

Supporting Information

Mann et al. 10.1073/pnas.1202490109

SI Materials and Methods

Histology and Immunohistochemistry. For the tissue microarray, immunohistochemistry was visualized using the Universal LSAB+ Kit (Dako) after deparaffinization, heat-induced antigen retrieval with vegetable steamer for 20 min in citrate buffer (pH 6.0), and 1 h of incubation with primary antibodies against δ -1-catenin (1:100, ab92514; Abcam) and GNAQ (1:800, ab75825; Abcam). Three separate 1.0-mm cores for each tumor in the tissue microarray were independently scored by a blinded observer using semi-quantitative histoscores (range = 0–300) representing the product of staining intensity (0–3) and percentage of tumor cells staining at that intensity (0–100). Membrane staining was scored for δ -1-catenin, whereas cytoplasmic staining was scored for GNAQ. For dichotomization, each tumor was assigned into a low- or high-level staining group based on its average histoscore. Survival estimates were generated using the Kaplan–Meier method and compared using log-rank tests (1). Overall survival time was measured from the date of surgery to the date of death because of any cause or last follow-up.

Mouse tissues were fixed in 10% formalin overnight and embedded in paraffin. For immunohistochemistry of Pdx1 and cytokeratin-19, 5- μ m sections were deparaffinized in xylene and rehydrated in descending percentages of ethanol to water. Endogenous peroxidase was inhibited using 1% H₂O₂ in methanol. The sections were then washed in distilled water and heated in a pressure cooker for 45 min in pH9 for epitope retrieval followed by 10% goat serum block for 40 min. The slides were then incubated for 1 h with monoclonal antibodies Troma III (Ck19; dilution 1:10) or anti-Pdx1 (clone F109-D12; dilution 1:100) followed by incubation with HRP polymer anti-rat Envision System from Dakocytomation for 30 min. Envision 3, 3' Diaminobenzidine (DAB) substrate was used to visualize, and counterstaining was performed with hematoxylin. Primary antibodies were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the National Institute of Child Health and Human Development and maintained by the Department of Biology, University of Iowa, Iowa City, IA.

Immunohistochemistry for smooth muscle actin was performed using the Leica Bondmax[®] autostainer; 5- μ m formalin-fixed paraffin sections were deparaffinized in BOND Dewax solution and rehydrated in descending percentages of ethanol to water. The slides were then heat-exposed with BOND Retrieval solution 2 (pH 9), and endogenous peroxidase was inhibited using 3% H₂O₂ in methanol for 15 min. The slides were incubated for 45 min with monoclonal mouse anti-human actin (smooth muscle actin, clone 1A4, dilution 1:200, M0851; Dako) and then incubated with HRP polymer anti-mouse for 10 min. DAB substrate was applied for 3 min to visualize, and counterstaining was performed with hematoxylin. Masson's Trichome Staining was performed as previously described (2).

Gaussian Kernel Convolution Method for Common Insertion Site Determination. Any insertion sites on the transposon donor chromosomes, mouse chromosome 1 for tumors derived from T2/Onc2, and mouse chromosome 9 for tumors derived from T2/Onc3 were excluded from common insertion site (CIS) analysis. The likelihood of local hopping of the transposon is increased where the transposon array is located (3). This phenomenon can significantly increase the background level of transposon insertion events, thereby complicating CIS analysis; 19,927 nonredundant insertion sites were used to identify CISs using a Gaussian kernel convolution (GKC) statistical framework (4). The previous GKC

analysis approach (5) was enhanced by using multiple kernel scales (widths of 30-, 50-, 75-, 120-, and 240-K nt). CISs predicted across multiple scales and overlapping in their genomic locations were clustered together such that the CIS with the smallest genomic footprint was reported as the representative CIS. For highly significant CISs with narrow spatial distributions of insertion sites, the 15-K kernel is typically the scale on which CISs are identified. The *P* value for each CIS was adjusted by chromosome, and a cutoff of *P* < 0.05 was used.

Gene centric CIS Computational Method. Gene centric CISs (gCISs) were analyzed using the methods published in the work by Brett et al. (6) with slight modifications. We only considered transposon insertions at uniquely mappable TA dinucleotides within the coding regions of all RefSeq genes. A mappable TA is defined by the presence of a uniquely mapped 40-bp junction on either side of the TA in the mouse genome.

Pathway Analysis Using Ingenuity Pathway Analysis and DAVID. GKC candidate cancer genes (CCGs; 136) and gene centric CCGs (gCCGs; 653) from *Sleeping Beauty* (SB)-driven pancreatic tumors and the sensitizing oncogene *Kras* were considered for placement in canonical signaling pathways and cellular functions (functional analysis) using ingenuity pathway analysis (7); 132 GKC CCGs and 645 gCCGs were associated with biological functions and/or diseases in the ingenuity knowledge base and were considered for the analysis. Gene ontology terms and signaling pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (8, 9) were queried using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (10, 11); 135 GKC CCGs and 653 gCCGs were associated with biological functions and/or diseases in this database and were considered for the analysis. Right-tailed Fisher exact test was used to calculate a *P* value determining the probability that each biological function and/or disease assigned to that dataset was caused by chance alone. *P* values for pathway enrichment were adjusted for multiple testing using the Benjamini–Hochberg method for control of the false discovery rate (12).

Statistical Analysis. Enrichment of CISs in human cancer mutation datasets. The hypergeometric distribution was used to test for overrepresentation of the SB CCG and SB gCIS in the Cancer Genome Consensus database with somatic mutations identified in the works by Jones et al. (13) and Campbell et al. (14). A *P* value < 0.05 was considered to represent a significant association.

Gene expression and patient survival analysis. Publicly available pancreatic cancer microarray data were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus database (GSE 21501) (15). These data comprised 132 samples from primary pancreatic ductal adenocarcinomas; overall survival data were available for 102 patients (16). Gene expression data relating to human homologs of SB pancreatic candidate cancer genes were extracted by matching on gene symbol. We found 124 of the GKC CCGs to be represented.

For each gene, Cox proportional hazards regression was used to examine the association between gene expression level and overall survival. *P* values were corrected for multiple testing to control the false discovery rate, and Cox regression *P* values from all genes were used to estimate the proportion of genes for which the null hypothesis was true (12). Genes were considered to be significantly associated with overall survival if at least one probe for

that gene had an adjusted P value of <0.05 . To determine whether the list of concordant GKC CCGs was significantly enriched for genes associated with survival, a resampling approach was taken. Random gene lists of 124 genes were generated and analyzed using the Cox regression method described above. This procedure was repeated 1,000 times, and for each random gene set, the number of genes with at least one probe significantly associated

with overall survival was recorded. These values were then compared with the values that were observed for the gene lists derived from the GKC CCGs, with P values calculated based on how many times the random gene lists showed equal or greater numbers of genes associated with survival. A resampling-based P value of <0.05 was considered to provide evidence of significant enrichment for survival-associated genes.

- Manuyakorn A, et al. (2010) Cellular histone modification patterns predict prognosis and treatment response in resectable pancreatic adenocarcinoma: Results from RTOG 9704. *J Clin Oncol* 28:1358–1365.
- Prophet EB (1992) *AFIP Laboratory Methods in Histotechnology* (American Registry of Pathology, Washington, DC), 1st Ed, pp 132–133.
- Morrow M, Samanta A, Kiuoussis D, Brady HJ, Williams O (2007) TEL-AML1 preleukemic activity requires the DNA binding domain of AML1 and the dimerization and corepressor binding domains of TEL. *Oncogene* 26:4404–4414.
- de Ridder J, Uren A, Kool J, Reinders M, Wessels L (2006) Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput Biol* 2:e166.
- March HN, et al. (2011) Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat Genet* 43:1202–1209.
- Brett BT, et al. (2011) Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS One* 6:e24668.
- IPA Ingenuity Systems. Available at <http://www.ingenuity.com>. Accessed August 17, 2011.
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D355–D360.
- Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37:1–13.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300.
- Jones S, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321:1801–1806.
- Campbell PJ, et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467:1109–1113.
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210.
- Stratford JK, et al. (2010) A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS Med* 7:e1000307.

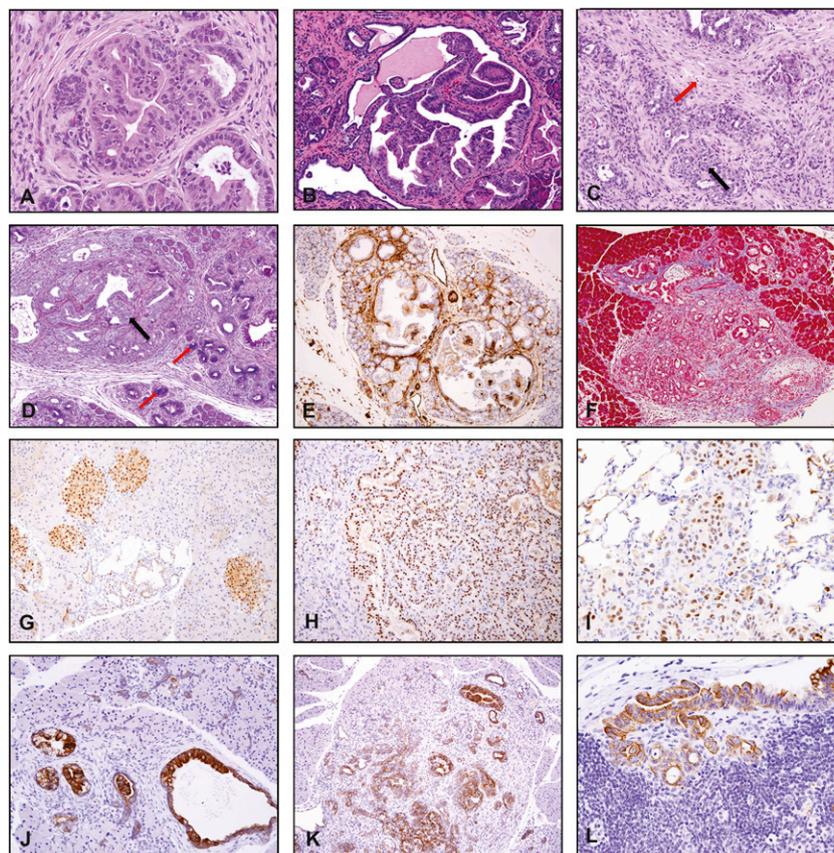


Fig. S1. Histological and immunohistochemical analysis of oncogenic $Kras^{G12D}$ -induced pancreatic tumors. Preinvasive early adenocarcinomas representing several different subtypes were observed in oncogenic $Kras^{G12D}$ -induced pancreatic tumors. Early adenocarcinomas with clear ductal morphology (A) were observed in several $Kras^{G12D}$ -induced pancreatic tumors. In some cases, early adenocarcinomas with papillary characteristics (B) or early adenocarcinoma (black arrow; C) with a prominent stromal component (red arrow; C) diagnosed as scirrhous adenocarcinoma were also observed. Staining sections for mucin using Alcian blue (D) showed high mucin content in early progressive neoplastic lesions (red arrows) that was reduced or absent in early adenocarcinomas (black arrow). Many adenocarcinomas had a prominent stromal component containing smooth muscle, which was shown by staining for smooth muscle actin (brown staining; E) and collagen as shown by Masson's trichrome (blue staining; F). Primary adenocarcinomas and metastatic lesions were characterized for the expression of Pdx1 (G–I) and cytokeratin-19 (J–L). G shows expression of Pdx1 in islets of the normal adult pancreas. Pdx1 is reexpressed in pancreatic adenocarcinomas (H) and lung metastases (I). *SB*-driven pancreatic adenocarcinomas are of ductal origin as shown by positive staining for cytokeratin-19, a ductal marker (J) in primary adenocarcinomas (K) and lymph node metastases (L).

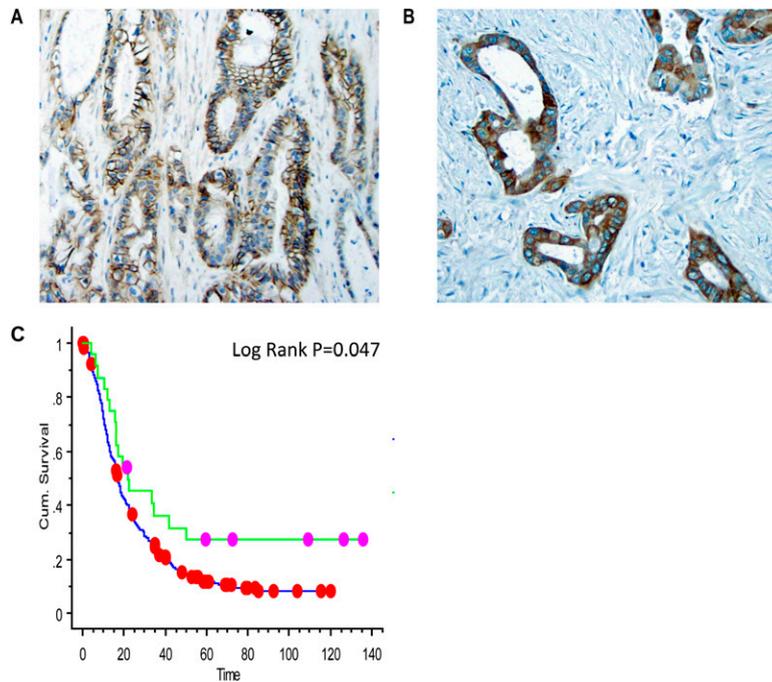


Fig. S3. CTNND1 and GNAQ staining in pancreatic adenocarcinoma with intact ductal morphology. Immunohistochemistry performed on a pancreatic adenocarcinoma tissue microarray shows that CTNND1 stains positive in regions of tumor with intact ductal membranes (A). Cytoplasmic staining of GNAQ is also observed in areas of tumor with intact ductal morphology (B). Kaplan–Meier plot of patient survival from a second independent 275-patient pancreatic cancer tissue microarray stained for CTNND1 (C). Two patient groups dichotomized into the top two tertiles (green) vs. bottom tertile (blue) using the histoscore for high and low staining.

Dataset S1. CISs identified by the GKC method

[Dataset S1 \(XLSX\)](#)

One hundred thirty-three CIS loci were identified from twenty-one *SB*-driven pancreatic tumors using the GKC statistical framework to analyze uniquely mapped *SB* transposon integrations sites (*SI Materials and Methods*). The GKC output merges the data from five kernel widths (30, 50, 75, 120, and 240 K) and reports the smallest kernel width that contains the CIS region. The gene name is listed for each CIS along with the chromosome and nucleotide position of the CIS peak. The calculated raw *P* value for each CIS is shown alongside the corrected *P* value for multiple testing on a chromosome basis. Note that, for many loci, the CIS width is smaller than the kernel width owing to the density and clustering of transposon integrations. The predicted mutagenic consequence of the insertions (i.e., activation of an oncogene or disruption of a tumor suppressor gene) is also listed. The number of tumors contributing to CISs on chromosome 9 is derived only from tumors with T2Onc2 for which the donor site is located on chromosome 1.

Dataset S2. Pancreatic CCGs mutated in human cancers identified by exon resequencing or whole-genome sequencing

[Dataset S2 \(XLSX\)](#)

Human homologs of the CCGs identified in *SB*-driven pancreatic tumors that are mutated in human cancer as determined by exon resequencing or genome sequencing are shown on a heat map with mutated genes denoted by blue rectangles. We considered all somatic mutations present in at least one sample from each of 20 reports from either exon resequencing or whole-genome sequencing (1–15). The different sequencing studies are listed at the top of the table; the mutated genes are listed on the left-hand column. We also queried the Cancer Gene Census (16) and Memorial Sloan Kettering Cancer Center database (17) for genes that overlap in our CCGs dataset (first two columns).

- Jones S, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321:1801–1806.
- Kan Z, et al. (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466:869–873.
- Sjöblom T, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314:268–274.
- Wood LD, et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318:1108–1113.
- Ding L, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455:1069–1075.
- Dalgliesh GL, et al. (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 463:360–363.
- Varela I, et al. (2011) Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 469:539–542.
- Parsons DW, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321:1807–1812.
- Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455:1061–1068.
- Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474:609–615.
- Campbell PJ, et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467:1109–1113.
- Shah SP, et al. (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 461:809–813.
- Parsons DW, et al. (2011) The genetic landscape of the childhood cancer medulloblastoma. *Science* 331:435–439.

14. Mardis ER, et al. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 361:1058–1066.
15. Pleasance ED, et al. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463:191–196.
16. Futreal PA, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4:177–183.
17. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: A gene selection resource for cancer genome projects. *Nucleic Acids Res* 35:D721–D726.

Dataset S3. Pairwise co-occurrences of pancreatic GKC CCGs in tumors

[Dataset S3 \(XLS\)](#)

Twelve significant co-occurrence pairs of CCGs were identified in our dataset. Fisher exact test was used to determine the probability of co-occurrence. The raw (uncorrected) *P* value is shown.

Dataset S4. CISs identified by the gCIS computational method

[Dataset S4 \(XLSX\)](#)

Six hundred seventy-one gCIS loci were identified from twenty-one *SB*-driven pancreatic tumors using the gCIS statistical framework to analyze uniquely mapped *SB* transposon integrations sites (*SI Materials and Methods*). Intragenic gCISs are defined by calculating the probability of transposon insertion events within a gene given the number of available TA dinucleotides compared with the number of TA dinucleotides within the entire genome. The gene name is listed for each gCIS along with the chromosome, total number of intragenic insertions, and number of tumors containing insertions. The *P* value for each gCIS is based on χ^2 analysis. The *P* value threshold using the Bonferroni correction (0.05/21508 RefSeq genes) is 2.32×10^{-6} . The number of tumors contributing to CISs on chromosome 9 is derived only from tumors with T2Onc2 for which the donor site is located on chromosome 1.

Dataset S5. Overlap of pancreatic CCGs identified by the GKC and gCIS methods with human pancreatic cancer mutations

[Dataset S5 \(XLSX\)](#)

Pancreatic CCGs identified by the GKC and gCIS methods with human homologs containing somatic mutations either listed in the Cancer Gene Census database or identified from sequencing studies of human pancreatic cancer are plotted as a heat map. The GKC CCGs are in red. There is statistically significant enrichment of both sets of CCGs for all three comparisons [$P = 1.67E-12$, Cancer Gene Census; $P = 2.11E-09$ (1); $P = 1.80E-21$ (2)].

1. Jones S, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321:1801–1806.
2. Campbell PJ, et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467:1109–1113.

Dataset S6. Human pancreatic cancer coding mutations in *SB* CCGs

[Dataset S6 \(XLSX\)](#)

Nonsilent coding mutations for 11 homologs and 4 orthologs of *SB* CCGs were validated using targeted PCR and Ion Torrent sequencing of a collection of 60 pancreatic cancer patient DNAs with matched normal controls from the Australian Pancreatic Cancer Genome Initiative. The point mutations are listed by gene name and patient sample. Genes highlighted in red are orthologs of *SB* CCGs that encode proteins similar in function to the *SB* CCG.

Other Supporting Information Files

[SI Appendix \(DOCX\)](#)