

A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns

Faiyaz Notta¹, Michelle Chan-Seng-Yue^{1*}, Mathieu Lemire^{1*}, Yilong Li^{2*}, Gavin W. Wilson¹, Ashton A. Connor¹, Robert E. Denroche¹, Sheng-Ben Liang³, Andrew M. K. Brown¹, Jaeseung C. Kim^{1,4}, Tao Wang^{4,5}, Jared T. Simpson^{1,7}, Timothy Beck¹, Ayelet Borgida⁸, Nicholas Buchner¹, Dianne Chadwick³, Sara Hafezi-Bakhtiari^{1,3}, John E. Dick^{1,6,9}, Lawrence Heisler¹, Michael A. Hollingsworth⁸, Emin Ibrahimov¹, Gun Ho Jang¹, Jeremy Johns¹, Lars G. T. Jorgensen¹, Calvin Law¹⁰, Olga Ludkovski⁹, Ilinca Lungu¹, Karen Ng¹, Danielle Pasternack¹, Gloria M. Petersen¹¹, Liran I. Shlush⁹, Lee Timms¹, Ming-Sound Tsao^{4,9}, Julie M. Wilson¹, Christina K. Yung¹, George Zogopoulos¹², John M. S. Bartlett¹, Ludmil B. Alexandrov¹³, Francisco X. Real¹⁴, Sean P. Cleary^{15,16}, Michael H. Roehr^{3,9}, John D. McPherson^{1,4}, Lincoln D. Stein^{1,6}, Thomas J. Hudson^{1,6}, Peter J. Campbell^{2,17} & Steven Gallinger^{1,15,16}

Pancreatic cancer, a highly aggressive tumour type with uniformly poor prognosis, exemplifies the classically held view of stepwise cancer development¹. The current model of tumorigenesis, based on analyses of precursor lesions, termed pancreatic intraepithelial neoplasm (PanINs) lesions, makes two predictions: first, that pancreatic cancer develops through a particular sequence of genetic alterations^{2–5} (*KRAS*, followed by *CDKN2A*, then *TP53* and *SMAD4*); and second, that the evolutionary trajectory of pancreatic cancer progression is gradual because each alteration is acquired independently. A shortcoming of this model is that clonally expanded precursor lesions do not always belong to the tumour lineage^{2,5–9}, indicating that the evolutionary trajectory of the tumour lineage and precursor lesions can be divergent. This prevailing model of tumorigenesis has contributed to the clinical notion that pancreatic cancer evolves slowly and presents at a late stage¹⁰. However, the propensity for this disease to rapidly metastasize and the inability to improve patient outcomes, despite efforts aimed at early detection¹¹, suggest that pancreatic cancer progression is not gradual. Here, using newly developed informatics tools, we tracked changes in DNA copy number and their associated rearrangements in tumour-enriched genomes and found that pancreatic cancer tumorigenesis is neither gradual nor follows the accepted mutation order. Two-thirds of tumours harbour complex rearrangement patterns associated with mitotic errors, consistent with punctuated equilibrium as the principal evolutionary trajectory¹². In a subset of cases, the consequence of such errors is the simultaneous, rather than sequential, knockout of canonical preneoplastic genetic drivers that are likely to set-off invasive cancer growth. These findings challenge the current progression model of pancreatic cancer and provide insights into the mutational processes that give rise to these aggressive tumours.

Pancreatic cancer will be the second leading cause of cancer-related death in a decade and the biological basis for the aggressive nature of this disease is largely undefined. Motivated by this, we explored the pancreatic cancer genome to address this concern. These genomes are highly unstable¹³, as evidenced by the marked modifications to

the DNA copy number landscape. Although this instability is further exacerbated with metastatic progression¹⁴, it remains unclear when the instability begins relative to the key genetic alterations that give rise to the invasive clone. Also, whether this instability propagates through single copy number changes that accumulate one after another or through large numbers of concurrent changes has not been fully addressed. These questions have important basic and translational implications. As a first step, the mechanisms at the root cause of this instability need to be identified. Mutational phenomena such as chromothripsis and polyploidization have been linked to unstable tumours^{15,16} and aggressive tumour behaviour¹⁷, indicating that they play a role in pancreatic cancer development. These particular phenomena are considered to accelerate cancer evolution because the DNA damage that ensues from such mitotic errors must be resolved in one or few rounds of cell division; otherwise the cell would die. To date, the extensive fibrosis in pancreatic cancer has obstructed the sequencing resolution needed to clearly decipher these events. In this study, we performed an in-depth analysis of more than 100 whole genomes (Extended Data Fig. 1) from purified primary and metastatic pancreatic tumours (referring to ductal adenocarcinoma only), focussing on the mutational phenomena linked to rapid tumour progression.

To evaluate polyploidization, we developed and validated a new informatic tool, termed CELLULOID, which estimates tumour ploidy and copy number from whole-genome data (Fig. 1a and Extended Data Fig. 2). We found that 45% (48/107) of tumours displayed changes in copy number consistent with polyploidization (ploidy solutions can be found in Supplementary Information). Of the polyploid tumours, 88% (42/48) were tetraploid and the rest were hexaploid. The mean ploidy of diploid tumours was 1.95, whereas those tumours that underwent genome duplication and triplication was 3.38 and 5.40 (relative to 4 and 6), indicating that a larger proportion of the genome was lost in the latter subgroup (Extended Data Fig. 3a, b), consistent with previous data¹⁶. Polyploid tumours had higher incidences of mutation in *TP53* ($P = 0.02$, Fisher's exact test; Extended Data Fig. 1e) and harboured 1.5-fold more copy number alterations compared to diploid tumours (median value of 112 versus 77, $P = 0.003$, t -test; Extended Data Fig. 3c).

¹Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ²Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ³UHN Program in BioSpecimen Sciences, Department of Pathology, University Health Network, Toronto, Ontario M5G 2C4, Canada. ⁴Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁵Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁶Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁷Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada. ⁸Eppley Institute for Research in Cancer, Nebraska Medical Center, Omaha, Nebraska 68198, USA. ⁹Princess Margaret Cancer Centre, University Health Network (UHN), Toronto, Ontario M5G 2M9, Canada. ¹⁰Division of Surgical Oncology, Sunnybrook Health Sciences Centre, Odette Cancer Centre, Toronto, Ontario M4N 3M5, Canada. ¹¹Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹²Research Institute of the McGill University Health Centre, Montreal, Québec, Canada, H3H 2L9. ¹³Theoretical Biology and Biophysics (T-6) and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 87545. ¹⁴Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain. ¹⁵Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. ¹⁶Department of Surgery, University Health Network, Toronto, Ontario M5G 2C4, Canada. ¹⁷Department of Haematology, University of Cambridge, Cambridge CB2 0XY, UK. *These authors contributed equally to this work

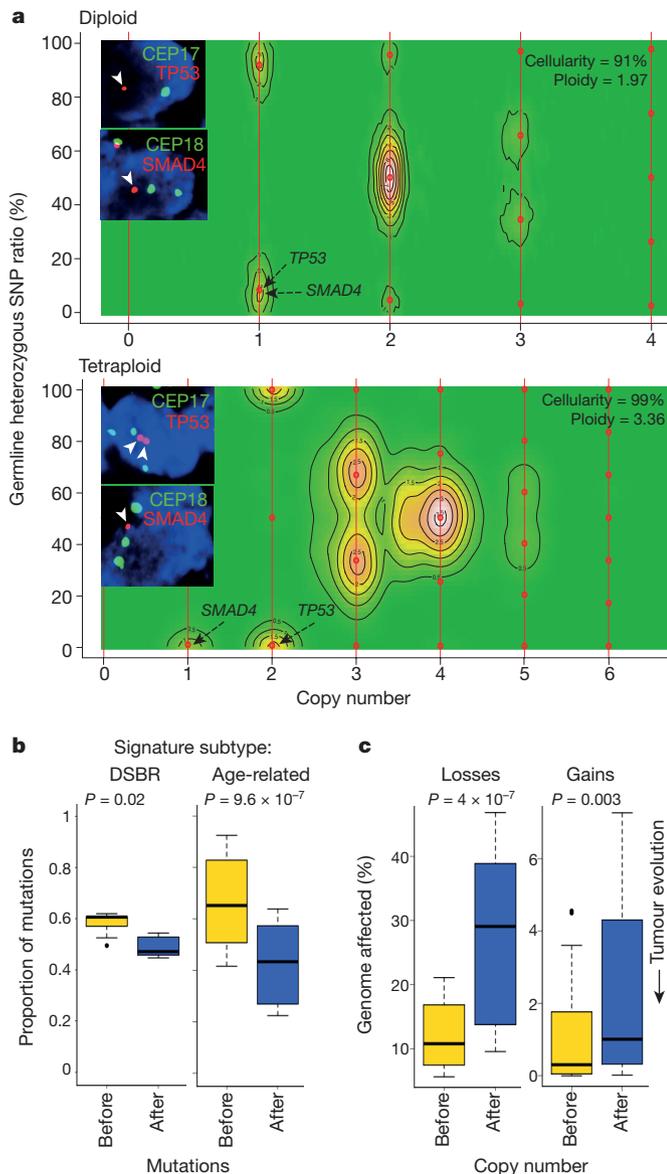


Figure 1 | Polypliodization in pancreatic cancer. **a**, CELLULOID profiles of a diploid (Ashpc_0008) and a tetraploid (Ashpc_0005) tumour. The predicted copy number of *SMAD4* and *TP53* genes is indicated with black arrows. Inset shows a FISH validation of the predicted copy number of *SMAD4* and *TP53* genes. CEP, centromeric probes. **b**, Proportion of mutations that occurred before (yellow) or after (blue) polypliodization. Cases were segregated based on mutational signature subtype: DSBR ($n = 5$; left) and age-related ($n = 32$; right). Owing to the increased genetic instability in polyploid cells, mutations in regions of copy number of 4 in tetraploids were used in this analysis. **c**, Fraction of the genome lost and gained either before (yellow) or after (blue) polypliodization. Box and whisker plots depict median and 10–90 percentile ranges. P values are indicated and were derived using a t -test. A detailed description of these data is given in Supplementary Results.

The marked loss of genomic material relative to baseline ploidy and increased amount of copy number alterations in polyploids demonstrates that these genomes are highly unstable.

We then used mutation data to infer the timing of the polypliodization event in tumour evolution (Supplementary Results). All cases were first categorized according to their dominant mutational signature, since specific aetiologies drive mutation accrual¹⁸. Two subgroups were evident: one where $C > T$ transitions dominated, linked to the process of cytosine deamination (approximately 80% Age-related, Extended Data Fig. 3d) and another where all six classes of base substitutions

were more-or-less balanced—a phenomenon associated with defects in double-strand break repair (DSBR, 17%; Extended Data Fig. 3d). Accordingly, half of the DSBR cases carried germline or somatic mutations in *BRCA1/2* (ref. 13). The remaining cases were comprised of heterogeneous signatures previously identified by Alexandrov *et al.*¹⁸ (Extended Data Fig. 3d).

We found that most mutations preceded polypliodization in both mutational subgroups (Fig. 1b). By contrast, most copy number losses and gains occurred after polypliodization, an effect that was markedly magnified when the size of the copy number change was taken into account (losses: $P = 4.3 \times 10^{-7}$; gains: $P = 0.003$, t -test; Fig. 1c and Extended Data Fig. 3e). This implies that changes in copy number that precede polypliodization were smaller and focal whereas those that come after are larger and more structurally damaging to the genome. Some of these larger changes are likely to be a consequence of the improper segregation of chromosomal material gained during polypliodization. Copy number alterations corresponding to the polypliodization event were commonly seen at integer values and indicate that such events are mostly or fully clonal (CELLULOID solutions in Supplementary Information). Two conclusions emerge from these data: first, polypliodization occurs after an extended diploid phase of mutation accrual; and second, changes in copy number related to polypliodization come to rapidly dominate in the tumour within a shorter timeframe, suggesting they are relevant to disease progression.

Many diploid and polyploid tumours harboured focal copy number alterations that oscillated between a few DNA copy-states, characteristic of chromothripsis¹⁵. We developed a sensitive algorithm, termed ChromAL (see Methods and Supplementary Results), to differentiate chromothripsis from localized gradual events that accumulate over time. We found that 65.4% (70/107) of tumours harboured at least one chromothripsis event (solutions provided in Supplementary Information). A similar frequency was observed in an independent genome cohort (60%, $n = 50$ out of 84, Supplementary Results). Of all chromothripsis events, 11% occurred on chromosome 18 (Extended Data Fig. 4a), resulting in the loss of the key tumour suppressor gene *SMAD4*. By comparing the consensus copy number profiles of tumours with and without chromothripsis, we found that *SMAD4* loss was accompanied by a gain in a region of chromosome 18 that harbours *GATA6*, an oncogene implicated in pancreatic cancer development (Extended Data Fig. 4b, top panel and Supplementary Fig. 1). Furthermore, 8% of events were observed on chromosome 12. The consensus copy number profile of these cases revealed a focal amplification in the region of *KRAS* (Extended Data Fig. 4b, middle panel). These amplifications commonly affected the mutant *KRAS* allele either directly, when chromothripsis and breakage–fusion–bridge (BFB) cycles were combined (Extended Data Fig. 4c, tumour Pcsi_0290), or indirectly, when polypliodization was subsequent to a chromothripsis event that removed the wild-type copy (Extended Data Fig. 4c, Pcsi_0356). There was significantly more chromothripsis in polyploid tumours than in diploid tumours, confirming the greater genetic instability in the former subgroup ($P = 0.013$, Fisher's exact test; Extended Data Fig. 4d). We observed worse overall survival in patients whose tumours had such an event ($P = 0.025$, log-rank test; Supplementary Fig. 2). The high prevalence of chromothripsis in pancreatic cancer, together with previously established links between chromothripsis and aggressive tumour behaviour in other cancers^{15,17}, strongly implicate this mutational processes as a key part of pancreatic cancer development. Notably, these data directly support the ‘catastrophic’ model of pancreatic cancer progression proposed by Real¹⁹ more than a decade ago.

We next performed a series of focused analyses, using individual tumours to illustrate the broad principles of the approach applied to the genome cohort. The data presented above raises an important question: how much of the overall genetic instability in these tumours can be attributed to a single chromothripsis event? In Pcsi_0082, a tetraploid tumour, 63% of all copy number alterations could be attributed to five

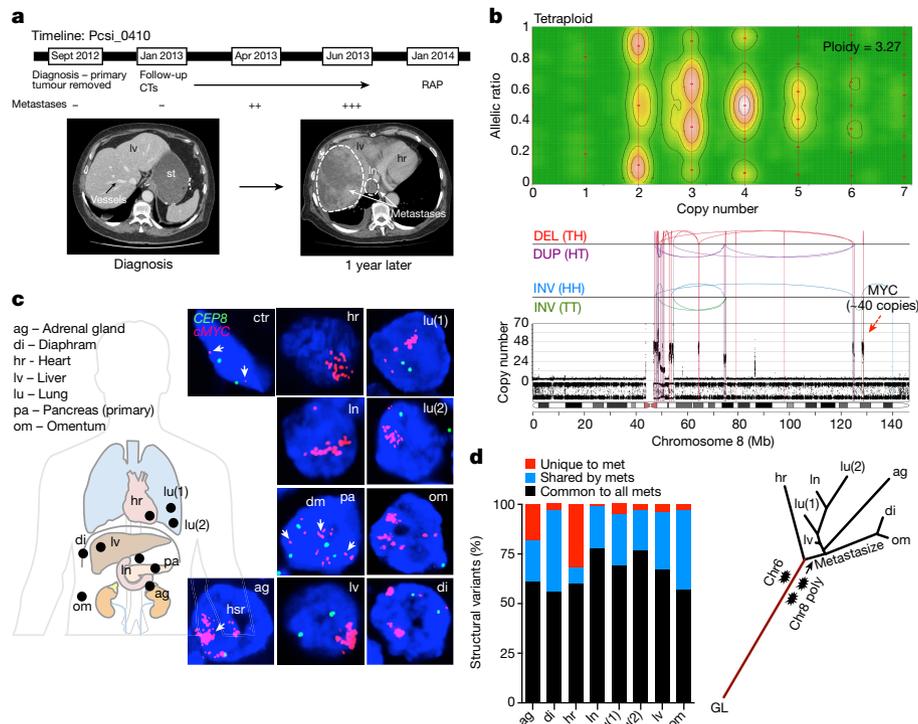


Figure 2 | Chromothripsis and polyploidization in a patient with metastatic progression. **a**, Timeline (top) and computerized tomography scan (CT; bottom) images of Pcsi_0410. White dashed lines indicate metastases. Eight distinct metastases from Pcsi_0410 (see image in **c**) were sequenced. RAP, rapid autopsy. **b**, Polyploidization (top) and chromothripsis (bottom) event from the adrenal gland metastasis (see also Extended Data Fig. 7). **c**, FISH analysis of *MYC* amplification in primary tumour and all metastases. ctr, control (fibroblasts). **d**, Left, the proportion of structural variants common to all (black), shared by two or more

(blue), or unique to each metastasis is shown. Right, copy numbers and structural variants were used to reconstruct phylogenetic tree of metastatic progression. The primary tumour was surgically removed one year before autopsy and fresh-frozen material was not available for whole-genome sequencing. It is possible that branch lengths of the phylogenetic tree would vary if the primary tumour were included in this analysis. Lines are to scale with the copy-number-based clustering dendrogram presented in Supplementary Fig. 15, with the exception of germline origin (GL), which is half the length.

distinct chromothripsis events, on chromosomes 8, 13, 15, 16 and 18 (Extended Data Fig. 5a). As chromothripsis is sustained and resolved in a single cell-division cycle^{20,21}, we can approximate that more than half of the genomic damage in Pcsi_0082 was incurred from approximately five aberrant mitoses. Because Pcsi_0082 had undergone polyploidization, we were able to infer the timing of chromothripsis events relative to the genome doubling using the magnitude of the copy number changes. As chromothripsis occurs on one copy of DNA, the events sustained on chromosomes 13, 16 and 18 must have occurred after polyploidization because the copy number changes on these chromosomes mostly vary by one (Extended Data Fig. 5a, events 2, 4 and 5). By contrast, the chromothripsis on chromosomes 8 and 15 occurred while the tumour was still diploid, since these copy number changes vary in multiples of two, a result of genome doubling (Extended Data Fig. 5a, events 1 and 3). Across all polyploid tumours, we observed that more than half (59%) of all chromothripsis events transpired before polyploidization (ChromAL solutions). This suggests that polyploidization further exacerbates the pre-existing genetic instability in these tumours. Overall, many copy number alterations in pancreatic cancer are acquired through rapid bursts of genetic change from a single or few mitotic events (Extended Data Fig. 5b) rather than a set of gradual events that accumulate over time.

To investigate the role of these mitotic events in disease progression, we analysed the genomes of 15 distinct metastases from six patients (Extended Data Fig. 6 and Supplementary Results). In one case of fulminant metastatic progression (Pcsi_0410), eight distinct metastases were sequenced (Fig. 2a shows the progression timeline). All metastases were polyploid and also carried two distinct chromothripsis events, one on chromosome 6 and another on chromosome 8, that resulted in the marked amplification of *MYC* (20–40 copies), resembling a double

minute (Fig. 2b, c and Extended Data Fig. 7a). The final copy number in areas of loss of heterozygosity (LOH) in both chromothripsis events is two, indicating that both chromothripsis events occurred before polyploidization (Extended Data Fig. 7b). Using fluorescence *in situ* hybridization (FISH), we confirmed that the primary tumour was also polyploid and harboured chromothripsis (Fig. 2c and Supplementary Fig. 3a, b). Thus, we can infer that both chromothripsis events preceded polyploidization and that the systemic spread of the disease occurred after polyploidization by a clone that harboured all three mitotic events (Fig. 2d). An additional chromothripsis event was detected on chromosome 13 in the adrenal gland metastasis (Supplementary Fig. 3c), consistent with previous data on ongoing genetic instability with metastatic progression¹⁴. Overall, we observed that chromothripsis was maintained in metastases if it was present in the primary tumour (Extended Data Fig. 6d). These data support the notion that the majority of genetic instability precedes metastases and is fostered early in tumorigenesis. If the dominant clonal lineage of the primary tumour arises from these types of mitotic events, it suggests that intra-tumoural heterogeneity in pancreatic cancer¹⁰ follows this event, akin to the ‘big-bang’ model proposed for colon cancer²².

The central tenet of the PanIN progression model posits that alterations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4* are acquired as part of a consecutive series of events in tumour evolution. To directly test this model, we used DNA rearrangements to reconstruct the evolutionary history of allelic losses of tumour suppressors based on evidence that allelic alterations are early events in tumorigenesis (Supplementary Results and Luttes *et al.*⁵). Ashpc_0005, a tetraploid tumour, had a complex pattern of rearrangements involving chromosomes 9, 17 and 18, where *CDKN2A*, *TP53* and *SMAD4* are found (Fig. 3a). Several features of this rearrangement pattern facilitate the reconstruction of

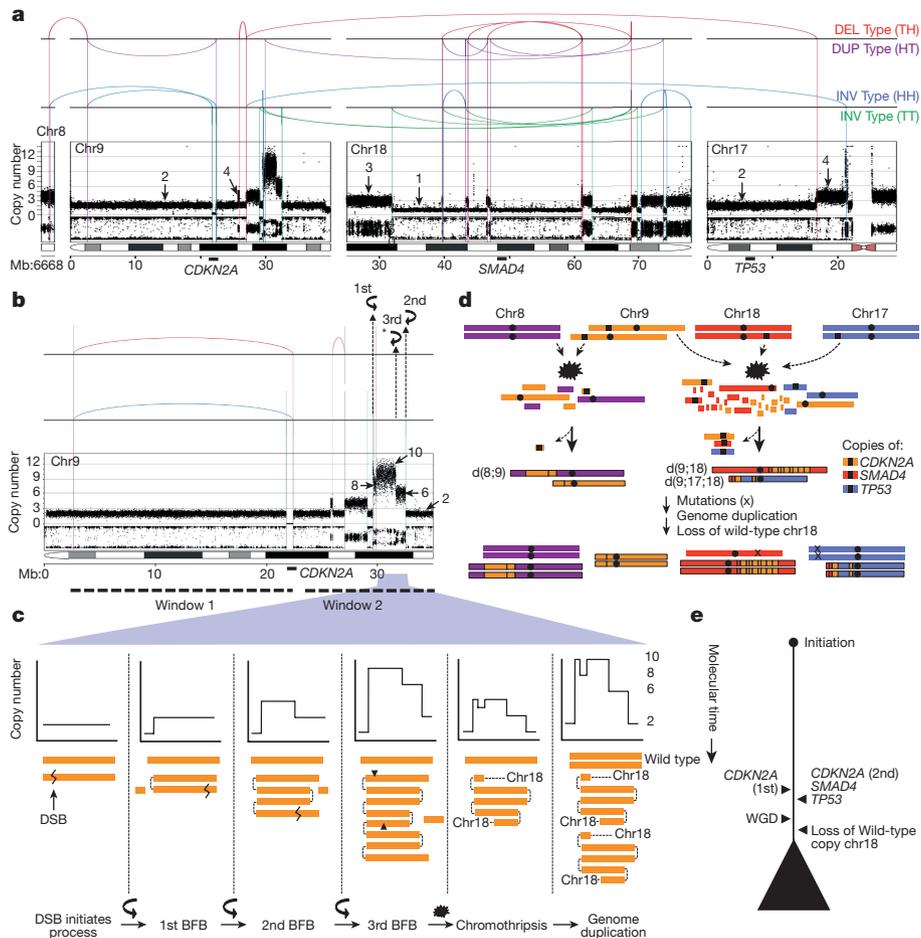


Figure 3 | Simultaneous knockout of pancreatic cancer driver genes.

a, Rearrangement profile of chromothripsis in Ashpc_0005. Positions of key genes (*CDKN2A*, *TP53*, *SMAD4*) are shown at the bottom. DEL, deletion; DUP, duplication; INV, inversion. **b**, Two distinct rearrangements windows on chromosome 9. In window 2, three fold-back inversions (two mapped and one unmapped, marked with an asterisk) are highlighted with curved black arrows. The copy number state of segments as a result of BFB cycles is shown with straight black arrows. **c**, Schematic depiction of the three cycles of BFB that generated the final copy number state in window 2 (**b**).

the mutational events in this tumour. First, there are two independent sets of rearrangements on chromosome 9 that flank *CDKN2A* (Fig. 3b, windows 1 and 2), indicating that the two copies of this gene were lost as part of independent chromothripsis events. Second, there are distinct amplified DNA segments in window 2 (Fig. 3c) that are bounded by a specific type of rearrangement referred to as a fold-back inversion, an alteration that leaves behind steep copy number drops (>2) indicative of a cycle of BFB¹⁴. Three steep copy number drops in window 2 are evidence of three cycles of BFB (Fig. 3c). Third, the intervening change in copy number (from 10 to 8) on one of these amplified segments suggests that a chromothripsis event followed three cycles of BFB and was likely to be the final major event that stabilized the derivative chromosome²³ (Fig. 3c, penultimate panel). Fourth, all copy number changes in the event are in multiples of two, indicating that polyploidization followed the BFB cycles and chromothripsis (Fig. 3a). Finally, the copy number change on chromosome 18 from 3 to 1 (rather than 4 to 2) indicates that one wild-type copy of this chromosome was lost after polyploidization (Fig. 3a). The relative order of the first and the second copy losses of *CDKN2A* cannot be deciphered, but a single event involving BFB and chromothripsis knocked out a single copy of *CDKN2A*, *TP53* and *SMAD4* in synchronized fashion (Fig. 3d, e). Using rearrangements to reconstruct the

Zigzag symbol indicates DNA double-strand break to initiate BFB.

d, Temporal order of events based on rearrangement profile. The leftover *TP53* and *SMAD4* alleles carry inactivating mutations (x). As both *TP53* alleles carry the mutations (ploidy >1), this mutation was acquired before genome duplication. Relative timing of the *SMAD4* mutation cannot be inferred because there is only one copy of this allele and the mutation is fully clonal. d(8;9), d(9;18) and d(9,17;18) refer to candidate derivative chromosomes based on DNA rearrangement profiles. **e**, Summary of tumour evolution in Ashpc_0005. WGD, whole-genome duplication.

sequence of events in a second case (Pcsi_0171) demonstrated that a single chromothripsis event simultaneously knocked out *CDKN2A* and *SMAD4* (Extended Data Fig. 8). Notably, rearrangement patterns in 16% of cases (17/107) combined allelic alterations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4* genes, predominantly as double knockouts (14% if only tumour-suppressor genes are considered; Supplementary Fig. 4). In a proof-of-principle experiment using single-cell sequencing in a tumour where rearrangements did not span these genes, we found an ancestral clone that harboured a *SMAD4* loss but retained *TP53* and *CDKN2A* (Extended Data Fig. 9). These data provide direct evidence that a number of cases do not conform to the accepted mutational hierarchy predicted by the PanIN progression model and warrant future investigation into the sequence of mutational events that give rise to these aggressive tumours.

Studies dating back two decades have been critical in moulding the current perspective of how pancreatic cancer develops¹. Key features of our data provide a framework to broaden this view. First, analysis of polyploid tumours revealed that most mutations accumulate when these tumours are still diploid. Assuming that preneoplastic cells are diploid, a fraction of these mutations must be preneoplastic. In line with this reasoning, Murphy *et al.* have demonstrated that preneoplasms in pancreatic cancer acquire an extensive mutation burden but remain

non-invasive²⁴. This suggests a prolonged preneoplastic phase predates the onset of invasive disease and that copy number events are crucial for transformation (Extended Data Fig. 10). These data carry implications for the design of future studies on the early detection of pancreatic cancer¹¹. Second, copy number changes from chromothripsis are essentially clonal, suggesting that these events are sustained early in tumorigenesis. The inactivation of well-known preneoplastic drivers (*CDKN2A*, *TP53*, *SMAD4*) *en bloc* strongly supports this notion and implies that chromothripsis can be a transforming event under the right gene context^{17,23}. Our data also raise the possibility that some pancreatic cancers may not progress through a linear series of PanIN lesions¹⁹. Why catastrophic mitotic phenomena are so frequent in pancreatic cancer cannot be easily answered. Perhaps the extensive fibrosis in these tumours, known to suppress tumour development^{25,26}, apply a selective pressure that favours punctuated events over gradual ones. Lastly, pancreatic cancer is well known for its proclivity to metastasize. In mouse models of pancreatic cancer, genetic instability contributes to metastatic progression²⁷. If chromothripsis is indeed the transforming event in some tumours, as our data suggest, a single event could thus confer a cell with both invasive and metastatic properties. In this scenario, there would be a very short latency period between the birth of the invasive clone and the ability of that clone to metastasize^{28,29}. This supposition is consistent with the observation that 80% of pancreatic cancer patients present with advanced disease at diagnosis. How these mutational processes contribute to disease progression and metastatic phenotype is therefore a critical topic of investigation; such knowledge will be essential to guide more effective screening and therapeutic strategies, both for pancreatic cancer and other aggressive tumour types.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2015; accepted 2 September 2016.

Published online 12 October 2016.

- Hruban, R. H., Goggins, M., Parsons, J. & Kern, S. E. Progression model for pancreatic cancer. *Cancer Res.* **6**, 2969–2972 (2000).
- Moskaluk, C. A., Hruban, R. H. & Kern, S. E. *p16* and *K-RAS* gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res.* **57**, 2140–2143 (1997).
- Wilentz, R. E. *et al.* Inactivation of the *p16* (*INK4A*) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer Res.* **58**, 4740–4744 (1998).
- Wilentz, R. E. *et al.* Loss of expression of *Dpc4* in pancreatic intraepithelial neoplasia: evidence that *DPC4* inactivation occurs late in neoplastic progression. *Cancer Res.* **60**, 2002–2006 (2000).
- Lüttges, J. *et al.* Allelic loss is often the first hit in the biallelic inactivation of the *p53* and *DPC4* genes during pancreatic carcinogenesis. *Am. J. Pathol.* **158**, 1677–1683 (2001).
- Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
- Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–372 (2015).
- Ross-Innes, C. S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat. Genet.* **47**, 1038–1046 (2015).
- Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).
- Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
- Chari, S. T. *et al.* Early detection of sporadic pancreatic cancer: summative review. *Pancreas* **44**, 693–712 (2015).
- Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J. M.) 82–115 (Freeman, Cooper and Company, 1972).
- Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).

- Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148**, 59–71 (2012).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Real, F. X. A. A "catastrophic hypothesis" for pancreas cancer progression. *Gastroenterology* **124**, 1958–1964 (2003).
- Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
- Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).
- Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
- Li, Y. *et al.* Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
- Murphy, S. J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145**, 1098–1109.e1 (2013).
- Özdemir, B. C. *et al.* Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* **25**, 719–734 (2014).
- Rhim, A. D. *et al.* Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25**, 735–747 (2014).
- Hingorani, S. R. *et al.* *Trp53^{R172H}* and *Krac^{G12D}* cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Haeno, H. *et al.* Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* **148**, 362–375 (2012).
- Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank N. Simard, S. Zhao and members of the SickKids-UHN Flow facility for technical support. Funding sources for this study include grants to the Pancreatic Cancer Sequencing Initiative program from the Ontario Institute for Cancer Research (OICR), through support from the Ontario Ministry of Research and Innovation, the Canada Foundation for Innovation; research award to F.N. from the OICR and the Canadian Institutes for Health Research (CIHR); Canadian Friends of the Hebrew University, the SMGS Family Foundation, NCI grant P50 CA102701 (Mayo Clinic SPORÉ in Pancreatic Cancer) and NCI grant R01 CA97075 (Pancreatic Cancer Genetic Epidemiology Consortium). F.N. is supported by a fellowship award from CIHR and is a recipient of a scholar's research award from the Ontario Institute of Cancer Research (OICR), through support from the Ontario Ministry of Research and Innovation. G.Z. is a Clinician–Scientist of the Fonds de la Recherche en Santé du Québec. P.J.C. is a Wellcome Trust Senior Clinical Fellow. T.J.H., L.D.S., J.D.M. and S.G. are recipients of Senior or Clinician–Scientist Awards from the Ontario Institute for Cancer Research.

Author Contributions Data analysis and interpretation was performed by F.N., M.L., Y.L., M.C.-S.-Y., G.W.W., A.A.C., F.X.R., P.J.C., S.G. and T.J.H.; tumour enrichment by S.-B.L., I.L. and F.N.; pathological assessment by T.W., M.-S.T., J.M.S.B., M.H.R. and S.H.-B.; genomics by R.E.D., A.M.K.B., K.N., J.C.K., L.T., N.B., D.P., L.H., E.I., G.H.J., J.J., L.G.T.J., J.D.M., L.D.S., L.I.S., L.H., J.E.D., C.K.Y., T.B. and L.B.A.; FISH by O.L.; CELLULOID analysis by M.L. and single-cell analysis by G.W.W., J.T.S. and F.N. Sample acquisition, annotation and collection from institutes external to University Health Network, was performed by G.M.P., M.A.H., G.Z. and C.L. Sample acquisition, annotation and collection from the University Health Network was performed by J.M.W., A.B., S.G. and S.P.C. The study was designed by F.N., T.J.H. and S.G.; F.N. prepared and wrote the manuscript; the manuscript was edited by M.L., F.X.R., J.E.D., P.J.C., T.J.H. and S.G.

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 the paper. Correspondence and requests for materials should be addressed to F. N. (faiyaz.notta@oicr.on.ca) or T. J. H. (tom.hudson@oicr.on.ca).

Reviewer Information *Nature* thanks S. Chanock, M. Rossi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Ethical approval and sample acquisition. A total of 107 surgically resectable samples of pancreatic ductal adenocarcinoma tissue were obtained from collaborating hospitals in Canada and the United States from patients that gave informed consent under the ICGC protocol. 84 samples were obtained from the University Health Network (Toronto, Canada), 14 samples from the Mayo Clinic, 3 samples from the University of Nebraska as part of a rapid autopsy program, 5 samples from Sunnybrook Health Sciences Centre (Toronto, Canada), and 1 sample from McGill University (Montreal, Canada). Consent for WGS was obtained locally at each institute. At the Ontario Institute for Cancer Research, approval was obtained through the University Health Network Research Ethics Board (08-0767-T) and University of Toronto Research Ethics Board (30024). Pre-operatively, blood samples were collected for germline DNA. Where blood was not collected, duodenal mucosa or other non-cancerous tissue was collected post-operatively to obtain germline DNA. Tumours were sectioned to confirm the diagnosis of ductal adenocarcinoma and pieces were snap-frozen in liquid nitrogen and stored at -80°C or -150°C before proceeding with laser capture microdissection (LCM). For 21 cases (17 UHN, 4 Sunnybrook), fresh tumour material was dissociated and viably sorted at -150°C (below). We obtained clinical follow-ups on the majority of cases.

Sample dissociation and cell sorting. Freshly resected tumours were minced into fine pieces in 10-cm tissue culture dishes using a razor blade. After mechanical dissociation, 9 ml of RPMI supplemented with 1% FBS was added. 1 ml of $10\times$ collagenase/hyaluronidase mix (Stemcell technologies) was added to bring the volume to 10 ml and the sample was placed in a 37°C incubator. Every 20 min, the tissue pieces in the culture dish were pipetted through narrowing orifices (for example, a 10 ml then 5 ml then 1 ml pipette) for a total of 60–120 min. The sample was then passed through a 70–150- μm nylon mesh, centrifuged and resuspended in DMSO (Sigma) based cryopreservation media (20% FBS/10% DMSO final) and placed at -150°C for long-term storage.

For cell sorting, frozen vials of viable cells were thawed via dropwise addition of RPMI solution (IMDM + 20% FBS + DNaseI). Final concentration of DNaseI (Roche Applied Science, 10104159001) in RPMI solution was $200\mu\text{g ml}^{-1}$. After thawing, cells were spun at a low r.p.m. ($\sim 1,000$) for 20 min at 4°C . After spinning, the thawing solution was removed and cells were resuspended in $100\mu\text{l}$ of PBS + 5% FBS for antibody staining and cell sorting. The following antibodies were used for cell sorting: GlyA FITC (BD bioscience, clone HIR2), CD140b PE (BD bioscience, clone 28D4), CD45 PC5 (Beckman Coulter, clone IM1833), EpCAM PerCP-eFluor710 (eBioscience, clone 1B7), CD31 PC7 (eBioscience, clone WM-59), CD90 (BD Biosciences, clone 5E10), CD34 APC7 (BD bioscience, clone 581, custom conjugation). Cell sorting was performed on the BD FACSAria III using 4-laser configuration.

Laser capture microdissection. Snap-frozen tumour tissue embedded in optimal cutting temperature compound was cut into $8\mu\text{m}$ sections and mounted on PEN-Membrane Slides (Leica). Sections were stained with diluted haematoxylin to distinguish tumour epithelium from stroma. A staff pathologist marked tumour sections and LCM was performed according to manufacturer's protocol on the Leica LMD7000 system. Specimens were collected by gravity, contact-free and contamination-free, and directly placed in DNA lysis buffer.

Whole-genome sequencing was performed on DNA from tumour-enriched material. Details of sequencing protocols are included in the Supplementary Methods.

CELLULOID: evaluation of tumour cellularity, tumour ploidy, and absolute copy number profiles. After alignment, reads are counted in 1-kb bins using functions from the R package 'HMMcopy'. These counts are then adjusted for the GC content of each bin using LOESS (local) regression and scaled to the mean (scaled GC-corrected read count (SRC)). Segmentation of the data in both tumour and normal tissue (say, from matched non-malignant tissue or from blood) is performed using penalized least squares, as implemented in the R package 'copynumber'. Each segment is assigned the mean SRC value, which is calculated from the bins within the segment. SRC is proportional to the mean number of chromosomes (copies), averaged over all sequenced cells.

Germline heterozygous positions are extracted in the autosome, except in regions of the genome where duplication or deletion events are observed in the normal tissues. The number of reads supporting each allele (the reference allele—the one observed on the reference human assembly—and the alternate allele) is recorded from the tumour data and the allelic ratio (AR; the proportion of reads supporting the reference allele) calculated. Each heterozygous position is also paired with the SRC value of the segment it belongs to, evaluated from the tumour

data, to form pairs of values (SRC, AR). These pairs of points are represented in a three-dimensional graph as a contour (elevation) plot (Fig. 1). This figure is a visual representation of the autosomal-wide copy-number profile of the tumour. Each peak (or pair of peaks since the graph is reflected around $\text{AR} = 0.5$) corresponds to a specific copy number state that summarizes both the total copy number (on the x axis, once appropriately scaled) and the ratio of relative abundance of maternal and paternal copies (on the y axis, once contamination from normal tissues—or tumour cellularity—is accounted for). The relative positions of these peaks can be mathematically derived in the following way.

Let us define the autosomal ploidy of a sequenced sample (that includes both tumour and possibly contamination from normal cells) as:

$$P = \frac{1}{N_B} \sum_b c_b$$

where c_b represents the mean number of chromosomal copies at base b , averaged over all cells, and N_B is the number of autosomal bases. This can be interpreted as the relative abundance of autosomal DNA in the sequenced sample compared to a normal (reference) haploid autosomal genome. We aim to use the SRC values to estimate the ploidy. Re-writing the above as:

$$P \approx K \times \frac{1}{N_{\text{bins}}} \sum_{\text{bin}} \text{SRC}_{\text{bin}}$$

(where K is a scaling constant) is not informative since the SRC values are scaled and relative, making this expression trivial. However, because SRC are scaled to the mean, bins that fall in regions of exactly P copies (averaged over all cells) are expected to display SRC values of 1. Let S be the value of SRC that would be expected in regions where all cells display 2 copies of chromosomes (such regions do not need to actually exist in the sequenced sample). Because of proportionality, we have the relationship:

$$P = \frac{2}{S}$$

thus, ploidy can be evaluated by finding S .

Consider the more general case of a sequenced sample that consists of a proportion n of normal cells and t of tumour cells ($n + t = 1$). Because ploidy may differ in normal and tumour cells, these percentages are not equivalent to percentages of reads originating from normal or tumour cells. Consider a segment in the genome that is present in 2 copies in the normal cells and an average of T copies in the tumour cells. The tumour cells can be further broken down in subclones, in proportions t_1, t_2, \dots ($t = t_1 + t_2 + \dots$), each subclone displaying a different number of copies (T_1, T_2, \dots). Then, by proportionality, the SRC of bins in that segment are expected to take the value:

$$\frac{S}{2} \times (2n + T_1 t_1 + T_2 t_2 + \dots) \quad (1)$$

To determine the expected AR of heterozygous positions in that segment, the number of copies need to be further broken down into number of maternal and paternal copies: $T_i = M_i + P_i$. Normal cells are assumed to have one maternal chromosome and one paternal chromosome. In a segment that displays M_i maternal and P_i paternal copies in subclone i , the AR is expected to take the value:

$$\frac{n + M_1 t_1 + M_2 t_2 + \dots}{2n + T_1 t_1 + T_2 t_2 + \dots} \quad (2)$$

if, say, the maternal chromosome carries the reference allele, and reflected around 0.5 otherwise. Let:

$$\text{EP}(S, n, t, M_1, P_1, M_2, P_2, \dots)$$

represent the (x, y) coordinates described in equations (1) and (2). Let $\text{OP} = \{\text{OP}_i\}$ be the set of observed contour plot peaks (or subset of peaks deemed of particular interest by the user). The algorithm used to estimate S , n and t finds parameters that minimize the total distance between the observed peaks and the expected peak (EP) coordinate closest to each. In other words, if:

$$d_i(S, n, t) = \min_{M_1, P_1, M_2, P_2, \dots} |\text{OP}_i - \text{EP}(S, n, t, M_1, P_1, M_2, P_2, \dots)|,$$

then the algorithm consists on finding S , n and t that minimize:

$$\sum_i d_i(S, n, t).$$

In practice, the number of expected peak locations grows exponentially with the number of subclones and the number of maternal/paternal configurations. The algorithm further depends on a set of allowed copy number configurations (a set of M_i and P_i) that can be set by the user. For example, the user might want to ignore configurations where the number of maternal chromosomes is smaller than the number of paternal chromosomes in one subclone but higher in another; this would reduce the number of possible ARs. Other restrictions may include situations where the number of copies between different subclones cannot differ (by difference or by ratio) by more than some specified threshold.

The objective function to be minimized is not convex and multiple local minima exist. Optimization is done either by simulated annealing if a global minimum is desired (using the R package GenSA) or using the R built-in function 'optim' with grid-defined starting points to survey and inspect a set of local minima.

Once values for S , n and t are obtained, the ploidy in the tumour cells (P_T) can then be calculated as:

$$P_T = \frac{P - 2n}{1 - n}$$

where $P = 2/S$ is the ploidy of the whole sample that was sequenced. The SRC values can be rescaled into their corresponding integer copy number in tumours using equation 1 above.

The above describes the current implementation of an R package named CELLULOID, which can be obtained from <http://github.com/mathieu-lemire>.

Chrom-AL: detecting catastrophic mitotic events. Chrom-AL is an in-house tool developed to standardize the detection of complex rearrangement patterns linked to chromothripsis²⁰. Chrom-AL applies a series of statistical tests and thresholds at the level of the chromosome and also within the windows of the structural events to infer a call. We inspected 80 genomes manually and estimate that the false-positive and false-negative rate of Chrom-AL is $\sim 7\%$ and $\sim 8\%$, respectively, in our dataset. The tool is designed based on the chromothripsis criteria presented by Korbel and Campbell³⁰. Complex rearrangement patterns can often involve multiple distinct types of mitotic errors (for example, FoSTeS, MMBIR) including a chromothripsis event^{20,21,23}. Chrom-AL is not designed to distinguish chromothripsis from other replication-based mitotic errors, which can also be catastrophic within one or few cell divisions. As such, we use the term chromothripsis to broadly refer to a 'one-off' mitotic catastrophe.

As chromothripsis events typically increase the number of structural variants in a genome, there is a correlation between tumours with increased numbers of structural variants and rate of chromothripsis. Thus, proper structural variant calling becomes a critical parameter in implementation of any algorithm to call chromothripsis. Despite this correlation, a high rate of structural variants does not necessarily imply a chromothripsis event. Thus, the false-positive and false-negative rates of Chrom-AL will probably vary with the overall rate of structural variants that differs amongst tumour types. For this reason, visual inspection still remains a critical tool in evaluating such events. Chrom-AL does not detect chromothripsis events that are predominately driven by a single type of structural variation. For example, on rare occasions we observed the typical copy number oscillation hallmark of chromothripsis that was connected mostly by head-to-head (HH) or tail-to-tail (TT) inversions. Whether such rearrangements were indeed accumulated over time or all at once is not known. To remain consistent with the criteria discussed below, we excluded these events from the analysis. Below, we describe the criteria and conditions used to detect cataclysmic events by Chrom-AL. Chrom-AL was implemented in R.

Threshold for number of structural variants and copy number alterations at the (chromosomal level); test 1. Catastrophic events typically have large numbers of structural variations and copy numbers changes. Only events with at least 7–8 structural variants and 8 copy number segments were considered in the analysis.

Clustering of break points (chromosomal level); test 2. Catastrophic events are typically localized to particular genomic regions that can be assessed statistically. To do this, we ordered the break points sequentially and calculated the distances between each break point. The distribution of distances was compared against the exponential distribution as described by Korbel and Campbell³⁰ using a Kolmogorov–Smirnov (KS) test and followed by Bonferroni correction. Regions with a $q \leq 0.1$ were considered to display evidence of break-point clustering.

Chromosomal break-point enrichment (chromosomal level); test 3. We observed several instances where structural variants comprising a catastrophic event were scattered chromosome-wide and did not cluster within a particular region of a chromosome. Thus, they failed the KS test described above. To account for this shortcoming, we performed an additional test to determine if structural variants were enriched on any particular chromosome than would be expected by chance. To identify chromosomes enriched for structural variants, a hypergeometric test was run on each chromosome based on all the breakpoints identified

in the tumour. This was followed by a Bonferroni correction. Chromosomes with a $q \leq 0.1$ were identified as having a high rate of break points.

Join distribution (chromosomal and window level); test 4. In paired-end sequencing, all structural variants can be categorized into four read-pair orientations based on the direction of the + or – reads: tail-to-head (+/–, TH), head-to-head (–/–), tail-to-tail (+/+) or head-to-tail (–/+, HT). Pairs in standard orientation (+/–) are considered to be a deletion-type structural variant with a TH join. Duplication-type structural variants are in the reverse-orientation –/+ and defined by a HT join. Inversions can be both in the forward (+/+) orientation or reverse (–/–) orientation. In the forward orientation, they were defined as TT and in the reverse orientation they were defined as HH. Using read-pair information for structural variants, we classified each structural variant based on their segment joins. In a catastrophic event, we expect structural variants of all four types to be present. For each region we tested this hypothesis. To initially run the test, we required at least one type of read-pair join from each of the four subtypes to be present. A multinomial test, from the EMT v1.1 package, was run to test the distribution of segment joins against an equal distribution. The regions with $P > 0.05$ were considered to show evidence of equal distributions of segment joins.

Copy-number oscillations (chromosomal and window level); test 5. Catastrophic events typically display oscillations in copy number that vary between a few states. However, when chromothripsis is co-opted with BFB cycles as part of a single catastrophic event, there will be some segments in the event that will oscillate between limited copy number states and other segments that may appear to increase in a stepwise manner. To be categorized as a bona fide one-off event, there must be some sequential segments that retain an oscillation pattern. We required at least 4 sequential segments in any catastrophic event must oscillate between two different states. Due to polyploidization, the amplitude of the copy number step was defined as variable (1, 2 or more).

Interspersed LOH (chromosomal and window level); test 6. Chromothripsis drives copy number losses, and thus copy number oscillations should correspond to interspersed loss of heterozygosity (LOH). To test for LOH, we identified all the high confidence germline heterozygous SNPs in the genome and determined the allelic ratio in the tumour sample. The distributions of allelic ratios between each sequential copy number segments were compared using a t -test. A minimum of ten positions had to be identified within each copy number segment to be processed otherwise those segments were excluded from the analysis. A Bonferroni correction test was run. Those segments in which $q \leq 0.1$ were considered significantly different. To show evidence of interspersed LOH, at least four comparisons had to be made (thus at least five copy number segments had to be present in the region). At least 50% of the compared segments had to show some significant difference in the distribution of allele ratio to be classified as showing interspersed LOH.

Chromosome-level analysis. Genomic regions were first evaluated at the chromosome level. For each sample, all chromosomes were independently evaluated for the above tests. For tests 2 and 3, we used copy number break points for segments where a matching structural variant could not be mapped. The importance of this point is shown in Extended Data Fig. 5c (bottom left panel; Ashpc_0008, event 2). In this case, there was a chained chromothripsis event connecting chromosomes 3 and 20. On chromosome 3, the left edge at 42.8 Mb was part of the chromothripsis event but the corresponding structural variant to this copy number loss is not mapped. This was also the case for the right edge of the chromothripsis event on chromosome 20 (7.1 Mb). In this scenario, utilization of the copy number break point was critical in the tests to decipher whether this was indeed a chromothripsis event. If copy number break points are not integrated into the analysis, such events would go undetected or be misclassified. We found that including the copy number break point was necessary to properly establish the DNA windows of chromothripsis events, especially when structural variants could not be properly mapped (discussed below).

Identification of DNA rearrangement windows. The next step was to identify the borders of the catastrophic event on each chromosome. Catastrophic events typically display overlapping structural variants throughout the region of the event. To localize the chromosomal window where the catastrophic event occurred, we selected the left and right borders of overlapping structural variant break points. Structural variants resulting in translocations were used to establish the rearrangement window when at least two independent translocations were detected between the same two chromosomes. In this manner, we could establish inter- and intra-chromosomal windows to facilitate the segregation of multi-chromosome events from single-chromosome events. Each window was flanked with 6 kb on either end. The windows that define each candidate catastrophic event were used for downstream analysis.

Window-level analysis. A window was first scored on whether there were at least eight structural variants present within the window. Each window was then evaluated for tests 4–6.

Classification of single-chromosome versus multi-chromosome catastrophic events. Single-chromosome catastrophic events were classified when all structural variants within a window occurred on the same chromosome. In the case of translocations, at least two structural variants had to have occurred between the same two chromosomes to be considered a multi-chromosome event.

Event (criterion 1 versus criterion 2). Each window was independently scored. If a window was classified as a catastrophic and was involved in a multi-chromosome event, both windows on either side of the translocation were considered to be catastrophic at the chromosomal level but were counted as a single catastrophic event. Through a large number of iterations, in which the tests described above were iteratively optimized, we established two distinct criteria: 'maximize sensitivity' and 'specificity of detection'.

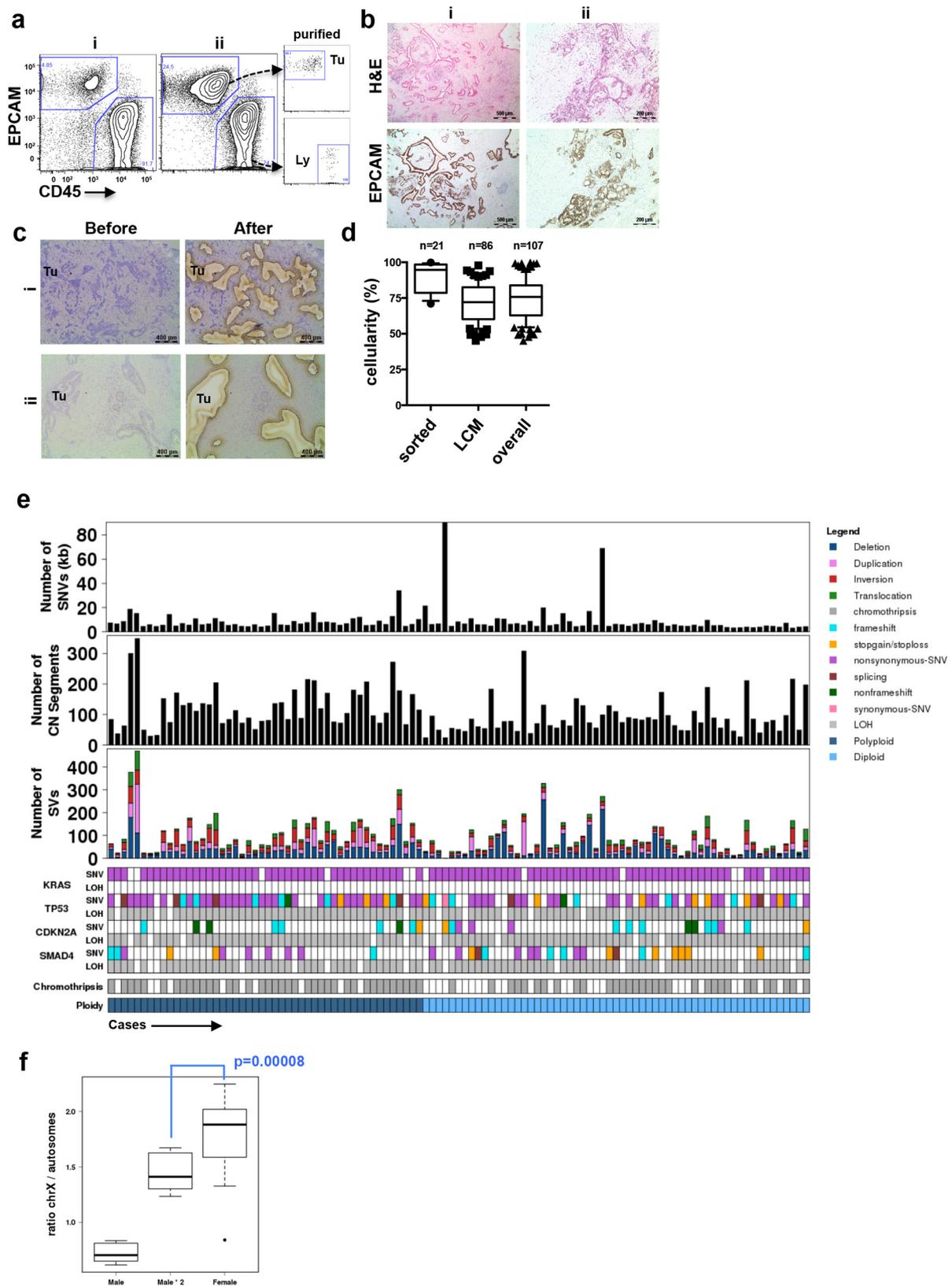
Criterion 1 (CR1). To be classified as an event under CR1, a region had to pass at least five of the six chromosomal level tests (test 1–6). A window had to be

identified with at least eight structural variants and the window had to pass the segment-join and the interspersed LOH test (test 4 and 6).

Criterion 2 (CR2). To be classified as event under CR2, a region had to pass at least 5 of the following tests: the 6 chromosomal level tests (test 1–6), the identification of a window with at least 8 structural variants, the window segment join and the window interspersed LOH test (test 4 and 6). In addition to these conditions, the window had to have at least 7 structural variant events and had to pass window oscillation (criterion 5).

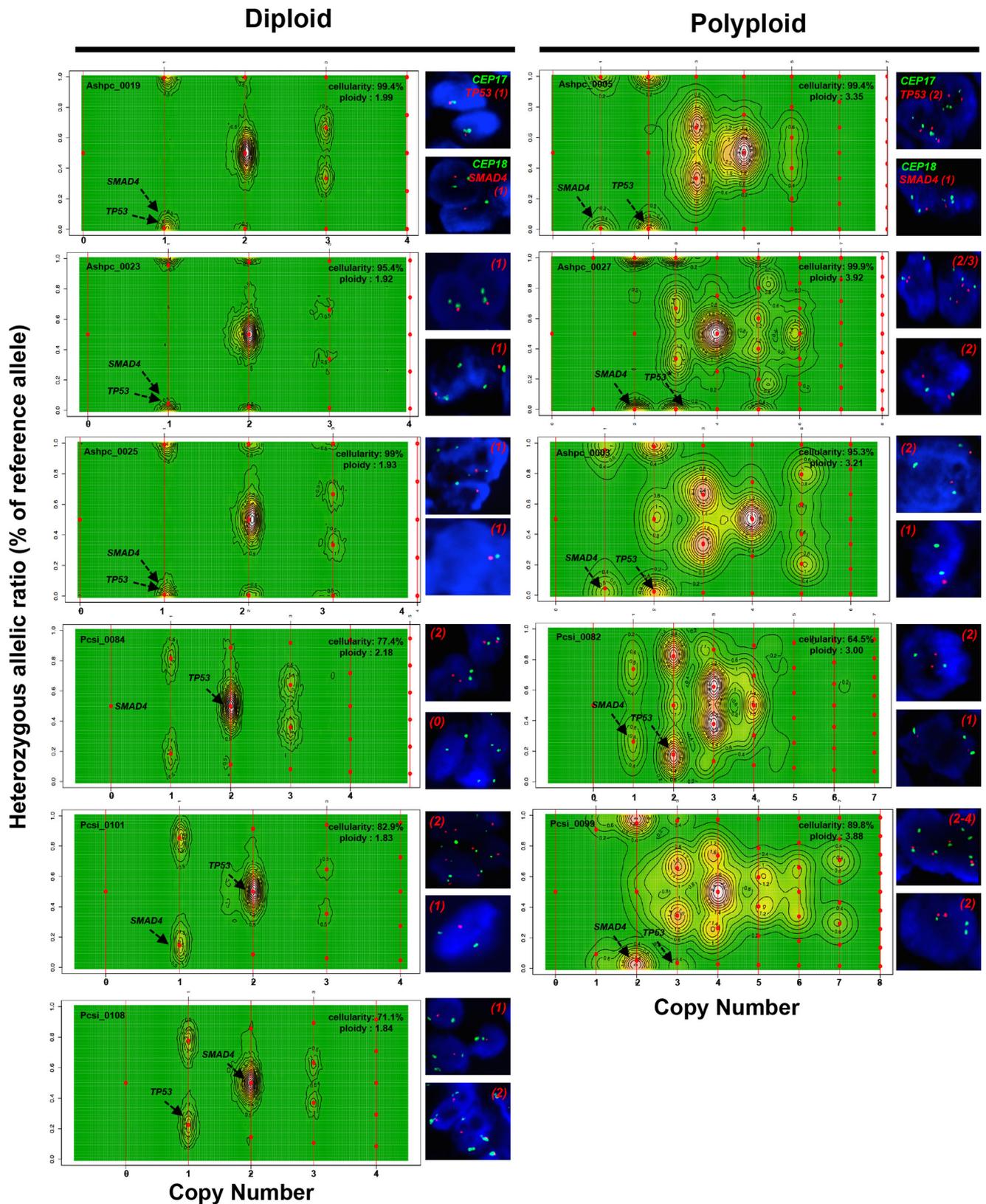
Data availability. Raw data (fastq files) and clinical information on the patient cohort are available from the International Cancer Genome Consortium (ICGC) data portal at <http://dcc.icgc.org>. DNA sequencing data have also been deposited in the European Genome-phenome Archive (EGA): EGAD00001001956.

30. Korbelt, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).



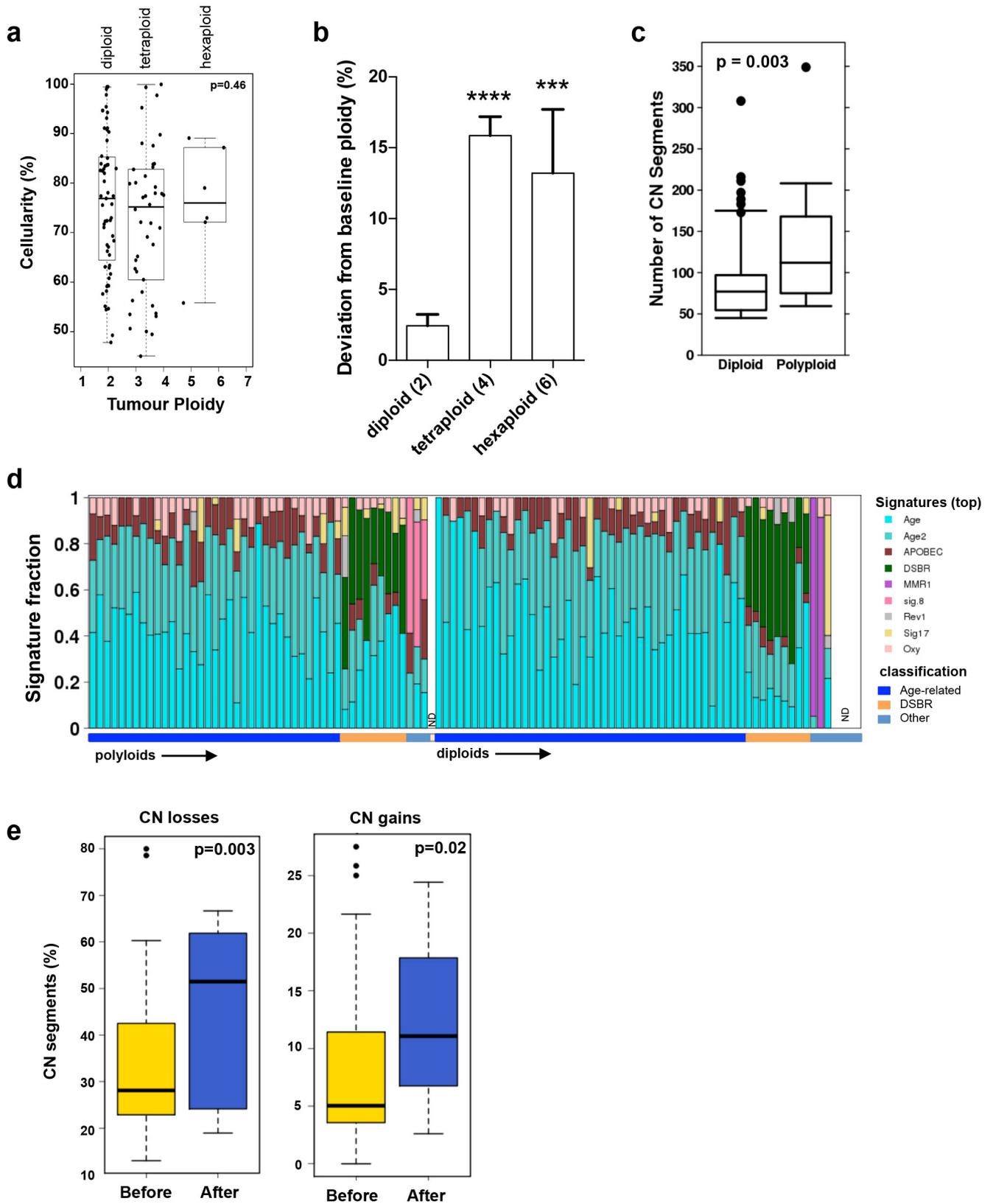
Extended Data Figure 1 | Tumour enrichment and overview of somatic alterations in the cohort used in this study. **a**, Flow cytometry profiles of EpCAM and CD45 from 2 (of 21) representative cases of pancreas ductal adenocarcinoma (PDA) (i,ii). On the right, post-sort analysis of EpCAM⁺ cells (Tu) and CD45⁺ lymphocytes (Ly) demonstrates the high level of purity obtained from flow-sorting. **b**, Immunohistochemical analysis of formalin-fixed tumours using the EpCAM clone for flow sorting in a (H&E, haematoxylin and eosin). Two independent cases are shown (i, ii). **c**, Profiles of haematoxylin-stained sections of PDA before and after LCM from two representative cases (of 86) (i, ii). **d**, Box and whisker

plots represent median and 10–90 percentile ranges of tumour cellularity of flow-sorted ($n = 21$), LCM ($n = 86$) and the total cohort ($n = 107$) of tumours. Dashed line depicts cellularity of bulk tumours that have not undergone enrichment. **e**, Overview of somatic alterations of the cohort used in the study. **f**, X-chromosome mutation ratio in diploid pancreatic cancer genomes showing hypermutation on this chromosome in females. Males were corrected for single copy of the X chromosome by doubling the raw value. P values were derived from t -tests. A more detailed description of these data is provided in Supplementary Results.



Extended Data Figure 2 | CELLULOID validation. The copy number for common alterations (*TP53*, *SMAD4*; shown by black arrow) was derived from ploidy estimates generated by CELLULOID. Six diploid and five polyploid tumours were analysed by FISH (shown on the right of each contour plot). In all cases, the copy number from CELLULOID ploidy estimates was confirmed. In Pcsi_0084 (diploid), CELLULOID predicted zero copies of *SMAD4*. The allelic ratio in this region was 50%

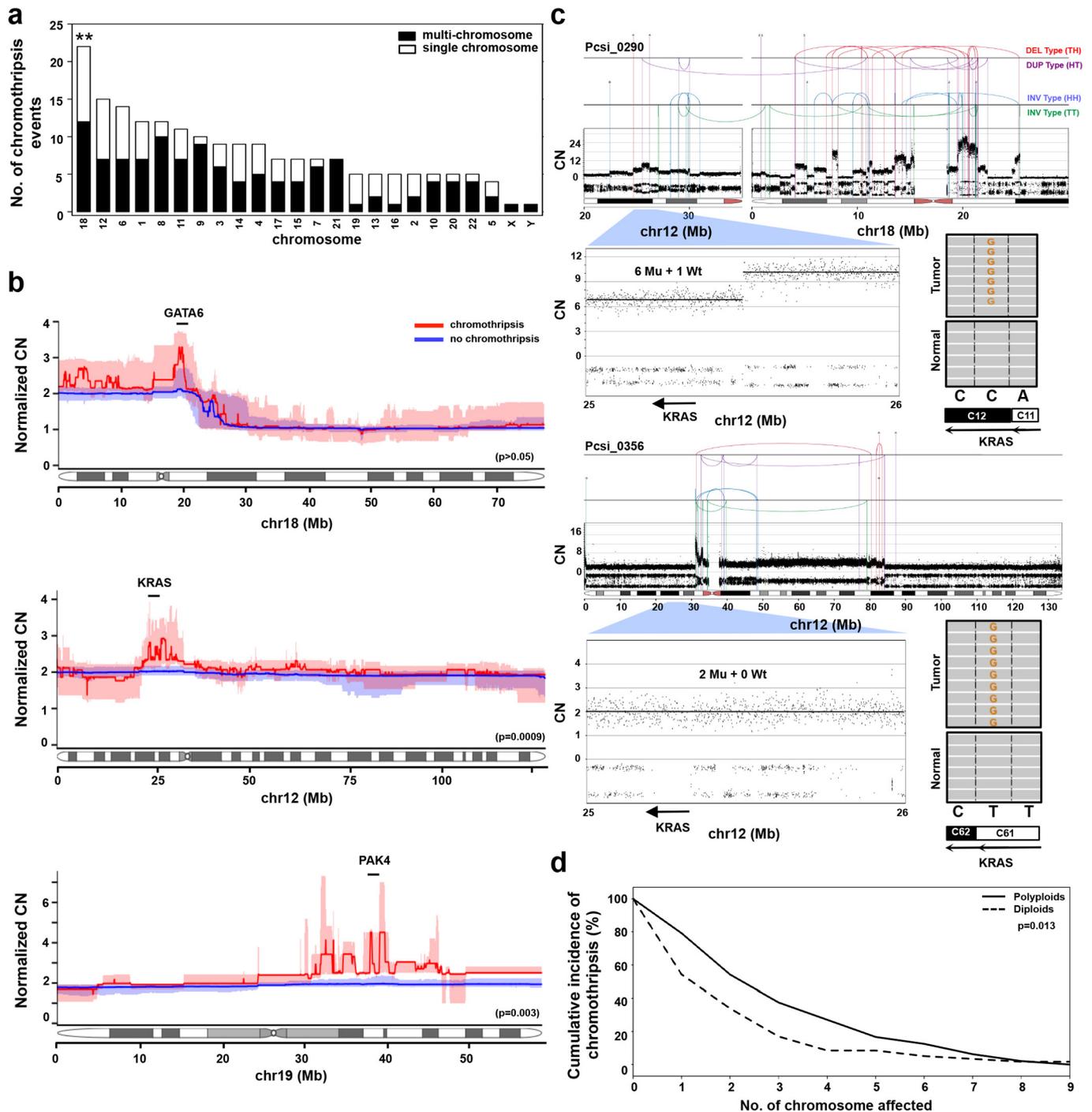
(heterozygous) as only reads from normal cells spanned this region. In Ashpc_0027, both CELLULOID and FISH indicate that this tumour is polyploid. The CELLULOID plot demonstrates that there is a further subclonal amplification in *TP53* from polyploid clone (copy state = 3.2 derived from one allele). FISH analysis shows tumour cells with two or three copies of *TP53* supporting this is subclonal. Copy number by FISH for *SMAD4* and *TP53* is indicated in red at the top right of each plot.



Extended Data Figure 3 | See next page for caption.

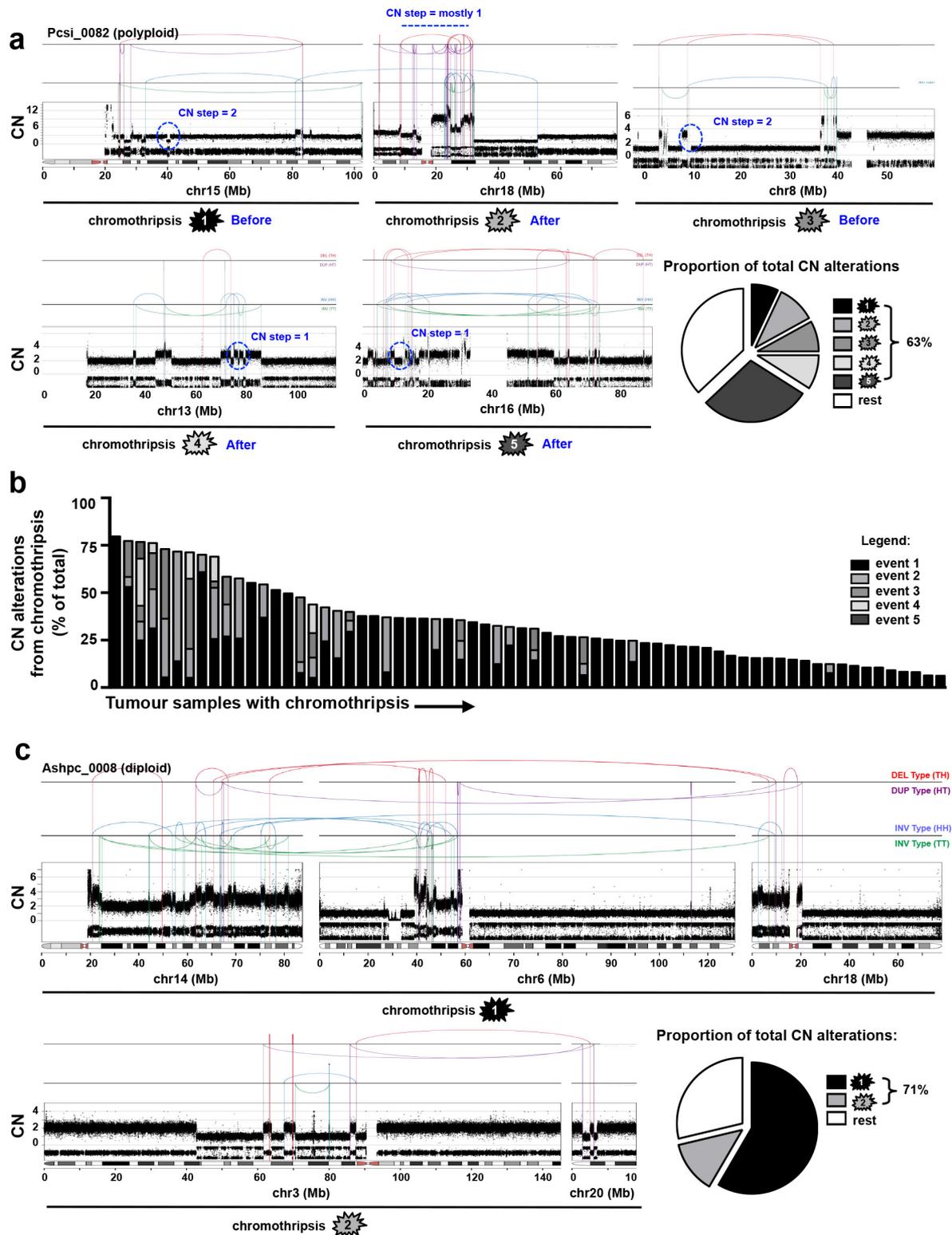
Extended Data Figure 3 | Tumour ploidy and genetic instability in pancreatic cancer. **a**, Tumour ploidy and sample cellularity estimates are interconnected: although the ploidy of a tumour can always be doubled and still provide copy number segments at integer levels (albeit only at even values), the estimate of cellularity would have to decrease. To maintain an allelic ratio at a given value, the proportion of tumour cells has to be reduced to compensate for the higher copy numbers in them (from a cellularity value t to a value $t/(2 - t)$ in the case of a doubling of the ploidy). A test can thus be designed to verify that ploidy estimates have not been systematically over- or underestimated, simply by comparing the distribution of cellularity estimates stratified by ploidy. P value was derived using Kruskal–Wallis test. **b**, Deviation from baseline ploidy in diploids, tetraploids and hexaploids indicates a marked loss of genomic material in polyploids. **c**, Box and whisker plots (showing the median

and 10th–90th percentile ranges) of the total copy number alterations in polyploid and diploid tumours. **d**, Mutational signatures of the 107 genomes used in this study. The signatures were derived using the trinucleotide mutation context as previously published¹⁸. The proportion of individual signature operative in each tumour is shown in the bar plot. The overall classification of each case is indicated below. Signatures of polyploidy tumours is shown on the left, diploids is shown on the right. ND, not done; $n = 1$ polyploid and 4 diploid patient samples. Detailed analysis of mutational signatures in PDA is covered elsewhere (Connor *et al.*, manuscript under review) **e**, Percentage of copy number losses (left) and gains (right) that occurred before (yellow) or after (blue) genome duplication for each polyploid tumour. Box and whisker plots depict median \pm 10th–90th percentile range. P values were derived using a t -test.



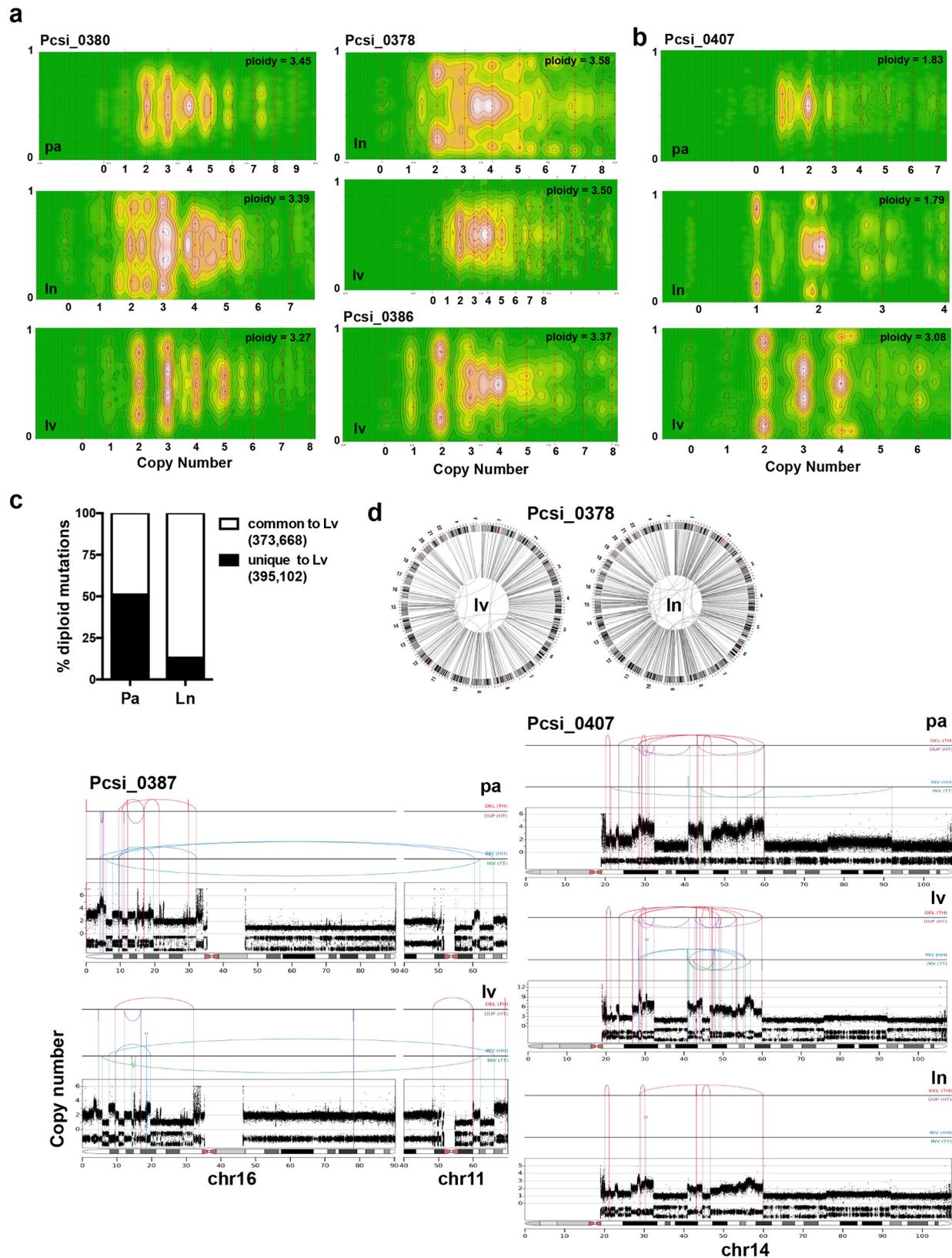
Extended Data Figure 4 | Characterization of chromothripsis events in pancreatic cancer. **a**, The distribution of chromothripsis events across the genome (single-chromosome, white; multi-chromosome, black). $**P < 0.001$ (Monte Carlo sampling, Supplementary Methods). **b**, The specific effects of chromothripsis on the copy number of chromosome 18 (top, $n = 22$), chromosome 12 (middle, $n = 15$), and chromosome 19 (bottom, $n = 5$). Statistical differences in copy number between the groups were performed using Wilcoxon test using 10-kb bins that covered *GATA6* (chromosome 18), *KRAS* (chromosome 12) and *PAK4* (chromosome 19) genes (description of *PAK4* event is covered in supplementary results). Copy number profiles of polyploids were adjusted according to tumour

ploidy to allow comparison against diploids (referred to as 'Normalized copy number' on the y axis). Interquartile ranges for chromothripsis cases are indicated in pale red and for non-chromothripsis cases in pale blue. **c**, Two cases of chromothripsis resulting in the amplification of the mutant *KRAS* allele. In Pcsi_0290, the mutant allele was amplified as part of a multi-chromosomal event involving chromothripsis and BFB with chromosome 18 (top). In Pcsi_0356, the chromothripsis event was co-opted with cycles of BFB to knock out the wild-type allele (bottom). The absolute copy number of the locus encompassing *KRAS* and mutation is shown for each case. **d**, Cumulative incidence of chromothripsis events in polyploid and diploid tumours ($P = 0.013$, Fisher's exact test).



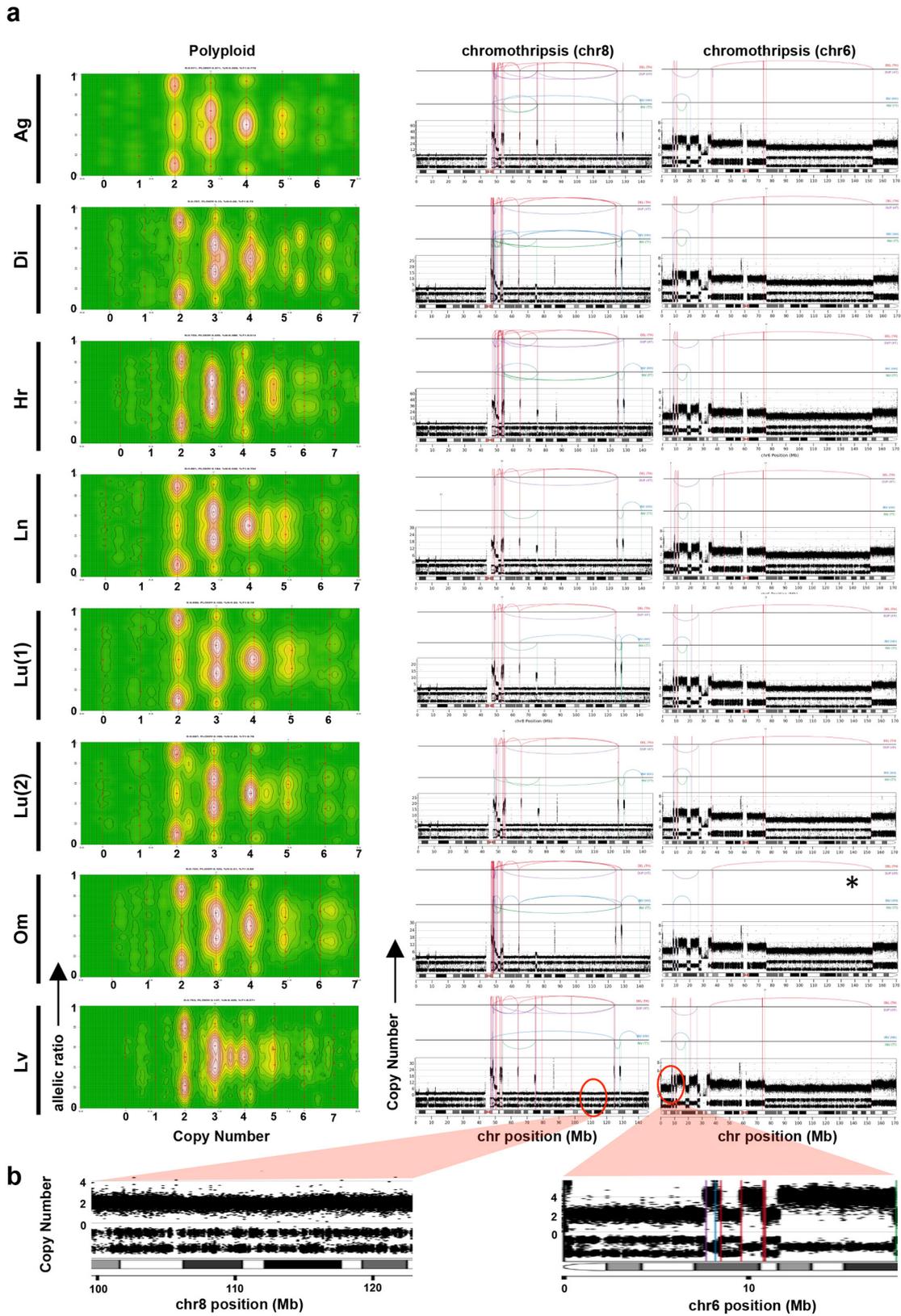
Extended Data Figure 5 | Most copy number alterations arise from individual chromothripsis events. **a**, In Pcsi_0082, five distinct chromothripsis events on chromosome 15 (top, 1), chromosome 18 (top, 2), chromosome 8 (top, 3), chromosome 13 (bottom, 4), and chromosome 16 (bottom, 5) are displayed. Copy number steps on chromosome 15 (1), chromosome 8 (2) are 2 or greater indicating that these events occurred before polyploidization. Single copy number steps on chromosome 18 (2), chromosome 13 (4) and chromosome 16 (5) indicate that these events were sustained after polyploidization. The single rearrangement between

chromosome 15 and chromosome 18 appears to be independent from the chromothripsis on chromosome 18. Pie charts depict the proportion of copy number alterations derived from each chromothripsis event. **b**, Distribution of copy number alterations due to chromothripsis for all cases where such an event was detected by ChromAL. **c**, In Ashpc_0008, two multi-chromosomal chromothripsis events, joining chromosome 14, chromosome 6, chromosome 18 (top, 1), and chromosome 3, chromosome 20 (bottom, 2), are shown (discussed in Supplementary Results).

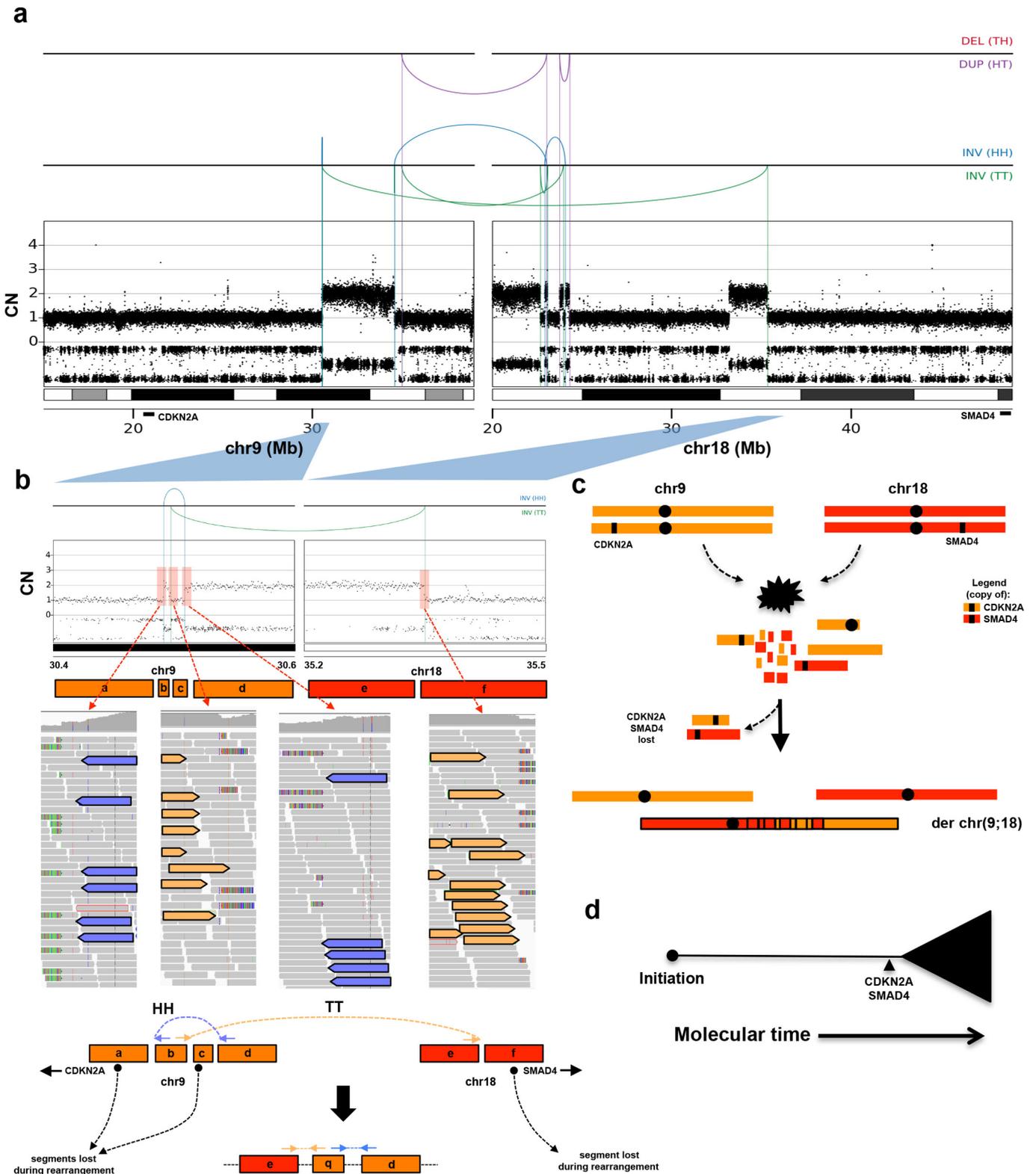


Extended Data Figure 6 | Characterization of chromothripsis and polyploidy in metastases. **a**, CELLULOID plots illustrating polyploidy in metastases. In Pcsi_0380, the primary tumour was directly available for analysis. Similarly to Pcsi_0378, multiple metastases were polyploid suggesting the primary tumour was also polyploid. The primary tumour was unavailable for sequencing in this case. **b**, A case (Pcsi_0407) with

discordant ploidy amongst different metastases. **c**, Percentage of diploid mutations from liver metastases that are shared (white) or unique (black) when compared to the primary tumour or the lymph node metastasis. **d**, Plots of chromothripsis events in metastases. ln, lymph node; lv, liver; pa, primary tumour.

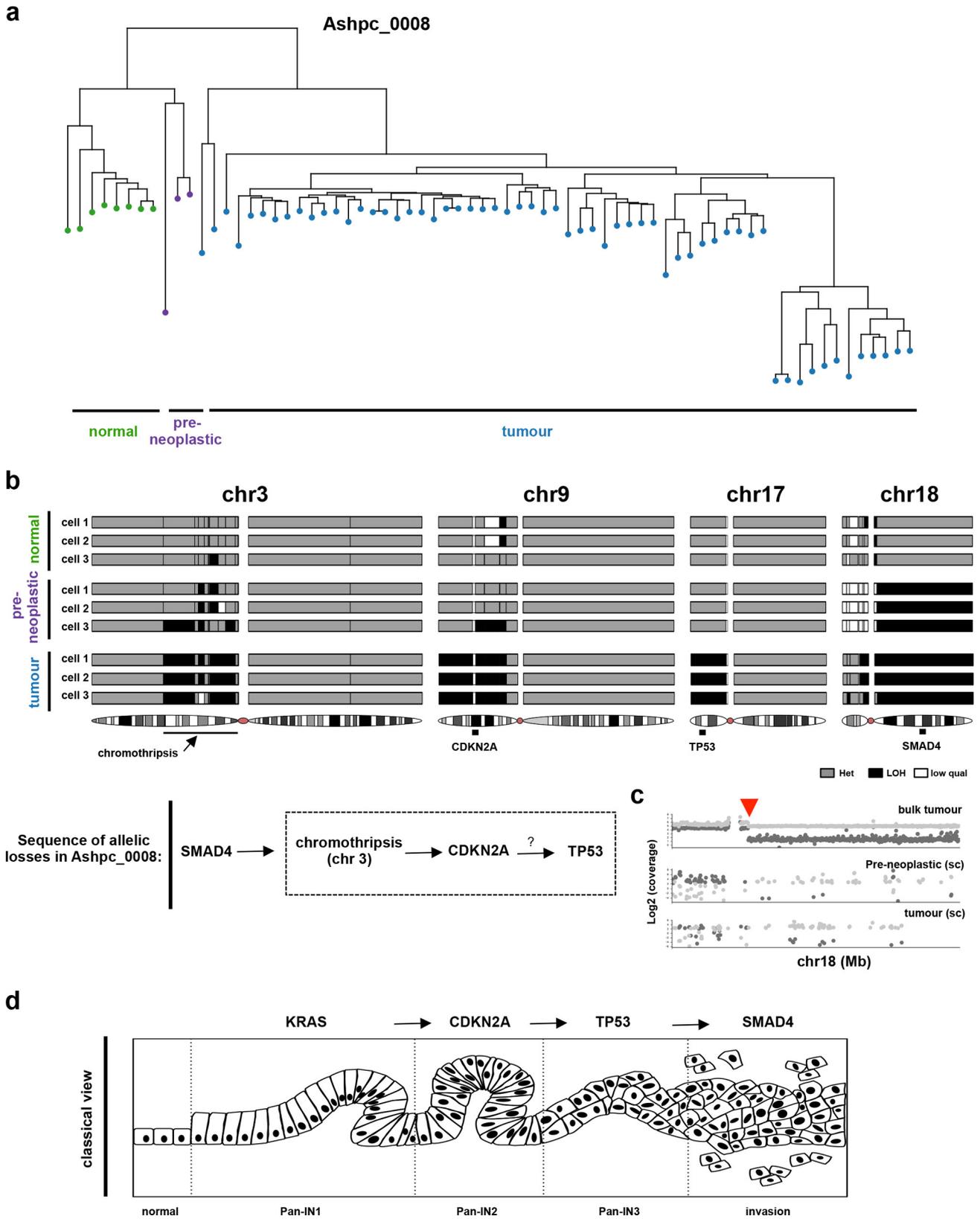


Extended Data Figure 7 | Chromothripsis and polyploidy in Pcsi_0410. This figure accompanies Fig. 2. **a**, CELLULOID (left panel) and chromothripsis plots (middle and right) of the different metastases from a patients with fulminant metastatic progression. **b**, Copy number and LOH from chromosome 8 (left) and chromosome 6 (right) chromothripsis events indicate that these events were sustained before polyploidization.



Extended Data Figure 8 | Case of a simultaneous loss of *CDKN2A* and *SMAD4* due to a chromothripsis event. a, Rearrangement and copy number profile of a multi-chromosome chromothripsis event between chromosome 9 and chromosome 18 (Pcsi_0171). **b**, Detailed view of the two inversions (one in the head-to-head orientation (HH), the other in tail-to-tail orientation (TT) for more detail, see Methods) in the

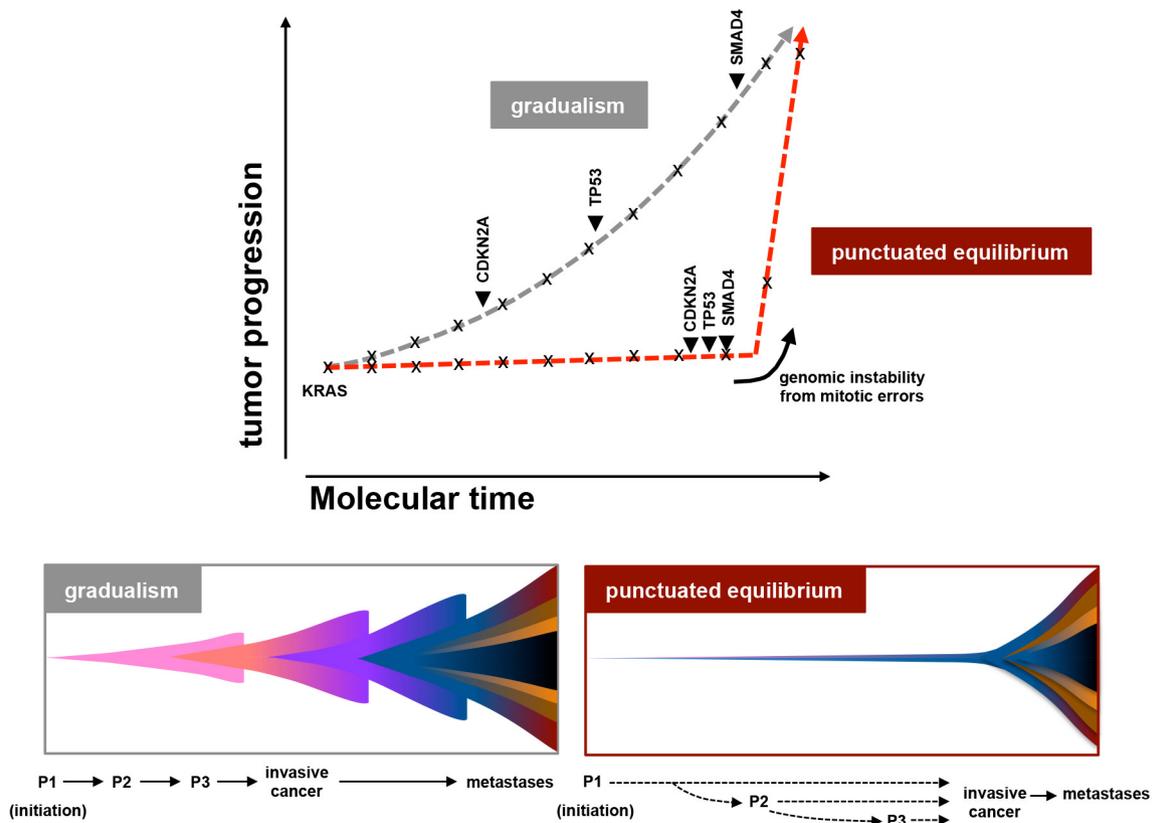
chromothripsis event that resulted in the concurrent loss of *CDKN2A* and *SMAD4*. **c**, Schematic depiction of the temporal order of events derived from the rearrangement profile shown in **a**. **d**, Summary of tumour evolution in Pcsi_0171. A more detailed description of Pcsi_0171 is provided in Supplementary Results.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Single-cell sequencing reconstruction of the evolutionary events when rearrangements did not span the classical pancreatic cancer drivers. **a**, A fresh tumour specimen (Ashpc_0008) was dissociated and single tumour cells were deposited using flow sorting. The whole genomes of 96 single cells were amplified using REPLI-g and paired-end whole-genome sequencing was performed using an Illumina HiSeq 2500 system. Single cells were sequenced to a median whole-genome depth of $3.9\times$ (Supplementary Fig. 18). Only cells with enough whole-genome coverage ($n = 70$) were used in the analysis. This sequencing depth allowed us to track heterozygous SNPs across the whole genome in single cells. Using this methodology, we were able to follow LOH events across the whole genome in single cells that show high concordance with

bulk tumour tissue (Supplementary Fig. 18). Hierarchical clustering based on LOH events across the whole genome was performed and found four independent cell clusters. **b**, Specific LOH events on chromosome 3, chromosome 9, chromosome 17 and chromosome 18 are shown from single cells in **a**. The chromothripsis event on chromosome 3 is shown in greater detail in Extended Data Fig. 5c. A summary of the sequence of allelic losses is shown below. Supportive data that allelic losses precede mutational inactivation is shown in Supplementary Figs 13, 14. **c**, Plot of the shared chromosomal break point on chromosome 18 on the bulk (top), preneoplastic single cell (middle) and tumour single cell (bottom). **d**, The classical model of pancreatic tumour progression.



Extended Data Figure 10 | Theoretical model of pancreatic cancer tumour progression. Shown is the classical model of tumour evolution driven at a gradual pace (grey) and an alternate model driven at punctuated equilibrium (red). In the classical model, there is a period of latency between the driver mutations that lead to tumour development and the multiple, independent, transforming events are required for tumour development (top, grey dashed line; bottom-left schematic). In the punctuated equilibrium model, tumour development can be divided into two major events, the cancer-initiating event and cancer-transforming

event (top, red dashed line; bottom-right schematic). Under this model, most mutations (indicated by \times) would accrue in an extended phase of preneoplastic tumour development. Transformation, probably due to genetic instability from copy number changes (arrow heads) ensuing from a cataclysmic event, would rapidly lead to invasive cancer and metastases. Classical drivers (*KRAS*, *CDKN2A*, *TP53*, *SMAD4*) from the PanIN progression model are overlaid onto these models. Theoretical PanIN stages are shown as P1–P3.