LETTER

doi:10.1038/nature09515

Distant metastasis occurs late during the genetic evolution of pancreatic cancer

Shinichi Yachida[1]*, Siân Jones[2]*, Ivana Bozic[3], Tibor Antal[3,4], Rebecca Leary[2], Baojin Fu[1], Mihoko Kamiyama[1], Ralph H. Hruban[1,5], James R. Eshleman[1], Martin A. Nowak[3], Victor E. Velculescu[2], Kenneth W. Kinzler[2], Bert Vogelstein[2] & Christine A. Iacobuzio-Donahue[1,5,6]

Metastasis, the dissemination and growth of neoplastic cells in an organ distinct from that in which they originated[1,2], is the most common cause of death in cancer patients. This is particularly true for pancreatic cancers, where most patients are diagnosed with metastatic disease and few show a sustained response to chemotherapy or radiation therapy[3]. Whether the dismal prognosis of patients with pancreatic cancer compared to patients with other types of cancer is a result of late diagnosis or early dissemination of disease to distant organs is not known. Here we rely on data generated by sequencing the genomes of seven pancreatic cancer metastases to evaluate the clonal relationships among primary and metastatic cancers. We find that clonal populations that give rise to distant metastases are represented within the primary carcinoma, but these clones are genetically evolved from the original parental, non-metastatic clone. Thus, genetic heterogeneity of metastases reflects that within the primary carcinoma. A quantitative analysis of the timing of the genetic evolution of pancreatic cancer was performed, indicating at least a decade between the occurrence of the initiating mutation and the birth of the parental, non-metastatic founder cell. At least five more years are required for the acquisition of metastatic ability and patients die an average of two years thereafter. These data provide novel insights into the genetic features underlying pancreatic cancer progression and define a broad time window of opportunity for early detection to prevent deaths from metastatic disease.

We performed rapid autopsies of seven individuals with end stage pancreatic cancer (Supplementary Table 1). In all patients, metastatic deposits were present within two or more anatomic sites in each patient, most often the liver, lung and peritoneum, as is typical for this form of neoplasia[4].

Low passage cell lines (six patients) or first passage xenografts (one patient) were created from one of the metastases present at each patient's autopsy. These samples comprised seven of the 24 pancreatic cancers which previously underwent whole exome sequencing and copy number analysis, as described in a mutational survey of the pancreatic cancer genome[5]. In this earlier study, a total of 426 somatic mutations in 388 different genes were identified among 220,884,033 base pairs (bp) sequenced in the seven index metastatic lesions, corresponding to an average of 61 mutations per index metastatic lesion (range 41–77). In all samples, the vast majority of mutations were represented by missense or silent single base substitutions (Supplementary Fig. 1 and Supplementary Table 2).

For each of the somatic mutations identified in the seven index metastasis lesions, we determined whether the same somatic mutation was present in anatomically distinct metastases harvested at autopsy from the same patients. We also determined whether these mutations were present in the primary pancreatic tumours from which the metastases arose. A small number of these samples of interest were cell lines or xenografts, similar to the index lesions, whereas the majority were fresh-frozen tissues that contained admixed neoplastic, stromal, inflammatory, endothelial and normal epithelial cells (Fig. 1a). Each tissue sample was therefore microdissected to minimize contaminating non-neoplastic elements before purifying DNA.

Two categories of mutations were identified (Fig. 1b). The first and largest category corresponded to those mutations present in all samples from a given patient ('founder' mutations, mean of 64%, range 48–83% of all mutations per patient; Fig. 1b, example in Supplementary Fig. 2a). These data indicate that the majority of somatically acquired mutations present in pancreatic cancers occur before the development of metastatic lesions. All other mutations were characterized as 'progressor' mutations (mean of 36%, range 17–52% of all mutations per patient;

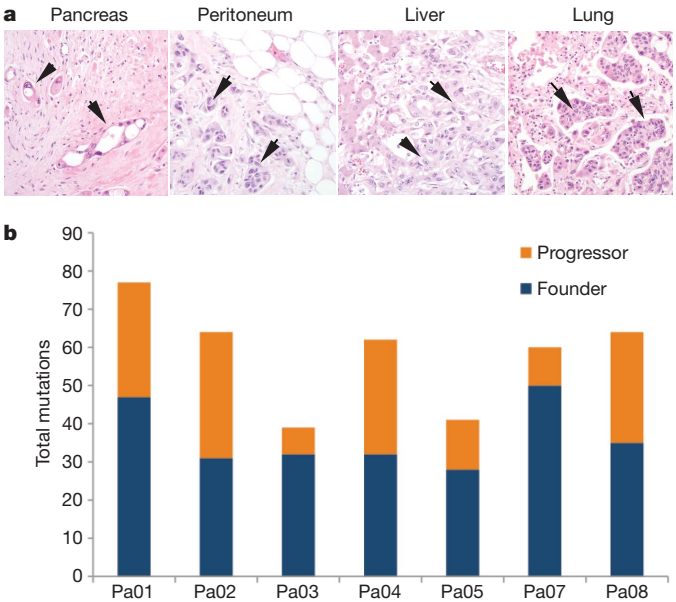

Figure 1 | Summary of somatic mutations in metastatic pancreatic cancers. **a**, Histopathology of primary infiltrating pancreatic cancer and metastatic pancreatic cancer to the peritoneum, liver and lung. In addition to infiltrating cancer cells in each lesion (arrows), non-neoplastic cell types are abundant. **b**, Total mutations representing parental clones (founder mutations), and clonal evolution (progressor mutations) within the primary carcinoma based on comparative lesion sequencing. Mutations common to all samples analysed were the most common category identified.

[1]Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. [2]The Ludwig Center for Cancer Genetics and Therapeutics and The Howard Hughes Medical Institute at The Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland 21231, USA. [3]Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]School of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, UK. [5]Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. [6]Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA.
*These authors contributed equally to this work.

Fig. 1b, example in Supplementary Fig. 2b). These mutations were present in one or more of the metastases examined, including the index metastasis, but not the parental clone.

These mutation types were used to classify the lesions that contained them into parental clones (containing only founder mutations) and subclones (containing both founder and progressor mutations). By definition, there could be only one parental clone in a patient, although there could be many different subclones. Parental clones tended to contain more deleterious mutations (nonsense, splice site or frameshift mutations) than subclones (12.6% of the mutations in the parental clones versus 8.1% of the mutations in subclones, Supplementary Table 2). The parental clones had already accumulated mutations in all driver genes (*KRAS*, *TP53* and *SMAD4*) previously shown to drive pancreatic tumorigenesis[6]. Through combined analysis of high-density single nucleotide polymorphism (SNP) chip data on the index lesion (Supplementary Table 3) plus the sequencing data on all lesions (Supplementary Table 2) we found that the vast majority of homozygous mutations (51 mutations, representing 89% of all homozygous mutations) in the index lesion were already present in the parental clones. Homozygous mutations are characteristic of tumour suppressor genes such as *SMAD4* and *CDKN2A* and often occur in association with chromosomal instability[7]. In sum, the parental clones harboured the majority of deleterious genetic alterations and chromosomal instability, upon which were superimposed an accumulation of progressor mutations associated with clonal evolution and metastasis.

Evolutionary maps were constructed for each patient's carcinoma based on the patterns of somatic mutation and allelic losses and the locations of individual metastatic deposits (Fig. 2 and Supplementary Figs 3–8). These maps showed that, despite the presence of numerous founder mutations within the parental clones, the cells giving rise to the metastatic lesions had a large number of progressor mutations. For example, in Pa01 the parental clone contained 49 founder mutations, yet a clonal expansion marked by the presence of mutations in six additional genes was present in the lung and peritoneal metastases (Supplementary Fig. 3). Moreover, 22 more mutations were found in the liver metastasis. Note that all mutations in the metastatic lesions were clonal, that is, present in the great majority if not all neoplastic cells of the metastasis, as assessed by Sanger sequencing. Thus, these mutations were present in the cell that clonally expanded to become the metastasis. Similarly, large numbers of progressor mutations were generally observed in the metastases from each of the seven cases examined (Fig. 2 and Supplementary Figs 3–8).

To distinguish between the possibilities that clonal evolution occurred inside the primary cancer versus within secondary sites, we sectioned the primary tumours from two patients into numerous, three-dimensionally organized pieces (Fig. 2a, b) and examined the DNA

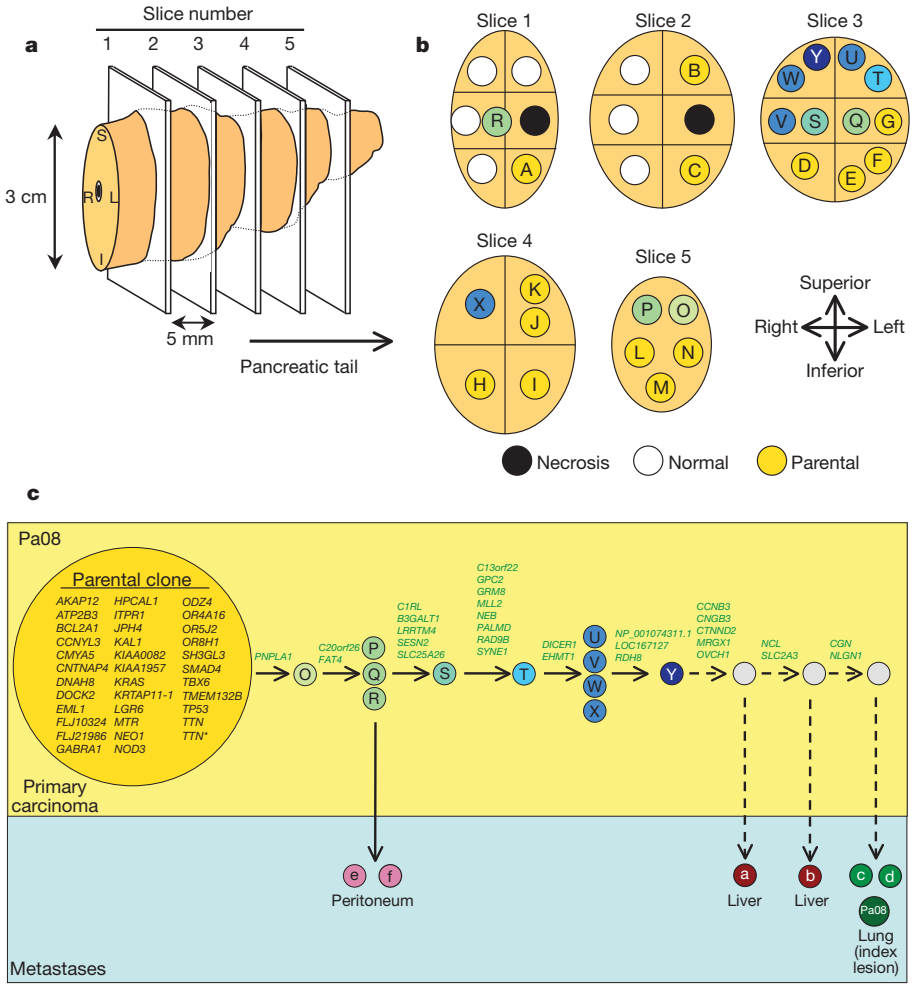

Figure 2 | Geographic mapping of metastatic clones within the primary carcinoma and proposed clonal evolution of Pa08. **a**, Illustration of the pancreatic specimen removed from Pa08 at rapid autopsy, and the planes of sectioning of the specimen. **b**, Mapping of the parental clone and subclones identified by comparative lesion sequencing within serial sections of the infiltrating pancreatic carcinoma. Metastatic subclones giving rise to liver and lung metastases are non-randomly located within slice 3, indicated by blue circles. These clones are both geographically and genetically distinct from clones giving rise to peritoneal metastases in this same patient, indicated in green. **c**, Proposed clonal evolution based on the sequencing data. In this model, after development of the parental clone, ongoing clonal evolution continues within the primary carcinoma (yellow rectangle), and these subclones seed metastases in distant sites. *Two mutations were found in the *TTN* gene.

from each piece for each of the founder and progressor mutations. In Patient Pa08, there were three progressor mutations present in two independent peritoneal metastases (defining one subclone) and 23, 25 or 27 additional progressor mutations present in liver and lung metastases (defining three additional subclones; Fig. 2c). Through the analysis of distinct regions of the primary tumour, it was clear that subclones giving rise to each of these metastases were present in the primary tumour. Moreover, these subclones were not small; from the size of the pieces (Fig. 2a) and the amounts of DNA recovered, each subclone must have contained in excess of 100 million cells. In addition, more than four different subclones, each containing a similarly large number of cells, could be identified through the analysis of other pieces of the same tumour. These subclones could be put into an ordered hierarchy establishing an evolutionary path for tumour progression (Fig. 2c). Analysis of multiple primary tumour pieces and metastatic lesions from patient Pa04 revealed a similar clonal evolution, with distinct, large subclones within the primary tumours giving rise to the various metastases (Supplementary Fig. 8).

To clarify further clonal evolution within the primary site, we attempted to correlate the mutation signatures representing the subclones of Pa08 (Fig. 2c) with the geographic location of the pieces of the primary tumour used to define them (Fig. 2a, b). Samples representative of the parental clone were located throughout the primary carcinoma. By contrast, samples representing subclones were non-randomly located in proximity to each other, within which the subclones specifically giving rise to peritoneal versus distant metastases were seen. Thus, we conclude that the genetic heterogeneity of metastases reflects heterogeneity already existing within the primary carcinoma, and that the primary carcinoma is a mixture of numerous subclones, each of which has independently expanded to constitute a large number of cells.

This data set could also be used to infer the timing of the development of the various stages of pancreatic tumour progression[8]. We assume that the tumour is initiated by a genetic event that confers a selective growth advantage to the cell that goes on to become the founder cell of the tumour. To estimate the timing, we first used Ki-67 labelling to determine the proliferation rate of seven samples of normal duct epithelium from surgically resected pancreata of individuals without pancreatic cancer as well as of each index metastasis. Ki-67-positive nuclei constituted an average of 0.4% of normal ductal cells, whereas an average of 16.3% of cancer cells within the index metastasis lesions were Ki-67-positive, consistent with prior estimates[9,10] (Supplementary Table 4). Based on these data plus that from sequencing of the index lesions, we derived estimates for three critical times in tumour evolution: T_1, the time between tumour initiation and the birth of the cell giving rise to the parental clone; T_2, the subsequent time required for the birth of the cell that gave rise to the index metastasis; and T_3, the time between the dissemination of this cell and the patients' death (Fig. 3). In other words, there is a time point, t_0, when the tumour was initiated, and a time point t_1 when a cell is born that has all mutations that exist in the parental clone. Similarly, there is a time point in tumour evolution, t_2, when a cell is born that has all the mutations that exist in the index metastasis. T_1 is given by $t_1 - t_0$ and T_2 is given by $t_2 - t_1$. If we denote t_3 as the time of patient's death, then $T_3 = t_3 - t_2$.

Using the mathematical model described in the Methods, we were able to conservatively estimate an average of 11.7 years from the initiation of tumorigenesis until the birth of the cell giving rise to the parental clone, an average of 6.8 years from then until the birth of the cell giving rise to the index lesion, and an average of 2.7 years from then until the patients' death (see Supplementary Discussion and Supplementary Table 5).

We show, for the first time, that primary pancreatic cancers contain a mix of geographically distinct subclones, each containing large numbers (hundreds of millions) of cells that are present within the primary tumour years before the metastases become clinically evident. The features of these metastatic subclones that promote metastasis formation have yet to be discerned, because no consistent genetic signature of metastatic subclones could be identified. We did identify several genes

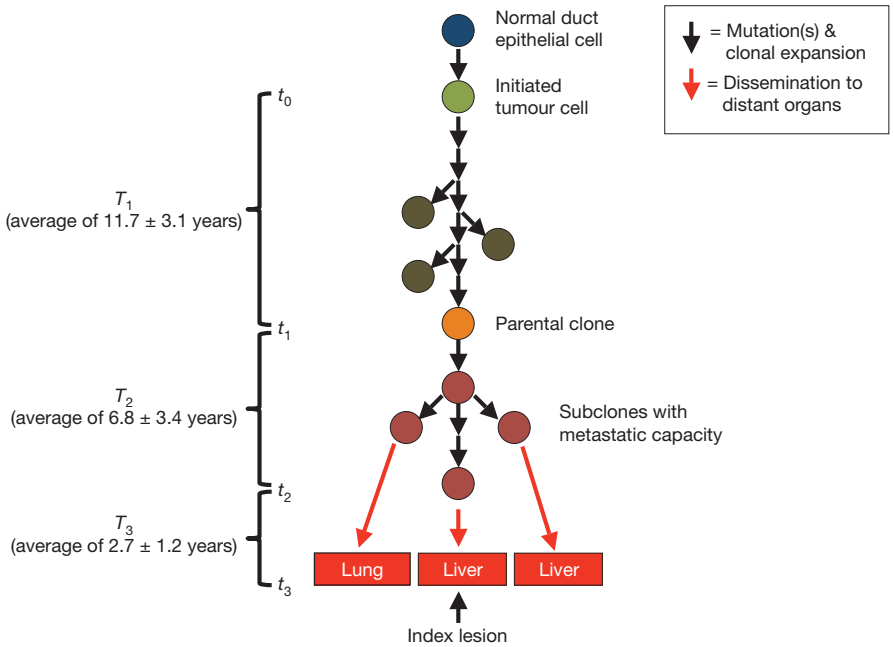

Figure 3 | Schema of the genetic evolution of pancreatic cancer. Tumorigenesis begins with an initiating mutation in a normal cell that confers a selective growth advantage. Successive waves of clonal expansion occur in association with the acquisition of additional mutations, corresponding to the progression model of pancreatic intraepithelial neoplasia (PanIN) and time T_1. One founder cell within a PanIN lesion will seed the parental clone and hence initiate an infiltrating carcinoma (end of T_1 and beginning of T_2). Eventually, the cell that will give rise to the index lesion will appear (end of T_2 and beginning of T_3). Unfortunately, most patients are not diagnosed until well into time interval T_3 when cells of these metastatic subclones have already escaped the pancreas and started to grow within distant organs. The average time for intervals T_1, T_2 and T_3 for all seven patients is indicated in the parentheses at left (see also Supplementary Table 6).

that were mutated in one or more of the index metastatic lesions from these seven patients with Stage IV disease, but not in the primary pancreatic index lesions from 17 patients with Stage II disease (Supplementary Table 2). These genes include those that may have a role in invasive or metastatic ability through heterotypic cell adhesion (*CNTN5*), motility (*DOCK2*), proteolysis (*MEP1A*) and tyrosine phosphorylation (*LMTK2*). However, these mutations were not metastasis-specific *per se* as all but one were present in the matched primary carcinoma of those same seven patients, and there is no evidence that the mutations we observed endowed these genes with metastagenic activity. These data also do not reveal the selective pressures within the primary carcinoma that led to the formation of progressor mutations. In light of recent findings indicating that pancreatic cancers are poorly vascularized[11], one possibility is that intratumoural hypoxia creates a fertile microenvironment for the formation of additional mutations beyond that of the parental clone.

One of the major implications of these data is their implication for screening to prevent pancreatic cancer deaths. Quantitative analysis indicated a large window of opportunity for diagnosis while the disease was still in the curative stage—at least a decade. Our model also predicts an average of 6.8 years between the birth of the cell giving rise to the parental clone and the seeding of the index metastasis. Unfortunately, the great majority of patients are not diagnosed until the last 2 years of the entire tumorigenic process. The challenge is to detect these tumours during time T_1, or even after T_1 but before seeding of metastases. Advanced imaging methods, as well as blood tests to detect cancer-specific proteins, transcripts or genes[12], offer hope for such non-invasive early detection.

METHODS SUMMARY

Rapid autopsies were performed on seven individuals with Stage IV pancreatic cancer[13]. Genomic DNA was extracted from cell lines or xenografts established from one metastasis of each patient and used for exomic sequencing as described previously[5]. The Illumina Infinium II Whole Genome Genotyping Assay using the BeadChip platform was also used to analyse each sample at 1,072,820 (1M) SNP loci as described previously[5]. Samples of snap-frozen pancreatic cancer tissue were microdissected using a PALM MicroLaser System (Carl Zeiss MicroImaging) and DNA extracted using QIAamp DNA Micro Kits (Qiagen). Genomic DNA was quantified by calculating long interspersed nuclear elements (LINE) by real-time PCR. Whole genome amplification (WGA) was performed using 10 ng total template DNA and an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). Ki-67 immunolabelling (Clone MIB-1, Dako Cytomation) was performed on formalin-fixed, paraffin-embedded sections of normal pancreatic ducts and metastatic pancreatic cancer tissues for each patient using the Ventana Discovery staining system (Ventana Medical Systems), and this information was used to inform computational models of the timing of clonal evolution of each patient's pancreatic cancer (full details of these models are available in Full Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 15 September 2010.

1. Fidler, I. J. Critical determinants of metastasis. *Semin. Cancer Biol.* **12,** 89–96 (2002).
2. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nature Rev. Cancer* **9,** 274–284 (2009).
3. Stathis, A. & Moore, M. J. Advanced pancreatic carcinoma: current treatment and future challenges. *Nature Rev. Clin. Oncol.* **7,** 163–172 (2010).
4. Yachida, S. & Iacobuzio-Donahue, C. A. The pathology and genetics of metastatic pancreatic cancer. *Arch. Pathol. Lab. Med.* **133,** 413–422 (2009).
5. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–1806 (2008).
6. Maitra, A. & Hruban, R. H. pancreatic cancer. *Annu. Rev. Pathol.* **3,** 157–188 (2008).
7. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
8. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105,** 4283–4288 (2008).
9. Terada, T. *et al.* Cell proliferative activity in intraductal papillary-mucinous neoplasms and invasive ductal adenocarcinomas of the pancreas: an immunohistochemical study. *Arch. Pathol. Lab. Med.* **122,** 42–46 (1998).
10. Hisa, T. *et al.* Growth process of small pancreatic carcinoma: a case report with imaging observation for 22 months. *World J. Gastroenterol.* **14,** 1958–1960 (2008).
11. Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324,** 1457–1461 (2009).
12. Sidransky, D. Emerging molecular markers of cancer. *Nature Rev. Cancer* **2,** 210–219 (2002).
13. Embuscado, E. E. *et al.* Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol. Ther.* **4,** 548–554 (2005).
14. Fu, B. *et al.* Evaluation of GATA-4 and GATA-5 methylation profiles in human pancreatic cancers indicate promoter methylation patterns distinct from other human tumor types. *Cancer Biol. Ther.* **6,** 1546–1552 (2007).
15. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16,** 1136–1148 (2006).
16. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
17. Sasaki, K. *et al.* Relationship between labeling indices of KI-67 and brdurd in human malignant tumors. *Cancer* **62,** 989–993 (1988).
18. Rew, D. A. & Wilson, G. D. Cell production rates in human tissues and tumours and their significance. Part II: clinical data. *Eur. J. Surg. Oncol.* **26,** 405–417 (2000).
19. Steel, G. G. *The Growth Kinetics of Tumours.* (Clarendon Press, Oxford, 1977).
20. Amikura, K., Kobari, M. & Matsuno, S. The time of occurrence of liver metastasis in carcinoma of the pancreas. *Int. J. Pancreatol.* **17,** 139–146 (1995).
21. Naumov, G. N. *et al.* A model of human tumor dormancy: an angiogenic switch from the nonangiogenic phenotype. *J. Natl. Cancer Inst.* **98,** 316–325 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by National Institutes of Health grants CA106610 (C.A.I.-D.), CA62924 (C.A.I.-D., M.A.N.), CA43460 (B.V.), CA57345 (K.W.K. and V.E.V.), CA121113 (V.E.V. and K.W.K.), GM078986 (M.A.N.), the Bill and Melinda Gates Foundation Grand Challenges Grant 37874, the Uehara Memorial Foundation (S.Y.), the AACR-Barletta Foundation (C.A.I.-D.), the John Templeton Foundation, Pilot Funding from the Sol Goldman Pancreatic Cancer Research Center, the Michael Rolfe Pancreatic Cancer Foundation, the George Rubis Endowment for Pancreatic Cancer Research, the Joseph C. Monastra Foundation for Pancreatic Cancer Research, the Alfredo Scatena Memorial Fund, the Virginia and the D.K. Ludwig Fund for Cancer Research, The Joint Program in Mathematical Biology and J. Epstein. We would like to acknowledge T. C. Cornish, C. Henderson, N. Omura and S.-M. Hong for their technical assistance in selected aspects of this work.

Author Contributions Sample collection and processing was performed by C.A.I.-D., S.Y., B.F. and M.K. Microdissection, DNA extractions and whole genome amplification reactions were performed by S.Y. Sequencing was performed by S.J. Copy number analyses were performed by R.L. Computational models and estimates of clonal evolution were performed by I.B., T.A. and M.A.N.; C.A.I.-D., S.Y., S.J., R.H.H., J.R.E., M.A.N., I.B., T.A., V.E.V., K.W.K. and B.V. directed the research. C.A.I.-D., B.V., S.Y., I.B. and T.A. wrote the manuscript, which all authors have approved.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.I.-D. (ciacobu@jhmi.edu).

METHODS

Patients and tissue samples. Tissue samples from seven patients with pancreatic ductal adenocarcinoma were collected in association with the Gastrointestinal Cancer Rapid Medical Donation Program (GICRMDP). This programme was approved by the Johns Hopkins institutional review board and deemed in accordance with the Health Insurance Portability and Accountability Act. Details of the programme have been described in detail previously[13]. The tissue harvesting protocol consists of the following; after opening of the body cavity using standard techniques, the whole pancreas including the pancreatic cancer and each grossly identified metastasis were sampled using a sterile blade and forceps. The whole pancreas was sliced into $1 \times 1 \times 0.4$ cm sections for overnight fixation in 10% buffered-formalin, for freezing in Tissue-Tek OCT compound (Sakura Finetechnical) in liquid nitrogen and for snap-freezing in liquid nitrogen in 1.7 ml cryovials and storage at $-80\,^{\circ}$C. Xenograft enriched or low passage cell lines were generated from the post mortem cancer tissues of these seven patients as described previously[13,14].

Laser capture microdissection (LCM). Frozen tissue sections of autopsy tissues were cut into 7 μm sections using a cryostat and embedded onto UV-treated PALM membrane slides (Carl Zeiss MicroImaging) and the slides were stored immediately at $-80\,^{\circ}$C until subsequent fixation. Tissue sections that underwent LCM were defrosted, fixed in 100% methanol for 3 min, and stained with toluidine blue before microdissection to remove contaminating stromal elements. Sections were dissected using a PALM MicroLaser System (Carl Zeiss MicroImaging). Dissected tissues were catapulted into adhesive caps. Generally, >20,000 cells were obtained from 5–10 serial sections by LCM to obtain sufficient quantity and quality of genomic DNA for subsequent amplification and sequencing.

Genomic DNA extraction and whole genome amplification. Genomic DNA from microdissected tissues was extracted using a QIAmp DNA Micro Kit (Qiagen) according to the manufacturer's protocol. Genomic DNA was quantified by calculating long interspersed nuclear elements (LINE) by real-time PCR. The LINE primer set 5′-AAAGCCGCTCAACTACATGG-3′ (forward) and 5′-TGCTTTGA ATGCGTCCCAGAG-3′ (reverse) was designed. The real-time PCR conditions were 95 °C for 10 min; 40 cycles of 94 °C for 10 s, 58 °C for 15 s and 70 °C for 30 s. PCR was carried out using Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). To minimize sequencing bias from using low-copy starting templates, only samples for which the measured concentration by LINE assay was ≥ 3.3 ng μl^{-1} (1,000 genome equivalents) were used as a starting template for whole genome amplification (WGA). WGA was performed using 10 ng total template DNA and an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare), following the manufacturer's protocol. WGA products were purified using a Microspin G-50 system (GE Healthcare). The purified WGA products were quantified by NanoDrop spectrophotometer (Thermo Fisher Scientific) and diluted to 20 ng μl^{-1} for sequencing analysis. Using these methods and quality controls, there was complete concordance in the mutational signatures obtained of cultured cell lines/xenografts versus WGA materials prepared from their matched frozen tissues.

Sanger sequencing. PCR amplification and sequencing was performed using the conditions and primers described previously[5]. A small number of sequencing reactions failed (<2% of the total reactions) and these corresponding genes were not included in progression models or quantitative time estimates of clonal evolution.

Genotyping. The Illumina Infinium II Whole Genome Genotyping Assay using the BeadChip platform was used to analyse tumour samples at 1,072,820 (1M) SNP loci as previously described[5]. Briefly, all SNP positions were based on the hg18 (NCBI Build 36, March 2006) version of the human genome reference sequence. The genotyping assay begins with hybridization to a 50-nucleotide oligonucleotide, followed by a two-colour fluorescent single-base extension. Fluorescence intensity image files were processed using Illumina BeadStation software to provide normalized intensity values and allelic frequency for each SNP position. For each SNP, the normalized experimental intensity value (R) was compared to the intensity values for that SNP from a training set of normal samples and represented as a ratio (called the 'log R ratio') of $\log_2$ ($R_{\text{experimental}}/R_{\text{training set}}$). For each SNP, the normalized allele intensity ratio (theta) was used to estimate a quantitative allelic frequency value (called the 'B allele frequency') for that SNP[15]. Using Illumina BeadStudio software, log R ratio and B allele frequency values were plotted along chromosomal coordinates and examined visually. Regions of loss of heterozygosity (LOH) were identified as genomic regions >2 megabases (Mb) with consecutive homozygous genotype calls (B allele frequency near 0 or 1). Smaller (<2 Mb) regions of LOH were identified by requiring co-occurrence of decreased log R ratio scores in regions of consecutive homozygous genotype calls (B allele frequency near 0 or 1). Visual analysis of these data plotted along chromosomal coordinates was followed by manual analysis of the data for selected genes of interest.

Estimations of proliferation rates. To estimate the cell division rate, the Ki-67 labelling index (LI) in the proband lesion for each case was calculated. The Ki-67 LI on the pancreatic ducts in the histologically normal pancreas parenchyma was also calculated. Normal pancreas was collected from two autopsied patients who died of causes other than pancreatic cancer and five patients who underwent distal pancreatectomy for a serous cystadenoma or an islet cell tumour at The Johns Hopkins Hospital. Paraffin blocks were cut into sections 4-μm thick for Ki-67 immunostaining with all staining processes from deparaffinization to counterstaining with haematoxylin being performed automatically with the Ventana Discovery staining system (Ventana Medical Systems). An anti-human Ki-67 mouse monoclonal antibody (Clone MIB-1, Dako Cytomation) was used. At least 12 randomly selected high-power fields containing a minimum of 2,000 cells were evaluated for each case, and the labelling index (LI) was calculated as the percentage of positive cell nuclei. Reactive small lymphocytes in each case were regarded as internal positive controls for Ki-67. Equal or more intense nuclear staining in comparison with the internal positive controls was considered to indicate positivity.

Modelling tumour evolution. Passenger mutations were defined as those unlikely to drive tumorigenesis. To be conservative, we considered passenger mutations as those not included as candidate cancer genes in a recent study based on whole exome sequencing of 24 pancreatic cancers[5]. As the great majority of mutations identified in cancers are believed to be passengers, the results of the model are not highly dependent on the model used to estimate the relatively small number of drivers[16].

Because passenger mutations are neutral and do not affect the evolution in any way, they are accumulated independently in each cell lineage. Following the lineage of the founder cell of the parental clone back in time, we can assume that it acquired a new neutral mutation with rate r at each cell division, with r being the product of the mutation rate per base pair per cell division and the number of base pairs sequenced. The accumulation of neutral mutations in a cell lineage can be well-described by a Poisson process with rate r per cell division. We are interested in the number of cell divisions in the single lineage between tumour initiation and birth of the founder cell of the parental clone during which N_1 passenger mutations accumulate. On the other hand, N_1 is also the number of mutations that are found in all tumour samples from one patient. Since we sequenced at least one sample from the primary tumour and at least three samples from different metastases from each patient, these specific N_1 mutations had to be present in the founder cells of all three metastases and in cells in the primary tumour. Thus there was a cell in the tumour that had these N_1 specific mutations for the first time, and that is, by definition, the founder cell of the parental clone. Since we can neglect the accumulation of mutations before the onset of the tumour, these N_1 mutations are accumulated along the single lineage from the tumour initiator cell to the founder cell of the parental clone. As the number of cell divisions between two subsequent mutations is distributed according to an exponential distribution with mean $1/r$, the required number of cell divisions is the sum of N_1 independent exponentially distributed random variables with mean $1/r$, and is distributed according to a Gamma distribution with shape parameter N_1 and scale parameter $1/r$. The mean of this distribution is N_1/r and the standard deviation is $\sqrt{N_1}/r$ (see Supplementary Table 6). Because the number of base pairs sequenced in the study is 31.7×10^6, and the mutation rate per base pair per generation is estimated at 5×10^{-10}, $r = 31.7 \times 10^6 \times 5 \times 10^{-10} \approx 0.016$ per generation[8].

Using our measurements of Ki-67 labelling index of the seven index lesions (average 16.3%), we were able to estimate the S-phase fraction of cells in the seven index lesions (average LI = 9.5%)[17]. Assuming a median value for the S-phase duration in human tissues and tumours, T_s, of 10 h (ref. 18) and using the formula for the potential cell doubling time $T_{\text{pot}} = \lambda T_s/\text{LI}$, we get an estimate for T_{pot} of 3.5 days. Here λ is a correction factor for the nonlinear age distribution of cells through the cell cycle, which was assumed to be 0.8 (ref. 19). This estimate is consistent with the average cell doubling time in pancreatic cancer from ref. 20 of 2.3 days. We use this latter estimate in our analysis, as we believe it is more accurate for pancreatic cancer.

Our model works very well for estimating the number of cell divisions between discrete events in tumour evolution. In order to go from number of cell divisions to actual time we need to have an estimate for the average rate of cell division. The accuracy of our predictions regarding actual time therefore depends on the accuracy of that estimate. If we let T_{gen} denote the average time between subsequent cell divisions in a cell lineage, we arrive at the expression for time T_1:

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

We therefore estimate the number of cell divisions, and hence the time T_1 between tumour initiation and birth of the founder cell of the parental clone, to be proportional to the number of passenger mutations, N_1, that the tumour acquired during that time. In our calculations, we use the estimate for cell doubling time in pancreatic cancer from the literature[20] as the value of T_{gen}.

T_2 is determined analogously, with N_2 defined as the number of passenger mutations present in the index lesion but not in the parental clone. T_3 is

determined from literature-based estimates of the tumour and cell doubling times, and the size of the index lesions at autopsy[20].

The median doubling time of pancreatic cancer metastases was reported as 56 days[20]. To estimate the age of the index metastasis, we used a two stage model. We estimated the tumour doubling time was equal to the cell doubling time (T_{gen}) until the tumour size reached 100 μm in diameter at which time angiogenesis is required[21]. Thereafter, we used the median doubling time described above.