



Published in final edited form as:

Nat Med. 2015 September ; 21(9): 1060–1064. doi:10.1038/nm.3919.

Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma

Nemanja Rodi¹, Jared P. Steranka¹, Alvin Makohon-Moore^{1,7}, Allison Moyer¹, Peilin Shen¹, Reema Sharma¹, Zachary A. Kohutek⁸, Cheng Ran Huang², Daniel Ahn¹, Paolo Mita^{3,4}, Martin S. Taylor⁴, Norman J. Barker¹, Ralph H. Hruban^{1,5,6}, Christine A. Iacobuzio-Donahue^{1,5,6,7}, Jef D. Boeke^{2,3,4,5,9}, and Kathleen H. Burns^{*,1,2,4,5}

¹Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

²McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

³Department of Molecular Biology & Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

⁴High Throughput (HiT) Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA

⁵Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Hospital, Baltimore, MD, USA

⁶The Sol Goldman Center for Pancreatic Cancer Research, Johns Hopkins Medical Institutions, Baltimore, MD, USA

⁷Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

⁸Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

⁹Institute for Systems Genetics, New York University Langone School of Medicine, New York, NY, USA

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 a protein, ORF1p, encoded by Long INterspersed Element (LINE-1) retrotransposon, in PDAC¹. 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

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence Kathleen H. Burns, M.D., Ph.D., kburns@jhmi.edu.

Author Contributions

N. Rodic, R. Hruban, C. Iacobuzio-Donahue, J. Boeke, and K. Burns conceived of the project; N. Rodic, A.M.-M., and C. Iacobuzio-Donahue obtained tissues and reviewed histology; J. Steranka, A. Moyer, P. Shen, and P. Mita designed and performed molecular biology assays; N. Rodic, R. Sharma, M. Taylor, and N. Barker performed and reviewed immunostains; Z. Kohutek, C. R. Huang, and D. Ahn designed and performed sequence analysis; N. Rodic, J. Steranka, J. Boeke, and K. Burns interpreted data; J. Steranka summarized data for the Supplementary Tables; N. Rodic and K. Burns wrote the manuscript. All authors contributed edits and approved of the final manuscript.

in 20 PDAC genomes absent from corresponding normal samples. In cases where matched normal tissue, primary PDAC, and metastatic disease sites were available, insertions were found in primary and metastatic 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 when the disease is recognized clinically. Somatic 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 *KRAS*^{G12D} or *KRAS*^{G12V}, and many inconsequential (passenger) mutations^{2,3}. Most mutations occur while the neoplasm is localized to the pancreas, before metastasis⁴.

Acquired structural alterations to PDAC genomes are less well characterized. In this study, we focused on Long INterspersed element (LINE-1, L1) retrotransposon insertions. L1 is a mobile genetic sequence comprising about 17% of our genome. Most copies are invariant and not competent to transpose. However, some L1Hs or L1PA1 sequences are a significant source of inherited genetic variation and encode endonuclease and reverse transcriptase activities that promote their propagation in genomes^{5,6,7}. We recently described expression of a LINE-1 encoded protein, open reading frame 1 protein (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 and tumor samples were collected postmortem. 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 reported⁸⁹. Twenty-two individuals donated samples used in this study (Supplementary Table 1). All were initially diagnosed with PDAC; this was confirmed in 20 cases. Two 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 for the 3' UTR of L1Hs repeats (Fig. 1b). Amplicons were sheared and sequenced to an average read count of 43 million and sequences aligned to the reference genome assembly. We identified most 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, significant read pile-ups (peaks) from normal tissue TIP-seq indicated novel constitutional L1Hs insertions. We examined these locations for potential roles in cancer risk. 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 (*PRSSI*)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 primary PDAC cases (Fig. 2a, Supplementary Tables 3 and 4). There was significant variability between cases in the numbers of somatic insertions identified (range 0–65; average 15 insertions/case).

To identify somatic insertions at sites of metastatic PDAC, we assayed 15 metastases in 15 cases and found 242 insertion candidates not present in corresponding normal tissue samples (range 0–42, average 16 insertions/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 (Fig. 2b, Fig. 3a, Supplementary Tables 3 and 4).

LINE-1 ORF1p expression level scored by immunohistochemistry correlated with numbers of somatically acquired LINE-1 insertions albeit not to statistical significance ($p=0.07$). Survival after diagnosis was inversely correlated with somatically acquired LINE-1 insertions found in the primary tumor and with LINE-1 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 flanking predicted insertion sites. We Sanger sequenced amplicons larger than those corresponding to the preinsertion or 'empty' site present in tumor and absent from normal. In total, 117 loci were considered; 92 were in a region allowing primer design and had short-read sequences consistent with a LINE-1 insertion by manual review. Of these, 81 (88%) could be PCR validated; for 75 (81%), we recovered Sanger reads of the insertion. Validation rates were similar regardless of tissue of origin.

Structures of sequenced insertions are provided (Supplementary Table 5). These show expected hallmarks of somatic retrotransposition (Fig. 1c)¹⁴. All insertions are 5' truncated with an average size approximating 1 kilobase; the average target site duplication (TSD) is 11bp. Inverted or reversed orientation of the 5' L1Hs segment was commonly seen¹⁵ (Supplementary Figure 1, Supplementary Table 6). Transduction events of sufficient size to identify the templating element¹⁶ were recovered in two instances. Both were found in the same PDAC case and reflect activity of a known active element on 14q23.1¹⁷, which copied a 324bp segment 3' of its insertion site to positions on chromosomes 9 and 16 during development of the primary tumor (Supplementary Table 5).

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 relating mutation rate to the rate of cell division, numbers of acquired neutral

that LINE-1 insertions are not commonly canonical ‘driver’ mutations, and that any one insertion is likely dispensable for tumorigenesis.

To explore the possibility that insertions influence gene expression, we prioritized insertions in close proximity of exons or potential regulatory sites as indicated by ENCODE project annotations. One example is a somatic LINE-1 detected in an intron of apoptotic protease activating factor 1 (*APAF1*); it is a 443bp insertion with an inverted and truncated 5' end. The L1 is located in a small intron (1.6kb) 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), while 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 APAF-1 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 lung⁸, colon¹⁸, and hepatocellular¹⁹ carcinoma, 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, where 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 cancer²². Cellular mechanisms restricting somatic retrotransposition are not well understood, and perhaps the discontinuity we see in the timing of insertions is owed to discrete breaches in host defenses. Tumors with many insertions may have acquired these in series of breaches or even in isolated catastrophic episodes that can be likened to chromothripsis in cancer genomes²³.

Although many LINE-1 insertions can occur in metastatic progenitor lineages, they appear dispensable for PDAC. Some cases had no evidence of retrotransposition, and we have no evidence of recurrent insertion sites or exonic insertions. While we cannot definitively assign function to acquired LINE-1 insertions, our findings leave open the possibility that somatically acquired LINE-1 influence local gene regulation and contribute to cancer cell phenotypes. Emerging technologies such as single cell LINE-1 insertion site sequencing²⁴

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

Tissue samples

All participants were consented as part of the Gastrointestinal Cancer Rapid Medical Donation 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) prior to 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 insertions¹. For each sample, we split 10µg of genomic DNA into six parallel restriction enzymes digests: *AseI*, *BspHI*, *BstYI*, *HindIII*, *NcoI*, *PstI* (NEB). We ligated vectorette oligonucleotide adapters 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'-AGATATACCTAATGCTAGATGACACA-3') initiates extension through the LINE-1 3' UTR and polyA 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 300bp. We then performed end-repair, dA-tailing, and index-specific adapter 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 adapter-ligated DNA at approximately 450bp before performing a final PCR amplification. After purifying the PCR amplified libraries we submitted them for quality control and Illumina HiSeq 100bp paired-end sequencing at HudsonAlpha Institute for Biotechnology (HudsonAlpha; Huntsville, AL). Similar methods for LINE-1 insertion site amplification for next generation sequencing have been described previously^{2,3}.

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 BOWTIE⁴ and called peaks using CisGenome⁵. Consensus sequences of the insertion sites were created by multi-sequence alignment of initially unaligned reads which when trimmed to 35bp 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 1000; 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 unambiguously alignable to 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 LATaq (Takara Bio; Shiga, Japan) on both normal and tumor DNA and amplicons run on agarose gels. When we observed an amplicon 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 PDAC⁶. This model assumes a constant mutation rate summarized by the following relationship :

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

where r is the mutation rate per generation, T_{gen} is the cellular doubling time, which we took as 2.3 days⁷, 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 and 6.8 years, respectively⁶.

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 website^{8,9}.

Immunohistochemistry

APAF1 Slides were deparaffinized and heated at 98°C for 20 minutes citrate buffer (pH 6) for antigen retrieval. Endogenous peroxidase was blocked using 3% hydrogen peroxide for 5 minutes at room temperature. A rabbit polyclonal antibody (Cat # HPA031373, Sigma-Aldrich, St. Louis, MO) was used as the primary antibody at 1:60 for 30 minutes. The reaction was developed with a biotin-free Bond polymer detection system on a Bond-Leica 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 minutes citrate buffer (pH 6) for antigen retrieval. After blocking, we incubated with a 1:1000 dilution of a monoclonal LINE-1 ORF1p antibody targeting amino acids 35 to 44 of LINE-1 ORF1p¹⁰ 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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was started by funding from the Sol Goldman Pancreatic Cancer Research Center (KHB and NR) and supported also by the Fred and Janet Sanfilippo Fund in the Department of Pathology at the Johns Hopkins University School of Medicine (NR); a Burroughs Wellcome Fund Career Award for Biomedical Scientists Program (KHB); and NIH awards F31CA180682 (AM-M), R01CA163705 (KHB), R01GM103999 (KHB), P50CA62924 (RHH, CI-D), R01CA179991 (CI-D), as well as the NIGMS Center for Systems Biology of Retrotransposition P50GM107632 (KHB and JDB). Computational resources were provided through the National Science Foundation funded MRI-R2 project #DBI-0959894. The authors would like to thank Haig Kazazian, Szilvia Solyom, and Adam Ewing for helpful discussions. This work is dedicated to Dr. Frank Kretzer.

References

1. Rodi N, et al. Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American journal of pathology*. 2014
2. Vogelstein B, et al. Cancer genome landscapes. *Science*. 2013; 339:1546–1558. [PubMed: 23539594]
3. Jones S, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008; 321:1801–1806. [PubMed: 18772397]
4. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010; 467:1114–1117. [PubMed: 20981102]
5. Burns KH, Boeke JD. Human transposon tectonics. *Cell*. 2012; 149:740–752. [PubMed: 22579280]
6. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet*. 2011; 12:187–215. [PubMed: 21801021]
7. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev*. 2012; 22:191–203. [PubMed: 22406018]
8. Iskow RC, et al. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*. 2010; 141:1253–1261. [PubMed: 20603005]
9. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res*. 2010

10. Murphy KM, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. *Cancer Res.* 2002; 62:3789–3793. [PubMed: 12097290]
11. Jones S, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science.* 2009; 324:217. [PubMed: 19264984]
12. Whitcomb DC, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet.* 1996; 14:141–145. [PubMed: 8841182]
13. Whelan AJ, Bartsch D, Goodfellow PJ. Brief report: a familial syndrome of pancreatic cancer and melanoma with a mutation in the CDKN2 tumor-suppressor gene. *N Engl J Med.* 1995; 333:975–977. [PubMed: 7666917]
14. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet.* 2001; 35:501–538. [PubMed: 11700292]
15. Ostertag EM, Kazazian HH Jr. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* 2001; 11:2059–2065. [PubMed: 11731496]
16. Goodier JL, Ostertag EM, Kazazian HH Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet.* 2000; 9:653–657. [PubMed: 10699189]
17. Beck CR, et al. LINE-1 retrotransposition activity in human genomes. *Cell.* 2010; 141:1159–1170. [PubMed: 20602998]
18. Solyom S, et al. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 2012; 22:2328–2338. [PubMed: 22968929]
19. Shukla R, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell.* 2013; 153:101–111. [PubMed: 23540693]
20. Lee E, et al. Landscape of Somatic Retrotransposition in Human Cancers. *Science.* 2012; 337:967–971. [PubMed: 22745252]
21. Helman E, et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* 2014; 24:1053–1063. [PubMed: 24823667]
22. Tubio JM, et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science.* 2014; 345:1251343. [PubMed: 25082706]
23. Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell.* 2011; 144:27–40. [PubMed: 21215367]
24. Evrony GD, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell.* 2012; 151:483–496. [PubMed: 23101622]

Methods References

1. Huang CR, et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell.* 2010; 141:1171–1182. [PubMed: 20602999]
2. Iskow RC, et al. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell.* 2010; 141:1253–1261. [PubMed: 20603005]
3. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* 2010
4. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
5. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol.* 2008; 26:1293–1300. [PubMed: 18978777]
6. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature.* 2010; 467:1114–1117. [PubMed: 20981102]
7. Amikura K, Kobari M, Matsuno S. The time of occurrence of liver metastasis in carcinoma of the pancreas. *International journal of pancreatology : official journal of the International Association of Pancreatology.* 1995; 17:139–146. [PubMed: 7622937]
8. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102:15545–15550. [PubMed: 16199517]

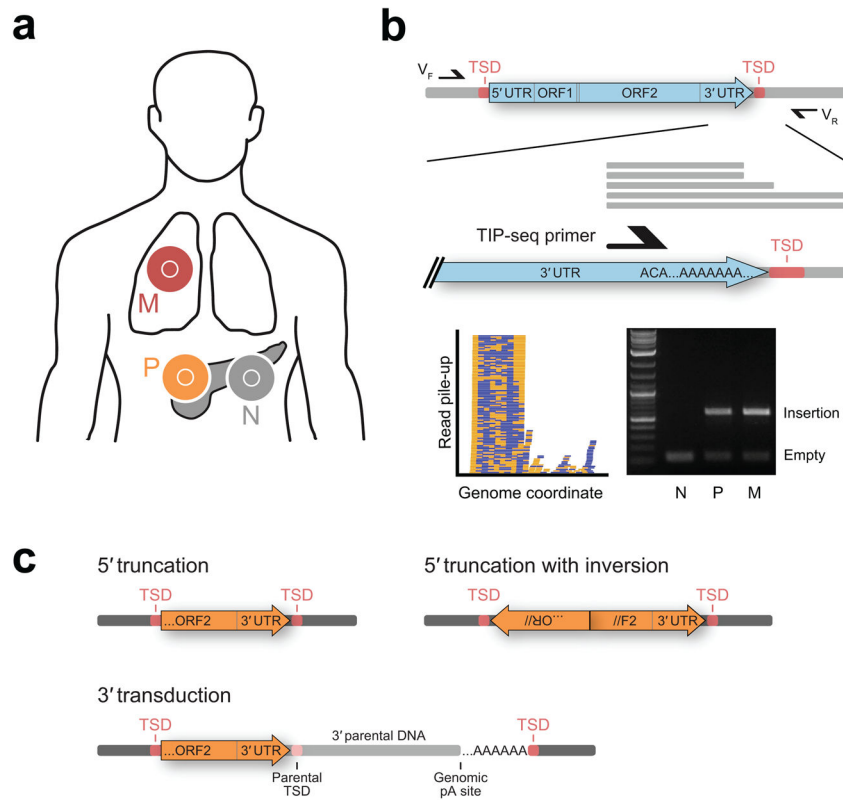
9. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003; 34:267–273. [PubMed: 12808457]
10. Rodi N, et al. Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American journal of pathology.* 2014

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Figure 1.**

Experimental approach. (a) Normal (N); primary (P) tumor samples; and metastatic (M) samples were collected. Sections were evaluated histologically to assess viability and composition prior to DNA extraction. (b) Transposon Insertion Profiling (TIP)-seq. A full-length (6kb) LINE-1 is diagrammed (blue). TSD, target site duplication; UTR, untranslated region; ORF, open reading frame. PCR amplicons initiate from an oligonucleotide complementary to the 3' UTR and extend to a restriction enzyme site. These are sheared and sequenced. Alignment to the reference genome shows read pileups (peaks, lower left). There are no alignments 5' of the insertion; reads align to the plus (blue) and minus (orange) strands 3' of the insertion. Peaks found in neoplastic samples and not matched normal samples indicate potential somatic insertions. Reads spanning the polyadenylated tail and unique DNA indicate the point of insertion. Insertions are validated by spanning PCR (lower right). The gel shows four lanes with a marker in the leftmost; amplicons from normal (N), primary (P) tumor, and metastasis (M) are shown. The “empty” or pre-insertion site is amplified in all samples, whereas the insertion is recovered from primary and metastatic disease sites. (c) Structure of LINE-1 insertions. LINE-1 is frequently 5' truncated (upper left). Insertions can have inverted 5' ends (upper right). The diagram shows a sequence breakpoint (//); the 3' LINE-1 sequence is in the regular orientation (rightward arrow), and 5' LINE-1 sequence is in the opposing orientation (leftward arrow). Finally, 3' transduction events incorporate DNA downstream of the parental LINE-1 at the new insertion site.

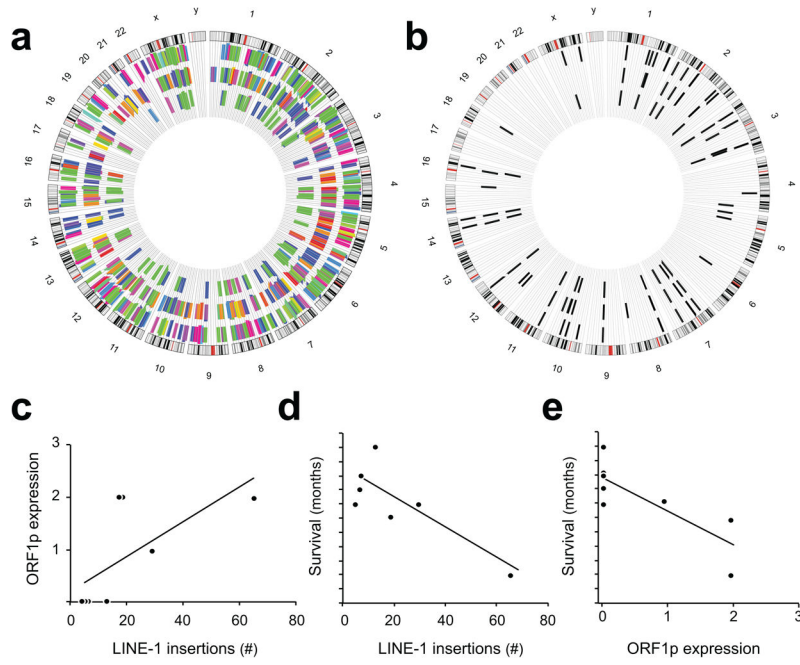
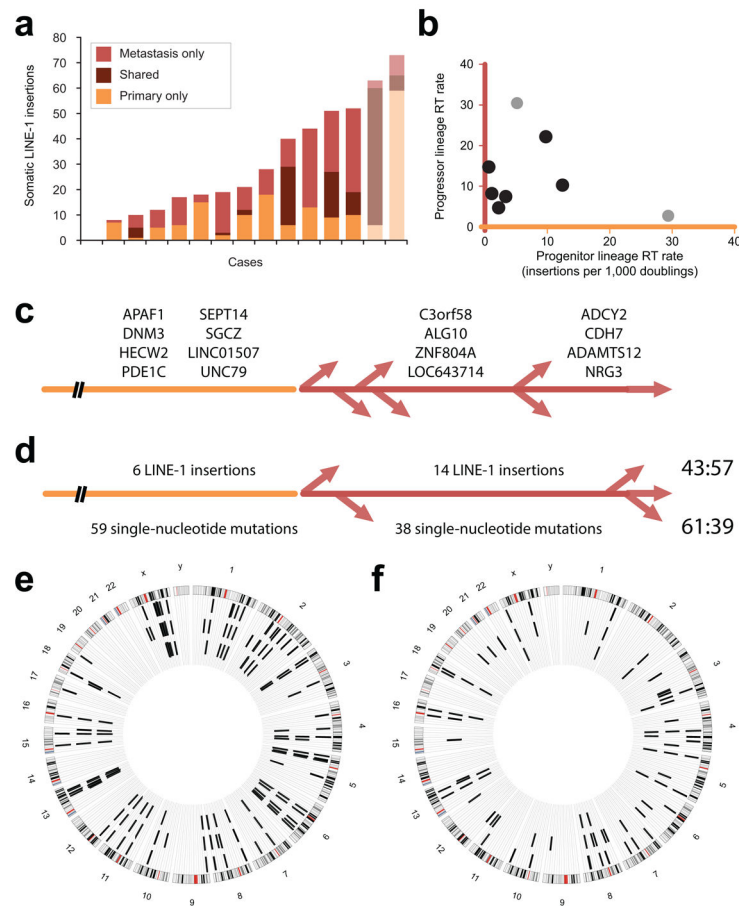


Figure 2.

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)** Image b. 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 ($p=0.07$, 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 ($p=0.03$).

**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 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 progressor lineage rate (red, y-axis) reflects insertions found only in the metastatic sample. PDAC cases are indicated with black dots; other adenocarcinomas are in grey. **(c)** Somatic LINE-1 in PDAC. Twenty insertions were detected in the primary and metastatic samples for this case and not in normal DNA. Gene names are listed for 8 intronic insertions. These occurred in the progenitor lineage of the primary tumor (leftmost, orange); they were PCR amplified from all metastases. Eight insertions were found by TIP-seq at a metastatic site and not in the primary tumor; nearest gene names are given for intergenic insertions. Four were present in two additional metastases and occurred in a progressor clone of the primary tumor (middle, red). The remaining four were unique to the metastasis profiled by TIP-seq (rightmost) and occurred later in the primary tumor or at that metastatic site. **(d)** Somatic LINE-1 and single-nucleotide mutations in PDAC. Numbers

of events occurring in the progenitor (orange) and progressor (red) lineages are shown; ratios are on the right. **(e)** LINE-1 insertions in a gastric adenocarcinoma. Sixty 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, and 42 in the metastasis. Eighteen were shared between the two sites.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

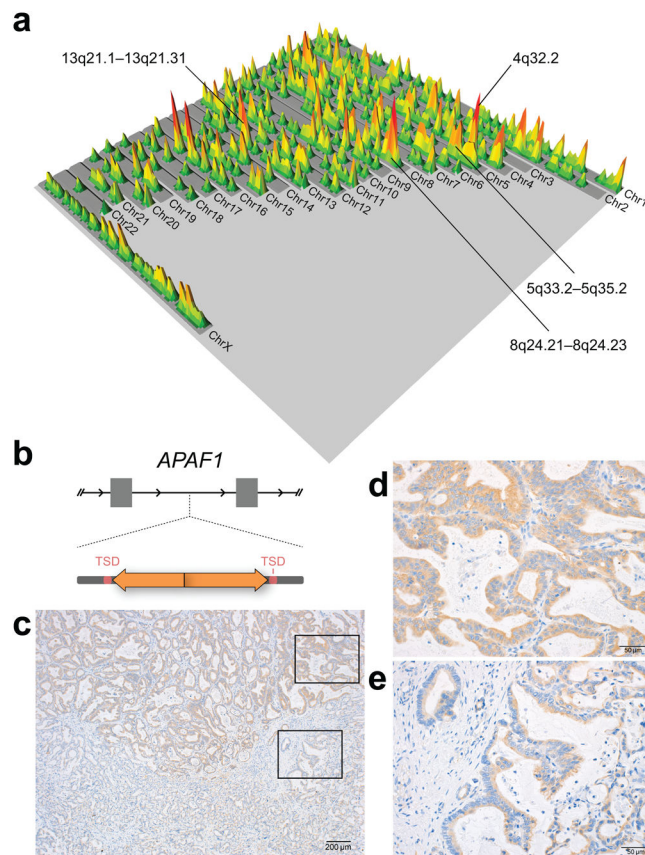


Figure 4.

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 more independent insertions are reflected by peak height and color. The pattern is distributed; highly recurrently mutated loci do not stand out. Areas accruing the most LINE-1 insertions are marked. Somatic LINE-1 insertions occurred over a 17MB region on 4q32.2 in 6 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 19MB interval on 5q33.2-5q35.2 showed somatic L1 insertions in 11 cases. An 11MB interval spanning 8q24.21-8q24.23 showed insertions in 6 cases with some cases showing multiple independent insertions. Seven cases had evidence for insertions over a 6MB 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 likely artifact. **(b)** The schematic illustrates an intronic LINE-1 in apoptotic protease activating factor 1 (*APAF1*). The L1 is a 443bp insert with an inverted and truncated 5' end; it is flanked by 16bp TSDs. **(c)** An immunohistochemistry slide (20x) shows two regions 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, **(d)** 100x magnification) and reduced immunoreactivity in a less

differentiated region (lower right inset, **(e)**). Other tumors showed uniform APAF1 expression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript