SO, Bamlet WR, Chaffee KG, DiCarlo J, Wu Z, Samara R, Kasi PM, McWilliams RR, Petersen GM, Couch FJ. 2018. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. *JAMA* **319**:2401–2409. doi:10.1001/jama.2018.6228

Jaffe A, Wojcik G, Chu A, Golozar A, Maroo A, Duggal P, Klein AP. 2011. Identification of functional genetic variation in exome sequence analysis. *BMC Proc* **5**:9–13. doi:10.1186/1753-6561-5-S9-S13

Kimura H, Klein AP, Hruban RH, Roberts NJ. 2021. The Role of Inherited Pathogenic CDKN2A Variants in Susceptibility to Pancreatic Cancer. *Pancreas* **50**:1123–1130. doi:10.1097/MPA.0000000000001888

Kimura H, Paranal RM, Nanda N, Wood LD, Eshleman JR, Hruban RH, Goggins MG, Klein AP, Brand R, Cote ML, Du M, Gallinger S, Goggins M, Kurtz RC, Petersen GM, Rustgi AK, Schwartz AG, Stoffel EM, Syngal S, Zogopoulos G, Roberts NJ. 2022. Functional CDKN2A assay identifies frequent deleterious alleles misclassified as variants of uncertain significance. *Elife* **11**:1–16. doi:10.7554/eLife.71137

Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**:9530–9535. doi:10.1073/pnas.1105422108

Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**:310–315. doi:10.1038/ng.2892

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**:1073–1082. doi:10.1038/nprot.2009.86

McWilliams RR, Wieben ED, Chaffee KG, Antwi SO, Raskin L, Olopade OI, Li D, Highsmith WE, Colon-Otero G, Khanna LG, Permuth JB, Olson JE, Frucht H, Genkinger J, Zheng W, Blot WJ, Wu L, Almada LL, Fernandez-Zapico ME, Sicotte H, Pedersen KS, Petersen GM. 2018. CDKN2A germline rare coding variants and risk of pancreatic cancer in minority populations. *Cancer Epidemiol Biomarkers Prev* **27**:1364–1370. doi:10.1158/1055-9965.EPI-17-1065

Muddabhaktuni BMC, Koyyala VPB. 2021. The Cancer Genome Atlas. *Indian J Med Paediatr Oncol* **42**:353–355. doi:10.1055/s-0041-1735440

Ng PKS, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, Sengupta S, Wang Z, Bhavana VH, Tran R, Soewito S, Minussi DC, Moreno D, Kong K, Dogruluk T, Lu H, Gao J, Tokheim C, Zhou DC, Johnson AM, Zeng J, Ip CKM, Ju Z, Wester M, Yu S, Li Y, Vellano CP, Schultz

N, Karchin R, Ding L, Lu Y, Cheung LWT, Chen K, Shaw KR, Meric-Bernstam F, Scott KL, Yi S, Sahni N, Liang H, Mills GB. 2018. Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell* **33**:450-462.e10. doi:10.1016/j.ccell.2018.01.021

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. 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* **17**:405–424. doi:10.1038/gim.2015.30

Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, Kurtz RC, Olson SH, Rustgi AK, Schwartz AG, Stoffel E, Syngal S, Zogopoulos G, Ali SZ, Axilbund J, Chaffee KG, Chen YC, Cote ML, Childs EJ, Douville C, Goes FS, Herman JM, Iacobuzio-Donahue C, Kramer M, Makohon-Moore A, McCombie RW, Wyatt Mcmahon K, Niknafs N, Parla J, Pirooznia M, Potash JB, Rhim AD, Smith AL, Wang Y, Wolfgang CL, Wood LD, Zandi PP, Goggins M, Karchin R, Eshleman JR, Papadopoulos N, Kinzler KW, Vogelstein B, Hruban RH, Klein AP. 2016. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov* **6**:166–175. doi:10.1158/2159-8290.CD-15-0402

Ruas M, Peters G. 1998. The p16(INK4a)/CDKN2A tumor suppressor and its relatives. *Biochim Biophys Acta - Rev Cancer* **1378**. doi:10.1016/S0304-419X(98)00017-1

Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, Siddiqui A, Witmer PD, Tamura K, Song TJ, Almario JAN, Brant A, Borges M, Ford M, Barkley T, He J, Weiss MJ, Wolfgang CL, Roberts NJ, Hruban RH, Klein AP, Goggins M. 2017. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. *J Clin Oncol* **35**:3382–3390. doi:10.1200/JCO.2017.72.3502

Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS, Sabatini DM, Chen ISY, Hahn WC, Sharp PA, Weinberg RA, Novina CD. 2003. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *Rna* **9**:493–501. doi:10.1261/rna.2192803

Stoffel EM, Mckernin SE, Brand R, Canto M, Goggins M, Moravek C. 2019. Evaluating Susceptibility to Pancreatic Cancer : ASCO Provisional Clinical Opinion. *J Clin Oncol* **37**:153–164. doi:10.1200/JCO.18.01489

Sun P, Nallar SC, Raha A, Kalakonda S, Velalar CN, Reddy SP, Kalvakolanu D V. 2010. GRIM-19 and p16INK4a synergistically regulate cell cycle progression and E2F1-responsive gene expression. *J Biol Chem* **285**:27545–27552. doi:10.1074/jbc.M110.105767

Tarumoto Y, Lin S, Wang J, Milazzo JP, Xu Y, Lu B, Yang Z, Wei Y, Polyanskaya S, Wunderlich M, Gray NS, Stegmaier K, Vakoc CR. 2020. Salt-inducible kinase inhibition suppresses acute myeloid leukemia progression in vivo. *Blood* **135**:56–70. doi:10.1182/blood.2019001576

Tung NM, Robson ME, Ventz S, Santa-Maria CA, Nanda R, Marcom PK, Shah PD, Ballinger TJ, Yang ES, Vinayak S, Melisko M, Brufsky A, DeMeo M, Jenkins C, Domchek S, D'Andrea A, Lin NU, Hughes ME, Carey LA, Wagle N, Wulf GM, Krop IE, Wolff AC, Winer EP, Garber JE. 2020. TBCRC 048: Phase II Study of Olaparib for Metastatic Breast Cancer and Mutations in Homologous Recombination-Related Genes. *J Clin Oncol* **38**:4274–4282. doi:10.1200/JCO.20.02151

Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, Gelber RD, de Azambuja E, Fielding A, Balmaña J, Domchek SM, Gelmon KA, Hollingsworth SJ, Korde LA, Linderholm B, Bandos H, Senkus E, Suga JM, Shao Z, Pippas AW, Nowecki Z, Huzarski T, Ganz PA, Lucas PC, Baker N, Loibl S, McConnell R, Piccart M, Schmutzler R, Steger GG, Costantino JP, Arahmani A, Wolmark N, McFadden E, Karantza V, Lakhani SR, Yothers G, Campbell C, Geyer CE. 2021. Adjuvant Olaparib for Patients with BRCA1 - or BRCA2 -Mutated Breast Cancer . *N Engl J Med* **384**:2394–2405. doi:10.1056/nejmoa2105215

Wilcox EH, Sarmady M, Wulf B, Wright MW, Rehm HL, Biesecker LG, Abou Tayoun AN. 2022. Evaluating the impact of in silico predictors on clinical variant classification. *Genet Med* **24**:924–930. doi:10.1016/j.gim.2021.11.018

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**:614–620. doi:10.1093/bioinformatics/btt593

Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, Hruban RH, Cote ML, Mcwilliams RR, Roberts NJ, Cannon-Albright LA, Li D, Moyes K, Wenstrup RJ, Hartman A-R, Seminara D, Klein AP, Petersen GM, Author GM. 2015. BRCA1, BRCA2, PALB2, and CDKN2A Mutations in Familial Pancreatic Cancer (FPC): A PACGENE Study HHS Public. *Genet Med* **17**:569–577. doi:10.1038/gim.2014.153

Acknowledgments

Funding

National Institutes of Health grant P50CA62924 (NJR)

Susan Wojcicki and Dennis Troper (NR)

The Sol Goldman Pancreatic Cancer Research Center (NJR)

The Rolfe Pancreatic Cancer Foundation (NJR)

The Japanese Society of Gastroenterology Support for Young Gastroenterologists Studying in the United States (HK)

The Japan Society for the Promotion of Science Overseas Research Fellowships (HK)

Author contributions

Conceptualization: HK, KL, CT, NJR

Resources: CT, NJR

Data curation: HK, KL, CT, NJR

Formal analysis: HK, KL, CT, NJR

Investigation: HK, KL, CT, NJR

Visualization: HK

Methodology: HK, KL, CT, NJR

Writing-original draft: HK, NJR

Project administration: NJR

Writing-review and editing: HK, KL, CT, NJR

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

All data are available in the main text or the supplementary materials.

Figures

Figure 1. Pooled analysis of *CDKN2A* variants at two residues with previously reported pathogenic and benign variants.

Figure 2. Functional characterization of all possible *CDKN2A* missense variants.

Figure 3. Comparison of functional classifications and in silico variant effect predictions for all possible CDKN2A missense variants.

Figure 4. Functional classification of missense somatic mutations in *CDKN2A*.

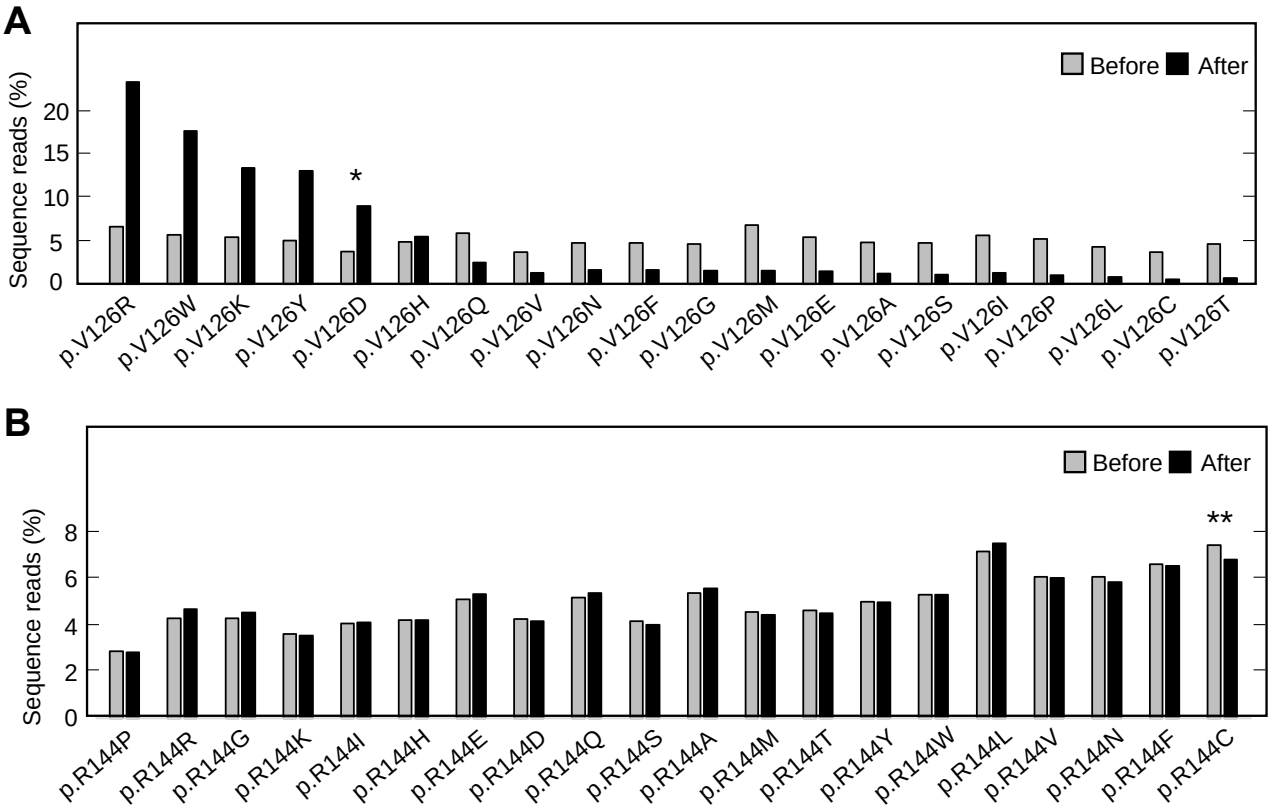

Figure 1. Pooled analysis of *CDKN2A* variants at two residues with previously reported pathogenic and benign variants. PANC-1 cell stably expressed with a total of 20 *CDKN2A* variants, 19 missense variants and 1 synonymous variant, at resides were cultured. Variant representation, as percent of reads supporting the variant sequence, before and after a period in vitro cell growth determined by next generation sequencing for two residues, p.V126 (**A**) or p.R144 (**B**). *CDKN2A* variant p.V126D (*) was previously reported as pathogenic and increased representation during in vitro growth. CDKN2A variant p.R144C (**) was previously reported as benign variant and maintained representation during in vitro growth.

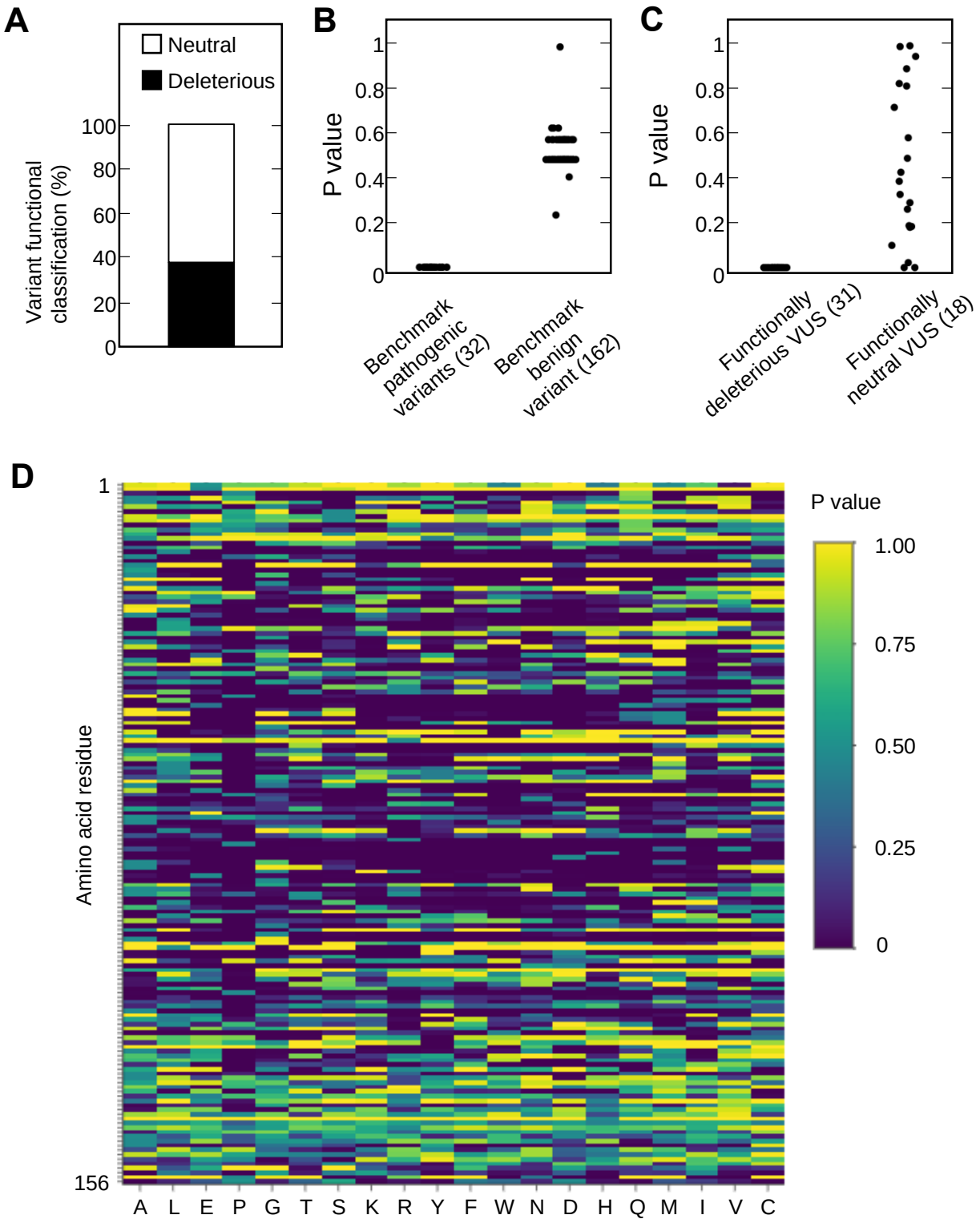
A
Neutral
Deleterious
Variant functional classification (%)
100
80
60
40
20
0
B
P value
1
0.8
0.6
0.4
0.2
0
Benchmark pathogenic variants (32)
Benchmark benign variant (162)
C
P value
1
0.8
0.6
0.4
0.2
0
Functionally deleterious VUS (31)
Functionally neutral VUS (18)
D
Amino acid residue
1
156
A L E P G T S K R Y F W N D H Q M I V C
P value
1.00
0.75
0.50
0.25
0

Figure 2. Functional characterization of all possible *CDKN2A* missense variants. (**A**) Functional classifications of 3,120 CDKN2A variants, including all possible 2,964 missense variants and 156 synonymous variants. Variants were classified as functionally deleterious or neutral based on P value. 1,182 (39.9%) of variants were classified as functionally deleterious. (**B**) P values for 32 benchmark pathogenic variants and 162 benign variants. All pathogenic variants were classified as functionally deleterious, and all benign variants were classified as functionally neutral, based on P values. (**C**) High-throughput functional assay P values for 32 *CDKN2A* VUSs previously reported to have functionally deleterious effects and 18 VUSs previously reported to have functionally neutral effects. (**D**) Heat map with P values for all 3,120 CDKN2A variants assayed.

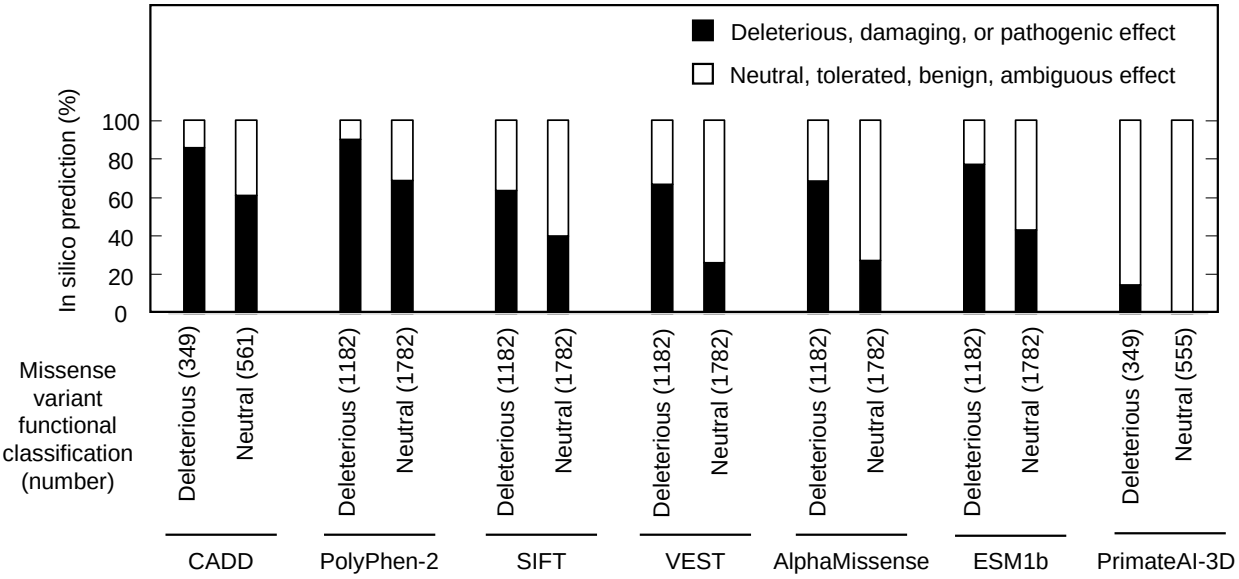

Figure 3. Comparison of functional classifications and in silico variant effect predictions for all possible CDKN2A missense variants. Variant effect predictions for CDKN2A missense variants using CADD, PolyPhen-2, SIFT, VEST, AlphaMissense, ESM1b, and PrimateAI-3D. Predicted deleterious, damaging, or pathogenic effects (black box) and predicted neutral, tolerated, benign, or ambiguous effects (white box) presented as percent of missense variants with available prediction. Number of missense variants with available prediction for each in silico model given in parentheses. CADD; Combined Annotation Dependent Depletion, PolyPhen-2; Polymorphism Phenotyping v2, SIFT; Sorting Intolerant From Tolerant, VEST; Variant Effect Scoring Tool score.

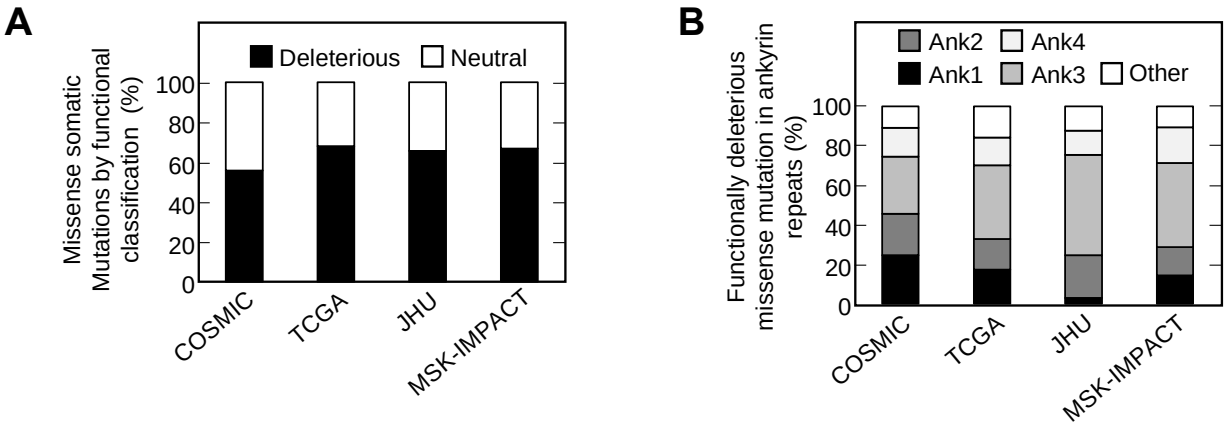

Figure 4. Functional classification of missense somatic mutations in *CDKN2A*. (**A**) Missense somatic missense variants in *CDKN2A* reported in COSMIC, TCGA, JHU, or MSK-IMPACT, by functional classification (deleterious – black box; neutral – white box). (**B**) Distribution of functionally deleterious missense somatic mutations *CDKN2A* reported in COSMIC, TCGA, JHU, or MSK-IMPACT by ankyrin (ANK) repeat.

Supplementary Information

Figures

Figure 1-figure supplement 1. Development and validation of high-throughput CDKN2A functional assay.

Figure 2-figure supplement 1. Functional characterization of all possible *CDKN2A* missense variants.

Figure 4-figure supplement 1. Functional classification of missense somatic mutations in *CDKN2A*.

Figure 4-figure supplement 2. Aggregate functional classifications for missense somatic mutations in *CDKN2A*

Tables

Appendix 1-table 1. Codon optimized *CDKN2A* sequence.

Appendix 1-table 2. Assay outputs and functional classifications for all possible *CDKN2A* missense and synonymous variants.

Appendix 1-table 3. Benchmark pathogenic variants, benchmark benign variants, and variants of uncertain significance with previously reported function data.

Appendix 1-table 4. In silico variant effect predictions for *CDKN2A* missense variants.

Appendix 1-table 5. Assessment of in silico variant effect prediction models.

Appendix 1-table 6. Missense somatic mutations in *CDKN2A* reported in COSMIC, TCGA, JHU, MSK-IMPACT.

Appendix 1-table 7. Sequences of primers used in study.

Source data

Figure 1-source data

Raw data in Figure 1

Figure 1-figure supplement 1-source data 1

Raw data in Figure 1-figure supplement 1A

Figure 1-figure supplement 1-source data 2

Raw data in Figure 1-figure supplement 1B

Figure 2-source data 1

Raw data in Figure 2A

Figure 2-source data 2

Raw data in Figure 2B and C

Figure 2-source data 3

Raw data in Figure 2D

Figure 2-figure supplement 1-source data 1

Raw data in Figure 2-figure supplement 1B

Figure 2- figure supplement 1-source data 2

Raw data in Figure 2- figure supplement 1C

Figure 3-source data 1

Raw data in Figure 3

Figure 4-source data 1

Raw data in Figure 4

Figure 4-figure supplement 1-source data 1

Raw data in Figure 4-figure supplement 1

Figure 4-figure supplement 2-source data 1

Raw data in Figure 4-figure supplement 2

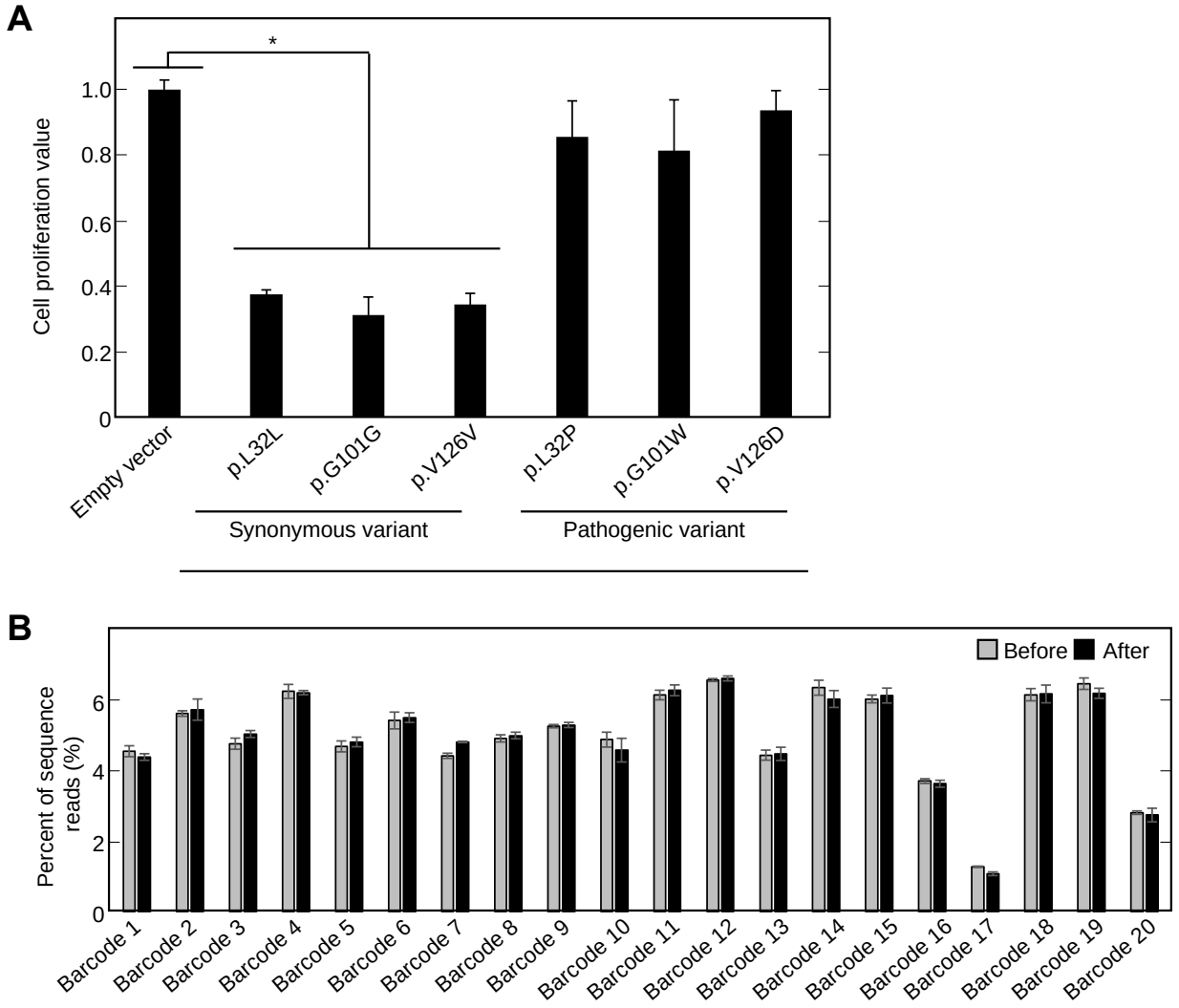

Figure 1-figure supplement 1. **Development and validation of high-throughput CDKN2A functional assay.** (A) Cell proliferation of PANC-1 cells stably expressing empty expression vector, one of three synonymous variants (p.L32L, p.G101G, p.V126V), or one of three pathogenic variants (p.L32P, p.G101W, p.V126D) over 14 days in culture. Cell proliferation values are given as mean of three repeats ± standard deviation normalized to PANC-1 cells that stably express empty vector. Statistically significant inhibition of cell proliferation inhibition in PANC-1 cells that stably express synonymous variants *; Student's t test, P value < 0.001). (B) PANC-1 cells stably express codon optimized CDKN2A transduced with a CellTag lentiviral library of 20 nonfunctional barcodes were cultured and representation (percent of reads supporting each barcode) before and after a period of in vitro cell proliferation was measured determined by next generations sequencing. Percent values are given as the mean of three repeats ± standard deviation.

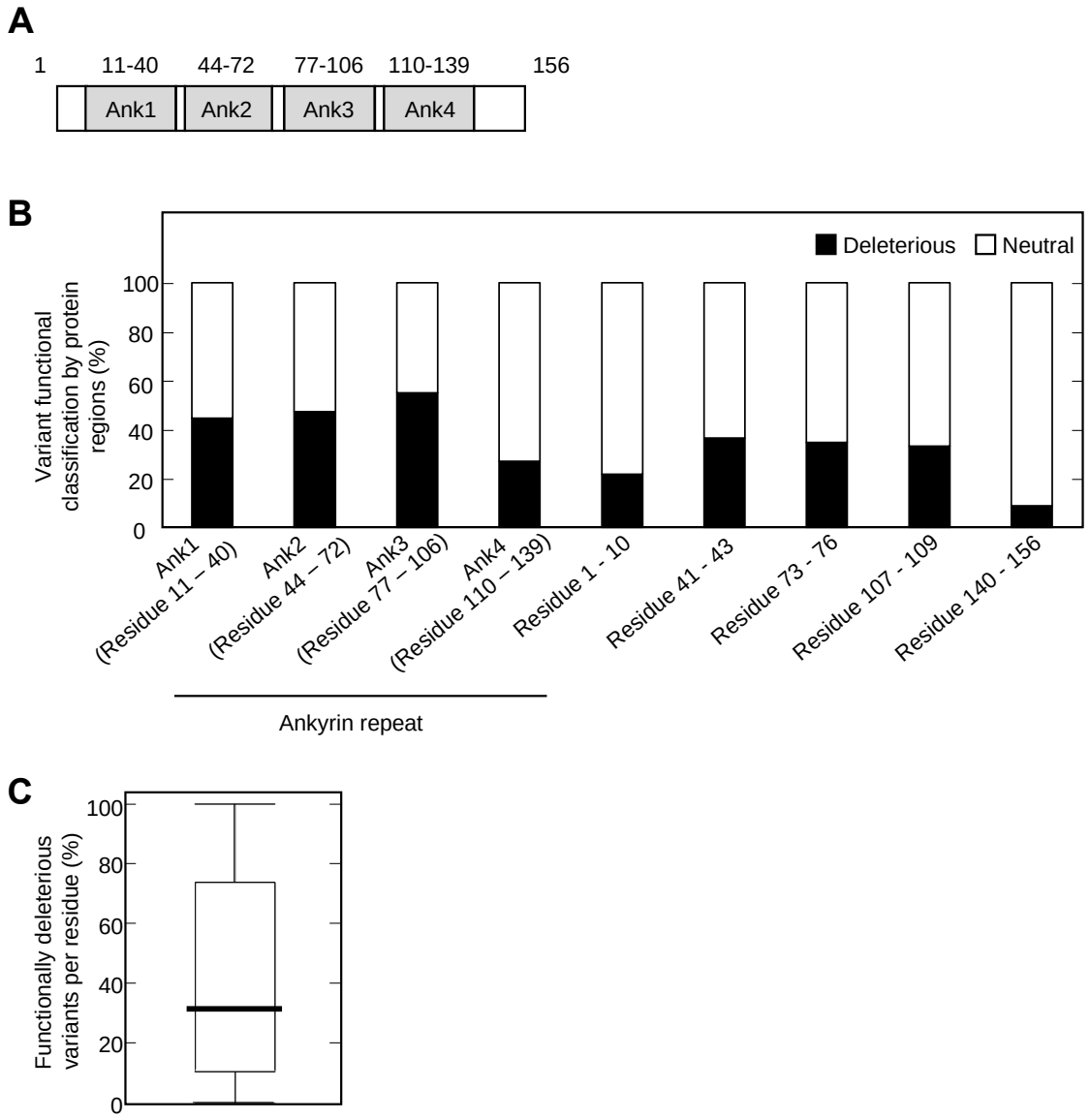

Figure 2-figure supplement 1. Functional characterization of all possible *CDKN2A* missense variants. (**A**) Schematic representation of CDKN2A ankyrin repeats. (**B**) Percent of functionally deleterious (black box) and functionally neutral variants (white box) within ankyrin repeats and non-ankyrin repeat regions of CDKN2A. Ank; Ankyrin repeat. (**C**) Box plot showing distribution of percent functionally deleterious missense variants per residue.

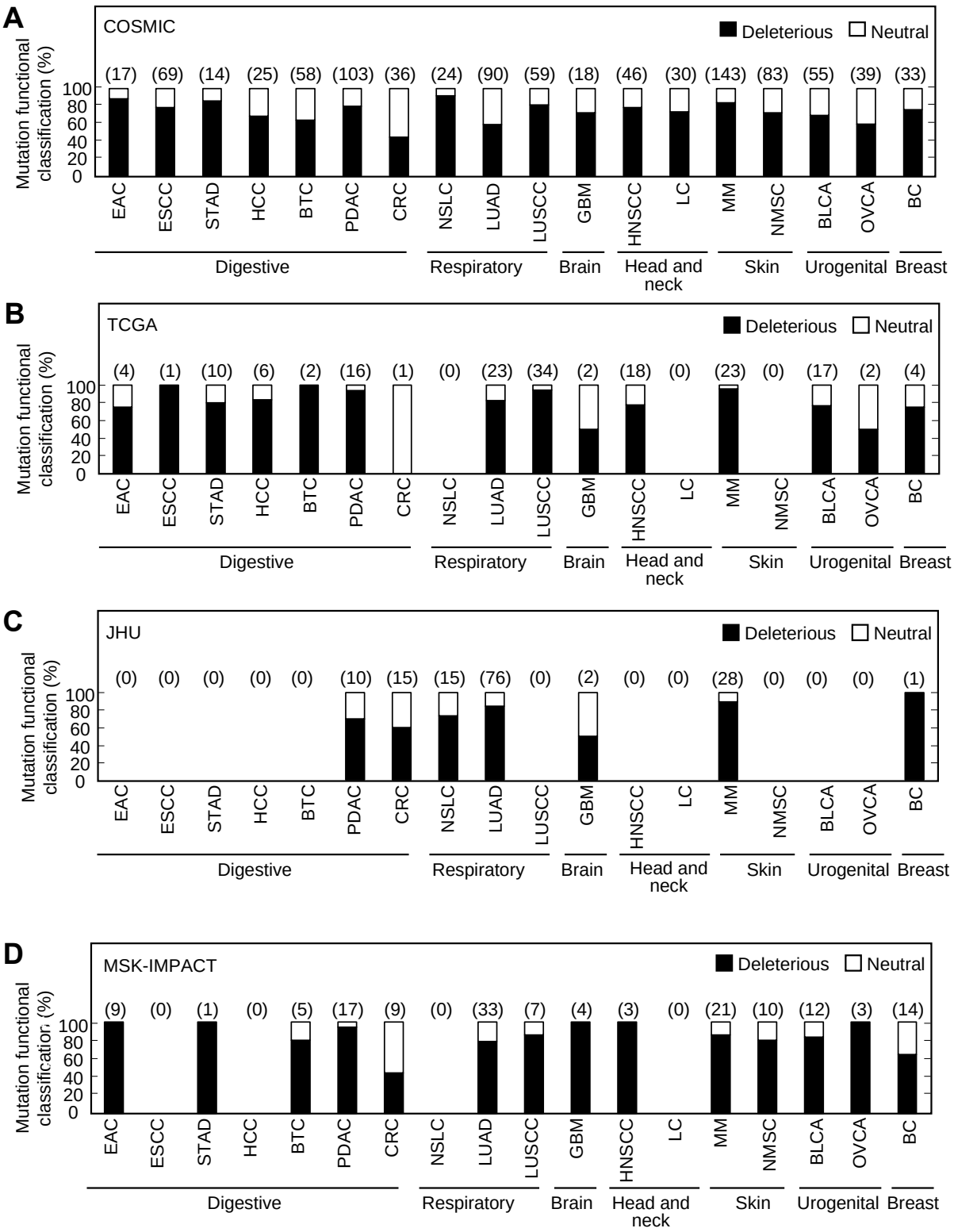
A
COSMIC
Deleterious
Neutral
Mutation functional classification (%)
(17) (69) (14) (25) (58) (103) (36) (24) (90) (59) (18) (46) (30) (143) (83) (55) (39) (33)
EAC ESCC STAD HCC BTC PDAC CRC NSLC LUAD LUSCC GBM HNSCC LC MM NMSC BLCA OVCA BC
Digestive
Respiratory
Brain
Head and neck
Skin
Urogenital
Breast
B
TCGA
Deleterious
Neutral
Mutation functional classification (%)
(4) (1) (10) (6) (2) (16) (1) (0) (23) (34) (2) (18) (0) (23) (0) (17) (2) (4)
EAC ESCC STAD HCC BTC PDAC CRC NSLC LUAD LUSCC GBM HNSCC LC MM NMSC BLCA OVCA BC
Digestive
Respiratory
Brain
Head and neck
Skin
Urogenital
Breast
C
JHU
Deleterious
Neutral
Mutation functional classification (%)
(0) (0) (0) (0) (0) (10) (15) (15) (76) (0) (2) (0) (0) (28) (0) (0) (0) (1)
EAC ESCC STAD HCC BTC PDAC CRC NSLC LUAD LUSCC GBM HNSCC LC MM NMSC BLCA OVCA BC
Digestive
Respiratory
Brain
Head and neck
Skin
Urogenital
Breast
D
MSK-IMPACT
Deleterious
Neutral
Mutation functional classification (%)
(9) (0) (1) (0) (5) (17) (9) (0) (33) (7) (4) (3) (0) (21) (10) (12) (3) (14)
EAC ESCC STAD HCC BTC PDAC CRC NSLC LUAD LUSCC GBM HNSCC LC MM NMSC BLCA OVCA BC
Digestive
Respiratory
Brain
Head and neck
Skin
Urogenital
Breast

Figure 4-figure supplement 1. Functional classification of missense somatic mutations in *CDKN2A*. (**A - D**) Percent of missense somatic mutations in CDKN2A reported in COSMIC (**A**), TCGA (**B**), JHU (**C**), or MSK-IMPACT (**D**) that were classified as functionally deleterious (black box) or functionally neutral (white box) group by tumor type. Cancer types with 10 or more missense somatic mutations in COSMIC are presented. The number of missense somatic mutations for each tumor type given in parentheses. COSMIC; the Catalogue Of Somatic Mutations In Cancer, TCGA; The Cancer Genome Atlas, JHU; The Johns Hopkins University School of Medicine, MSK-IMPACT; Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets.

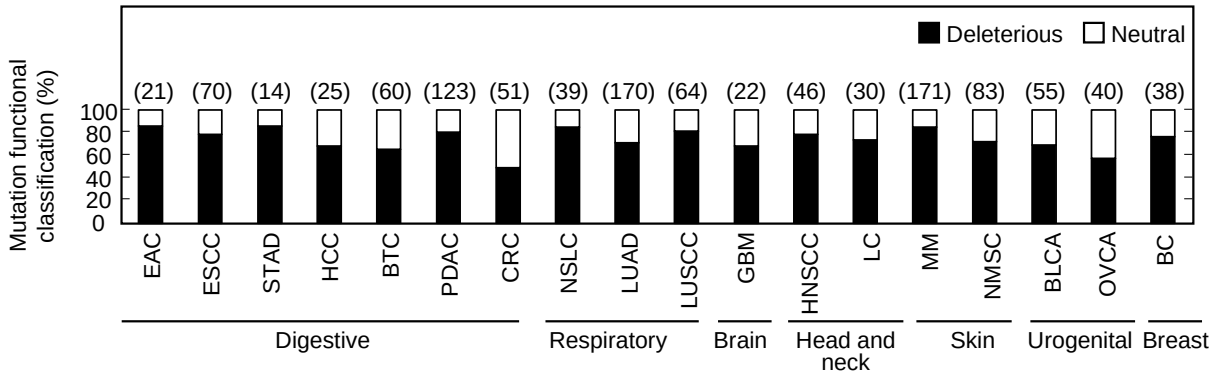

Figure 4-figure supplement 2. Aggregate functional classifications for missense somatic mutations in *CDKN2A*. Percent of missense somatic mutations in *CDKN2A* that were classified as functionally deleterious (black box) or functionally neutral (white box) group by tumor type. Missense somatic mutations reported in COSMIC, TCGA, JHU, and MSK-IMPACT were combined. Cancer types with 10 or more missense somatic mutations in COSMIC are presented. The number of missense somatic mutations for each tumor type given in parentheses. COSMIC; the Catalogue Of Somatic Mutations In Cancer, TCGA; The Cancer Genome Atlas, JHU; The Johns Hopkins University School of Medicine, MSK-IMPACT; Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets.

Appendix 1-table 1. Codon optimized *CDKN2A* sequence.

Sequence
ATGGAACCCGCCGCTGGCTCATCAATGGAACCCTCTGCAGATTGGCTCGCTACCGCA GCCGCGCGAGGACGAGTGGAAGAAGTCCGAGCGCTCCTTGAAGCCGGTGCATTGCC AAATGCGCCTAACTCCTATGGACGCCGCCCAATACAAGTTATGATGATGGGATCTGC AAGGGTAGCTGAACTTCTCCTCTTGCATGGAGCAGAACCTAATTGTGCTGATCCAGC GACCCTGACTAGACCTGTACATGATGCCGCGCGTGAAGGGTTTCTCGATACCCTTGT CGTCCTTCATCGAGCTGGTGCCCGCCTCGATGTCCGGGACGCATGGGGACGGCTCCC AGTCGATCTCGCAGAAGAACTTGGGCACAGGGACGTAGCCAGATATTTGCGTGCAG CCGCCGGTGGTACGAGGGGATCAAATCACGCTAGAATCGACGCTGCCGAGGGCCCA AGCGACATACCAGACTAA

Appendix 1-table 5. Assessment of in silico variant effect prediction models.

In silico model	Accuracy (%)	Sensitivity	Specificity	Positive predictive value	Negative predictive value
CADD	56.8	0.86	0.39	0.47	0.81
Polyphen-2	54.6	0.90	0.31	0.46	0.82
SIFT	61.4	0.64	0.60	0.51	0.71
VEST	70.8	0.67	0.74	0.63	0.77
AlphaMissense	70.9	0.69	0.72	0.62	0.78
ESM1b	64.9	0.77	0.57	0.54	0.79
PrimateAI-3D	66.7	0.15	0.99	0.93	0.65