# Genetic Alterations Associated With Progression From Pancreatic Intraepithelial Neoplasia to Invasive Pancreatic Tumor

STEPHEN J. MURPHY,[1] STEVEN N. HART,[2] JOEMA FELIPE LIMA,[1] BENJAMIN R. KIPP,[3] MITCHELL KLEBIG,[3] JENNIFER L. WINTERS,[3] CSILLA SZABO,[4] LIZHI ZHANG,[3] BRUCE W. ECKLOFF,[5] GLORIA M. PETERSEN,[2] STEVEN E. SCHERER,[6] RICHARD A. GIBBS,[6] ROBERT R. MCWILLIAMS,[7] GEORGE VASMATZIS,[1,§] and FERGUS J. COUCH[2,3,§]

[1]*Department of Molecular Medicine,* [2]*Department of Health Sciences Research,* [3]*Department of Laboratory Medicine and Pathology,* [5]*Medical Genomics Facility,* [7]*Department of Oncology, Mayo Clinic, Rochester, Minnesota;* [4]*Department of Biological Sciences, University of Delaware, Newark, Delaware;* [6]*Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas*

**BACKGROUND & AIMS:** Increasing grade of pancreatic intraepithelial neoplasia (PanIN) has been associated with progression to pancreatic ductal adenocarcinoma (PDAC). However, the mechanisms that control progression from PanINs to PDAC are not well understood. We investigated the genetic alterations involved in this process. **METHODS:** Genomic DNA samples from laser-capture microdissected PDACs and adjacent PanIN2 and PanIN3 lesions from 10 patients with pancreatic cancer were analyzed by exome sequencing. **RESULTS:** Similar numbers of somatic mutations were identified in PanINs and tumors, but the mutational load varied greatly among cases. Ten of the 15 isolated PanINs shared more than 50% of somatic mutations with associated tumors. Mutations common to tumors and clonally related PanIN2 and PanIN3 lesions were identified as genes that could promote carcinogenesis. *KRAS* and *TP53* frequently were altered in PanINs and tumors, but few other recurrently modified genes were detected. Mutations in DNA damage response genes were prevalent in all samples. Genes that encode proteins involved in gap junctions, the actin cytoskeleton, the mitogen-activated protein kinase signaling pathway, axon guidance, and cell-cycle regulation were among the earliest targets of mutagenesis in PanINs that progressed to PDAC. **CONCLUSIONS: Early stage PanIN2 lesions appear to contain many of the somatic gene alterations required for PDAC development.**

*Keywords:* Pancreas; Tumorigenesis; LCM; Whole-Genome Amplification.

BASIC AND TRANSLATIONAL PANCREAS

Pancreatic cancer is the fourth leading cause of cancer-associated mortality, accounting for more than 35,000 deaths each year. Heterogeneous in form, 90% of pancreatic cancers are pancreatic ductal adenocarcinomas (PDACs) that present generally in the seventh decade of life.[1] Most cases have mild symptoms before diagnosis at late-stage, locally advanced, or metastatic disease.[2] Carcinoma of the exocrine pancreas is associated with a median survival of 6 months and a 5-year survival rate of 5%.[3] Approximately 15%–20% of patients present with resectable disease and only 15%–25% of surgically resected patients survive to 5 years.[4,5]

Despite the high incidence and poor survival associated with PDAC, few advances have been made in understanding the etiology and basic biology of pancreatic cancer and the mechanisms by which precursor lesions become early stage invasive tumors. Precursor lesions in the form of noninvasive pancreatic intraepithelial neoplasia (PanIN) are grouped into 3 grades according to increasing degree of cytologic and architectural atypia.[6,7] PanIN1 lesions are subdivided further into flat (PanIN1A) and papillary types (PanIN1B). Additional loss of polarity, nuclear crowding, cell enlargement, and hyperchromasia are present in PanIN2s. Advanced PanIN3 lesions have severe nuclear atypia, luminal necrosis, and manifest epithelial cell budding into the lumen of ducts.[7]

Evidence is strong for PanIN involvement in a cancer progression model. Although PanIN1 lesions frequently are observed in normal pancreatic autopsy tissues, PanIN2 lesions are more common in tissue derived from neoplastic pancreata. PanIN3 lesions rarely are observed in pancreatic tissues in the absence of cancer. In addition, the full spectrum of PanINs has been observed before tumorigenesis in mouse models of pancreatic cancer.[8] Tumor recurrence at surgical margins containing unresected PanIN3 lesions further supports this model.[9] Although it is thought that PDAC develops by stepwise progression through increasing grades of precursor lesions, the early genetic events that promote the development of PanINs and progression to PDAC are not well defined. Identification of these early driver genes and pathways is expected to lead to improved selection of therapeutic targets and may result in improved early diagnosis of early stage PDAC or advanced precursor lesions. Here, we report on exome sequencing of DNA

[§]Authors share co-senior authorship.

*Abbreviations used in this paper:* AAF, alternate allele frequency; ATM, ataxia telangiectasia mutated; LCM, laser capture microdissection; P, pancreatic intraepithelial neoplasia alone; PanIN, pancreatic intraepithelial neoplasia; PDAC, pancreatic ductal adenocarcinoma; PxT, tumors and adjacent pancreatic intraepithelial neoplasia; SNV, single-nucleotide variant; T, tumor; WGA, whole-genome amplification.


0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2013.07.049

isolated from pancreatic tumors and adjacent PanINs and on the identification of genes and pathways that contribute to PanIN development and progression to PDAC.

## Materials and Methods

### *Tissue Selection and Laser Capture Microdissection*

H&E-stained sections of pancreatic cancer tissues were reviewed by pathologists and 10 cases containing high-grade PanINs (P2 or P3) adjacent to tumor were selected. Ten frozen pancreatic tissue sections (10-$\mu$m) were cut and stained with Cresyl Violet (AM1935 LCM Staining Kit; Ambion, Foster City, CA). PanIN, tumor, and histologically normal regions were isolated individually by laser capture microdissection (LCM) using the Arcturus PixCell II microscope and CapSure Macro caps (LCM 0211; Arcturus, Carlsbad, CA).

### *Direct DNA Extraction and Amplification*

Whole-genome amplification was performed directly on LCM captured cells using a single-step procedure.[10] LCM cells were incubated for 10 minutes in 0.5× Repli-g D2 buffer (6.5 $\mu$L) (Qiagen, Valencia, CA) and then in Repli-g Stop Solution (3.5 $\mu$L). Cells then were mixed with Repli-g mini kit Master Mix (40 $\mu$L) and incubated at 30°C for 16 hours. Four individual 50-$\mu$L whole-genome amplification (WGA) reactions were pooled for each sample. DNA was quantified by Quant-iT PicoGreen analysis (P7581; Invitrogen, Eugene, OR) and qualitative multiplex polymerase chain reaction was performed (P0982; Sigma-Aldrich, St. Louis, MO).

### *Exome Sequencing*

DNA (3 $\mu$g) was fragmented to approximately 200 bp (Covaris E210, Woburn, MA) before assembly of adapter-flanked Illumina-indexed paired-end libraries (NEB Next DNA Kit) using Illumina adapters (Illumina, San Diego, CA). Exome capture was performed using the SureSelect Human All Exon 50-Mb kit (Agilent, Santa Clara, CA). Two indexed libraries were sequenced per lane on the Illumina HiSeq platform. A total of 100-bp paired-end reads were aligned to the human genome, hg19, using Novoalign (v2.07.13; Novocraft, Selangor, Malaysia).[11] Local sequence realignment was performed by GATK version 1.6-7[12] within the context of the Targeted RE-sequencing Annotation Tool framework.[13]

### *Somatic Single-Nucleotide Variants, Insertion, and Deletion Calling*

Each PanIN and tumor sample was compared with a corresponding normal sample using SomaticSniper[14] for somatic single-nucleotide variants (SNVs) or GATK's somatic insertion or deletion detector.[15] A minimum somatic score of 20 and more than 8 times coverage was required in the reference normal sample. Genotypes were re-coded to take advantage of the multiple samples from the same individuals. For somatic variants with less than 30 times read depth, 3 or more alternate reads supporting the variant call were required. For somatic variants sequenced fewer than 30 times, alternate alleles exceeding 4% of all reads were required. Functional significance of mutations was predicted using SNPEffect (SnpEFF; available: http://snpeff.sourceforge.net/), and sorting intolerant from tolerant (SIFT) and PolyPhen. Genes were categorized into pathways using the model-based gene set analysis (mgsa).[16] Analyses were performed using R-package version 1.2.0. (Vienna, Austria) with annotation from MSigDB.

## Results

### *Enrichment of Genomic DNA From Tumors and Their Adjacent PanINs*

Evaluation of frozen pancreatic tumor tissue by a study pathologist identified specimens containing

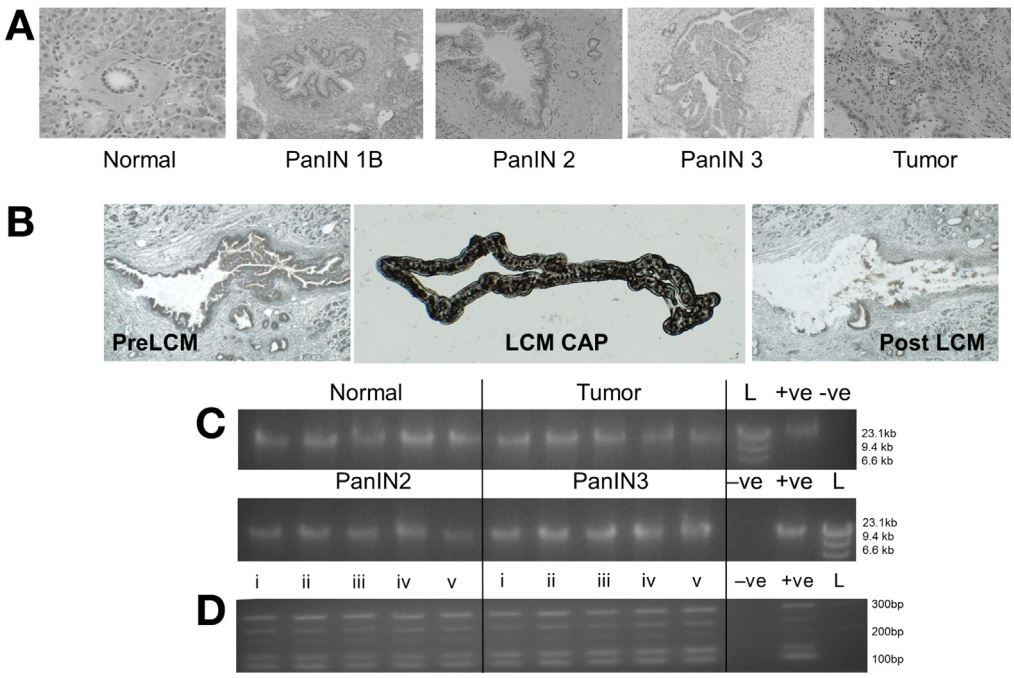



**Figure 1.** LCM of PanIN and tumor yields high-quality DNA after WGA. (*A*) Representative examples of PanIN and tumor pathology. (*B*) LCM of ductal epithelial cells from PDAC. (*C*) Reproducibility of in situ WGA. Five replicate amplifications from normal, PanINs, and tumor LCM samples yielded equivalent size and quantities of DNA, compared with non–LCM-positive controls (+ve). –ve, phosphate-buffered saline–negative control; L, size marker. (*D*) Multiplex polymerase chain reaction of 5 targets from 5 chromosomes. Five replicate WGA (i–v) from representative normal (*left panel*) and tumor (*right panel*) samples are shown.

adjacent PanIN2 and/or PanIN3 lesions (Figure 1*A*). Epithelial cells from these lesions were purified by LCM (Figure 1*B*). The heterogeneity in shape and size of PanINs resulted in substantial variation in the number of cells collected. In general, the number of cells constituting a PanIN lesion increased with grade and ranged from 20 to more than 100 cells per 10-$\mu$m section. To obtain sufficient genomic DNA for exome sequencing, a direct WGA protocol was used in which laser captured cellular material was lysed on the cap membrane and DNA was amplified directly with no intermediate DNA extraction.[10] This direct WGA method was applied to all tumor and PanIN tissues, yielding reproducible and consistent qualities of WGA DNA (Figure 1*C*). Minimal amplification bias was observed using a qualitative multiplex polymerase chain reaction assay (Figure 1*D*), indicating that the WGA DNA from small PanIN lesions was of sufficient quality for exome capture and sequencing. Germline DNA derived from peripheral blood mononuclear cells from the same pancreatic cancer cases was not amplified for this study.

### *Exome Capture and Sequence Analysis*

DNA from tumors, adjacent PanINs, and normals from 10 PDAC cases were exome-captured and sequenced. On average, more than 80% of baits yielded more than 20 times sequence depth and more than 70% had 40 times coverage. Variation in coverage did not correlate with the number of LCM purified cells and the read duplication rate was only moderately higher for WGA material ($\sim$20% vs $\sim$15%, data not shown). To evaluate drop-out caused by nonlinear WGA, the mean alternate allele frequency (AAF) for each sample, relative to corresponding non-amplified DNA from blood, was calculated using all heterozygous germline variants with more than 50 sequence reads (Supplementary Figure 1). WGA DNA from PanINs, tumors, and non-WGA blood DNA all showed a mean AAF of 46%, consistent with limited allele drop-out. Based on these results, this direct in situ WGA methodology allows for comprehensive genomic interrogation of lesions with limiting cell numbers.

In total, 1053 somatic mutations altering protein-coding sequences were detected in 10 tumors and 15 adjacent PanIN specimens. These included 845 non-synonymous, 121 frame shifts, 51 nonsense, and 36 mutation types. In total, 937 genes contained at least one variant, many of which were reported previously mutated in cancer (catalogue of somatic mutations in cancer [COSMIC] cancer database). Figure 2*A* shows the total numbers of SNVs and insertion or deletion variants for each case. Case 6 presented the most mutations in this study, with more than 200 mutations present in each PanIN2 and the associated tumor. A total of 902 high-confidence somatic SNVs were detected in the 25 samples. An average of 69.2 somatic SNVs was observed per sample (48.3, omitting the outlier case 6), in line with previous studies in pancreatic cancer.[17] Excluding case 6 again, a trend toward fewer alterations in early stage disease was observed, with an average of 30.2 mutations per PanIN2 compared with 49.3 in tumors. Conversely, late-grade PanIN3s presented with increased numbers of mutations (62.6) compared with tumors. All somatic mutations from each case are shown in Supplementary Table 1. A total of 151 insertions or deletions were observed within the 25 samples, for an average of 6 insertions or deletions per specimen (Figure 2*A* and Supplementary Table 2).

### *Commonality of Tumors and Adjacent PanINs*

Of the somatic mutations detected, approximately 66% were common to tumors and adjacent PanINs (PxT) (Figure 2*B*). Mutations found only in tumor or only in PanIN totaled 10% and 24% of all variants, respectively. Commonality between PanINs and tumors was assessed for each PanIN by calculating the percentage of variants also present in the adjacent tumor (Figure 2*C*). Although overlap between PanINs and associated tumor ranged from 34% to 96%, greater than 50% commonality with tumor was observed for 10 of the 15 sequenced PanINs. The PanIN3s of cases 41 (41P3) and 30 (30P3) showed the highest commonality with adjacent tumors, with 96% and 87% overlap, respectively. The PanIN2/3 pairs of cases 12 and 37 displayed the least commonality with just 34% to 40% homology with adjacent tumors. Conversely, PanIN 37P3 had much greater commonality with the adjacent 37P2 (77%), than either had with the associated tumor (Figure 2*B*). Separately, lineage was assessed using hierarchical clustering (Figure 2*D*). PanINs and tumors from the same cases were related most closely overall. However, distal branching of the 37T and 41P2 specimens from adjacent PanINs and tumors was observed, fitting with the differential commonality shown in Figure 2*B*. Overall, these results suggest that all of the adjacent lesions arose from common ancestral lineages and that the majority of somatic mutations in tumors arose early in the progression at the PanIN2 stage or earlier.

### *KRAS and TP53 Somatic Mutations*

As expected, *KRAS* and *TP53* were the most commonly mutated genes, observed in 9 and 7 cases, respectively, with 60% of the cases containing mutations in both genes. Almost all PanINs (13 of 15) and tumors (9 of 10) presented with G12 KRAS mutations, which result in constitutively active forms of the guanosine triphosphatase.[18] Table 1 shows the algorithmic read depths and alternative allele frequencies at G12 for all samples studied. KRAS G12 mutations initially were detected in 16 of 25 samples (*bold text*) using our conservative algorithmic settings. However, visualization of all reads at the G12 position identified 6 additional low-frequency KRAS mutations, which all were confirmed by Sanger sequencing using locked nucleic acids to ensure good analytic sensitivity (data not shown). Six cases had the relevant *KRAS* mutation in all samples, confirming *KRAS* mutations early in progression. G12V and G12D mutations were most abundant, observed in 6 and 5 cases, respectively. Multiple different G12 mutations were

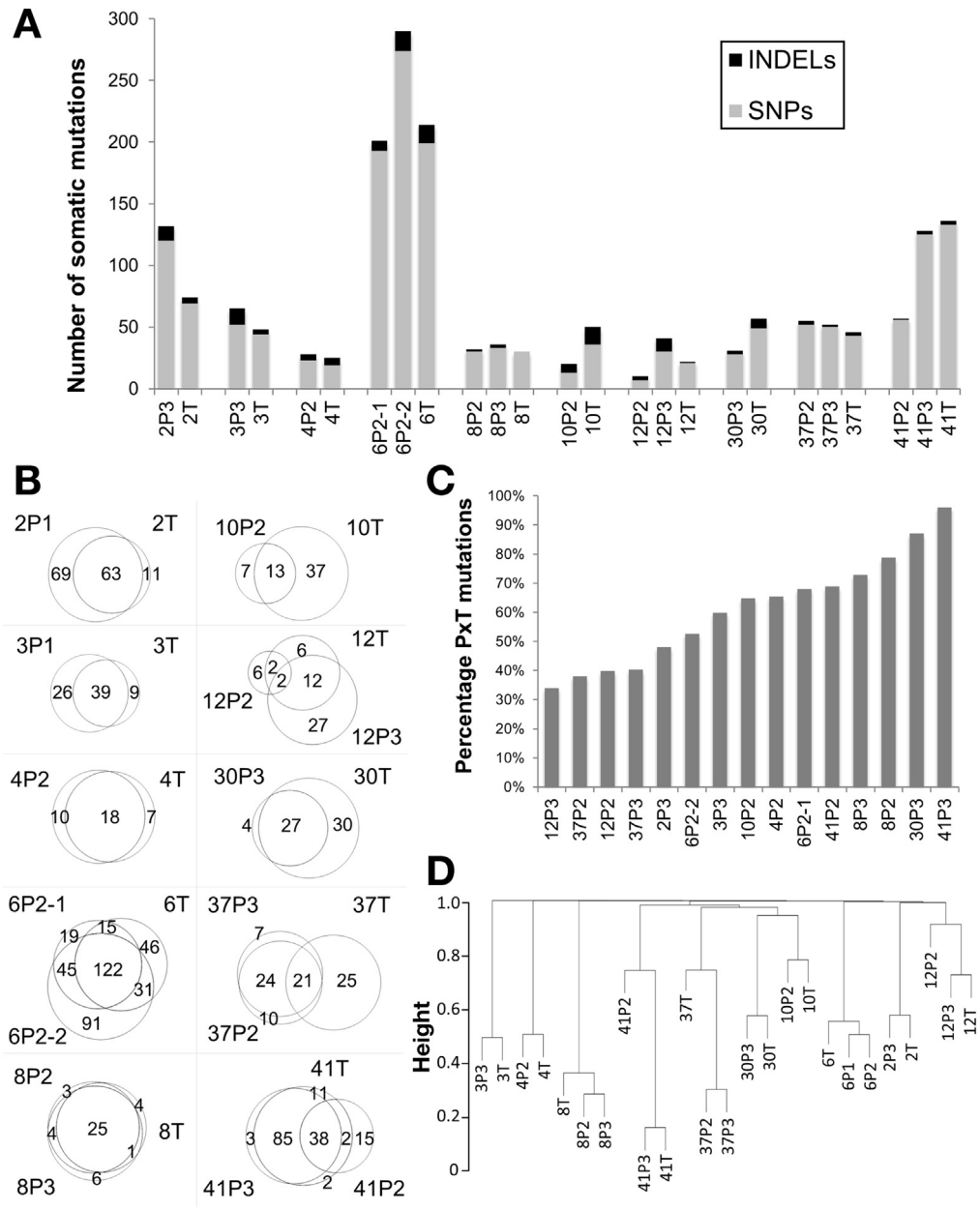



**Figure 2.** Numbers of somatic mutations and relatedness of tumors and adjacent PanINs. (*A*) Numbers of somatic mutations (gray) and insertions and deletions (black) per sample (grouped by case). P2, PanIN2; P3, PanIN3. (*B*) Percentage of commonality of PanINs with associated tumors. (*C*) Venn diagrams of somatic variants in each case. (*D*) Hierarchical clustering of tumors and PanINs using Euclidean distance measures for each possible comparison. INDEL, insertion and deletion.

observed within 3 cases (cases 12, 30, and 37), and a KRAS Q61H mutation also was identified in the 6P2-2 of case 6. Three different G12 mutations were observed in case 37, consistent with the relatively low clonality between these lesions, and predicting early divergence of the tumor and adjacent PanINs before KRAS mutagenesis (Figure 2). In addition, the presence of 2 independent *KRAS* mutations in 37P2, 12P2, and 6P2-2 suggest that heterogeneity exists within the premalignant lesions.

In the 7 tumors with *TP53* mutations, 6 displayed SNVs and 1 (40T) contained a 1-bp deletion (Table 1). Sanger sequencing positively confirmed all 7 mutations. *TP53* mutations were observed only in the tumors from 4 of the 7 cases (Table 2), but were present in both tumor and adjacent PanINs (2 PanIN2s and 2 PanIN3s) from the 3 other cases. The presence of *TP53* mutations in PanIN2 lesions, especially 4P2 and 8P2, which are highly clonal with the associated tumors, suggests that *TP53* mutations may occur early in progression to tumor at a relatively high frequency. Case 4 was the only case for which no KRAS mutation was detected, but a damaging missense R175H TP53 mutation was observed, which also was present in case 2. Two unique missense mutations (R282G and S241F) and 2 nonsense mutations (R306X and S945X) also were identified, each of which is associated with *TP53* inactivation (www.p53.iarc.fr).

### *Commonly Mutated Genes*

Sixty-seven genes were mutated in 2 or more cases (Supplementary Table 1). The large mucosal membrane protein genes *MUC16* and *FCGBP* each were recurrent in 4 cases, whereas *OTOF*, *PABPC1*, *RBMX*, and *SPTA1* each

**Table 1.** KRAS and TP53 Mutations

| | | KRAS | | | | | TP53 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | chr12: 25398284C>A | chr12: 25398284C>T | chr12: 25398285C>G | chr12: 25398285C>A | chr12: 25380275T>G | chr17: 7574020TC>T | chr17: 7577022G>A | chr17: 7577094G>C | chr17: 7577559G>A | chr17: 7578406C>T | chr17: 7579406G>T |
| | G12 read depth | Alternative allele reads | | | | | | | | | | |
| | | G12V | G12D | G12R | G12C | Q61H | E204fs | R174[a] | R150P | S109F | **R136H** | **S55*** |
| 2P3 | 33 | | **9** | | | 43, 0 | 13, 0 | 66, 0 | 37, 0 | 80, 0 | **3, 14** | 32, 0 |
| 2T | 28 | | **11** | | | 39, 0 | 27, 0 | 81, 0 | 55, 0 | 86, 0 | **14, 7** | 29, 0 |
| 3P3 | 38 | 3 | | | | 36, 0 | 23, 0 | 84, 0 | 52, 0 | 85, 0 | 23, 0 | 36, 0 |
| 3T | 29 | 4 | | | | 30, 0 | 16, 0 | 60, 0 | 35, 0 | 80, 0 | 25, 0 | 26, 0 |
| 4P2 | 38 | | | | | 57, 0 | 16, 0 | 71, 0 | 41, 0 | 78, 0 | **7, 14** | 26, 0 |
| 4T | 11 | | | | | 29, 0 | 19, 0 | 76, 0 | 46, 0 | 88, 0 | **9, 7** | 35, 0 |
| 6P2-2 | 43 | **3** | | | | **40, 13** | 24, 0 | 78, 0 | 60, 0 | 87, 0 | 25, 0 | 34, 0 |
| 6P2-1 | 38 | 3 | | | | 45, 1 | 16, 0 | 71, 0 | 46, 0 | 81, 0 | 24, 0 | 30, 0 |
| 6T | 29 | **7** | | | | 51, 0 | 18, 0 | 72, 0 | 51, 0 | **23, 50** | 14, 0 | 31, 0 |
| 8P2 | 54 | | | **14** | | 82, 0 | 33, 0 | 65, 0 | **9, 20** | 42, 0 | 28, 0 | 98, 0 |
| 8P3 | 54 | | | **8** | | 52, 0 | 24, 0 | 62, 0 | **19, 17** | 31, 0 | 30, 0 | 102, 0 |
| 8T | 34 | | | **4** | | 61, 0 | 35, 0 | 88, 0 | **34, 13** | 43, 0 | 43, 0 | 137, 0 |
| 10P2 | 18 | | | | | 38, 0 | 19, 0 | 56, 0 | 26, 0 | 23, 0 | 16, 0 | 56, 0 |
| 10T | 13 | | 3 | | | 19, 0 | 20, 0 | 48, 0 | 21, 0 | 21, 0 | 12, 0 | 59, 0 |
| 12P2 | 40 | 4 | **5** | | | 42, 0 | 21, 0 | 56, 0 | 28, 0 | 3, 0 | 25, 0 | 79, 1 |
| 12P3 | 55 | | **19** | | | 54, 0 | 10, 0 | 26, 0 | 12, 0 | 5, 0 | 8, 0 | 47, 0 |
| 12T | 52 | | **8** | | | 57, 0 | 17, 0 | 64, 0 | 49, 0 | 30, 0 | 31, 0 | **77, 17** |
| 30P3 | 18 | 2 | | | | 22, 0 | 13, 0 | 38, 0 | 24, 0 | 20, 0 | 20, 0 | 30, 0 |
| 30T | 17 | | 3 | | | 29, 0 | **5, 6** | 22, 0 | 7, 0 | 24, 0 | 16, 0 | 30, 0 |
| 37P2 | 40 | | **12** | | **7** | 42, 0 | 19, 0 | 80, 0 | 36, 0 | 34, 0 | 27, 0 | 80, 0 |
| 37P3 | 39 | | | | **15** | 40, 0 | 10, 0 | 62, 0 | 24, 0 | 28, 0 | 23, 0 | 88, 0 |
| 37T | 8 | **5** | | | | 20, 0 | 4, 0 | **22, 32** | 30, 0 | 11, 0 | 8, 0 | 23, 0 |
| 41P2 | 27 | **9** | | | | 46, 0 | 27, 0 | 88, 0 | 52, 0 | 39, 1 | 42, 1 | 120, 1 |
| 41P3 | 27 | **10** | | | | 56, 0 | 24, 0 | 60, 0 | 35, 0 | 31, 0 | 13, 0 | 81, 0 |
| 41T | 28 | **3** | | | 1 | 48, 0 | 46, 0 | 122, 0 | 50, 0 | 55, 0 | 62, 0 | 136, 0 |

NOTE. Positions are listed as chromosome:position reference > alternate. Commas separate the following numbers: reference allele, alternative allele reads.
P2, PanIN2; P3, PanIN3; Indel, insertion or deletion. Bolded numbers are events called by initial conservative algorithmic calling.
[a]New stop site.

**Table 2.** Commonly Mutated Genes

| Tissue[a] | Gene | Cases, n | Impact[b] | Exons, n | Amino acid[c] (exon) | | | | COSMIC[d] |
|---|---|---|---|---|---|---|---|---|---|
| T | SMAD4 | 2 | D/D | 12 | R361H(9) | A452fs(11) | | | 417/9350 |
| | PTPN5 | 2 | D/D | 15 | T125M(6) | R438fs(13) | | | 27/4418 |
| T, PxT | TP53 | 7 | All D | 12 | Table 1 | | | | 22871/77279 |
| | ATM | 2 | D/T | 63 | S1004N(20) | C2801Y(57) | | | 384/8337 |
| | CTCF | 2 | D/D | 12 | R187C(4) | E291G(9) | | | 71/4425 |
| | MAST4 | 2 | D/D | 29 | D626Y(19) | E763V(22) | | | 9/4860 |
| | RNF43 | 2 | D/D | 10 | D140E(4) | A169T(5) | | | 56/4240 |
| | C9orf174 | 2 | D/T | 51 | R1324W(42) | S1520R(47) | | | 16/4148 |
| | NEB | 2 | D/T | 182 | E3400Q(74) | H4431R(117) | | | 201/4283 |
| | RPGR | 2 | D/n | 19 | D668Y(15) | Q1104K(16) | | | 24/4265 |
| PxT | KRAS | 9 | All D | 6 | Table 1 | | | | 24608/114312 |
| | MUC16 | 4 | D/n/n/n | 84 | P6997R(3) | P7895T(3) | S11132G(5) | Q13573K(56) | 392/4696 |
| | OTOF | 3 | D/T/T | 47 | P490fs(14) | M607T(16) | M607T(16) | E1323K(32) | 99/4381 |
| | PABPC1 | 3 | D/D/T | 15 | D165G(3) | K231E(5) | E345(8)[e] | | 32/4216 |
| | RBMX | 3 | D/D/T | 9 | A78T(4) | P167A(5) | G105fs(4) | | 30/4265 |
| | SPTA1 | 3 | D/T/T | 52 | R468H(11) | E846D(18) | I2265T(49) | | 232/4287 |
| | C15orf39 | 2 | D/D | 3 | G411fs(2) | D575fs(2) | | | 16/4238 |
| | OSBPL9 | 2 | D/D | 23 | V51D(2) | F233C(14) | | | 17/4216 |
| | VAV3 | 2 | D/D | 27 | E95K(2) | G639fs(21) | | | 53/5465 |
| | GPX5 | 2 | D/T | 5 | A77V(2) | T203M(5) | | | 15/4238 |
| | HLA-DRB1 | 2 | D/T | 6 | K118Q(3) | S124A(3) | | | 9/4148 |
| | KIF4B | 2 | D/T | 1 | R742(1)[e] | R762H(1) | | | 52/4248 |
| | KLHDC3 | 2 | D/T | 11 | F49C(2) | N357T(11) | | | 7/4238 |
| | SH3RF1 | 2 | D/T | 12 | M518I(9) | RP827S(11) | | | 30/4718 |
| | TCF7L2 | 2 | D/T | 13 | E344(11)[e] | K385R(13) | | | 38/4734 |
| | SCLT1 | 2 | D/n | 21 | R417(15)[e] | Splice Site(2) | | | 28/4671 |
| | SKA3 | 2 | D/n | 13 | K386R(8) | Splice Site(7) | | | 19/4239 |
| PxT, P | FCGBP | 4 | D/D/D/T | 36 | E390(2)[e] | A1108T(6) | S4284(24)[e] | K3848E(28) | 139/4242 |
| | ATP8B1 | 2 | D/D | 28 | D554E(16) | D554E(16) | | | 50/4922 |
| | FN1 | 2 | D/D | 46 | E888V(18) | R1207G(24) | | | 94/4730 |
| | LZTS1 | 2 | D/D | 3 | R36W(1) | L113P(1) | | | 26/4238 |
| | OBSCN | 2 | D/D | 81 | C1188F(12) | R4593C(52) | | | 184/4531 |
| | PAK2 | 2 | D/D | 15 | Q101H(4) | K128R(4) | | | 12/5091 |
| | PGAP1 | 2 | D/D | 27 | F565C(18) | I606N(20) | | | 37/4238 |
| | RBFOX1 | 2 | D/D | 14 | S27L(3) | T118M(4) | | | 53/4238 |
| | APOB | 2 | D/T | 29 | N886H(18) | V4227L(29) | | | 220/4306 |
| | GABRA5 | 2 | D/T | 11 | R221T(8) | T412I(11) | | | 47/4238 |
| | KCNJ12 | 2 | D/T | 3 | (R261H(3) | I262S(3) | L211F(3) | E289Q(3) | 59/4239 |
| | LRP2 | 2 | D/T | 79 | L3587V(55) | F4300I(70) | | | 237/4312 |
| | MST1 | 2 | D/T | 18 | P19S(1) | T104S(3) | | | 33/4426 |
| | RYR2 | 2 | D/T | 105 | R389C(13) | D3203E(68) | | | 352/4284 |

fs, frame shift.

[a]Tissue distribution of mutated genes. *Gene* refers to genes selected based on somatic mutation in 2 or more cases, predicted damaging (D) in one mutation, and listing in the COSMIC cancer database.

[b]Impact: summary of SIFT/PolyPhen analysis: D, predicted damaging; T, predicted tolerated.

[c]Amino acid alterations plus exon involved is indicated in parentheses.

[d]The frequencies of observed mutations in COSMIC for any cancer type (COSMIC all cancers).

[e]New stop site.

were mutated in 3 cases. PABPC1 and RBMX are involved in messenger RNA regulation, and SPTA1 directs cell shape and axon guidance. None have been linked previously with pancreatic cancer. An additional 58 genes were mutated in 2 cases. Table 2 lists 42 recurrently mutated genes previously implicated in cancer (COSMIC database) that had at least one predicted damaging mutation in the 25 samples. Only 2 of the commonly mutated genes (*SMAD4* and *PTPN5*) were altered exclusively in tumors. Although SMAD4 has been heavily implicated late in pancreatic cancer development,[17] the protein tyrosine phosphatase PTPN5[19] has not been linked previously to pancreatic cancer. The majority of the recurrently mutated genes carried PxT variants (common to tumor and PanIN of the same case). Seven recurrently altered genes were mutated in single tumors and also in adjacent tumor and PanIN combinations (T, PxT), but never in PanINs alone (P) (Table 2). ATM, which regulates the cellular response to DNA damage and recently was implicated as a pancreatic cancer predisposition gene,[20] was mutated in 2 cases, both of which were validated by Sanger sequencing. Because the cases with *ATM* alterations did not contain any of the 7 *TP53* mutations, 9 of the 10 tumors in this study contained aberrations in DNA damage response pathways. The CTCF transcriptional repressor/chromatin binding factor and the RNF43 E3-ubiquitin-protein ligase

regulator of the WNT signaling pathway that has been linked to cell growth promotion and cancer[21] also were in this group. PxT mutated genes included 2 Rho family regulators *VAV3* and *SH3RF1*, the *TCF7L2* WNT signaling factor, the oxidative damage protection gene *GPX5*, and 3 chromosomal structural regulating genes, *KIF4B*, *KLHDC3*, and *SKA3*. Additional PxT:P mutated genes included the *LZTS1* tumor suppressor and the *PAK2* Ser/Thr kinase, as well as the structural regulatory proteins fibronectin (FN1) and obscurin (OBSCN), which also have been linked to cancer progression.[22,23]

### *Common Driver Genes in Cancer*

Mutations that were retained in progression from PanIN to tumor likely are enriched for alterations that drive tumor formation, whereas mutations specific to tumor are more likely drivers of progression. Table 3 lists the genes found to contain high-confidence mutations in this study that commonly are mutated in cancer and are listed in the COSMIC gene census database as cancer drivers. In addition to *SMAD4*, 7 other known driver genes were mutated only in tumor samples including the transcription factors *MYCL1*, *PAX8*, and TRIP11,[24–26] the tyrosine receptor kinase *NTRK3*, and the mixed-lineage leukemia family member *MLL2*, previously linked with pancreatic cancer.[27] Nine PxT mutated genes were observed in addition to *KRAS*, including 4 transcription factors, *DAXX*, *NCOA1*, *TCF7L2*, and *ZNF521*. Of these genes, TCF7L2 is associated with Wnt signaling and an increased risk of type 2 diabetes, and increased expression has been reported in pancreatic cancer.[28] In contrast, mutations in the established *GNAS* and *CHEK2* cancer drivers[29,30] were detected in PanIN lesions but not in adjacent tumors, highlighting independent progression of the PanINs and the tumors. Additional mutations involved the DNA replication repair gene *BLM* and the tumor suppressor *CCDC6*. Sanger sequencing was used to validate these genes further (Table 3). In addition to the multiple *KRAS* and *TP53* mutations, alterations in *MLL2*, *PAX8*, *TRIP11*, *LCP1*, *MALT1*, *GNAS*, and 2 each in *ATM* and *SMAD4* faithfully validated. However, mutations in *MLL3*, *CHEK2*, and *NSD1* failed to validate. The limited sensitivity of Sanger sequencing for low alternative allele frequencies resulting from heterogeneity in these lesions could explain the failure of these events to validate.

### *Commonly mutated genes in pancreatic cancer*

A comparison of the 937 genes with mutations in this study with genes frequently mutated in pancreatic cancer from 2 other pancreatic exome sequencing studies[17,31] identified 40 genes in common (Table 4). Mutations in the key pancreatic cancer driver genes *KRAS*, *TP53*, *SMAD4*, and *ATM* were observed in the 3 studies. However, apart from *KRAS* and *TP53*, these mutations were not identical. The majority of the recurrently mutated genes contained PxT mutations, suggesting that many of these commonly mutated genes may be early targets of mutation during tumor development. Of these, the potential tumor suppressor gene *CSMD1*, linked with aggressive carcinomas,[32] had not been linked previously to pancreatic cancer. Two unrelated motor protein genes, *DNAH8* and *MYO1E*, were mutated only in individual tumors in this study and also were observed in 2 cases each in the previous studies. Although no link to pancreatic cancer was reported, MYO1E has been linked to TP53-associated DNA damage response.[33] A number of common mutated genes, including the mucin (MUC) and RYR family's members, are more likely passenger events owing to the large size of the associated genes, rather than common driver events in pancreatic cancer. A further 6 genes in Table 4 were mutated solely in PanINs. This group may be enriched for passenger genes that do not contribute to tumor development or progression. However, no absolute conclusions can be drawn on the basis of a single case. For instance, *NAV3* is a putative tumor suppressor previously linked to pancreatic and other cancers.[34]

### *Pathway Analysis*

To identify commonly mutated pathways involved in the development of pancreatic cancer, genes with mutations occurring in PxT or tumor only were annotated by pathway (Supplementary Table 3). The most significant pathways are summarized in Supplementary Table 4. neural cell adhesion molecule, insulin, and platelet-derived growth factor signaling pathways were implicated in the early development of pancreatic cancer in part because of the involvement of *KRAS* mutations. Gap junction signaling, chemokine signaling, and regulation of the actin cytoskeleton were altered predominantly by PxT mutations in each of the 10 cases. Similarly, mitogen-activated protein kinase signaling and cell-cycle regulation pathways were modified in all cases, but equal numbers of PxT and T mutations were detected. The recently highlighted Axon guidance pathway in the International Cancer Genome Consortium (ICGC) study[31] also was observed early in 2 separate pathway sets (Kegg and Reactome). PxT mutations were observed in 17 genes in the neuroactive ligand receptor-interaction pathway in 8 cases and in 12 independent genes in olfactory signaling in 7 cases. Mechanistically, genes in these pathways are major components of the superfamily of rhodopsin-like G-protein–coupled receptors, which can induce cascades of responses related to carcinogenesis.[35] Because of the limited number of genes mutated only in tumors, few pathways were associated only with tumors. However, mutations unique to tumors were enriched in genes involved in Wnt and transforming growth factor-$\beta$ signaling, consistent with the loss of response to these ligands late in tumor progression.[36]

## Discussion

A detailed understanding of the genes and signaling pathways that influence pancreatic cancer onset

**Table 3.** Common Driver Genes in Cancer

| | Gene | Cases, n | Case | Tissue | Chromosome | Position | Mutation[a] | Sanger validation[b] | Amino acid | Impact | Ref allele[c] | Allele 1[c] | Allele 2[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | GOLGA5 | 1 | 2 | T | 14 | 93276655 | T - C | / | F350S | D | T 22 | T 22 | C 8 |
| | MLL2 | 1 | 37 | T | 12 | 49433874 | TG - T | Yes | H2560fs | D | TG 7 | TG 2 | T 6 |
| | MYCL1 | 1 | 8 | T | 1 | 40363341 | C - A | / | Q266H | T | C 120 | C 88 | A 31 |
| | NTRK3 | 1 | 6 | T | 15 | 88476311 | G - T | / | C599[d] | D | G 80 | G 47 | T 16 |
| | PAX8 | 1 | 10 | T | 2 | 114002179 | G - A | Yes | R72W | D | G 197 | G 32 | A 11 |
| | SLC45A3 | 1 | 4 | T | 1 | 205628369 | GAG - G | / | Y551fs | D | GAG 16 | GAG 2 | G 4 |
| | SMAD4 | 2 | 37 | T | 18 | 48591919 | G - A | Yes | R361H | D | G 31 | G 16 | A 12 |
| | | | 10 | T | 18 | 48603054 | C - CT | Yes | A452fs | D | C 14 | C 8 | CT 6 |
| | TRIP11 | 1 | 30 | T | 14 | 92470155 | C - T | Yes | E1389K | D | C 110 | C 45 | T 15 |
| PxT, T | ATM | 2 | 10 | T | 11 | 108214082 | G - A | Yes | C2801Y | D | G 49 | G 33 | A 14 |
| | | | 41 | PxT | 12 | 108142067 | G - A | Yes | S1004N | T | G 146 | G 90 | A 44 |
| | TP53 | 7 | Table 1 | | 17 | | | Yes | | | | | |
| PxT | DAXX | 1 | 6 | PxT | 6 | 33288845 | A - T | / | F161Y | D | A 11 | A 10 | T 8 |
| | KRAS | 9 | Table 1 | | 12 | | | Yes | | D | | | |
| | LCP1 | 1 | 2 | PxT | 13 | 46730643 | G - A | Yes | R141W | D | G 72 | G 60 | A 26 |
| | MALT1 | 1 | 3 | PxT | 18 | 56376751 | A - G | Yes | E264G | T | A 77 | A19 | G 20 |
| | MLL3 | 1 | 2 | PxT | 7 | 151860002 | T - TA | ND | F3553fs | D | T 53 | T 8 | TA 15 |
| | NCOA1 | 1 | 6 | PxT | 2 | 24930303 | A - C | / | K655T | D | A 32 | A 23 | C 6 |
| | NUP98 | 1 | 6 | PxT | 11 | 3727776 | A - T | / | L942M | T | A 51 | A 35 | T 14 |
| | PDE4DIP | 1 | 4 | PxT | 1 | 144952220 | C - T | / | A135T | T | C 18 | C 9 | T 15 |
| | TCF7L2 | 2 | 6 | PxT | 10 | 114925320 | A - G | / | K385R | T | A 37 | A 11 | G 10 |
| | | | 2 | PxT | 10 | 114912131 | G - T | / | E344[d] | D | G 31 | G 60 | T 18 |
| | ZNF521 | 1 | 6 | PxT | 18 | 22804449 | C - T | / | V1145I | D | C 89 | C 64 | T 26 |
| P | BLM | 1 | 6 | P | 15 | 91358439 | A - T | / | K1395I | D | A 32 | A 18 | T 11 |
| | CCDC6 | 1 | 37 | P | 10 | 61572516 | G - A | / | P242S | D | G 37 | G 53 | A 37 |
| | CHEK2 | 1 | 2 | P | 22 | 29121277 | G - T | ND | T133K | T | G 41 | G 53 | A 24 |
| | CLTCL1 | 1 | 2 | P | 22 | 19209058 | C - A | / | A880S | D | C 43 | C 69 | A 27 |
| | GNAS | 1 | 41 | P | 20 | 57484421 | G - A | Yes | R186H | D | G 32 | G 57 | A 23 |
| | KDM6A | 2 | 37 | P | X | 44922970 | C - T | / | Q611[d] | D | C 9 | C 6 | T 7 |
| | | | 6 | P | X | 44910988 | AA - A | / | Q230fs | D | AA | AA 3 | A 6 |
| | NACA | 1 | 12 | P | 12 | 57114403 | GA - G | / | S304fs | D | GA 46 | GA 10 | G 19 |
| | NSD1 | 1 | 2 | P | 5 | 176719127 | C - T | ND | A1875V | D | C 47 | C 49 | T 14 |

NOTE. Cases refers to the number of cases with mutations in this gene, case refers to the pancreatic cancer case in this study, position refers to the nucleotide, impact (SIFT/PolyPhen). *Gene* refers to genes selected based on listing in the COSMIC census database of common driver genes.

ND, not detected/not evaluated.

[a]Mutation: reference – mutated nucleotide.

[b]Sanger validation was as follows: yes, mutations were confirmed by Sanger sequencing.

[c]Allelic nucleotide and read depth for the reference normal (ref allele) and the alternative alleles for the PanIN and tumor tissues (allele 1, allele 2).

[d]New stop site.

**Table 4.** Commonly Mutated Genes in Pancreatic Cancer

| | Gene | This study[a] | ICGC (99)[a] | Jones (24)[a] | Total | Common Site | Case | Tissue | Chromosome | Position | Amino acid | Impact | Allele 1 | Allele 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | SMAD4 | 2 | 16 | 8 | 26 | n | 37 | T | 18 | 48591919 | R361H | D | G 16 | A 12 |
| | | | | | | n | 10 | T | 18 | 48603054 | A452fs | D | C 6 | CT 8 |
| | DNAH8 | 1 | 1 | 1 | 3 | n | 10 | T | 6 | 38723768 | I476M | D | A 67 | G 26 |
| | MYO1E | 1 | 2 | 0 | 3 | n | 37 | T | 15 | 59564510 | I48V | D | T 60 | C 19 |
| | NTRK3 | 1 | 1 | 1 | 3 | n | 6 | T | 15 | 88476311 | C599[b] | D | G 47 | T 16 |
| T, PxT | TP53 | 7 | 33 | 18 | 58 | y | Table 1 | | | | | | | |
| | ATM | 2 | 5 | 0 | 7 | n | 10 | T | 11 | 108214082 | C2801Y | D | G33 | A 14 |
| | | | | | | n | 41 | PxT | 11 | 108142067 | S1004N | T | G 90 | A 44 |
| | NEB | 2 | 5 | 1 | 8 | n | 41 | PxT | 2 | 152474902 | E3400Q | D | C 143 | G 107 |
| | | | | | | n | 37 | T | 2 | 152421598 | H4431R | T | T 103 | C 32 |
| PxT | SLITRK3 | 1 | 1 | 1 | 3 | n | 6 | PxT | 3 | 164906758 | G621[b] | D | C 20 | A 10 |
| | ATP10A | 1 | 1 | 2 | 4 | n | 2 | PxT | 15 | 25959351 | T605M | D | G 18 | A 16 |
| | CDH4 | 1 | 2 | 0 | 3 | n | 6 | PxT | 20 | 60499411 | P476S | D | C 24 | T 8 |
| | CSMD1 | 1 | 5 | 1 | 7 | n | 2 | PxT | 8 | 4494898 | D90N | D | C 46 | T 26 |
| | DPP6 | 1 | 1 | 3 | 5 | n | 8 | PxT | 7 | 154593134 | R393[b] | D | C 31 | T 32 |
| | FNIP1 | 1 | 1 | 1 | 3 | n | 6 | PxT | 5 | 131008104 | L650S | D | A 39 | G 13 |
| | HMCN1 | 1 | 4 | 0 | 5 | n | 41 | PxT | 1 | 186063449 | C3413F | D | G26 | T 56 |
| | KRAS | 9 | 94 | 24 | 118 | y | Table 1 | | | | | | | |
| | LRFN5 | 1 | 0 | 2 | 3 | n | 2 | PxT | 14 | 42355997 | R57W | D | C 26 | T 18 |
| | MLL3 | 1 | 6 | 4 | 11 | n | 2 | PxT | 7 | 151860002 | F3553fs | D | T 11 | TA 18 |
| | MUC16 | 4 | 5 | 2 | 11 | n | 12 | PxT | 19 | 8993006 | Splice Site | N | C 21 | A 28 |
| | | | | | | n | 12 | P | 19 | 9066456 | P6997R | N | G 110 | C 110 |
| | | | | | | n | 6 | PxT | 19 | 9063763 | P7895T | N | G 34 | T 13 |
| | | | | | | n | 2 | PxT | 19 | 8999458 | Q13573K | D | G 62 | T 10 |
| | | | | | | n | 41 | PxT | 19 | 9048237 | S11132G | N | T 164 | C 86 |
| | ODZ4 | 1 | 1 | 1 | 3 | n | 41 | PxT | 11 | 78381167 | G2075R | D | C 158 | G 70 |
| | PKHD1L1 | 1 | 2 | 0 | 3 | n | 2 | PxT | 8 | 110454274 | N1415H | D | A 37 | C 8 |
| | RYR1 | 1 | 1 | 1 | 3 | n | 41 | PxT | 19 | 38991484 | M2490V | T | A 18 | G 4 |
| | RYR2 | 2 | 2 | 2 | 6 | n | 6 | PxT | 1 | 237870277 | D3203E | T | T 6 | A 9 |
| | | | | | | n | 37 | P | 1 | 237604778 | R389C | D | C 161 | T45 |
| | RYR3 | 1 | 4 | 0 | 5 | n | 10 | T | 15 | 34145258 | N4542K | T | T 98 | A 34 |
| | SCN2A | 1 | 1 | 1 | 3 | n | 41 | PxT | 2 | 166187962 | D758H | D | G 92 | C 54 |
| | SPATA17 | 1 | 1 | 1 | 3 | n | 41 | PxT | 1 | 217975170 | P328L | D | C 35 | T 33 |
| | SYCE1 | 1 | 3 | 1 | 5 | n | 2 | PxT | 10 | 135368531 | P341S | D | G48 | A 24 |
| | USP34 | 1 | 2 | 0 | 3 | n | 6 | PxT | 2 | 61622085 | K219T | D | T 12 | G 9 |
| PxT, P | APOB | 2 | 1 | 1 | 4 | n | 2 | P | 2 | 21225615 | V4227L | T | C 24 | A 14 |
| | | | | | | n | 6 | PxT | 2 | 21245863 | N886H | D | T 22 | G 19 |
| | FN1 | 2 | 1 | 1 | 4 | n | 4 | PxT | 2 | 216259428 | R1207G | D | T 36 | C 19 |
| | | | | | | n | 2 | P | 2 | 216271900 | E888V | D | T70 | A 25 |
| | OBSCN | 2 | 2 | 1 | 5 | n | 37 | P | 1 | 228433195 | C1188F | D | G 104 | T 20 |
| | | | | | | n | 4 | PxT | 1 | 228505380 | R4593C | D | C 42 | T 42 |
| P | CDH23 | 1 | 2 | 0 | 3 | n | 37 | P | 10 | 73553299 | P2205L | D | C 77 | T 32 |
| | GPR133 | 1 | 1 | 2 | 4 | n | 8 | P | 12 | 131456051 | K79M | D | A 50 | T 35 |
| | LAMA1 | 1 | 1 | 1 | 3 | n | 12 | P | 18 | 7042183 | K407fs | D | CAG 14 | C 7 |
| | NAV3 | 1 | 1 | 1 | 3 | n | 41 | P | 12 | 78594300 | V2233L | D | G 97 | T 35 |
| | NBAS | 1 | 3 | 0 | 4 | n | 6 | P | 2 | 15613348 | L575M | D | A 21 | T 10 |
| | PCDHGA1 | 1 | 1 | 1 | 3 | n | 2 | P | 5 | 140711164 | E305[b] | D | G 15 | T 32 |

NOTE. Cases refers to the number of cases with mutations in this gene, case refers to the pancreatic cancer case in this study, position refers to the nucleotide, impact (SIFT/PolyPhen). *Genes* refers to genes commonly mutated in this study (ICGC), Biankin et al[31] and Jones et al.[17]

[a]The number of cases mutated in each study is shown. *Common site* refers to identical mutations in the studies, and *case* refers to the specific case in this study with a mutation in the specific gene.

[b]New stop site.

and progression is required to improve early diagnosis, prevention, and therapy. However, identification of the important driver events has proven difficult because of the heterogeneous nature of somatic mutations and accumulation of many passenger mutations. Recent whole-exome studies have verified the prominent involvement of *KRAS* mutations, the late presentation of *TP53* mutations, and the frequent involvement of *SMAD4* in pancreatic tumor progression.[17,31] Although important, these studies were limited by the use of bulk extracted or macrodissected tumor tissues for genomic interrogation. Given the low cellularity of these tumors, this can lead to inclusion of adjacent nontumor tissues and subsequent incorporation of large numbers of passenger mutations in studies. In addition, because these studies focused on advanced-stage tumors, little information about genetic alterations involved in progression from premalignant lesions to invasive tumors was obtained.

In this study, we have addressed these issues by isolating pure populations of cells using LCM and by focusing on PanINs and adjacent tumors. Robust methodology was used to reproducibly amplify exome capture and sequence DNA from small numbers of purified cells. Although WGA can introduce mutations and sequence bias resulting in loss of certain genomic regions, we show here that the in situ WGA technique can result in representative somatic mutation profiles. Specifically, 80% of the exome consistently was captured to sufficient sequence depth from as few as 100 cells. Allelic drop-out after WGA also was shown to be limited based on comparable mean AAF of germline single-nucleotide polymorphisms in amplified DNA and related unamplified DNA from blood (Supplementary Figure 1). LCM guided by pathology review was expected to reduce normal cell contamination, increasing cellularity. In an attempt to calculate tumor cellularity for each sample we evaluated the mean AAF for all heterogeneous somatic variants with more than 50 times read depth. However, the level of cellularity predicted using this model varied substantially across samples, ranging from 25% to 72% (data not shown), and generally was much lower than expected for defined PanIN lesions (Figure 1*A*). Further analysis showed the mean AAF was highly correlated with commonality, suggesting that this measure more likely represents heterogeneity within each sample, consistent with the existence of multiple KRAS mutations in certain samples (Table 1). Thus, the inability to establish co-existence of somatic variants in individual cellular populations and the associated under-representation of the genome owing to a limited and varied number of somatic events present at 50 times read depth, make estimation of tumor cellularity for this study challenging.

The simultaneous analysis of tumors and adjacent precursor lesions from a series of pancreatic cases distinguishes this study from other pancreatic tumor exome sequencing studies. The identification of identical somatic mutations present in tumor and adjacent PanINs (PxT; 66% of SNVs) internally validates these events and provides strong evidence that these aberrations arose in common clonal ancestors. This also suggests that PxTs, especially those in highly clonal tumors and PanINs, are enriched for drivers of tumorigenesis. The high frequency of the commonly mutated genes (*KRAS*, *TP53*) in this study and the presentation of the mutations at different stages of PanIN progression (*KRAS* in early PanINs, *TP53*/*SMAD3/4* in tumors) are consistent with results from exome sequencing of nonamplified samples. The additional identification of *TP53* mutations in 2 PanIN2s also showed that driver mutations in *TP53* can occur early during the development of pancreatic tumors, emphasizing the benefit of evaluating PanIN/tumor combinations. Mutations presenting solely in PanINs (24%) could not be implicated in cancer development, whereas mutations found only in tumors (10%) either were involved in driving cancer progression or represent bystander events. This limited population of unique mutations in individual samples suggests that WGA did not introduce large numbers of nonspecific mutations.

Although every PanIN had some somatic mutations in common with associated tumor samples (range, 34%–96%), the commonality varied considerably. In particular, the PanINs of cases 41, 30, and 8 shared more than two thirds of their rare somatic mutations with adjacent tumor. This may reflect very recent divergence of the tumor and PanIN clones from a common ancestor. Alternatively, it is possible that the PanIN lesions are actually components of these tumors that have spread by cancerization of ducts, although there was no correlation between clonality and the proximity of tumor and adjacent PanIN. Conversely, the PanINs of cases 12 and 37 presented with least commonality with their adjacent tumors (34%–40%), predicting more distal divergence of the tumor or independent histologic progression from a mutated background. Importantly, highly clonal lesions were very informative, not only in identifying PxT mutations as candidate drivers, but also in further defining the roles of P-only and T-only mutations in tumor progression.

The development of pancreatic cancer involves a compendium of genetic mutations. Whether a stepwise accumulation of mutations is required for the transition from premalignant to malignant lesions is an area of great debate. In this study, we show that PanIN2 lesions often contain as many mutations as PanIN3 and invasive tumor samples, even when accounting for commonality/clonality. This raises the possibility that PanIN2 lesions may transition directly to tumor without forming PanIN3 lesions. Alternatively, because mutational load is similar, it is possible that premalignant lesions may require epigenetic modifications, aneuploidy, or expression-based alterations to progress to invasive tumor. Further studies of highly clonal lesions will provide further insight into these important issues.

Despite the high frequency of *KRAS* and *TP53* mutations in pancreatic cancer, mutations in these genes may not be necessary for the development of pancreatic

tumors. Although case 4 presented no detectable KRAS mutations, an early TP53 mutation was detected in the PanIN2 (4P2). Only 10P2 had no detectable KRAS or TP53 mutations. The coverage across these genes was sufficient to identify mutations present in a high proportion of cells in each sample, however, mutations present only in subclones within these lesions could have been missed. In addition, exome sequencing may have overlooked potential large deletions of *TP53*. One limitation of this study was that 7 of 24 G12 KRAS mutations were overlooked in the initial algorithmic calling because the variants did not pass quality filters because of lower sequence depth of the alternative alleles. However, the 7 overlooked mutations were detected on re-analysis of sequences at the G12 position and were validated by locked nucleic acid Sanger sequencing. Although this suggests that mutations in other genes may have been overlooked in this study, because of low-sequence coverage, the more conservative filtering applied in this study was preferred because the number of false-positive variants was controlled. Mutations in other genes in the *TP53* and *KRAS* signaling pathways may have accounted for the loss of cell-cycle control and the enhanced proliferation needed for PanIN and tumor growth. One potentially significant observation from this study was that additional key modulators of cellular response to DNA damage were mutated (*ATM* [cases 10 and 41] and *MDC1* [case 3]) in the 3 cases in which no *TP53* mutations were observed. Additional mutations in *TOP2A, CHEK2*, *FANCB*, *FAN1*, *POLH*, *RECQL4*, *APLF*, *DCLRE1B*, *BLM*, *HELQ*, *TOPBP1*, and *MRE11a*, which also influence DNA repair or the cell-cycle response to DNA damage, suggest that defective DNA damage response signaling is an early event in PanIN and tumor development. Pathways involving gap junction, actin cytoskeleton, chemokine, mitogen-activated protein kinase, cell cycle, and axon guidance signaling also impacted all cases in this study with a predominant early presentation.

Despite resection to clear surgical margins during tumor excision, locoregional recurrence is common in pancreatic cancer.[37] The clinical significance of clonality of PanINs with adjacent tumors may have implications for surgical resection of pancreatic cancers where there is persistence of PanIN2 lesions at the surgical margins. Current clinical guidelines do not necessitate the expanded resection of PanIN2 areas. However, the results of this study emphasize the high probability of clonality of PanIN2s with adjacent tumors, the accumulation of multiple predicted somatic driver mutations of pancreatic cancer in PanINs, and the potential for PanIN2s to progress directly to tumor. Further studies, on the basis of these findings, are needed to determine whether PanIN2 and PanIN3 precursors at margins should be resected routinely.

Our results show extreme heterogeneity of mutational burden across different patients. Excluding known factors such as *TP53* and *KRAS* mutations, very few genes were mutated in multiple tumors or individuals. Although this study may not have been large enough to identify other commonly mutated genes, alignment of our somatic mutation lists with those of previous data sets from global cancer (COSMIC) and specific pancreatic studies[17,31] allowed identification of a large number of genes that may contribute to pancreatic tumor formation. Additional studies are needed to identify specific genes and/or pathways that influence PanIN development and progression to tumor.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2013.07.049.

### References


1. Fitzgerald TL, Hickner ZJ, Schmitz M, et al. Changing incidence of pancreatic neoplasms: a 16-year review of statewide tumor registry. Pancreas 2008;37:134–138.
2. Cubilla AL, Fitzgerald PJ. Pancreas cancer (non-endocrine). Clin Bull 2008;8:91–99.
3. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225–249.
4. Smeenk HG, Tran TC, Erdmann J, et al. Survival after surgical management of pancreatic adenocarcinoma: does curative and radical surgery truly exist? Langenbecks Arch Surg 2005;390:94–103.
5. Sohn TA, Yeo CJ, Cameron JL, et al. Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators. J Gastrointest Surg 2000;4:567–579.
6. Kloppel G, Hruban RH, Longnecker DS, et al. Ductal adenocarcinoma of the pancreas, WHO classification of tumors: pathology and genetics of the digestive system. Lyon: IARC Press, 2000.
7. Hruban RH, Takaori K, Klimstra DS, et al. An illustrated consensus on the classification of PanIN and intraductal papillary mucinous neoplasms. Am J Surg Pathol 2004;28:977–987.
8. Hingorani SR, Petricoin EF, Maitra A, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell 2003;4:437–450.
9. Kim J, Reber HA, Dry SM, et al. Unfavourable prognosis associated with K-ras gene mutation in pancreatic cancer surgical margins. Gut 2006;55:1598–1605.
10. Murphy SJ, Cheville JC, Zarei S, et al. Mate-pair sequencing of WGA DNA following LCM of prostate cancer. DNA Res 2012;19:395–406.
11. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.
12. Asmann YW, Middha S, Hossain A, et al. TREAT: a bioinformatics tool for variant annotations and visualizations in targeted and exome sequencing data. Bioinformatics 2012;28:277–278.
13. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.
14. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012;28:311–317.
15. Ng PC, Henikoff S. Predicting the effects of amino-acid substitutions on protein function. Annu Rev Genomics Hum Genet 2006;7:61–80.
16. Bauer S, Going GJ. Bayesian: model-based gene set analysis of genome-scale data. Nucleic Acids Res 2010;38:3523–3535.
17. Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801–1806.
18. Janakiraman M, Vakiani E, Zeng Z, et al. Genomic and biological characterization of exon 4 KRAS mutations in human cancer. Cancer Res 2010;70:5901–5911.

19. Mukherjee S, Poddar R, Deb I, et al. Dephosphorylation of specific sites in the kinase-specificity sequence domain leads to ubiquitin-mediated degradation of the tyrosine phosphatase STEP. Biochem J 2011;440:115–125.
20. Roberts NJ, Jiao Y, Yu J, et al. ATM mutations in patients with hereditary pancreatic cancer cancer. Discovery 2012;2:41–46.
21. Sugiura T, Yamaguchi A, Miyamoto K. A cancer-associated RING finger protein, RNF43, is a ubiquitin ligase that interacts with a nuclear protein, HAP95. Exp Cell Res 2008;314:1519–1528.
22. Perry NA, Shriver M, Mameza MG, et al. Loss of giant obscurins promotes breast epithelial cell survival through apoptotic resistance. FASEB J 2012;26:2764–2775.
23. Waalkes S, Atschekzei F, Kramer MW, et al. Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer. BMC Cancer 2010;10:503–508.
24. Morton CC, Nussenzweig MC, Sousa R, et al. Mapping and characterization of an X-linked processed gene related to MYCL1. Genomics 1989;4:367–375.
25. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. Nat Rev Endocrinol 2011;7:569–580.
26. Chang KH, Chen Y, Chen TT, et al. A thyroid hormone receptor coactivator negatively regulated by the retinoblastoma protein. Proc Natl Acad Sci U S A 1997;94:9040–9045.
27. Ansari KI, Mandal SS. Mixed lineage leukemia: roles in gene expression, hormone signaling and mRNA processing. FEBS J 2010; 277:1790–1804.
28. Frietz S, Wang R, Yao L, et al. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. Genome Biol 2012;13:R52.
29. Wu J, Matthaei H, Maitra A, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med 2011;3:92ra66.
30. Antoni L, Sodha N, Collins I, et al. CHK2 kinase: cancer susceptibility and cancer therapy- two sides of the same coin? Nat Rev Cancer 2007;7:925–936.
31. Biankin AV, Waddell N, Kassahn KS, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491:399–405.
32. Kamal M, Shaaban AM, Zhang L, et al. Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. Breast Cancer Res Treat 2010; 121:555–563.
33. Adamsen BL, Kravik KL, Clausen OP, et al. Apoptosis, cell cycle progression and gene expression in TP53-depleted HCT116 colon cancer cells in response to short-term 5-fluorouracil treatment. Int J Oncol 2007;31:1491–1500.
34. Bleeker FE, Lamba S, Rodolfo M, et al. Mutational profiling of cancer candidate genes in glioblastoma, melanoma and pancreatic carcinoma reveals a snapshot of their genomic landscapes. Hum Mutat 2009;30:E451–E459.
35. Wei P, Tang H, Li D. Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. PLoS One 2012;7:e46887.
36. Simeone DM, Pham T, Logsdon CD. Disruption of TGFbeta signaling pathways in human pancreatic cancer cells. Ann Surg 2000; 232:73–80.
37. Kim J, Reber HA, Dry SM, et al. Unfavourable prognosis associated with K-ras gene mutation in pancreatic cancer surgical margins. Gut 2006;55:1598–1605.

**Received August 13, 2012. Accepted July 29, 2013.**

*Reprint requests*

**Address requests for reprints to: Fergus J. Couch, PhD, Mayo Clinic, Stabile Building, 200 First Street SW, Rochester, Minnesota 55905. e-mail: fergus.couch@mayo.edu; fax: (507) 538-1937.**

*Acknowledgments*

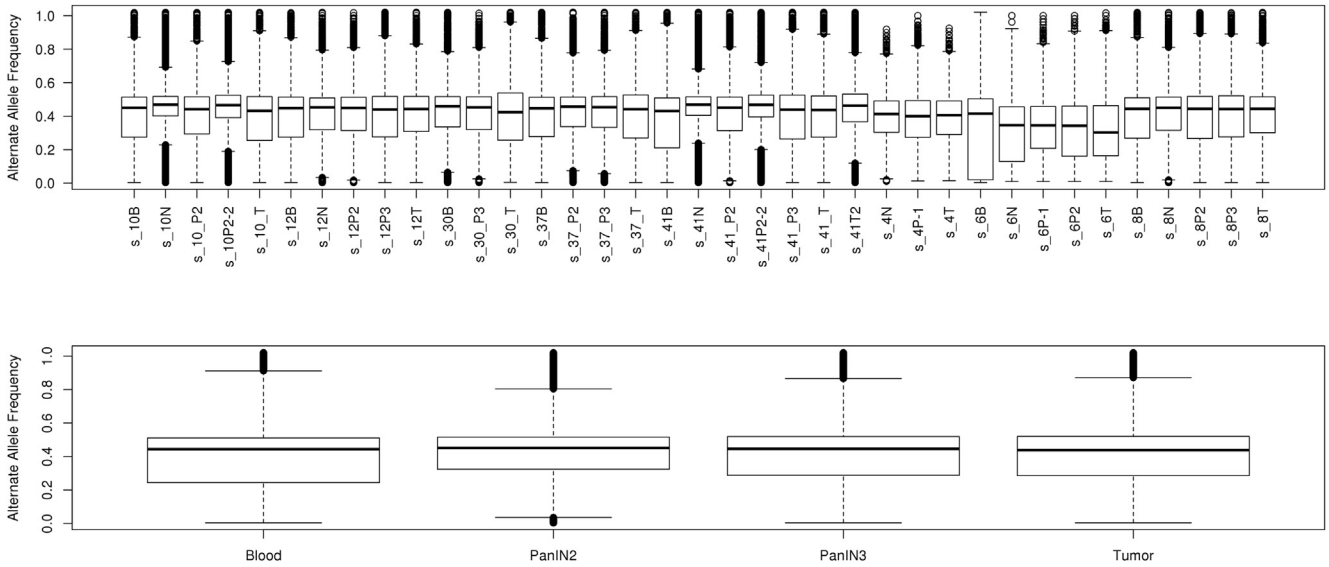
**The authors would like to acknowledge the contributions of Yean (Kit) Lee of the Mayo Medical Genome Facility for work on the exome capture and library (assembly protocols), and Donna Muzny, Jennifer Drummond, and David Wheeler of the Human Genome Sequencing Center at the Baylor College of Medicine for their help in initial exome sequencing.**

*Conflicts of interest*

**The authors disclose no conflicts.**

*Funding*

**Supported by a grant from The Lustgarten Foundation for Pancreatic Research, a National Institutes of Health Specialized Program of Research Excellence in Pancreatic Cancer (CA102701), the Mayo Clinic Center for Cell Signaling in Gastroenterology (NIDDK P30DK084567), the Minnesota Partnership for Biotechnology and Medical Genomics, and the Mayo Clinic Center for Individualized Medicine.**




**Supplementary Figure 1.** Alternative allele frequency ratios show that there is little bias for variant detection due to whole genome amplification (WGA). Top Panel: For each non-reference allele, the fractional composition of that variant is displayed. One would expect that heterozygous variants would be 50% reference and 50% alternate allele. Bottom Panel: Using the mean value from each of the samples, we show that there is no bias due to tissue histology.