Supplementary Experimental Methods

Biopsy protocol, patient consent and tissue acquisition:

Approval for the PancSeq protocol was obtained from the Dana-Farber/Harvard Cancer Center (DF/HCC) IRB (DF/HCC #14-408, #03-189) to obtain tumor tissue for genomic analyses in two situations: 1) Additional tissue obtained at the time of a biopsy performed for clinical purposes, and 2) Tissue obtained during a biopsy performed specifically for genomic evaluation, without an established clinical indication. Patients without prior histologic diagnosis of pancreatic cancer were eligible to enroll if PDAC was suspected based on clinical presentation and imaging studies. Patients must have met the following criteria: 1) ECOG performance status of 0-2, 2) ability to safely halt anticoagulation for the biopsy procedure, and 3) no concurrent chemotherapy treatment. A whole blood sample was obtained for paired germline DNA sequencing.

A median of 5 (range, 1-10) core needle biopsy specimens were collected per patient (Supplementary Table S2), with separate cores distributed by the following priority: core 1 for formalin fixed paraffin embedded (FFPE) histopathology, core 2 for whole exome sequencing (WES), core 3 for RNA-sequencing (RNA-seq), and any additional cores for patient-derived models or banking as fresh frozen tissue for future research. For clinically indicated biopsies, all specimens were held in a CLIA-certified facility until the diagnosis was confirmed, and then released for genomic analysis.

For comparison with advanced PDAC molecular information, genomic data from primary resection specimens have been shown at certain points in the manuscript. Primary surgical resection specimens were derived from patients treated at Dana-Farber Cancer Institute and the Brigham and Women's Hospital that were consented to IRB-approved protocols (DF/HCC #03-189, #11-104).

Whole exome sequencing (WES)

WES was performed in a laboratory at the Broad Institute that has been certified by the Clinical Laboratory Improvement Amendments (CLIA, #22D2055652). Starting with 250 ng input DNA, samples are quantified using a PicoGreen assay and diluted to a working stock volume and concentration (2 ng/uL in 50 uL), then libraries are constructed

and sequenced on Illumina HiSeq instruments with the use of 76-bp paired-end reads as previously described (1). Output from Illumina software is processed by the Picard data-processing pipeline to yield BAM files containing well-calibrated, aligned reads. All process steps are performed using automated liquid handling instruments, and all sample information tracking is performed by automated LIMS messaging. All data have been made publicly available through the NCBI database of Genotypes and Phenotypes (dbGaP).

WES data-processing pipeline ("Picard pipeline"):

The "Picard" pipeline (http://picard.sourceforge.net/) generates a BAM file (http://samtools.sourceforge.net/SAM1.pdf) for each sample and was developed by the Sequencing Platform at the Broad Institute. Picard pipeline aggregates data from multiple libraries and flow cell runs into a single BAM file for a given sample. This file contains reads aligned to the human genome with quality scores recalibrated using the Table Recalibation tool from the Genome Analysis Toolkit. Reads were aligned to the Human Genome Reference Consortium build 38 (GRCh38) using BWA v0.5.9 (2) (http://bio-bwa.sourceforge.net/). Unaligned reads that passed the Illumina quality filter (PF reads) were also stored in the BAM file. Duplicate reads were marked such that only unique sequenced DNA fragments were used in subsequent analysis. Sequence reads corresponding to genomic regions that may harbor small insertions or deletions (indels) were jointly realigned to improve detection of indels and to decrease the number of false positive single nucleotide variations caused by misaligned reads, particularly at the 3' end. All sites potentially harboring small insertions or deletions in either the tumor or the matched normal were realigned in all samples. Finally, the Picard pipeline provided summary QC metrics such as the target coverage and an estimated level of "oxo-G" artifacts (3) for each BAM that were used in subsequent processing.

Cancer genome analysis pipeline ("Firehose")

The Firehose pipeline (http://www.broadinstitute.org/cancer/cga/Firehose) performed additional QC on the BAM files, mutation calling, small insertion and deletion identification, rearrangement detection, coverage calculations, annotation of detected

mutations, filtering for OxoG artifacts and filtering by “panel-of-normals” and by Exome Aggregation Consortium (ExAC) dataset. The pipeline is an extensive series of tools for analyzing massively parallel sequencing data for both tumor DNA samples and their patient-matched normal DNA samples. The pipeline contains the following steps:

1. Quality control on BAM files: The sample cross-individual contamination levels were estimated using the ContEst program (4).
2. Somatic point mutation calling: The MuTect algorithm (5) was used to detect somatic single nucleotide variants (SNVs). SNVs were detected using a statistical analysis of the bases and qualities in the tumor and normal BAMs.
3. Small insertion and deletion detection: The Strelka algorithm (https://www.ncbi.nlm.nih.gov/pubmed/22581179) was used to detect small insertions and deletions (InDels).
4. SNVs and InDel annotations: SNVs and InDels detected by MuTect and Indelocator, respectively, were annotated using Oncotator (6). Oncotator mapped somatic mutations to respective genes, transcripts, and other relevant features. These annotations correspond to the fields in the Mutation Annotation Format (MAF) file provided for this manuscript.
5. Filtering for OxoG artifacts: G>T/C>A transversions that are a consequence of heating, shearing, and oxidative damage to the DNA during genomic library preparation (3). Eleven variants were filtered out of the call set for sample ID 0400100-T1.
6. Filtering by “panel-of-normals”: The sites of detected SNVs and InDels were examined against a panel of normal samples (PoN) as previously described (1). For a given SNV or InDel, a likelihood score that the allele counts are consistent with expectation of observed normals at the site is calculated. Candidate mutations with a likelihood score less than -2.5 were removed from subsequent analysis. We also removed variants outside coding regions.
7. Filtering by ExAC: Germline mutation calls from the ExAC database (http://exac.broadinstitute.org/) were used to screen for germline calls where coverage in normal was low. Only germline variants with population frequency of <1% were chosen to go through following steps in the interpretation pipeline.

Mutation Annotation Format (MAF) file

The MAF file was generated and has been made available as supplemental data along with the analyses contained within this manuscript.

Mutation Significance Analysis

Genes with a significant excess of the number of non-synonymous mutations relative to the estimated density of background mutations were identified using the MutSig algorithm (7, 8). MutSig has been previous used to identify significantly mutated genes (SMGs) in several tumor sequencing projects (9-12) and the algorithm's current version MutSig2CV (7) was used in this study to produce a robust list of significantly mutated genes. MutSig takes into account the background mutation rates of different mutation categories (*i.e.* transitions or transversions in different sequence contexts, the non-synonymous to synonymous mutation ratio for each gene, as well as the fact that different samples have different background mutation rates. It then uses convolutions of binomial distributions to calculate the p value for each gene, which represents the probability that we observe a certain configuration of mutations in a gene by chance, given the background model. Finally, it corrects for multiple hypotheses by calculating a q-value (False Discovery Rate) for each gene using the Benjamini & Hochberg procedure to produce the list of SMGs.

Copy number analysis

For copy number analysis based on exome sequencing, segmented copy data was obtained using copy number ratios. These were calculated as the ratio of tumor read depth to the average read depth observed in a panel of normal samples using the tool, RECAPSEG5. Allelic copy number analysis was done with Allelic-Capseg using B-allele frequencies from heterozygous sites. ABSOLUTE (13) was used to determine purity, ploidy, and whole genome doubling status using allelic copy number data along with the allelic fraction of all somatic mutations as input. In silico admixture removal (ISAR) was used to perform purity and ploidy correction of the RECAPSEG data. We used ABSOLUTE derived copy number from WES to identify genes with loss of heterozygosity

and homozygous deletions. High level amplifications were defined as those genes with three or more copies above baseline ploidy.

Significance of copy number alterations were assessed from the segmented data using GISTIC2.0 (Version 2.0.22) (14). Briefly, GISTIC2.0 deconstructs somatic copy-number alterations into broad and focal events and applies a probabilistic framework to identify location and significance levels of somatic copy-number alterations. For the purpose of this analysis, we defined an arm-level event as any event spanning more than 50% of a chromosome arm.

Clinical Interpretation and Reporting

Annotated whole exome sequencing data were cross-referenced with a curated list of 81 PDAC-relevant genes. These genes were chosen based on somatic or germline clinical relevance, therapeutic actionability and/or recurrent mutation across published PDAC genome sequencing cohorts. Most of these genes (n=69) were associated with clinical trial or off-label FDA-approved targeted therapies (Supplemental Table S3). Both somatic and germline variants in these genes were interpreted by an ABMG board certified clinical geneticist (AAG) through a previously published clinical interpretation pipeline (1, 15). The interpretation pipeline included the following steps: variants assessment, variant description and classification, variant return decision, and interpretation and report.

- Variants assessment – All variants in the 81 genes present in Supplemental Table S3 were assessed to exclude potential technical artifacts. The variants nomenclature was reviewed and rewritten if needed to standardize the transcript and amino acid designation for any given mutation. Germline variants in any of the genes in Table S3 were cross-referenced with the ExAC database to inquire the population frequency. Only germline variants with population frequency of <1% were chosen to go through following steps in the interpretation pipeline.
- Variant description and classification – Variants were fully delineated for the type, location, and putative functional consequence of the mutation, as well as role of the gene that harbors the mutation in pancreatic cancer. Based on the evidence to support

pathogenicity, variants were subsequently classified as pathogenic, likely pathogenic, VUS or benign according to the ACMG guidelines (16).

- Variant Return Decision – Variant return decision was made based on the functional relevance of the variant in a given gene. All variants in the categories of known functionally impaired, potentially functionally impaired, variant of unknown significance (VUS) were chosen to be included in the CLIA-certified report. Variants with known benign function were excluded in this step.
- Variant interpretations and Report – A report of clinically relevant events was returned to the referring clinician detailing somatic mutations, small insertions/deletions, and copy number alterations (CNAs) as well as pathogenic/likely pathogenic (P/LP) germline alterations. The report includes a first page summary that exhibits reported variants along with their specific relevant therapeutic agents or their biological classifications. The subsequent pages of the report provide full description of variants and genes, detail on classified therapies and biological function in pancreatic cancer.

Mutational signature analysis:

Mutational signature analysis was performed using a Bayesian variant of the non-negative matrix factorization (NMF) approach in a two-stage manner from the set of single nucleotide variants (SNV) in our dataset, as previously described (*SignatureAnalyzer*, (17-20)). First, we performed *de novo* signature discovery and our analytic pipeline identified three primary signatures: SigA best resembling COSMIC signature 3 with cosine similarity 0.87 (homologous recombination deficiency [HRD] or BRCA signature); SigB best resembling COSMIC signature 1 with cosine similarity 0.96 (C>T transitions at CpG dinucleotides, Aging); and SigC best resembling COSMIC signature 17 with cosine similarity 0.91 (etiology unknown) (Supplementary Figure S4A-B). In addition, we observed a relative elevation of C>G transversions and C>T transitions at TC[A/T] contexts in SigA corresponding to canonical hotspots of APOBEC mutagenesis (COSMIC signature 2 and 3), suggesting that an APOBEC-mediated mutational process is possibly operative in this cohort, but not cleanly separable due to an insufficient number of mutations (17). We identified four main mutational processes that were most active in the WES data (Aging, BRCA/HRD, APOBEC, COSMIC17), and performed a projection

analysis of these contributing signatures across the cohort to better evaluate discrete contributions of these mutational signature processes to the data.

RNA sequencing and data analysis

RNA-sequencing was performed at the Broad Institute. All analysis was performed on log-transformed RNA-seq gene expression data. Clustering analysis was performed in R, using the ConsensusClusterPlus package, and t-SNE was performed using the Rtsne package. Samples were classified into groups based on mRNA expression based on the results in Moffitt et al. (21). Gene expression signatures were obtained from Moffitt et al. (21), including the top 25 genes from basal-like, classical, activated stroma, normal stroma, liver, and exocrine gene sets. A composite tumor score was created by combining the 25 basal-like and classical genes. Similarly, a composite stromal score was created by combining the 25 activated stroma and normal stroma genes. Signature scores were defined as the mean expression of log-transformed gene expression for genes within the respective set. A subset of biopsy samples exhibited low composite tumor score, and were further investigated manually for possible explanations, resulting in specific "annotations." For example, the 7 samples defined as "High Liver" were found to have extremely high liver expression scores whereas the 2 samples defined as "High stroma" had correspondingly high composite stroma scores.

Analysis of cell-free DNA (cfDNA)

Plasma was collected in EDTA-containing tubes from patients for cfDNA analysis at baseline and throughout their treatment course. cfDNA was extracted from plasma and droplet digital PCR (ddPCR) was performed for target sequences using previously described methods (22). The ddPCR data were analyzed with QuantaSoft analysis software (Bio-Rad) to obtain Fractional Abundance of the mutant DNA alleles in the wild-type/normal background. The quantification of the target molecule was presented as the number of total copies (mutant plus wild-type) per sample in each reaction. Mutant Allele Fraction is calculated as follows: Mutant Allele Fraction (%) = (Nmut/(Nmut+Nwt))*100), where Nmut is the number of mutant events and Nwt is the number of wildtype events

per reaction. ddPCR analysis of normal control plasma DNA and no DNA template controls were always included. Probe and primer sequences are available upon request.

Evaluation of a more comprehensive panel of genes within cfDNA was pursued using the Guardant360 cfDNA assay (Guardant Health). The Guardant360 assay is a Clinical Laboratory Improvement Amendments (CLIA)–certified targeted digital sequencing panel designed to detect SNVs, as well as selected insertions/deletions, amplifications, and fusions (23, 24).

References

1. Cancer Genome Atlas Research Network aadhe. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer cell. 2017;32:185-203 e13.
2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589-95.
3. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013;41:e67.
4. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011;27:2601-2.
5. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213-9.
6. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat. 2015;36:E2423-9.
7. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505:495-501.

8. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214-8.
9. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, et al. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature. 2012;485:502-6.
10. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nature genetics. 2013;45:478-86.
11. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2012;109:3879-84.
12. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011;333:1157-60.
13. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012;30:413-21.
14. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome biology. 2011;12:R41.
15. Ghazani AA, Oliver NM, St Pierre JP, Garofalo A, Rainville IR, Hiller E, et al. Assigning clinical meaning to somatic and germ-line whole-exome sequencing data in a prospective cancer precision medicine study. Genet Med. 2017;19:787-95.
16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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;17:405-24.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415-21.

18. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nature genetics. 2017;49:1476-86.
19. Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nature communications. 2015;6:8866.
20. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nature genetics. 2016;48:600-6.
21. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nature genetics. 2015;47:1168-78.
22. Goyal L, Saha SK, Liu LY, Siravegna G, Leshchiner I, Ahronian LG, et al. Polyclonal Secondary FGFR2 Mutations Drive Acquired Resistance to FGFR Inhibition in Patients with FGFR2 Fusion-Positive Cholangiocarcinoma. Cancer discovery. 2017;7:252-63.
23. Zill OA, Greene C, Sebisanovic D, Siew LM, Leng J, Vu M, et al. Cell-Free DNA Next-Generation Sequencing in Pancreatobiliary Carcinomas. Cancer discovery. 2015;5:1040-8.
24. Strickler JH, Loree JM, Ahronian LG, Parikh AR, Niedzwiecki D, Pereira AAL, et al. Genomic Landscape of Cell-Free DNA in Patients with Colorectal Cancer. Cancer discovery. 2018;8:164-73.