Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma

Nemanja Rodić1, Jared P Steranka1, Alvin Makohon-Moore1,2, Allison Moyer1, Peilin Shen1, Reema Sharma1, Zachary A Kohutek3, Cheng Ran Huang4, Daniel Ahn1, Paolo Mita5,6, Martin S Taylor6, Norman J Barker1, Ralph H Hruban1,7,8, Christine A Iacobuzio-Donahue1,2,7,8, Jef D Boeke4–7,9 & Kathleen H Burns1,4,6,7

Pancreatic ductal adenocarcinoma (PDAC) is typically diagnosed after the disease has metastasized; it is among the most lethal forms of cancer. We recently described aberrant expression of an open reading frame 1 protein, ORF1p, encoded by long interspersed element-1 (LINE-1; L1) retrotransposon, in PDAC.1 To test whether LINE-1 expression leads to somatic insertions of this mobile DNA, we used a targeted method to sequence LINE-1 insertion sites in matched PDAC and normal samples. We found evidence of 465 somatic LINE-1 insertions in 20 PDAC genomes, which were absent from corresponding normal samples. In cases in which matched normal tissue, primary PDAC and metastatic disease sites were available, insertions were found in primary and metastatic tissues in differing proportions. Two adenocarcinomas secondarily involving the pancreas, but originating in the stomach and duodenum, acquired insertions with a similar discordance between primary and metastatic sites. Together, our findings show that LINE-1 contributes to the genetic evolution of PDAC and suggest that somatic insertions are acquired discontinuously in gastrointestinal neoplasms.

Pancreatic ductal adenocarcinoma (PDAC) affects about 270,000 people worldwide each year and is the fourth most common cause of cancer deaths in the United States. One- and five-year survival rates are 25% and 6%, respectively; most cases are locally advanced or metastatic by the time the disease is recognized clinically. Somatically acquired mutations are central to the development of PDAC. A typical case acquires 50 mutations in protein-coding sequences. These include a few causative (driver) mutations, like KRASG12D or KRASG12V, and many inconsequential (passenger) mutations.2,3 Most mutated genes occur while the neoplasm is still localized to the pancreas, before metastasis.4

Acquired structural alterations to PDAC genomes are less well characterized. In this study, we focused on long interspersed element-1 retrotransposition insertions. L1 is a mobile genetic sequence that comprises about 17% of the human genome. Most copies are invariant and not competent to transpose. However, some L1 sequences from the L1Hs or L1PA1 families are a significant source of inherited genetic variation and encode endonuclease and reverse-transcriptase activities that promote their propagation in genomes.5–7 We recently described expression of ORF1p in 89% of PDACs and 27% of pancreatic intraepithelial neoplasias (PanINs).1 In the current study, we tested whether somatic retrotransposition events would be evident in PDAC genomes.

Matched normal tissue and tumor samples were collected post-mortem. We extracted gDNA and used one-sided PCR to selectively amplify segments downstream of L1Hs insertions for sequencing, an approach we term transposon insertion profiling by sequencing (TIP-seq) (Fig. 1a). Similar approaches have been previously reported.8,9 Twenty-two individuals donated samples used in this study (Supplementary Table 1). All were initially diagnosed with PDAC; this condition was confirmed in 20 cases. Two tumors were reclassified as adenocarcinomas, likely originating in gastric or duodenal sites and secondarily involving the pancreas.

Genomic DNA was digested with restriction enzymes, ligated to vectorette oligonucleotides and amplified using a primer specific to the 3′ UTR of L1Hs repeats (Fig. 1b). Amplicons were sheared and sequenced to an average read count of 43 million, and sequences were aligned to the reference genome assembly. We identified most of the L1Hs in the reference genome in all samples (averaging 400 elements in the top 500 peaks, range 390–413; Supplementary Table 2). As expected, many known inherited polymorphisms were also seen.

After subtracting the known germline variants, read pile-ups (peaks) from normal tissue TIP-seq indicated novel constitutional L1Hs insertions. We examined these loci for potential roles in cancer risk.
Somatic LINE-1 insertions in PDAC. (a) Genomic positions are arrayed clockwise in a Circos plot. Positions of all somatic LINE-1 insertion sites are marked with radial hashes color-coded by case. Insertions found in the primary tumor are shown in the outermost wheel; insertions found in the metastasis are found in the middle wheel. Those shared by the two sites of disease are marked innermost. (b) Wheel shows insertions from a single case. Twenty-nine (29) LINE-1 insertions were identified in the primary tumor and 34 in the metastatic sample. Twenty-three (23) were shared between the two sites of disease. (c) LINE-1 ORF1p protein expression levels as scored by immunohistochemistry (0–3) corresponded to LINE-1 insertion events detected in primary PDAC tumors, though not to statistical significance (P = 0.07, one-way ANOVA). (d) Survival after diagnosis was inversely related to somatically acquired LINE-1 insertions found in the primary tumor (P = 0.025). Each hash on the y-axis indicates one month. (e) Survival after diagnosis was inversely related to LINE-1 ORF1p immunoreactivity (P = 0.03). Each hash on the y-axis indicates one month.

The expression level of the ORF1p LINE-1, as scored by immunohistochemistry, correlated with the number of somatically acquired LINE-1 insertions, albeit not to a statistically significant degree (P = 0.07). Survival after diagnosis was inversely correlated with somatically acquired LINE-1 insertions present in the primary tumor and with ORF1p immunoreactivity (P = 0.025 and P = 0.03, respectively) (Fig. 2c–e).

Two cases of adenocarcinoma that infiltrated the pancreas, but were later recognized to have originated in nearby stomach and small bowel, were also included. We found evidence for 51 and 63 somatic LINE-1 insertions in these PDAC mimics.

TIP-seq reads cover only the 3′ end of L1Hs insertions. To validate and sequence the entirety of somatically acquired insertions, we designed PCR primers for regions flanking predicted insertion sites. We Sanger sequenced amplicons that were larger than those corresponding to the pre-insertion, or 'empty', site present in tumor and absent from normal tissue. In total, 117 loci were considered; 92 were in a region that allowed primer design and had short-read sequences consistent with a LINE-1 insertion by manual review. Of these, 81 insertions (88%) could be PCR validated; for 75 (81%), we

No exonic LINE-1 could be sequence verified. No insertions were seen in the familial pancreatic cancer genes, breast cancer 2, early onset (BRCA2)10; partner and localizer of BRCA2 (PALB2)11; protease, Ser1 (PRSS1)12; or cyclin-dependent kinase inhibitor 2 (CDKN2)13.

To identify somatically acquired insertions, we compared TIP-seq profiles of matched tumor and normal gDNA samples. We found a total of 268 cancer-specific insertions in 18 cases of primary PDAC (Fig. 2a, Supplementary Tables 3 and 4). There was substantial variability between cases in the numbers of somatic insertions identified (range 0–65; average 15 insertions per case).

To identify somatic insertions at sites of metastatic PDAC, we assayed 15 metastases in 15 cases and found 242 insertion candidates that were not present in the corresponding normal tissue samples (range 0–42, average 16 insertions per site). In 13 cases with both primary and metastatic cancers available, we found 45 insertions shared between the two sites of disease. One case harbored 23 of these shared insertions (Figs. 2b and 3, Supplementary Tables 3 and 4).
**Figure 3** Retrotransposition (RT) events. (a) Somatic LINE-1 insertions. Stacked bars show insertions found in only the primary tumor (orange, bottom); those shared by the primary and metastatic samples (middle); and those only in the metastasis (red). Confirmed PDAC cases are shown in colors indicated by the legend; two other gastrointestinal adenocarcinomas are shown in lighter hues. (b) RT rate (insertions per 1,000 cell doublings) was calculated for each case during two phases of clonal evolution. The progenitor lineage rate (orange, x-axis) is proportional to the number of LINE-1 insertions acquired within a cellular lineage in the primary tumor antedating its seeding a metastatic site. These are shared by the primary and metastatic sites. The progenitor lineage rate (red, y-axis) reflects insertions found only in the metastatic sample. PDAC cases are indicated with black dots; other adenocarcinomas are in gray. (c) Soma-tically acquired LINE-1 insertions in PDAC. Twenty insertions were detected in the primary and metastatic samples for this case and not in DNA from normal samples. Gene names are listed for eight intronic insertions. The progenitor (orange) and progressor (red) lineages are shown; ratios are on the right. (e) LINE-1 insertions in a gastric adenocarcinoma. Sixty-four insertions were detected in the primary tumor; most (54) were shared with the metastatic site. (f) LINE-1 insertions in a duodenal adenocarcinoma. Twenty-seven insertions were identified in the primary tumor and 42 in the metastasis. Eighteen were shared between the two sites.

We next used somatic LINE-1 insertion numbers to infer rates of retrotransposition during the development and dissemination of PDAC. For this, we applied a model described for this disease that relates mutation rate to the rate of cell division, numbers of acquired neutral mutations and time. We expect that insertions shared between primary and metastatic samples of a case arose in the progenitor lineage of the primary tumor. These probably occurred after a genetic growth advantage was acquired by the primary tumor, but before a cell existed there with all of the insertions that would be carried to the metastatic site. Six cases, all of which had somatic LINE-1 insertions shared in primary and metastatic tissues, were considered. In aggregate, 259 independent insertions were found in these cases, with 45 LINE-1 insertions in the metastatic sites but not the corresponding primary tumors had occurred (average, 7.5), reflecting retrotransposition rates in progenitor lineages ranging from 0.5 to 12.4 insertions per 1,000 cell doublings (average, 4.0).

In contrast, we expected that the 82 insertions identified at metastatic sites but not the corresponding primary tumors had occurred later than had shared insertions. To test this in one case, we attempted to amplify eight of these insertions in six additional metastatic lesions. Four insertions (4 of 8) were uniquely amplified in the single metastasis (Fig. 3c). Four were found in two additional metastases (totaling 3 of 7 lesions). The latter insertions were acquired in the primary tumor in a subclone that would seed multiple metastases, though as a comparably late event (i.e., subsequent to the birth of the progenitor clone). None were detected in the primary tumor. Overall, these findings
Effects of somatic LINE-1 insertions in PDAC. (a) A landscape plot showing the human genome. Insertion sites are indicated as raised peaks; areas accumulating greater numbers of independent insertions are reflected by peak height and color. The pattern is distributed; highly recurrently mutated loci do not stand out. Areas that accrued the most LINE-1 insertions are marked. Somatic LINE-1 insertions occurred over a 17-Mb region on 4q32.2 in six cases. The nearest gene is follistatin-like 5 isoform (FSTL5), a putative tumor suppressor gene infrequently mutated in pancreatic cancer. None of the insertions were within the FSTL5 transcription unit. A 19-Mb interval on 5q33.2–5q35.2 showed somatic L1 insertions in 11 cases. An 11-Mb interval spanning 8q24.21–8q24.23 showed insertions in six cases, with some cases showing multiple independent insertions. Seven cases had evidence for insertions over a 6-Mb interval on 13q21.1–13q21.31. Insertions were broadly distributed within each of these regions, with no single gene locus recurrently involved. Peaks at 16p11.2 were precisely superimposed but, in satellite DNA, could not be PCR validated, and are thus probably artifact. (b) The schematic illustrates an intronic LINE-1 insertion in apoptotic protease activating factor 1 (APAF1). The L1 is a 443-bp insert with an inverted and truncated 5′ end; it is flanked by 16-bp TSDs. (c) An immunohistochemistry slide (20×) shows two regions (magnified in d and e) of the primary tumor with differential expression of APAF1 protein. Immunoreactivity is indicated in brown; the counterstain is blue. The tumor shows higher APAF1 expression in one sector (upper right inset) with an inverted and truncated 5′ insertion is probably dispensable for tumorigenesis.

validate the use of these insertions to estimate retrotransposition rates in what is termed the progressor clone. This ranges from 4.6 to 30 insertions per 1,000 cell doublings (average, 12.7).

Interestingly, within each PDAC case, we found no concordance between retrotransposition rates for these two phases of disease (Fig. 3b). This is also true of primary and metastatic diseases originating in the tubular gastrointestinal tract, assuming a similar pace of disease spread. The gastric adenocarcinoma case had many shared insertions (Fig. 3e) and disproportionately acquired these in the progenitor lineage of the primary tumor; its metastatic site had little evidence of continued retrotransposition. Conversely, the duodenal tumor accrued few somatic LINE-1 insertions in the progenitor lineage but produced a progressor clone that acquired many new insertions (Fig. 3f). These data suggest to us that retrotransposition occurs discontinuously during tumor evolution.

To test for concordance between acquired single-nucleotide changes and LINE-1 insertions, we compared proportions of shared (i.e., primary and metastasis) and metastasis-only alterations in a single PDAC case. Single-nucleotide mutations were detected by exome sequencing. For single-nucleotide mutations, we found a ratio of shared to metastasis-only mutations of approximately 60:40 (61:39). This is similar to ratios previously reported. The ratio is essentially reversed for LINE-1 insertions in the same PDAC case (43:57) (Fig. 3d). The discrepancy in the proportions of single-nucleotide mutations and LINE-1 insertions found provides additional evidence of asynchrony in the rates of these two mutational mechanisms.

Somatic LINE-1 insertions accumulate throughout PDAC genomes. Although loosely aggregated on individual chromosomes in some cases, hotspots for integration were not readily appreciated when all sites were considered (Fig. 4a). No exonic insertions were found.

Intronic insertions were seen in 202 genes. Gene set enrichment analysis does not clearly implicate specific pathways required for tumor development. Together, these data indicate that LINE-1 insertions are not commonly canonical ‘driver’ mutations and that any one insertion is probably dispensable for tumorigenesis.

To explore the possibility that insertions influence gene expression, we prioritized insertions that occurred in close proximity to exons or potential regulatory sites as indicated by ENCODE project annotations. One example is a somatic LINE-1 insertion that we detected in an intron of apoptotic protease activating factor 1 (APAF1); it is a 443-bp insertion with an inverted and truncated 5′ end. The L1 is located in a small intron (1.6 kb) of the APAF1 gene, adjacent to a DNase-hypersensitive region that can be bound by transcription factors. We used PCR to detect the insertion in the primary tumor and in all seven metastatic disease sites sampled in the same case. Immunohistochemistry to evaluate APAF1 protein showed a variegated pattern of expression in the primary tumor sample in this case (Fig. 4b–e), whereas other tumors displayed uniform APAF1 immunoreactivity. Repeated attempts to amplify the insertion from microdissected sectors of the tissue were unsuccessful. Our findings are consistent with an acquired event altering APAF1 expression in this case. However, they do not exclude other underlying mechanisms nor imply that this is a common feature in pancreatic cancers.

To date, targeted studies have demonstrated LINE-1 retrotransposition in lung cancer8, colon cancer18 and hepatocellular carcinoma19, and whole-genome sequencing from the Cancer Genome Atlas (TCGA)20,21 and the International Cancer Genome Consortium (ICGC)22 has shown more broadly that human malignancies of various origins are characterized by different levels of L1 activity. Our work adds pancreatic cancer to this picture. We show that PDAC is highly active for retrotransposition and, moreover, that single lineages acquire multiple LINE-1 integrations over their evolution.
The variability in numbers of acquired insertions between PDAC cases is similar to that previously described for other types of tumors. We have ascribed this to differences in each person’s inherited complement of active LINE-1 or to tumor-specific idiosyncrasies of retrotransposon control, but we expected new insertions to be acquired steadily during the course of disease. Our discovery of discordant overall rates of retrotransposition in matched progenitor and progressor subclones challenges this assumption. Similarly, in cases for which single-nucleotide mutation data are available, we find a lack of concordance between proportions of single-nucleotide mutations and LINE-1 insertions occurring in different phases of disease. Our findings suggest that LINE-1 insertions in gastrointestinal-tract cancers occur discontinuously. This is consistent with variable activity of individual LINE-1 elements previously reported in prostate cancer and lung cancer. Cellular mechanisms restricting somatic retrotransposition are not well understood, and perhaps the discontinuity that we see in the timing of insertions is a result of discrete breaches in host defenses. Tumors with many insertions could have acquired these in a series of breaches or even in isolated catastrophic episodes that can be likened to chromothripsis in cancer genomes.

Although many LINE-1 insertions can occur in metastatic progenitor lineages, they seem to be dispensable for PDAC. Some cases had no evidence of retrotransposition, and we have no evidence of recurrent insertion sites or exonic insertions. Although we cannot definitively assign a function to acquired LINE-1 insertions, our findings leave open the possibility that somatically acquired LINE-1 insertions influence gene regulation and contribute to cancer-cell phenotypes. Emerging technologies, such as single cell LINE-1 insertion site sequencing and phenotyping, should yield unprecedented opportunities to understand the timing of LINE-1 insertion events and how they contribute to genetic and phenotypic heterogeneity in cancer.

METHODS
Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
This work was started by funding from the Sol Goldman Pancreatic Cancer Research Center (K.H.B. and N.R.) and supported also by the Fred and Janet Sanfilippo Fund in the Department of Pathology at the Johns Hopkins University School of Medicine (N.R.); a Burroughs Wellcome Fund Career Award for Biomedical Scientists Program (K.H.B.); and US National Institutes of Health awards F31CA180682 (A.M.-M.), R01CA163705 (K.H.B.), R01GM103999 (K.H.B.), P50CA62924 (R.H.H. and C.A.I.-D.), R01CA179991 (C.A.I.-D.), as well as the National Institute of General Medical Sciences Center for Systems Biology of Retrotransposition P50GM107632 (K.H.B. and J.D.B.). Computational resources were provided through the National Science Foundation–funded MRI-R2 project PDBI-0959894. The authors would like to thank H. Kazazian, S. Solyom and A. Ewing for their helpful discussion. This work is dedicated to Dr. Frank Kretzer.

AUTHOR CONTRIBUTIONS
N.R., R.H.H., C.A.I.-D., J.D.B. and K.H.B. conceived of the project; N.R., A.M.-M. and C.A.I.-D. obtained tissues and reviewed histology; J.P.S., A.M., P.S. and F.M. designed and performed molecular-biology assays; N.R., R.S., M.S.T. and N.J.B. performed and reviewed immunostains; Z.A.K., C.R.H. and D.A. designed and performed sequence analysis; N.R., J.P.S., J.D.B. and K.H.B. interpreted data; J.P.S. summarized data for the supplementary tables; N.R. and K.H.B. wrote the manuscript. All authors contributed edits and approved of the final manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Rödil, N. et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am. J. Pathol. 184, 1280–1286 (2014).
2. Vogelstein, B. et al. Cancer genome landscapes. Science 339, 1546–1558 (2013).
3. Jones, S. et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801–1806 (2008).
4. Yachida, S. et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467, 1114–1117 (2010).
5. Burns, K. & Boeke, J.D. Human transposon tectonics. Cell 149, 740–752 (2012).
6. Beck, C.R., Garcia-Perez, J.L., Badge, R.M. & Moran, J.V. LINE-1 elements in structural variation and disease. Annu. Rev. Genomics Hum. Genet. 12, 187–215 (2011).
7. Hancks, D.C. & Kazazian, H.H. Jr. Active human retrotransposons: variation and disease. Curr. Opin. Genet. Dev. 22, 191–203 (2012).
8. Iskow, R.C. et al. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141, 1253–1261 (2010).
9. Ewing, A.D. & Kazazian, H.H. Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 20, 1262–1270 (2010).
10. Murphy, K.M. et al. Evaluation of candidate genes MAP2K4, MADH4, ACRVB1, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res. 62, 3789–3793 (2002).
11. Jones, S. et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 324, 217 (2009).
12. Whitcomb, D.C. et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat. Genet. 14, 141–145 (1996).
13. Whelan, A.J., Bartsch, D. & Goodfellow, P.J. Brief report: a familial syndrome of pancreatic cancer and melanoma with a mutation in the CDKN2A tumor-suppressor gene. N. Engl. J. Med. 333, 975–977 (1995).
14. Ostertag, E.M. & Kazazian, H.H. Jr. Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. 35, 501–538 (2001).
15. Ostertag, E.M. & Kazazian, H.H. Jr. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res. 11, 2059–2065 (2001).
16. Goodier, J.L., Ostertag, E.M. & Kazazian, H.H. Jr. Transduction of 3’-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. 9, 653–657 (2000).
17. Beck, C.R. et al. LINE-1 retrotransposition activity in human genomes. Cell 141, 1159–1170 (2010).
18. Solyom, S. et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 22, 2328–2338 (2012).
19. Shukla, R. et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 153, 101–111 (2013).
20. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. Science 337, 967–971 (2012).
21. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. 24, 1053–1063 (2014).
22. Tubio, J.M. et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science 345, 1251343 (2014).
23. Stephens, P.J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144, 27–40 (2011).
24. Ewry, G.D. et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell 151, 483–496 (2012).
ONLINE METHODS

Tissue samples. All participants were consented as part of the Gastrointestinal Cancer Rapid Medical Donations Program (GICRMDP) at the Johns Hopkins Hospital. The study was approved by the Johns Hopkins School of Medicine Institutional Review Board (IRB); protocol numbers are J03103 and NA_00036610. Each specimen was examined as a frozen section by a trained pathologist to ensure that (i) no cancer cells were identified in grossly normal tissue fragments and (ii) that viable cancer cells were present in the cancer specimens. K-ras mutation and immunohistochemistry (Cdx2, CK7, CK20) were used to identify two of the 22 cases that are more appropriately classified as adenocarcinomas originating in the nearby tubular gastrointestinal tract and secondarily involving the pancreas. Age at diagnosis for confirmed PDAC patients ranged from 36–85 years; average age was 65.

We incubated 30–40 mg of frozen tissue in a lysis buffer with RNase A (Life Technologies; Grand Island, NY) before an overnight proteinase K (NEB; Ipswich, MA) digestion. We extracted high molecular weight genomic DNA using phenol-chloroform and precipitated the DNA with ethanol. Quality was verified by spectrophotometry and agarose gel electrophoresis.

TIP-seq. Transposon insertion profiling (TIP-seq) uses a ligation mediated PCR to amplify genomic DNA 3′ of LINE-1 insertions25. For each sample, we split 10 μg of genomic DNA into six parallel restriction enzyme digests: AseI, BspHI, BstYI, HindIII, NcoI, PstI (NEB). We ligated vector backbone oligonucleotide adaptors designed with sticky-ends to match each restriction-enzyme–cut site to the digested DNA fragments. Touchdown PCR was run using ExTaq HS polymerase (Takara Bio; Shiga, Japan). An L1PA1-specific primer (5′-TACCAAATACTGACTGCTG-3′) initiates extension through the LINE-1 3′ UTR and poly(A) tail into adjacent DNA. The 3′-most ACA trinucleotide is selective for the youngest, active L1 subfamily (L1PA1).

We combined six PCR reactions for each sample and purified the DNA for sequencing library preparation. Using a Covaris E210 (Covaris; Woburn, MA) we sheared amplicons to an average size of 300 bp. We then performed end-repair, dA-tailing and index-specific adaptor ligation steps according to Illumina's TruSeq DNA Sample Prep v2 kit protocol (Illumina; San Diego, CA). Using 2% Size-Select E-gels (Life Technologies; Carlsbad, CA), we size-selected our adaptor-ligated DNA at approximately 450 bp before performing a final PCR amplification. After purifying the PCR amplified libraries, we submitted our adaptor-ligated DNA at approximately 450 bp before performing a final PCR amplification. Using a Covaris E210 (Covaris; Woburn, MA) we sheared amplicons to an average size of 300 bp. We then performed end-repair, dA-tailing and index-specific adaptor ligation steps according to Illumina's TruSeq DNA Sample Prep v2 kit protocol (Illumina; San Diego, CA). Using 2% Size-Select E-gels (Life Technologies; Carlsbad, CA), we size-selected our adaptor-ligated DNA at approximately 450 bp before performing a final PCR amplification. After purifying the PCR amplified libraries, we submitted them for quality control and Illumina HiSeq 100-bp paired-end sequencing at Hudson Alpha Institute for Biotechnology (HudsonAlpha; Huntsville, AL). Similar methods for LINE-1 insertion site amplification for next-generation sequencing have been described previously26,28.

Peak calling and filtering. We aligned demultiplexed sequencing reads to the human reference genome assembly (March 2006 human reference sequence NCBI Build 36.1, hg18) using BOWTIE (ref. 26) and called peaks using CisGenome27. Consensus sequences of the insertion sites were created by multi-sequence alignment of initially unaligned reads which, when trimmed to 35 bp, became alignable to a peak interval. We used Galaxy to identify differences in putative insertion sites across samples. We required peaks to rank in the top 1,000; insertions predicted by lower-ranking peaks were less commonly validated in subsequent PCRs. Known L1Hs (including polymorphic L1Hs) and L1PA2 insertions were removed. Consensus sequences were required to have an ‘A’ or ‘T’ homopolymer and sequence that could be unambiguously aligned with the reference assembly.

PCR validations. To validate candidate somatic insertions, we manually reviewed insertion-site consensus sequences and their alignment to the reference assembly by BLAT. We designed primers around the insertion-site coordinate using Primer3. Long-range spanning PCR was performed using LA-Taq (Takara Bio; Shiga, Japan) on both normal and tumor DNA, and amplicons were run on agarose gels. When we observed an amplicon that was larger than the ‘empty’ allele present only in tumor DNA, we excised the band and extracted the DNA for Sanger sequencing. LINE-1 length, target site duplications (TSD) and 5′ inverted structures are reported from the Sanger reads.

Retrotransposition rate estimates. We used a previously described model for clonal evolution in PDAC (ref. 4). This model assumes a constant mutation rate summarized by the following relationship:

$$r = \frac{T_{\text{gen}}}{T_1} (N_1 \pm \sqrt{N_1})$$

where $r$ is the mutation rate per generation; $T_{\text{gen}}$ is the cellular doubling time, which we took as 2.3 days24; $T_1$ is the time interval for the development of the metastatic progenitor cell in the primary tumor; and $N_1$ is the number of LINE-1 insertions detected that were shared by the primary tumor and the metastatic sample. An analogous relationship was used to estimate retrotransposition rates in the progressor subclone, with $T_2$ reflecting time for continued retrotransposition and $N_2$ signifying LINE-1 insertions detected uniquely in the metastatic sample. $T_1$ and $T_2$ were taken as 11.7 years and 6.8 years, respectively4.

Gene expression and gene set enrichment analysis. We overlapped peak intervals corresponding to somatic LINE-1 insertions ± 1 kb with RefSeq transcription units using Galaxy. This identified 248 unique annotations, 202 of which were Entrez genes. We performed gene set enrichment studies using the Broad Institute website29,30.

Immunohistochemistry. APAFI. Slides were deparaffinized and heated at 98 °C for 20 min in citrate buffer (pH 6) for antigen retrieval. Endogenous peroxidase was blocked using 3% hydrogen peroxide for 5 min at room temperature. A rabbit polyclonal antibody (Cat # HPA031373, Sigma-Aldrich) was used as the primary antibody at 1:60 for 30 min. The reaction was developed with a biotin-free Bond polymer detection system on a Bond-Link autostainer (Leica Microsystems, Bannockburn, IL); the 3,3-Diaminobenzidine chromogen-substrate was used for detection. Slides were counterstained with hematoxylin, dehydrated and coverslipped.

LINE-1 ORF1p. Slides were deparaffinized and heated at 98 °C for 20 min in citrate buffer (pH 6) for antigen retrieval. After blocking, we incubated with a 1:4,000 dilution of a monoclonal LINE-1 ORF1p antibody targeting amino acids 35 to 44 of LINE-1 ORF1p (ref. 1) overnight in a 4 °C humid chamber. Slides were then processed using the Histostain-SP Broad Spectrum kit according to the manufacturer’s instructions (Cat # 95-9643, Life Technologies).

25. Huang, C.R. et al. Mobile interspersed repeats are major structural variants in the human genome. Cell 141, 1171–1182 (2010).
26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
27. Ji, H. et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat. Biotechnol. 26, 1293–1300 (2008).
28. Amikura, K., Kobari, M. & Matsuno, S. The time of occurrence of liver metastasis in carcinoma of the pancreas. Pancreatol. 17, 139–146 (1995).
29. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005).
30. Mootha, V.K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267–273 (2003).