nature genetics

Article

https://doi.org/10.1038/s41588-025-02468-9

Genomic evolution of pancreatic cancer at single-cell resolution

Received: 29 September 2024

Accepted: 3 December 2025

Published online: 22 January 2026

Check for updates

Haochen Zhang[1,2,13,16], Palash Sashittal[3,14,16], Elias-Ramzey Karnoub[2,4,15], Akhil Jakatdar[3], Shigeaki Umeda[2,4], Jungeui Hong[2,4], Anne Marie Noronha[5], Agustin Cardenas III[4], Amanda Erakky[4], Caitlin A. McIntyre[6], Akimasa Hayashi[4,7], Nicolas Lecomte[2,4], Marc Hilmi[4,8], Wungki Park[4,9,10], Nan Pang[11], Eileen M. O'Reilly[4,9,10], Alice C. Wei[4,10,11], Benjamin J. Raphael[3] ✉ & Christine A. Iacobuzio-Donahue[2,4,12] ✉

Most evolutionary studies on pancreatic cancer rely on bulk sequencing, yet clonal evolution happens at the single-cell level. We used single-nucleus DNA sequencing to study 137,491 single nuclei from 24 pancreatic neoplasms reflecting various clinical scenarios. We found higher frequencies of somatic alterations to driver genes that bulk studies indicate; many manifest as copy number alterations and account for the majority of spatial heterogeneity. In pancreatic cancers with canonical *KRAS* oncogenic mutations, we found likely varied dependence on the genotype that may signify differential response to KRAS inhibition. In pancreatic cancers with germline heterozygous *BRCA2* mutations, we discovered varied mechanisms and timing of inactivation of the wild-type allele that sculpted differential evolutionary trajectories. Inactivation of tumor-intrinsic response to transforming growth factor-β happens through various mechanisms, takes place after oncogenesis and coincides with invasion and metastasis, reflecting increasing selective pressure for the phenotype later in pancreatic ductal adenocarcinoma development.

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal cancer types with a 5-year survival rate of 13%[1]. This rate has increased minimally over the past decades despite improvements in surgical and medical management, with PDAC projected to become the second most common cause of cancer-related deaths in the USA[2]. Numerous next-generation sequencing studies have revealed PDAC clonal evolution as a hybrid model of stepwise and punctuated events. Point mutations in multiple driver genes (such as *KRAS*, *CDKN2A* or *TP53*) followed by allelic loss of the wild-type alleles are gained through progressive waves of mutation and selection[3], while pervasive copy number variations (CNVs) or larger-scale genomic rearrangements that further grant growth advantages are acquired rapidly after polyploidization. Polyploidization coincides temporally with the loss of the wild-type *TP53* allele and the acquisition of invasive potential[4,5]. The identification of these genomic events has largely been accomplished by sequencing of bulk tissues; in some cases, resolution has been increased by the use

[1]Gerstner Sloan Kettering Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [2]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [3]Department of Computer Science, Princeton University, Princeton, NJ, USA. [4]David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [5]Marie-Josée & Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [6]Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA. [7]Department of Pathology, Kyorin University, Mitaka-shi, Tokyo, Japan. [8]Institut Curie, Paris, France. [9]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [10]Department of Medicine, Weill Cornell Medicine, New York City, NY, USA. [11]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [12]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York City, NY, USA. [13]Present address: Valar Labs, Palo Alto, CA, USA. [14]Present address: Department of Computer Science, Virginia Tech, Blacksburg, VA, USA. [15]Present address: Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA. [16]These authors contributed equally: Haochen Zhang (张皓辰), Palash Sashittal. ✉e-mail: braphael@princeton.edu; iacobuzc@mskcc.org

of highly purified microdissected samples and/or development of computational tools to deconvolve the complexity of subclonal mixtures[6–8].

Regardless, because clonal evolution occurs at the single-cell level, bulk sequencing would be unable to order successive events or to distinguish mutually exclusive events, indicating distinct populations. Thus, single-cell DNA sequencing (snDNA-seq) is required to further determine the clonal evolution and heterogeneity of PDAC. This goal is urgent because collaborative research efforts across the globe have revealed several opportunities for precision medicine. New therapeutic approaches such as poly-ADP ribose polymerase inhibitors, KRAS inhibitors and immunotherapies are being developed to target individual tumors' vulnerabilities with promising results[9–12]. Nonetheless, responses to these therapies have been mixed with some patients' showing no benefit[13].

We previously optimized a scalable workflow to apply high-throughput snDNA-seq to archival PDAC samples[14]. Here we used this workflow and developed an accompanying set of computational methods to generate the highest-resolution view of general genomic evolution patterns over the course of PDAC evolution, spanning early invasion to dissemination to secondary sites. Such high resolution that preserves information regarding co-occurring compared to mutually exclusive mutations has revealed insights into PDAC biology, ranging from innate mechanisms of KRAS resistance to genetic features of the most recent common ancestor (MRCA) tumors in patients with germline *BRCA2* mutations, and finally to molecular pathways that are under relatively constant selection.

Results

PDAC mutational and evolutionary landscape at single-nucleus resolution

Cohort overview. We analyzed single-cell genomic features in 24 patients with primary pancreatic neoplasms. The cohort included samples from patients with both early-stage disease collected from surgical resections ($n = 10$ patients) and late-stage unresectable disease from research autopsies ($n = 11$ patients; Fig. 1a and Supplementary Table 1). Matched longitudinal biopsies taken by 18-G/22-G core needle biopsies before and after treatment were collected from six patients (Supplementary Table 2). The final pathologic diagnosis for 23 patients was PDAC, and for 1 additional patient, it was an acinar cell carcinoma (ACC; patient PC22). We opted to study the latter patient, given their clinical-grade sequencing indicated a germline mutation in *BRCA2*; five of the PDAC patients had documented germline *BRCA2* mutations as well, providing an opportunity to understand genomic features of neoplasms that arise in the setting of this set of high-risk germline variants. Collectively, a total of 30 primary tumor samples and 42 metastases were studied (Supplementary Table 3). For 21 patients, at least two spatially distinct samples were studied per patient (range = 2–9). Metastases from 35 secondary sites were sampled in the autopsy participants, with a median of two metastases (range = 1–7) studied per patient.

Genetic features of cohort. To discern the genetic features of our cohort, we performed targeted snDNA-seq using a custom-designed 596-amplicon panel targeting known essential genes of primary pancreatic neoplasms; genes affected by both single-nucleotide variant (SNV), small insertion/deletion (indel) and/or CNVs were considered (Supplementary Tables 4 and 5 and Methods). With a custom-designed computational analysis pipeline (Supplementary Fig. 1a,b), in total 137,491 snDNA libraries were profiled by snDNA-seq with a mean of 5,729 per patient (range = 329–20,123) and 1,858 single nuclei studied per sample (range = 50–7,540; Supplementary Fig. 1c). We detected somatic SNVs, indels, monoallelic and/or biallelic losses in at least one known driver gene in all cases analyzed (Fig. 1b), even in low-cellularity biopsy samples (Supplementary Fig. 2a). Activating *KRAS* mutations were identified in all but two PDACs (21/23, 91%). Targeted clinical-grade sequencing performed for these two neoplasms revealed a *BRAF* fusion and an *NRG* fusion, respectively, that were not on our panel. No *KRAS* mutation was found in the ACC (PC22). Alterations in *TP53* (75%), *CDKN2A* (71%) and *SMAD4* (71%) were also observed at high frequency, as expected[15]. Copy number (CN) alterations identified included amplifications in *MYC* (46%), *GATA6* (17%) and *MTOR* (4%, one case), as well as biallelic loss (homozygous deletion) in genes such as *CDKN2A* (67%), *SMAD4* (25%) and *TGFBR2* (8%, two cases). A technical strength of snDNA-seq compared to bulk sequencing is that it enables direct calculation of each alteration's corresponding cancer cell fraction (CCF; Methods); as a result, we could distinguish among clonal drivers such as *TP53* (mean CCF = 0.85), *MYC* (0.83), *CDKN2A* (0.82), *KRAS* (0.76) and 'relative' subclonal drivers such as *SMAD4* (0.70), *ARID1A* (0.47) and *PIK3CA* (0.45) (Fig. 1b, square size), yet heterogeneity of clonality existed for almost every gene.

Subclonal and subgene (affecting only a part of the gene) monoallelic/biallelic losses may be difficult to detect with bulk-sequencing methods, particularly in PDACs with limited tumor purity. Our targeted snDNA-seq method evaluated each mutant locus's single-cell variant allele fraction (sc-VAF) and the read counts for single amplicons (200–300 bp) while accounting for amplicon dropout rates[14], to detect such events in small groups of cells (Methods). For instance, we could visualize a homozygous deletion confined to a single amplicon of *SMAD4* in the neoplastic clone of PC18 (Supplementary Fig. 2b). Across the cohort we detected more alterations (including both short mutation and monoallelic/biallelic loss) to *CDKN2A* and *SMAD4* (71% and 71%, respectively), many subclonal, than reported by The Cancer Genome Atlas (30%, 32%) and the International Cancer Genome Consortium (30%, 27%). Specifically, we detected homozygous deletion of *CDKN2A* in 67% cases and *SMAD4* in 25% cases, which are comparable to the numbers (~60% *CDKN2A*; 40% *SMAD4*) found in a mixed cohort of primaries/metastases using microdissection to enrich for tumor content[7]. Compared with previous reports of approximately 40% alteration frequency of p14ARF region (alternate transcript of *CDKN2A*)[16], we found that some part of p14ARF was deleted in all 16 cases with deletion in the region, 12 of which had apparent deletion to all three exons of both p16INK6a and p14ARF. Of the other four cases that showed subgene deletion in the region (PC07, PC09, PC14 and PC16), none preferentially inactivated p14ARF over p16INK6a. We also detected a high frequency of monoallelic losses (loss of heterozygosity, LOH) affecting *TGFBR1* (21%) and *TGFBR2* (29%). We interpret these findings to reflect the increase in sensitivity by our assay rather than more extensive sampling, as these rates are also higher than reported in a recent study of 91 multiregion sampled PDAC research autopsies[17].

While our samples included nuclei from non-neoplastic stromal cells, we were unable to discern if these nuclei were non-neoplastic epithelial, fibroblastic, neural, vascular or immune in origin. Nonetheless, mutations in lineages independent of the PDAC were noted in four cases. In PC05, a *DNMT3A* p.R882H mutation was found, mutually exclusive with the neoplastic cells (Supplementary Fig. 2c). This mutation is a common hotspot in clonal hematopoiesis[18]. PC14 had a *CTNNA2* p.R136H mutation that was also mutually exclusive with the neoplastic cells (Supplementary Fig. 2d). PC22 had an *RNF43* p.A349V mutation and an *MTOR* p.I313T mutation, both mutually exclusive with the neoplastic cells and each other (Supplementary Fig. 2e). We found only one case (PC11) with a secondary oncogenic *KRAS* mutation that did not belong to the tumor lineage (elaboration to follow).

Major patterns of evolutionary trajectory. Building upon our previous work[19], we developed fast Constrained Dollo Reconstruction (fast-ConDoR), a phylogenetic analysis pipeline for this snDNA-seq dataset (Supplementary Fig. 1a and Methods) and derived phylogenies for each patient (Fig. 1c). On average, 2.63 clonal driver SNV events (Methods) were covered by our panel, consistent with the number calculated unbiasedly by bulk whole-exome sequencing (WES)[17].

https://doi.org/10.1038/s41588-025-02468-9

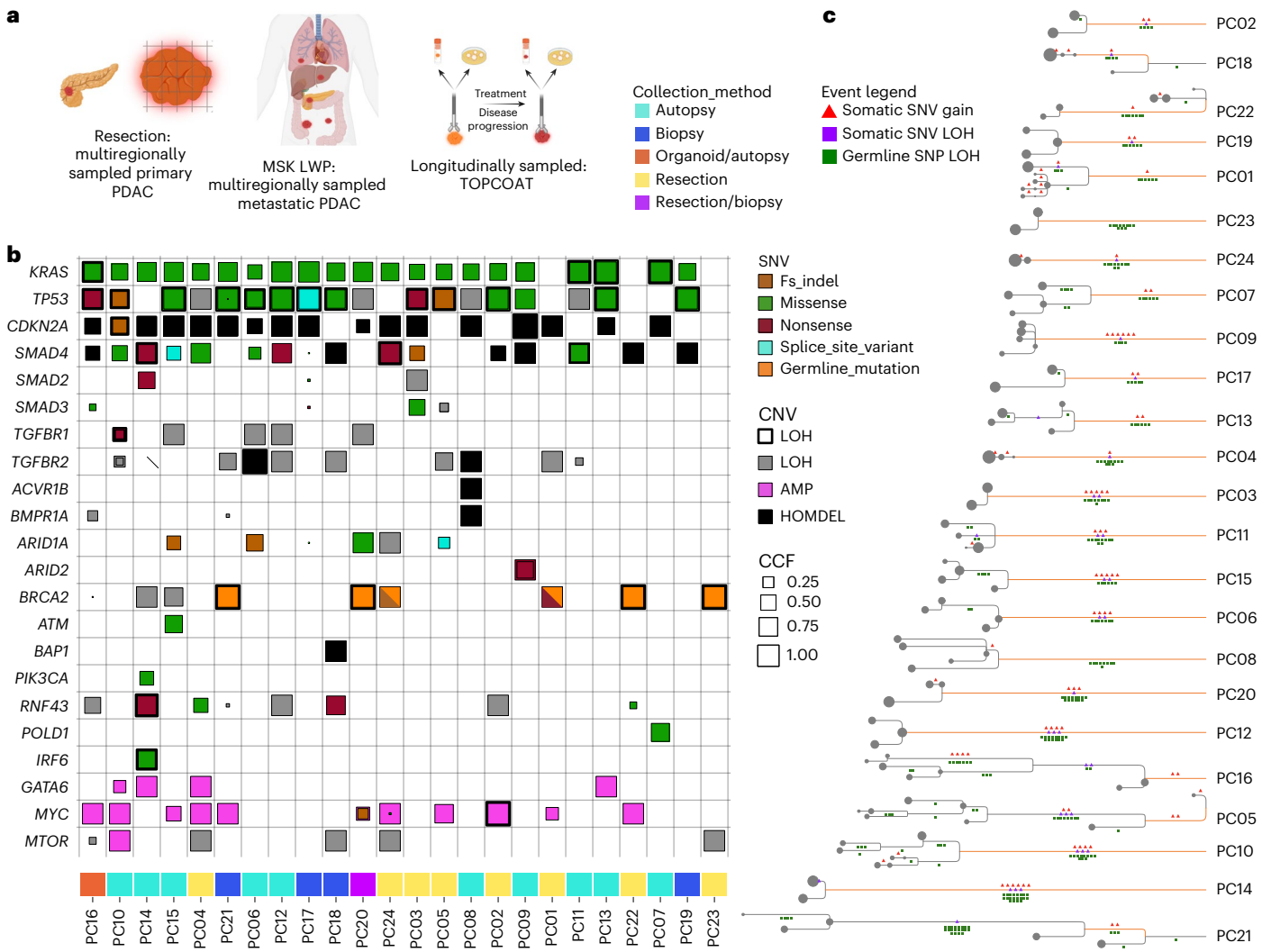

Fig. 1 | Cohort overview. a, Sample collection procedure of the study. Panel **a** was adapted from BioRender. Zhang, H. (2025) https://BioRender.com/nlt6c5v. **b**, Oncoprint depicting the type of mutations and their CCF in each pancreatic cancer case studied. On the *x* axis, cases are sorted by their respective number of mutational events in descending order from left to right. On the *y* axis, from top to bottom, genes are sorted by the number of alterations in the patient cohort in descending order, while clustering for biological pathways involved—TGFβ, homologous recombination repair and others. Within the grid, the square size is proportional to the CCF of each alteration. The square color corresponds to the alteration type, which is of two categories—SNVs and CNVs. The thickened border represents LOH status. **c**, Overview of single-cell clone phylogenies inferred for each case. The branch length on the *x* axis represents inferred phylogenetic distance among clones (gray circles). The orange branch is the inferred trunk of the phylogeny (Methods). The size of the circle is proportional to the number of cells in that clone. For events on the edge, red triangles indicate gain of somatic SNVs, purple triangles indicate LOH of somatic SNVs and green squares indicate LOH of germline SNPs. LWP, last wish program; TOPCOAT, tracking of pancreatic cancer regression and resistance.

Among all driver SNVs identified, on average, 74.4% were clonal in origin (that is, occurred before the MRCA of all clones identified) in keeping with prior estimates[17]. This indicates that the major contributor to intratumoral heterogeneity in PDAC is CNVs, as illustrated by the longitudinally sampled PC20—most of the SNV events and relevant LOHs were clonal, while total CNs varied across clones 1 and 2, the former dominating at baseline while the latter emerging at metastasis (Fig. 2a,b and Supplementary Fig. 3a,b). This pattern of early driver fixation, followed by generation of intratumoral heterogeneity via CNVs, is consistent with multiregional bulk sequencing of larger PDAC cohorts[20]. The number of subclones resolved per case was not correlated with the number of single nuclei analyzed, as demonstrated by two edge cases—PC11 had one of the highest patient-level library sizes (16,139 single nuclei) but only three tumor clones, while PC21 had only 329 single nuclei but four tumor clones.

As our panel targets driver mutations that are previously known and hereby confirmed to exhibit homogeneity across metastatic sites (two example cases shown in Supplementary Fig. 3c,d), our capacity to assess genetic heterogeneity and, in turn, metastatic seeding patterns was limited in most metastatic cases we sampled. However, PC10 demonstrated stark CN heterogeneity of *MYC*, *MTOR* and *GATA6* that allowed us to reaffirm the random metastatic patterns previously inferred by bulk WES[20,21]. The total CN of *MYC* stayed in the normal range in clones 1 and 2 in the upper branch, but uniformly amplified to over 20 in clones 3, 4 and 5 (Fig. 2c and Supplementary Fig. 4a,b). Such an amplification pattern was inferred to be caused by extrachromosomal DNA[22] (Supplementary Fig. 4c). Clones 1 and 2 in the upper branch had *MTOR* amplification instead. Intriguingly, clone 3, exclusively present in the primary tumor sampled, showed uniquely high CN state in *GATA6*; two populations with or without high CN state of *MYC* co-existed in clone 3 (Fig. 2d and Supplementary Fig. 4a).

As we mapped the clones to their anatomical location (Fig. 2d), we found little correlation between spatial proximity and clonal similarity. Notably, the two subclonal lineages defined by two *TGFBR2*

https://doi.org/10.1038/s41588-025-02468-9

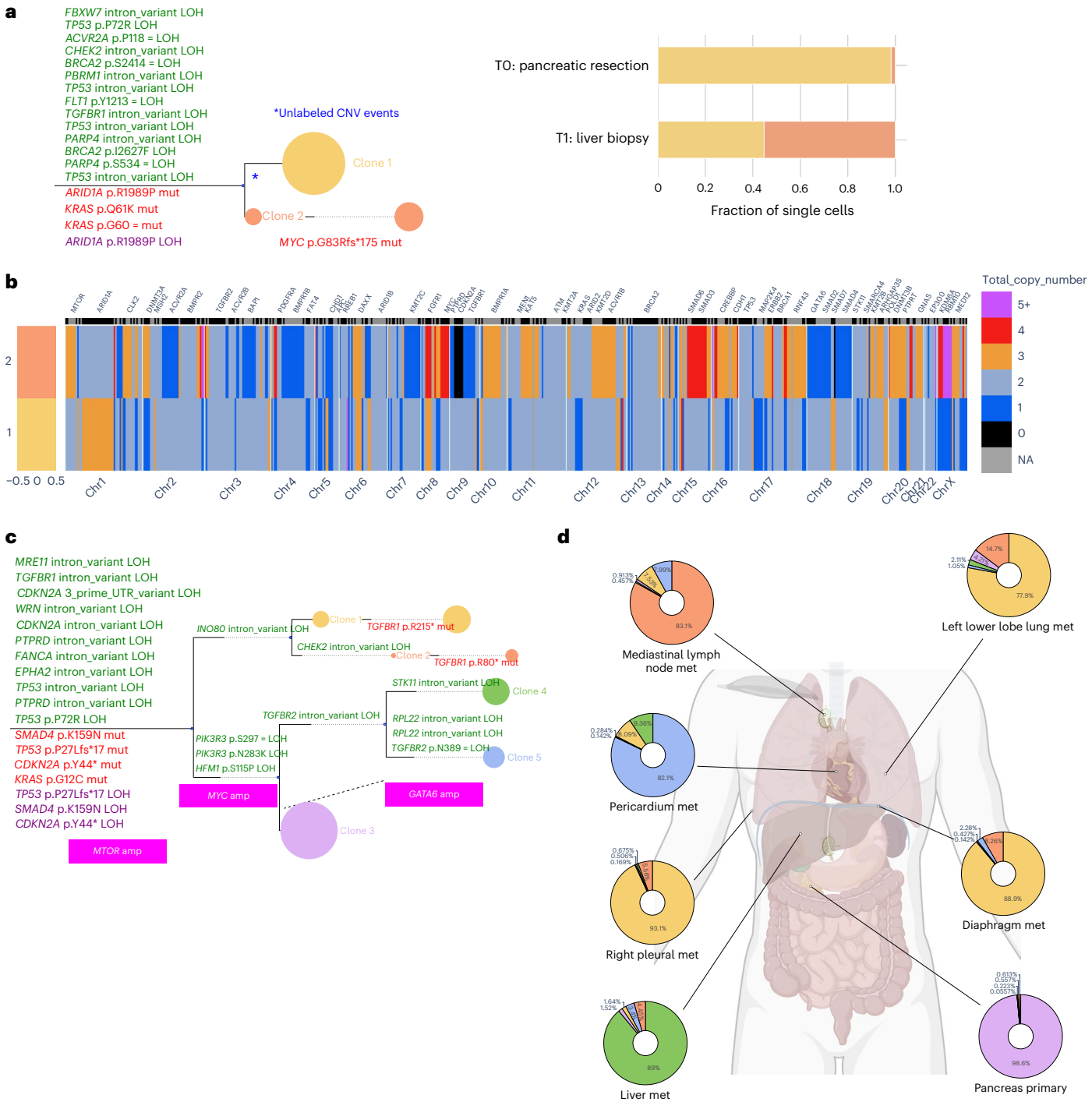

Fig. 2 | Refined evolution patterns of pancreatic cancer revealed by snDNA-seq. a, Single-cell clone phylogeny (left) and sample clonal proportion (right) of PC20. T0 sample was collected from pancreatic resection, and T1 was a liver biopsy of metastasis after 19.5 months. Mutation events are labeled on the trunk and branches of the phylogeny. Green events are LOHs of germline heterozygous SNPs. Red events are somatic SNVs, while purple events are LOH to them. CNVs that do not have corresponding SNP/SNV events are not labeled. Detailed views of the sc-VAF distributions of the *CHEK2* SNP and *MYC* SNV across the two time points are shown in Supplementary Fig. 2a,b. **b**, Clone total CN profiles of PC20. Clone colors correspond to those in **a**. The heat represents the total CN state of each clone at each genomic location. **c**, Single-cell clone phylogeny of PC10. **d**, Clonal proportions mapped to metastatic lesions sampled across PC10. Human anatomy in **d** was adapted from BioRender. Zhang, H. (2025) https://BioRender.com/e026ra1.

mutations migrated likely through hematogenous routes to the pleura, diaphragm, lung and its adjacent lymph node, while the two subclonal lineages defined by extrachromosomal DNA-cause *MYC* amplification migrated likely through similar means to the liver and pericardium. The polyclonal seeding and bifurcating migration correspond to patterns predicted by a single-stage evolution model, where founder cells of the metastatic lesions were apparently randomly sampled from the available lineages within the primary or secondary sites[23].

Heterogeneity of KRAS dependence

Oncogenic *KRAS* mutations have been well established as early drivers of PDAC through precursor lesion studies[24] and genetically engineered

mouse models[25]. Allelic imbalance or amplification of mutated *KRAS* alleles has also been described in association with progression and metastasis[26]. Across cases, we also observed a wide range of *KRAS* mutant VAFs, consistent with the presence of subclonal populations within each PDAC containing *KRAS* allelic imbalance or amplification (Fig. 3a, top). On average, the highest *KRAS* mutant sc-VAFs were found in advanced stage of PDACs ($\hat{\beta}$ = 0.13, s.e. = 0.08, one-sided P = 0.054)—for instance, PC12, PC13, PC16 and PC17—underscoring the role of increased KRAS-dependent signaling in PDAC progression. Unexpectedly, 9 out of 21 (42.9%) *KRAS*-mutant cases were noted to have subclones (>10% tumor cells) without any read of oncogenic *KRAS* mutation; this exceeded the technical dropout rate at 10% (Fig. 3a, bottom). For example, in PC19 (Fig. 3b–e) and PC03 (Fig. 3f–h), there existed large subclones carrying the other driver mutations (*TP53* for PC19, *MAP2K4* and *TP53* for PC03) but not the *KRAS* mutation. While it is possible that in these PDACs the *KRAS* mutation occurred later than other oncogenic mutations (that is, *TP53*) in the neoplasm, a more parsimonious explanation is that the *KRAS* mutation was lost in a subpopulation of the cancer cells. This is consistent with mouse models suggesting the loss of oncogenic *Kras* signaling after PDAC formation is tolerable[27,28]. Of note, we did not observe any examples of a *KRAS*-mutant PDAC carrying a second subclonal *KRAS* mutation, except for one patient (PC11), where a minor *KRAS* mutant subclone was identified across three spatially distinct samples of the primary tumor (single-cell prevalence = 0.0083). This minor clone was a distinct lineage with little correlation to the primary PDAC (Supplementary Fig. 5a,b), corresponding to one or more *KRAS*-mutant clones present within the pancreatic parenchyma that was infiltrated by the main PDAC[24,29].

In addition to allelic imbalance favoring *KRAS* mutation dosage, we also found independent mechanisms to activate oncogenic MAPK–ERK signaling that likely succeeded mutant *KRAS*. For example, PC03 had a clonal *MAP2K4* p.P232L mutation (Fig. 3f–h), while PC14 had a clonal *PIK3CA* p.Q545K mutation, and both colocalized with their *KRAS* mutations (Fig. 3h,i). In PC11 that had a clonal *KRAS* p.G12V mutation and was sampled from metastatic disease at autopsy, we noted two mutually exclusive subclonal *PIK3CA* mutations (p.E542K CCF = 0.009, p.K528E CCF = 0.002) that were scattered across metastatic sites (Fig. 3j and Supplementary Fig. 5c). These alternative mutations correspond to differential regulation of the MAPK–ERK signaling pathways favored later in cancer progression and could inform resistance mechanisms to *KRAS* inhibiting therapeutics.

Evolutionary patterns of PDACs with germline *BRCA2* mutations

Although patients with germline mutations in the *BRCA2* gene are associated with an inherited risk of developing PDAC as well as increased sensitivity to platinum salts, the evolutionary features of how pancreatic neoplasms arise in the setting of germline *BRCA2* mutations are poorly understood. Among six germline *BRCA2* mutation cases in our cohort, we noted biallelic *BRCA2* inactivation in all six cases through different mechanisms, indicating positive epistasis of somatic *BRCA2* inactivation in a germline *BRCA2* mutation background in PDAC (Fisher's exact test, P < 0.001). Tumor cells of PC20, PC21, PC22 and PC23 lost the wild-type *BRCA2* allele (LOH; Supplementary Fig. 5d), while PC01 and PC24 had inactivating somatic mutations to *BRCA2* (Fig. 1b). Mutation signature analysis confirmed homologous recombination repair deficiency (HRD) phenotype in all six cases (Supplementary Fig. 5e).

The median CCF for *BRCA2* somatic alterations was 1, suggesting their generally early occurrence that was followed by one or more clonal sweeps before sampling. Nevertheless, for four of these patients, phylogenetic analysis provided more profound insight into the timing of inactivation of the remaining wild-type allele relative to other somatic alterations in the same PDAC. The phylogeny of PC01 showed that the somatic *BRCA2* nonsense mutation was truncal, followed by several other LOH events and *CDKN2A* homozygous deletion (Fig. 4a, left). The phylogeny subsequently diverged into two mutually exclusive subclones carrying an *FGFR1* p.T17K mutation with LOH, or a *TGFBR2* p.K330Tfs*6 mutation with LOH. The subclones were separated spatially (Fig. 4a, right). Clinical-grade sequencing and whole-genome sequencing (WGS) independently performed on the same tissue identified a *TRIM24–BRAF* translocation, a KRAS-independent route toward oncogenic MAPK activation, suggesting that this PDAC developed through noncanonical mechanisms. Similarly, case PC23 showed a truncal *BRCA2* LOH (Fig. 4b), with no detectable alteration to other canonical PDAC driver genes; clinical-grade sequencing and WGS indicated an *NRG* fusion, another noncanonical oncogenic mechanism converging to the same pathway as KRAS. We interpret these findings to mean that biallelic inactivation of *BRCA2* was among the earliest events in these two PDACs, leading to HRD in the MRCA and, subsequently, KRAS-independent activation of MAPK signaling via gene rearrangements involving *TRIM24–BRAF* (PC01) and *NRG* (PC23).

By contrast, in case PC20, while the wild-type *BRCA2* allele was inferred to be lost along with occurrence of the other driver events (Fig. 4c), we observed a small population of cells containing a heterozygous germline *BRCA2* p.I2627F mutation, a heterozygous somatic *KRAS* p.Q61H mutation, a *CDKN2A* homozygous deletion and a homozygous somatic *ARID1A* p.R1989P mutation (Fig. 4d). The *ARID1A* p.R1989P homozygous genotype and germline single-nucleotide polymorphism (SNP) LOH validated this population as nondoublet and indicated that *BRCA2* LOH likely came after the canonical PDAC driver SNV and CNV events. The somatic *BRCA2* p.S1832Ifs*2 mutation of PC24 was not covered by our snDNA-seq panel; yet WGS suggested a lower CCF (0.80) than the two main drivers *KRAS* p.G12D (0.85) and *SMAD4* p.Q116* (1.00), categorizing it into a similar late-BRCA2 biallelic inactivation group. We interpret these findings to mean that, in these two PDACs, biallelic inactivation of *BRCA2*, and thus the HRD phenotype, was acquired after accumulation of canonical PDAC alterations.

In PC21, we observed in the liver metastasis a tumor subclone where *BRCA2* was homozygously deleted and *CDKN2A* was intact, in contrast to the main tumor population where *BRCA2* had LOH and *CDKN2A* had homozygous deletion (Fig. 4e,f). Both populations had the canonical drivers *KRAS* p.G12D and *TP53* p.L344Q; yet orthogonal statistical testing validated that the *BRCA2* homozygous deletion colocalized with a heterozygous *TP53* state (P = 0.0078; Supplementary Methods), while the main tumor population had *TP53* LOH. A parsimonious explanation is that *BRCA2* had LOH before *TP53* LOH and *CDKN2A* homozygous deletion, and that the mutated *BRCA2* allele was subsequently lost, resulting in the subclone in the liver biopsy.

In summary, although a somatic *BRCA2* second hit is associated with an HRD phenotype, the timing at which this second hit occurs can happen at different time points of PDAC development—in some patients, *BRCA2* biallelic inactivation defines the MRCA and drives PDAC in a noncanonical mutational route characterized by genomic rearrangements. By contrast, in other patients, the second hit occurs after accumulation of other canonical drivers such as *KRAS* and *TP53* and functions as an additive trait later in PDAC progression.

Convergent evolution toward TGFβ inactivation

PDAC genetics are notable for a high frequency of mutations in *KRAS*, *CDKN2A*, *TP53* and *SMAD4* (ref. 15). However, an unanswered question is the extent to which small populations of cells with additional mutations in known PDAC genes or pathways exist below the resolution of bulk-sequencing methods. The presence of such populations would indicate the pre-existence of subclones that are poised for selection when encountering new microenvironments or treatment bottlenecks, whereas the absence of them would favor strong selection and one or more clonal sweeps of the neoplasm after these mutations occurred.

The absence of rare subclones derived from the MRCA of each PDAC carrying secondary mutations to *KRAS*, *TP53* and *CDKN2A* strongly supports that mutations to these genes undergo strong

https://doi.org/10.1038/s41588-025-02468-9

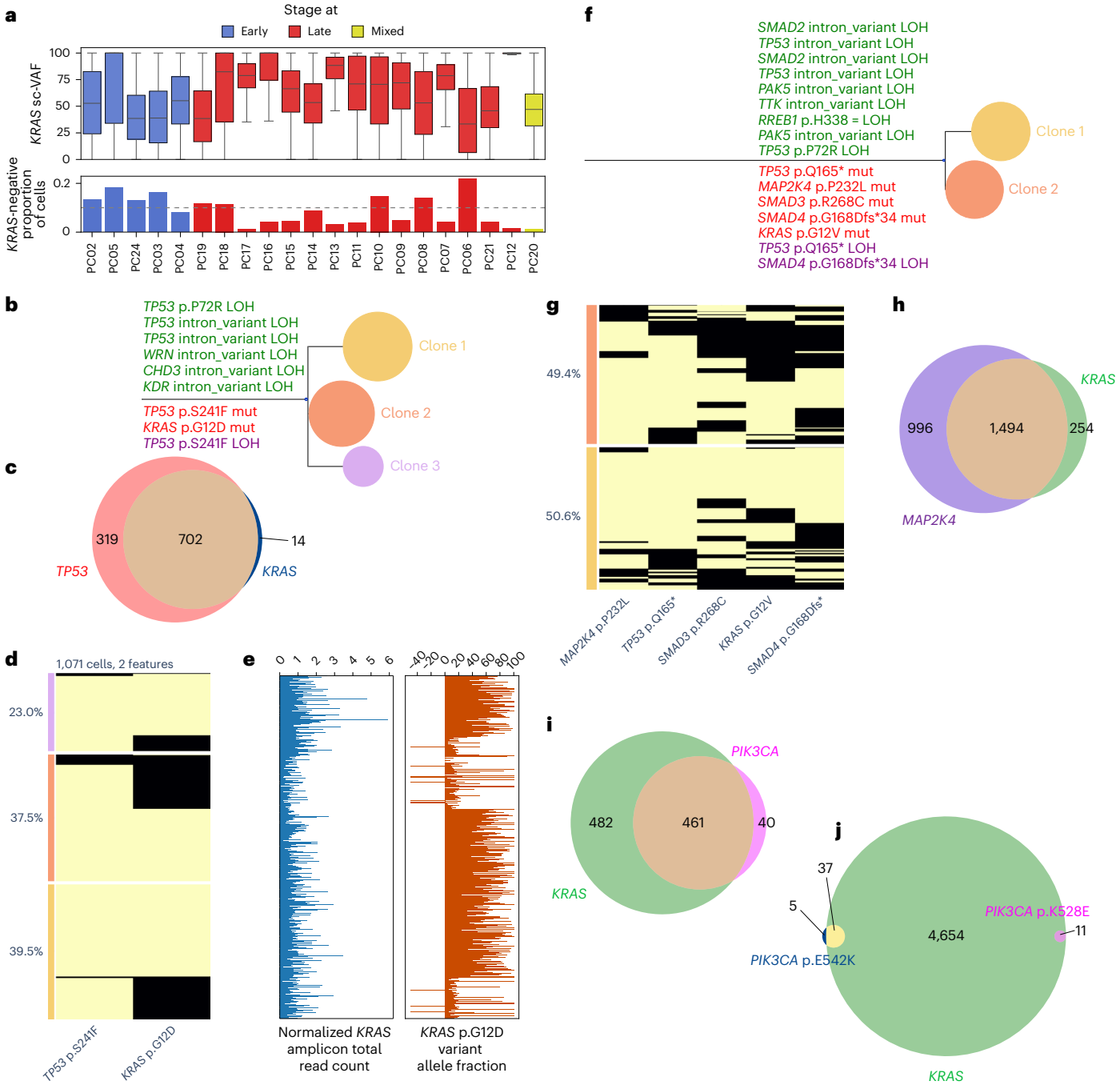

Fig. 3 | Heterogeneity of KRAS dependence in pancreatic cancer. a, Across tumors carrying *KRAS* mutation, box plot (top) showing cell-level distribution of sc-VAF of the mutated *KRAS* allele and histogram (bottom) showing proportions of cancer cells that do not have any reads of the mutated *KRAS* allele. Bottom, the value of 0.1 is labeled by the dashed line. Cases are grouped by stage at collection, with 'mixed' indicating that the case (PC20) had one sample at the early stage and another at the late stage. The box plot indicates the median, first and third quartiles (hinges); for each case, the number of cells/nuclei studied is provided in Supplementary Table 3. **b**, Single-cell clone phylogeny of PC19. **c**, For PC19, a Venn diagram showing the colocalization of *TP53* mutation and *KRAS* mutation-carrying cells. The numbers of cells carrying the corresponding mutations are labeled. **d**, For PC19, single-cell binarized genotype (yellow, presence; black, absence) heatmap of the two main somatic mutations across tumor clones. **e**, For each single cell in the genotype heatmap in **d**, normalized *KRAS* total CN (left) and the G12D sc-VAF (right) are shown. In the right panel, an sc-VAF of −50 indicates 0 read depth at the locus. For example, the top fraction of cells in clone 2 (orange) has reduced total CN and sc-VAF at *KRAS*-mutated locus, suggesting a likely loss of the mutated allele. **f**, Single-cell clone phylogeny of PC03. **g**, For PC03, single-cell binarized genotype heatmap of the five main somatic mutations across tumor clones. **h**, Venn diagram showing the colocalization of *MAP2K4* mutation and *KRAS* mutation-carrying cells. **i**,**j**, Venn diagrams demonstrating colocalization of *PIK3CA* mutations with *KRAS* mutations for PC14 (**i**) and PC11 (**j**).

selection before sampling. By contrast, in many PDACs the TGFβ pathway appears subject to continuous selection over the evolutionary lifetime of the PDAC, seen as multiple relevant mutations converging within the same tumors (Fig. 1b). Such convergence manifested as different genetic subclones in parallel—in PC01, three lineages each carrying (1) *TGFBR2* p.K300Tfs*6, (2) *TGFBR2* p.A426G/p.M425I and (3) *SMAD3* p.S425F mutations (Fig. 4a) co-existed before the early-stage tumor was resected. In other cases, this manifested as different events take place sequentially in a single lineage—PC08 had three TGFβ family receptors focally deleted, whereas PC06, PC10 and PC12

https://doi.org/10.1038/s41588-025-02468-9

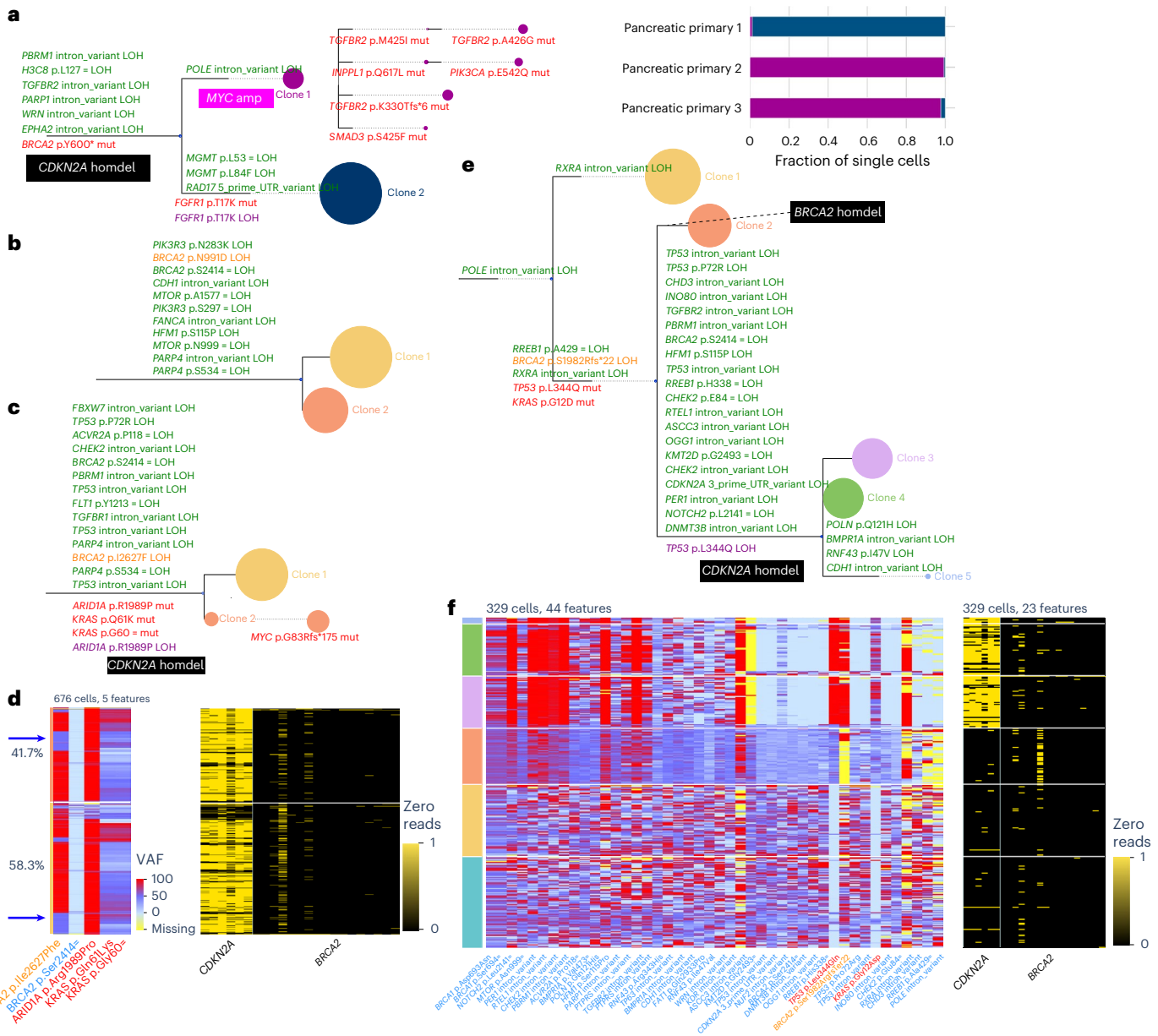

Fig. 4 | Genomic evolution of pancreatic cancers with germline *BRCA2* mutations. a, Single-cell clone phylogeny (left) and sample clonal proportions (right) of PC01. **b**, Single-cell clone phylogeny of PC23. **c**, Single-cell clone phylogeny of PC20. **d**, For PC20, paired single-cell mutation heatmap (the heat represents VAF of each mutation in each single cell, with 'missing' indicating zero depth at the locus) and single-cell amplicon heatmap (the heat is a binary indicator of the presence of any read of the amplicon) of mutations and genes of interest. For each heatmap, variants or amplicons are on the *x* axes, and single nuclei are hierarchically clustered on the *y* axes. Single-cell populations of interest are pointed with arrows. **e**, Single-cell clone phylogeny of PC21. **f**, Paired single-cell mutation heatmap and single-cell amplicon heatmap of mutations and genes of interest in PC21. The clone of interest is clone 2, where the germline *BRCA2* locus likely had homozygous deletion (homdel).

all had alteration to both *TGFBR1* and *TGFBR2* (Fig. 1b). PC11 (described as PA04 in previous work[14]) had two independent mutations to *SMAD4* on the same chromosome, which translate to 13 codons apart, followed by the loss of the wild-type allele (Fig. 5a, left) and finally homozygous deletion to 10 of 11 amplicons of *SMAD4* in the MRCA of all tumor clones (Fig. 5a, right). In the same patient, we observed a subclone with *TGFBR2* LOH exclusively present in the liver metastasis and one region of the primary, which further contributed to the phenotype of TGFβ inactivation[30] (Fig. 5b). In PC10, two subclonal *TGFBR1* nonsense mutations were noted in one of the two main lineages of metastatic disease in this patient (Fig. 5c). Focal analysis of the subclone suggested that the two mutations were present on different chromosomes in the same set of cells at one time in the tumor's evolution history (population ii in Fig. 5d), and then chromosomal loss likely resulted in the LOH of either mutation creating bifurcating lineages (populations iii and iv in Fig. 5d).

Notably, polyclonal, parallel convergence was observed only in one early-stage (resectable) PDAC case, whereas monoclonal, sequential convergence was observed repeatedly in late-stage disease. This suggests differential selective pressure across stages of tumor development, with potentially less selection at early stages, allowing for polyclonal evolution, whereas elevated selection forces monoclonal evolution. At the cohort level, in contrast to the model that places inactivation of tumor-intrinsic response to TGFβ before invasive tumor formation, we found that these alterations are mostly subtruncal in early-stage cases and truncal in late-stage cases (Fig. 5e).

https://doi.org/10.1038/s41588-025-02468-9

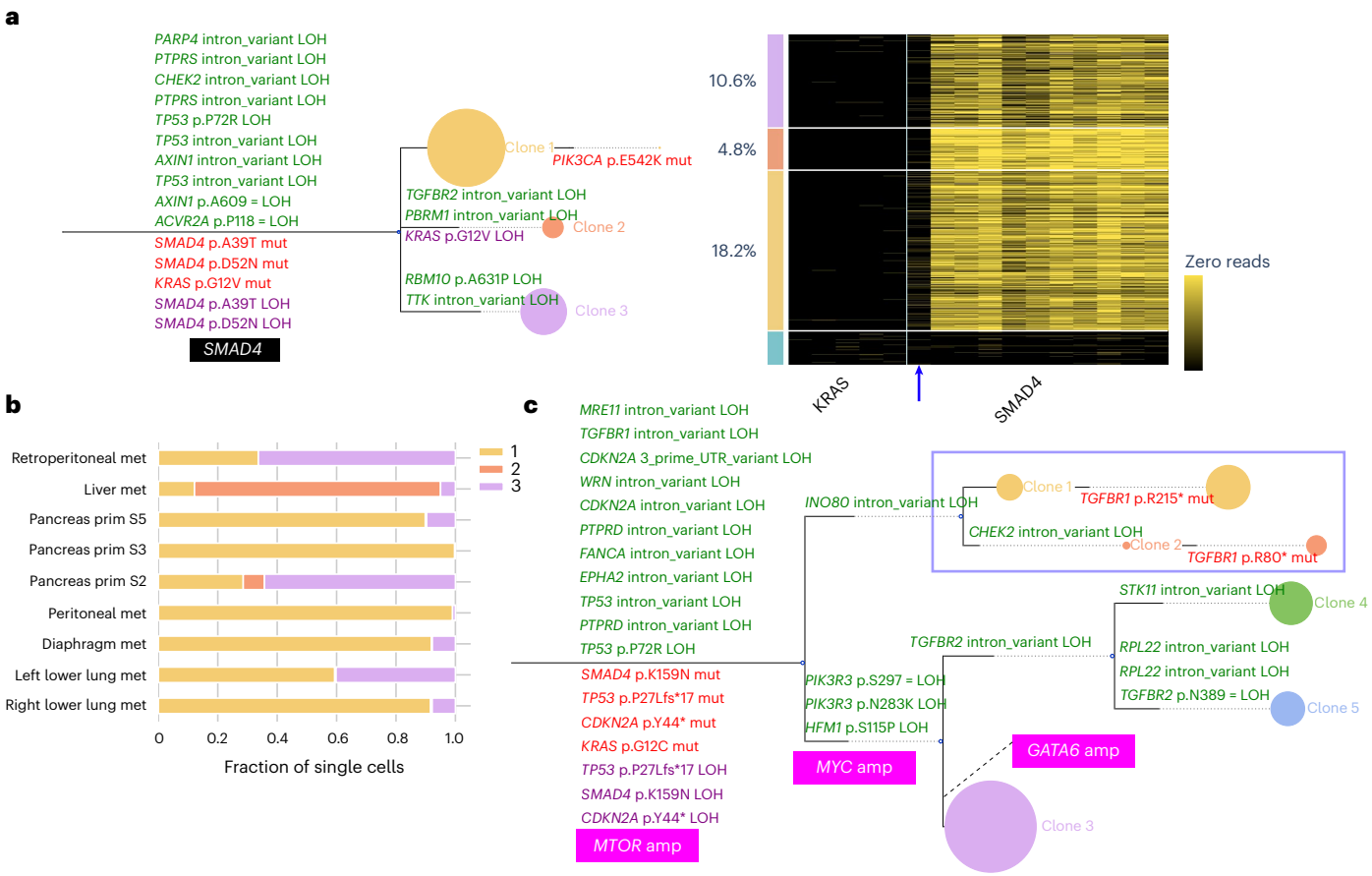

Fig. 5 | Convergent evolution in pancreatic cancer. a, Left, single-cell clone phylogeny of PC11. Right, paired single-cell amplicon heatmap of mutations and genes of interest in PC11, where the heat indicates zero reads of that amplicon in that cell. The one *SMAD4* amplicon that contained the two somatic mutations is indicated by the arrow. **b**, Clonal composition at each sample site of PC11. **c**, Single-cell clone phylogeny of PC10. **d**, Among clones 1 and 2 of **c** (circled), distribution of sc-VAF of the two *TGFBR1* mutations of interest. The density is labeled as colors and contours, with darker colors indicating higher density of single cells. The four major populations of interest are circled and labeled. **e**, The clonal status of TGFβ mutations across all cases studied.

Together, these results indicate that strong selective pressures imposed by the primary and secondary site microenvironments continuously sculpt for convergence toward failure to respond to TGFβ growth inhibitory signals.

Discussion

These data, based on high-resolution snDNA-seq, elucidate the closest-to-ground-truth pancreatic cancer genome in such contexts as early/late diagnosis, metastasis and nontargeted (that is, cytotoxic/cytostatic) treatments. Notably, we identified three features of the evolutionary life history of the pancreatic cancer genome (Fig. 6), two of which have direct implications for precision medicine.

In a sizable fraction of the PDACs analyzed, we found that the *KRAS*-mutant allele was lost in at least 10% tumor cells (Fig. 6a). This raises the possibility that resistance to allele-specific KRAS inhibitors may develop through loss of the mutant allele. Data from our institution found one colorectal cancer, harboring *KRAS* p.G12C mutation and treated with sotorasib, developed allelic imbalance favoring the wild-type *KRAS* allele through its amplification[31]. This reduction of the mutated allele's dosage post-treatment added plausibility of mutant

https://doi.org/10.1038/s41588-025-02468-9

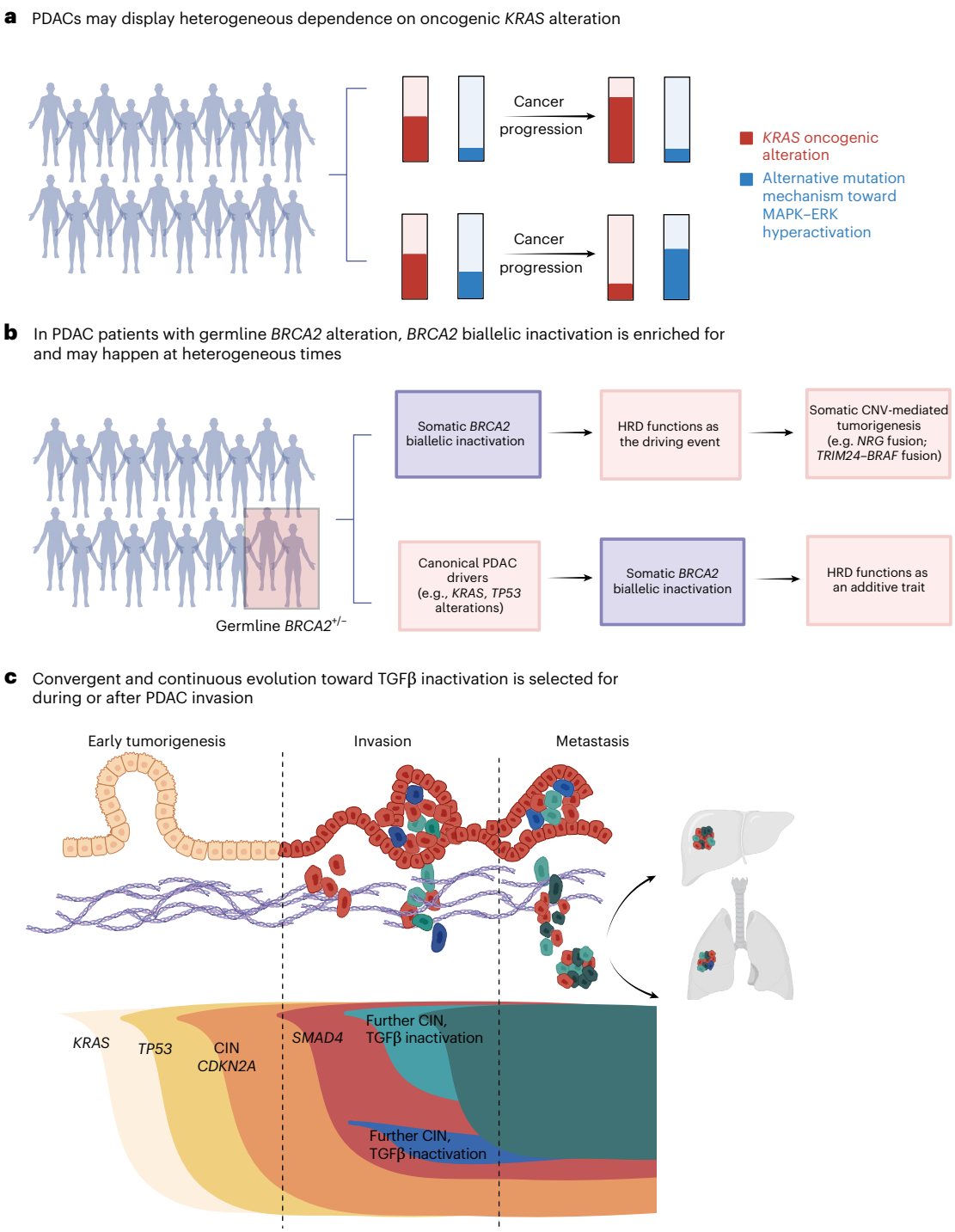

Fig. 6 | Summary of findings. a, In contrast to tumor suppressors such as *TP53* and *BRCA2* that almost always undergo multiple inactivation (for instance, one mutation + LOH) indicating ongoing selection for the corresponding phenotype, we found that oncogenic *KRAS* dosage varies across PDACs in our cohort, with some cases gaining/amplifying the mutated allele subclonally (top), while others gaining the mutation late or, more likely, losing the allele as the tumor progresses (as studies on precursors concluded that the rate of PDAC not initiated by a *KRAS* mutation is low). In the latter cases (bottom), alternative mutation mechanisms may contribute to a similar MAPK–ERK pathway hyperactivation phenotype. This is clinically important as PDACs showing less oncogenic *KRAS* dosage and potentially less dependence on it would be inherently insensitive to KRAS inhibitors. **b**, We showed that although somatic biallelic inactivation of *BRCA2* is preferred in the background of a germline *BRCA2* mutation, it could either take place early and drive PDAC in a noncanonical mutational route likely dominated by HRD-associated genomic instabilities, or take place later than canonical drivers such as oncogenic *KRAS* and *TP53* alterations and function as an additive trait. Although current clinical practices usually put patients with germline mutations in *BRCA2* in one group for treatment, such a finding could inform further stratification. **c**, Strong selective pressure, not at tumorigenesis but more likely after invasive tumor formation, selects for convergence toward mutation to the TGFβ pathway, and the pressure continues as the tumor continues to metastasize to secondary sites, resulting in continuous evolution to inactivate the pathway. This sometimes manifests as different tumor subclones carrying different mutations, but more often as one lineage carrying multiple mutations to TGFβ pathway, indicating strong selective pressure that favors rapid clonal sweeps. The figure was adapted from BioRender. Zhang, H. (2025) https://BioRender.com/y8ik2jm.

KRAS loss as a potential resistance mechanism to KRAS-targeted therapies. As expected, we also observed that oncogenic *KRAS* dosage varied across PDACs in our cohort, with several cases having allelic imbalance favoring mutant *KRAS* alleles; we also noted some PDACs with additive mechanisms to MAPK activation, such as *PIK3CA* mutations. These are known resistance mechanisms to targeted KRAS inhibitors and further indicate that KRAS inhibitor resistance mechanisms are pre-existent in PDACs that have not been exposed to these therapies. The extent to which *KRAS* allelic imbalance in favor of the mutant allele, the wild-type allele or convergent modes of MAPK activation leads to insensitivity to KRAS inhibition remains to be revealed by ongoing clinical trials[32].

Reversion mutations in *BRCA2*-mutant tumors are a known resistance mechanism to targeted therapies that exploit deficiencies in homologous recombination[33]. However, our findings suggest that the timing of biallelic inactivation of *BRCA2* provides an additional mechanism for consideration (Fig. 6b). In patients with germline *BRCA2* alterations, we discovered varied timing of somatic biallelic inactivation of *BRCA2* and the development of the HRD phenotype in relation to other oncogenic events, which may explain heterogeneous treatment responses in patients with similar germline *BRCA2* genotypes[34]. Still, our cohort size was too small to draw conclusions related to how these differences may affect outcome, particularly outside of a controlled clinical trial setting. Future studies to confirm or refute such a correlation are in progress with the goal of improved understanding of germline *BRCA2* mutant PDACs.

The approach also revealed more granular features of the genetic progression model, such as frequent subclonal deletions in *CDKN2A* and *SMAD4*, supporting their classification as later events in the genetic progression of PDAC, likely arising from increased genomic instability and selective pressures experienced during invasion of the pancreatic parenchyma. Compared to other driver events such as *KRAS* and *TP53* alterations, where a single event dominated in each PDAC, corresponding to a near-complete clonal sweep of such single events during tumorigenesis, multiple TGFβ pathway alterations clonally evolve in sequence or in parallel, signifying convergent evolution sculpted by continuous selective pressure during PDAC progression in the primary and secondary sites (Fig. 6c). This finding may explain the lack of clinical benefit when targeting TGFβ in patients based on numerous clinical trials conducted thus far[35].

Our findings are based on targeted snDNA-seq and a suite of accompanying computational analysis tools. Such a large-scale application of this technique on clinical solid tumor samples supports its further clinical utility—the wet-lab workflow is largely automated and accommodates challenging clinical samples such as small frozen biopsy samples. One exception is that it currently cannot perform on formalin-fixed tissues, possibly due to DNA damage caused by the fixation process. Furthermore, given single-cell studies' potential to bring about new biological insights[36], a targeted workflow enables high-depth sequencing of actionable genes/pathways and yields more clinically relevant information than low-depth pan-genome methods aimed more at biological discovery[37,38], as demonstrated by the success of Memorial Sloan Kettering Cancer Center Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT)[39]. We acknowledge the limitations of targeted snDNA-seq, which mostly allows for granular studies of already known targets instead of discovering new targets. Although we curated our targeted panel based on WES/WGS data from the selected patients, a small fraction of important genetic events remained unrepresented. For these reasons, we suggest that the current best practice is for snDNA-seq to be paired with highly purified bulk sequencing with the WES/WGS done first to explore and define the mutational landscape of interest, followed by snDNA-seq to further elucidate the clonal evolution of events of interest.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02468-9.

References

1. American Cancer Society. Survival rates for pancreatic cancer. www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html (2025).
2. Rahib, L., Wehner, M. R., Matrisian, L. M. & Nead, K. T. Estimated projection of US cancer incidence and death to 2040. *JAMA Netw. Open* **4**, e214708 (2021).
3. Yachida, S. et al. Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clin. Cancer. Res.* **18**, 6339–6347 (2012).
4. Notta, F. et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
5. Baslan, T. et al. Ordered and deterministic cancer genome evolution after p53 loss. *Nature* **608**, 795–802 (2022).
6. Aung, K. L. et al. Genomics-driven precision medicine for advanced pancreatic cancer: early results from the COMPASS trial. *Clin. Cancer. Res.* **24**, 1344–1354 (2018).
7. Connor, A. A. et al. Integration of genomic and transcriptional features in pancreatic cancer reveals increased cell cycle progression in metastases. *Cancer Cell* **35**, 267–282 (2019).
8. Chan-Seng-Yue, M. et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* **52**, 231–240 (2020).
9. Golan, T. et al. Maintenance olaparib for germline *BRCA*-mutated metastatic pancreatic cancer. *N. Engl. J. Med.* **381**, 317–327 (2019).
10. Hosein, A. N., Dougan, S. K., Aguirre, A. J. & Maitra, A. Translational advances in pancreatic ductal adenocarcinoma therapy. *Nat. Cancer* **3**, 272–286 (2022).
11. Wasko, U. N. et al. Tumour-selective activity of RAS-GTP inhibition in pancreatic cancer. *Nature* **629**, 927–936 (2024).
12. Rojas, L. A. et al. Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature* **618**, 144–150 (2023).
13. Dilly, J. et al. Mechanisms of resistance to oncogenic KRAS inhibition in pancreatic cancer. *Cancer Discov.* **14**, 2135–2161 (2024).
14. Zhang, H. et al. Application of high-throughput single-nucleus DNA sequencing in pancreatic cancer. *Nat. Commun.* **14**, 749 (2023).
15. Hayashi, A., Hong, J. & Iacobuzio-Donahue, C. A. The pancreatic cancer genome revisited. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 469–481 (2021).
16. Bardeesy, N. et al. Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl Acad. Sci. USA* **103**, 5947–5952 (2006).
17. Mullen, K. M. et al. The evolutionary forest of pancreatic cancer. *Cancer Discov.* **15**, 329–345 (2025).
18. Gozdecka, M. et al. Mitochondrial metabolism sustains DNMT3A-R882-mutant clonal haematopoiesis. *Nature* **642**, 431–441 (2025).
19. Sashittal, P., Zhang, H., Iacobuzio-Donahue, C. A. & Raphael, B. J. ConDoR: tumor phylogeny inference with a copy-number constrained mutation loss model. *Genome Biol.* **24**, 272 (2023).
20. Makohon-Moore, A. P. et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat. Genet.* **49**, 358–366 (2017).
21. Reiter, J. G. et al. Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).

22. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
23. Naxerova, K. Evolutionary paths towards metastasis. *Nat. Rev. Cancer* **25**, 545–560 (2025).
24. Braxton, A. M. et al. 3D genomic mapping reveals multifocality of human pancreatic precancers. *Nature* **629**, 679–687 (2024).
25. Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
26. Varghese, A. M. et al. Clinicogenomic landscape of pancreatic adenocarcinoma identifies KRAS mutant dosage as prognostic of overall survival. *Nat. Med.* **31**, 466–477 (2025).
27. Muzumdar, M. D. et al. Survival of pancreatic cancer cells lacking KRAS function. *Nat. Commun.* **8**, 1090 (2017).
28. Kapoor, A. et al. Yap1 activation enables bypass of oncogenic Kras addiction in pancreatic cancer. *Cell* **158**, 185–197 (2014).
29. Makohon-Moore, A. P. et al. Precancerous neoplastic cells can move through the pancreatic ductal system. *Nature* **561**, 201–205 (2018).
30. Ijichi, H. et al. Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-β signaling in cooperation with active Kras expression. *Genes Dev.* **20**, 3147–3160 (2006).
31. Yeh, C. et al. Acquired high tumor mutational burden and activity of immunotherapy after targeted therapy in microsatellite stable colorectal cancer. *Clin. Cancer Res*. https://doi.org/10.1158/1078-0432.CCR-25-2566 (2025).
32. Wolpin, B. M. et al. Trial in progress: RASolute 302—a phase 3, multicenter, global, open-label, randomized study of daraxonrasib (RMC-6236), a RAS(ON) multi-selective inhibitor, versus standard of care chemotherapy in patients with previously treated metastatic pancreatic ductal adenocarcinoma (PDAC). *J. Clin. Oncol.* **43**, TPS4230–TPS4230 (2025).
33. Pettitt, S. J. et al. Clinical *BRCA1/2* reversion analysis identifies hotspot mutations and predicted neoantigens associated with therapy resistance. *Cancer Discov.* **10**, 1475–1488 (2020).
34. Kindler, H. L. et al. Overall survival results from the POLO trial: a phase III study of active maintenance olaparib versus placebo for germline BRCA-mutated metastatic pancreatic cancer. *J. Clin. Oncol.* **40**, 3929–3939 (2022).
35. Ciardiello, D., Elez, E., Tabernero, J. & Seoane, J. Clinical development of therapies targeting TGFβ: current knowledge and future perspectives. *Ann. Oncol.* **31**, 1336–1349 (2020).
36. Lim, B., Lin, Y. & Navin, N. Advancing cancer research and medicine with single-cell genomics. *Cancer Cell* **37**, 456–470 (2020).
37. Wang, K. et al. Archival single-cell genomics reveals persistent subclones during DCIS progression. *Cell* **186**, 3968–3982 (2023).
38. Funnell, T. et al. Single-cell genomic variation induced by mutational processes in cancer. *Nature* **612**, 106–115 (2022).
39. Cheng, D. T. et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).

Methods

Ethics statement

The use of human samples in this study was approved by the Institutional Review Board at MSK (under protocols 06-107, 15-021 and 15-149).

Patient sample collection and preprocessing

Patient samples used in this study consist of the following three categories: multiregionally sampled surgical resection of primary pancreatic cancer, multiregionally sampled autopsy of metastatic pancreatic cancer and longitudinally sampled biopsies of pancreatic cancer. Each participating patient provided written and informed consent and was not compensated. Detailed patient and sample information is summarized in Supplementary Tables 1–3.

For multiregional sampled surgical resections, treatment of naive patients with tumors of ≥2 cm on cross-sectional imaging were identified preoperatively. A single cross-sectional piece of tumor was sampled sequentially using a Cartesian coordinate system with 0.6 × 0.6-cm grid, with three to five samples obtained from each tumor. An adjacent normal pancreas or duodenum was also collected. All samples were stored at −80 °C until use.

For multiregional sampled autopsies, all patients had a premortem diagnosis of PDAC based on pathological review of resected biopsy material and/or radiographic and biomarker studies. Rapid autopsy was performed after our previously described workflow[20].

The above two types of tissue sample were embedded in optimal cutting temperature compound, stained with hematoxylin and eosin, and reviewed by a gastrointestinal pathologist (S.U.) to ensure total cellularity and tumor purity. Normal samples were reviewed to confirm that no contaminating cancer cells were present. Tissues were sectioned to optimize a volume >8 mm^3 for nuclei extraction.

For longitudinally sampled biopsies, tissue cores were obtained using a 22-gauge needle at endoscopic ultrasound for primary tumors or an 18-gauge needle at interventional radiology for metastatic lesions and were flash-frozen until use. The biopsy samples were substantially smaller than the two sample types described above, often too small to be visible to the naked eye, precluding pathologic review for cellularity or tumor purity. Tissues stored in collection tubes were suspended in nuclei storage buffer (S2 Genomics) and then transferred directly into the Singulator machine (S2 Genomics) for single-nucleus extraction. Albeit the suboptimal input amount, even those biopsies with the lowest yields (that is, PC18) resulted in useful genotype information (Supplementary Fig. 1d), thus proving the feasibility of applying this snDNA-seq approach to routine clinical biopsies.

Additionally, because we were initially concerned that the small number of biopsies would be insufficient for snDNA library construction, we expanded four of six samples from one case (PC16) into organoids to increase cell numbers. The organoid workflow was as previously described[40].

Worth noting is that microdissection was omitted so that the resulting snDNA libraries were mostly mixes of tumor and normal cells.

Bulk-sequencing library preparation, sequencing and bioinformatics

Bulk WGS/WES data were collected for each tumor and matched normal sample simultaneously for quality control (QC) of germline/somatic and SNV event calling. Genomic DNA was extracted from each tissue using the phenol-chloroform extraction protocol or QIAamp DNA Mini Kits (Qiagen). DNA quantification, library preparation and sequencing were performed in the Integrated Genomics Operation, where Illumina HiSeq 2000, HiSeq 2500, HiSeq 4000, NovaSeq 6000 and NovaSeq X platforms were used to target sequencing coverages of >80× for WGS samples and >150× for WES samples.

Bioinformatics analyses, including calling germline/somatic SNVs, CNV and structural variants, estimating mutational signatures, and scoring HRDetect[41], were performed using the time-efficient mutational profiling in oncology research pipeline at the MSK Center for Molecular Oncology Computational Sciences. Its code repository is available at https://github.com/mskcc/tempo and documentation at https://cmotempo.netlify.app/.

snDNA-sequencing library preparation, sequencing and bioinformatics

Nuclei extraction from frozen tissue, counting, QC and cryopreservation. Single nuclei from optimal cutting temperature-embedded snap-frozen primary tissue samples were extracted, counted, quality-controlled and cryopreserved after our published protocol[14].

Panel design for single-cell targeted library preparation. The panel was designed as an expansion of our previous 186-amplicon panel[14] with the following goals:

1. To unbiasedly cover genes/genomic segments hypothesized to undergo convergent evolution in PDAC, constituting the following three main categories:
   a. Canonical PDAC drivers—*KRAS*, *TP53*, *SMAD4* and *CDKN2A*.
   b. Components of the TGFβ pathway, including receptors, mediators and downstream effectors—*TGFBR1*, *TGFBR2*, *ACVR1B*, *ACVR2A*, *BMPR1A*, *BMPR1B*, *SMAD2*, *SMAD3*, *SMAD7*, *ARID1A*, *ARID1B* and *ARID2*.
   c. The *BRCA2* gene, which is of particular interest to PDAC precision medicine.
2. To study genes frequently affected by CNV in PDAC, including *MYC*, *GATA6*, *BAP1*, *MUS81* and *KAT5*. Thus, over four amplicons were given per gene. At analysis, amplicons for the same gene were constrained to have the same ploidy to obviate the PCR amplification noise of single amplicons and get more confident absolute ploidy calls.
3. To study other recurrent mutations in PDAC, which came from the curation of the following two databases:
   a. Bulk WES/WGS of PDAC autopsy, resection, derived-organoid banks consisting of over 100 patients collected by the Iacobuzio lab and the MSK David M. Rubenstein Center for Pancreatic Cancer Research.
   b. MSKIntegratedMutationProfilingofActionableCancerTargets sequencing of PDAC attained from MSK cbioportal[42–44].

The final design was 596 amplicons covering 253 genes' UCSC canonical exon transcripts. The full list of amplicons and more details are provided in Supplementary Table 4.

Library preparation and sequencing. Library preparation and sequencing were done as described in our previous paper, except that for this batch of samples, we used the custom 596-amplicon panel described above.

QC and cell calling. FASTQ files for snDNA libraries were processed through Mission Bio's Tapestri pipeline with default parameters. Briefly, it trims adaptor sequences, aligns reads to the hg19 genome (UCSC) and assigns reads to cell barcodes. The CellFinder module then filtered for barcodes corresponding to 'complete cells/nucleus' based on total read completeness (>8 × number of amplicons) and per-amplicon read completeness (>80% data completeness for working amplicons, which are defined as amplicons with >0.2 × mean of all amplicon reads per qualified barcode).

Mutation calling. As a first pass, with aligned snDNA reads (binary alignment maps, BAMs), Mutect2 (GATK, v.4.2.5.0) was used for mutation calling and filtering in each single-cell BAM. Preliminary technical filters were applied, including 'base_qual/low_allele_frac/weak_evidence/slippage/multiallelic/clustered_events'. Required minimum depth was four

reads and VAF was 0.2. Still, we observed many called mutations present in small numbers of single barcodes, which are likely false positives from PCR/sequencing errors. We concatenated mutations called in each single cell, and with the belief that the false-positive rate of a mutation decreases exponentially with the number of barcodes it is detected in, we only considered mutations present in ≥3 single cells. It should be noted that the choice of three single cells was arbitrary and came from empirical testing, where a prevalence filter of three single cells filtered out good amounts of false-positive mutations while leaving enough mutations for downstream, more stringent filters to be described below.

As a second pass, we used the mutation list from the first pass for genotyping in each single-cell BAM with 'bcftools mpileup', outputting a mutation by single-cell matrix that has information for each mutation in each single cell.

Furthermore, to enable de novo mutation calling, we performed the following QC steps, inspired by bulk DNA-seq bioinformatics.

Detecting rare variants leveraging multiregional sampling. Sampling multiple regions of a tumor enables mutations that have a large clonal fraction in one sample but lower than detection threshold fraction in another to be captured in the latter. Therefore, mutations from all samples of the same tumor were pooled and genotyped in all single nuclei with 'bcftools mpileup'.

Filtering by relative prevalence and mutational signature. In both neoplastic and normal samples taken from different organ sites, we noticed a substantially similar mutational signature, signified by a large proportion of T > C variants, enriched in mutations of <0.5% single-cell prevalence detected by this snDNA-seq technology (Supplementary Fig. 1b). As different types of human tissues have been shown to display distinct mutational signatures and the signature did not seem to match any of them[45], we deemed it likely artifactual and decided to filter out all mutations below 0.5% single-cell prevalence. T > C variants in the 0.5–1% were also filtered.

Assembly of a panel of normal (PON). A PON, proven effective in eliminating library artifacts in bulk sequencing, was assembled for snDNA-seq from the following three sources:

1. Eight unrelated normal pancreas samples were subject to the same snDNA-seq and mutational calling process (until filtering by prevalence). Variants called in ≥4 unmatched normal samples were discarded.
2. Variants present in a bulk WES PON (gs://gatk-best-practices/somatic-b37/Mutect2-exome-panel.vcf) were discarded.
3. snDNA-seq SNV lists of all tumor samples studied here were concatenated, and variants called in >50% samples, except for those in known hotspot (*KRAS* codon 12), were discarded. Manual curation proved this effective in removing low-prevalence SNVs likely caused by library error, and confirmed that no driver mutations identified by matched bulk sequencing were discarded.

To retain real SNPs/SNVs with high prevalence in the population, variants with a frequency >0.01 in the 1000G Project were whitelisted.

Additional filters for phylogenetic analysis. As our phylogenetic analysis relies on high-density single-cell VAF signal of mutations, those with >0.3 biallelic dropout rate in any sample were discarded.

snDNA-seq genetics analysis

Calculating CCFs. With snDNA-seq, because we can observe VAF of each SNV in each single cell, we could calculate CCF directly by designating one clonal population as cancer, and measure the number of cells within that population that show a confident signal for each mutation.

We first genotyped each mutation in each cell based on hard thresholding—a mutation is considered positive in a cell given ≥3 mutant reads, ≥0.2 VAF and ≥8 total depth.

To define the cancer clonal population, tumor cells were defined as those classified as nondiploid by Tapestri-CN, a single-cell CN-calling tool to be described below. Using CN to define cancer cells was preferred over indexing on particular mutations, as other bulk-sequencing bioinformatics, for the following three reasons:

1. It does not make any assumption based on a priori knowledge.
2. Using CN to define cancer cells is prevalent in single-cell RNA-sequencing analysis of PDAC[46].
3. It orthogonally quantifies clonality of SNVs without using any SNV as reference, thereby providing an objective view of relative clonality among SNVs, which is the primary goal for CCF calculation.

Amplifications and homozygous deletions' CCFs were calculated as the proportion of their corresponding clones as inferred by Tapestri-CN. The CCFs of LOHs were calculated as the proportion of their corresponding clones as inferred by fast-ConDoR.

The CCF value of 0.75 to distinguish between clonal and subclonal drivers was chosen empirically (similar to prior literature using bulk sequencing[47]) as we observed a clear gap dividing two clusters (cluster 1, *TP53*, *CDKN2A*, *KRAS*; cluster 2, *SMAD4*, *ARID1A* etc.) at that value. It should be cautioned that this snDNA-seq CCF calculation was subject to noise from technical dropout and sampling errors, and the values should only be used for quantifying 'relative' clonality among SNVs.

snDNA-seq phylogenetics analysis

CN calling from snDNA-seq data. We developed a new tool, 'Tapestri-CN' to infer the CN profiles of individual cells from the snDNA-seq data. The tumor sample from each patient is composed of multiple cancer clones with distinct CN profiles in varying proportions. As such, we model each tumor sample as a mixture of multiple CN profiles with unknown proportions, where the read count data are given by a negative binomial model. We derive an expectation maximization algorithm to infer the CN clones, their proportions and the assignment of each cell to one of the CN clones in each tumor sample. Details of the model and testing/validation process are provided in Supplementary Methods.

Single-cell clone phylogeny inference and refinement. We modified our previously published method ConDoR[19] to make it scalable to our dataset, which had cases with over 10,000 single cells as all samples from the same patient were merged. The modifications required for scalability are described in the Supplementary Methods. The updated method fast-ConDoR takes the following as input: (1) variant and total read counts of mutation (SNPs or SNVs) in each cell, (2) CN clustering of cells into distinct CN clones and (3) annotation of the mutations into two groups—germline SNPs and somatic SNVs. The annotation of mutations is required because, while SNVs are somatic mutations that may be gained or lost during cancer evolution, SNPs are germline mutations that are affected only by LOH events.

We obtained the annotation of the mutation measured from snDNA-seq data by comparing them with mutations derived from bulk-sequencing data of matched normal and tumor samples. Because homozygous SNPs are not informative for cancer phylogeny inference, mutations with pseudobulk VAF of >0.9 in snDNA-seq were filtered out. Upon testing fast-ConDoR, we found a set of loci recurrently called as LOH in non-neoplastic clones (that is, cells without driver mutations such as *KRAS* and *TP53*) in multiple patients, likely arising due to artifactual allelic imbalance at corresponding amplicons caused by PCR. These loci were manually selected and filtered out.

Fast-ConDoR produces a tree where the leaves represent the sequenced cells and the edges indicate gain or loss of SNVs and SNPs. To avoid overfitting the data, we merge CN clones with low proportions (that is, comprising <0.1% total cells) and do not have any SNVs. This was implemented because, based on observation of our samples, such clones always had only likely false-positive LOHs called, thus could not be confirmed. We assign SNV and CN states to each internal vertex using the Sankoff algorithm. Then, the branch length of each edge in the phylogeny is calculated by combining the difference between the SNV and CN states of the source vertex u and target vertex v of the edge using the formula below. $\mathbf{CNV}(v)$ is a vector that describes the copy number states of the vertex v; N_{somatic} is the number of somatic SNVs, and N_{LOH} is the number of LOHs that occurred on the edge. p_1, p_2 and p_3 are normalization factors arbitrarily set to 100/3, 100/3 and 0.5, mostly for illustration purpose.

$$p_1 \times N_{\text{somatic}} + p_2 \times N_{\text{LOH}} + p_3 \times |\mathbf{CNV}(v) - \mathbf{CNV}(u)|$$

Defining driver, finding the trunk of the tree. Because we designed our panel to target only genes critical to pancreatic cancer development and progression, we considered all functional mutations (nonsynonymous, intron, etc.) detected by this panel to be driver events.

Based on the pancreatic cancer genetic progression model and our observation of the inferred phylogenies, none of which had two early-diverging tumor branches, we simply define the trunk to be the first edge that has a somatic SNV event when traversing the tree in preorder. For PC22 and PC24, which do not have somatic SNVs covered by our panel, their trunk could be trivially found by hand.

Statistical analysis. Statistical comparison of *KRAS* mutation sc-VAF distribution between early-stage and late-stage PDACs was performed using a linear mixed-effects model (statsmodels v0.14.4 in Python), with a random intercept for each sample to account for repeated measurements. The fixed effect of stage (early versus late) was tested for a one-sided alternative hypothesis (mean of late stage > mean of early stage, that is, the estimated effect size >0). The estimated effect size was $\hat{\beta} = 0.13$ (s.e. = 0.08), with a one-sided P value of 0.054. To compare the BRCA biallelic inactivation probability between germline *BRCA2* mutation cases (6/6) and other cases (2/18), Fisher's exact test was used to test the hypothesis that there is no association between gBRCA status and BRCA2 biallelic inactivation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Processed H5 files required for replicating relevant analysis using provided scripts (detailed in below section) are published on Zenodo (https://doi.org/10.5281/zenodo.17155593 (ref. 48)). Raw BAM files are deposited to European Genome–phenome Archive (EGA) under Study ID EGAS50000001351. The raw BAM data are available under restricted access, as required by the MSKCC Medical Donation Program Data Access Agreement (MSKCC MDP DAA). Readers interested in gaining access through EGA need to contact the data access committee (DAC) of this dataset and start an application. The DAC will try to respond within 2 weeks but may take longer in special conditions. The MSKCC MDP DAA, to be provided by the DAC and include guidelines and restrictions on data usage, must be signed. Once the application is approved, an EGA account will be provided for data access. The time length of access to the data will be determined by the DAC on a case-by-case basis. Further information about EGA can be found at https://ega-archive.org and 'the European Genome–phenome Archive of human data consented for biomedical research' (http://www.nature.com/ng/journal/v47/n7/full/ng.3312.html).

Code availability

The custom computational analysis pipeline consists of three modules—a bioinformatics pipeline for single-cell mutation calling and genotyping, available at https://github.com/haochenz96/TapVarCallSmk or https://doi.org/10.5281/zenodo.17309019 (ref. 49). CN clone calling pipeline, Tapestri-CN, available at https://github.com/haochenz96/tap_cn_calling or https://doi.org/10.5281/zenodo.17309013 (ref. 50). The fast-ConDoR pipeline for phylogenetic analysis, available at https://github.com/Bolladeen/full-ConDoR or https://doi.org/10.5281/zenodo.17309066 (ref. 51). The downstream analyses to reproduce figures in the manuscript are available at https://github.com/haochenz96/Tapestri_main_manuscript_analysis or https://doi.org/10.5281/zenodo.17309023 (ref. 52).

References

40. Hayashi, A. et al. A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. *Nat. Cancer* **1**, 59–74 (2020).
41. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
42. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
43. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1 (2013).
44. De Bruijn, I. et al. Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal. *Cancer Res.* **83**, 3861–3867 (2023).
45. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
46. Hwang, W. L. et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat. Genet.* **54**, 1178–1191 (2022).
47. Dentro, S. C. et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184**, 2239–2254 (2021).
48. Zhang, H. et al. Dataset for Zhang et al., 'genomic evolution of pancreatic cancer at single-cell resolution'. *Zenodo* https://doi.org/10.5281/zenodo.17155593 (2025).
49. Zhang, H. haochenz96/TapVarCallSmk: Zhang et al. Nature Genetics 2025. *Zenodo* https://doi.org/10.5281/zenodo.17309019 (2025).
50. Zhang, H., Sashittal, P. & Jakatdar, A. haochenz96/tap_cn_calling: Zhang et al. Nature Genetics 2025. *Zenodo* https://doi.org/10.5281/zenodo.17309013 (2025).
51. Zhang, H. & Jakatdar, A. Bolladeen/full-ConDoR: version 1.0.0—initial public release. *Zenodo* https://doi.org/10.5281/zenodo.17309066 (2025).
52. Zhang, H. & Sashittal, P. haochenz96/Tapestri_main_manuscript_analysis: Zhang et al. Nature Genetics 2025. *Zenodo* https://doi.org/10.5281/zenodo.17309023 (2025).

Acknowledgements

The authors thank the Integrated Genomics Operations, the Bioinformatics Core and the Center for Molecular Oncology of MSK for their technical support; the Iacobuzio lab at the David M. Rubenstein Center for Pancreatic Cancer Research of MSK for their work on sample collection and processing; and the Raphael lab at Princeton University for their contribution to computational analysis. The authors also thank A. Makohon-Moore, S. Shah, R. White, C. Delley, Y. Wang, M. Imielinski, M. Arribas-Layton, R. Chaligné and I. Masilionis for their input to the project. C.A.I.D. was supported by the NIH/NCI grant R35

(CA220508-03), U2C (CA233284-03) and Cycle for Survival. E.M.O. was supported by the American Association of Cancer Research, the American Society of Clinical Oncology, Imedex, Research to Practice, Stand Up To Cancer, the NIH/NCI Cancer Center Support Grant/Core P30 (CA008748) and the NCI/NIH P50 (CA257881-01A1). B.J.R. was supported by the NCI U24 (CA248453). A.H. was supported by Daiichi-Sankyo Foundation of Life Science Fellowship, Mochida Memorial Foundation for Medical and Pharmaceutical Research Fellowship. M.H. was supported by Ligue Contre le Cancer, Institut Curie, Fondation Nuovo Soldati, Fondation de France, Fondation Monahan and Institut Servier. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

H.Z. and C.A.I.D. conceptualized the study. H.Z., P.S., E.-R.K., A.J., B.J.R. and C.A.I.D. performed the methodology. S.U., A.C., A.E., C.A.M., A.H., N.L., M.H., W.P., N.P., E.M.O., A.C.W. and C.A.I.D. were responsible for sample collection and processing. J.H. and A.M.N. were responsible for computational processing. H.Z. was responsible for visualization. E.M.O., A.C.W., B.J.R. and C.A.I.D. supervised the study and were responsible for funding acquisition. H.Z., P.S. and C.A.I.D. contributed to writing the manuscript.

Competing interests

H.Z. owns equity in Revolution Medicines and engages in compensated professional activities with Valar Labs. M.H. receives research funding from Servier and Astellas, and receives honoraria from Pierre Fabre, Astrazeneca, Amgen and Viatris. W.P. receives research funding from the NIH/NCI, Merck, Astellas, Lepu Biopharma, Amgen, Revolution Medicines, Break Through Cancer, Parker Institute for Cancer Immunotherapy, Society for Immunotherapy of Cancer and The Society of MSK; is a consulting and advisory board member of Astellas, EXACT Therapeutics, Revolution Medicines, Innovent Biologics and Regeneron Pharmaceuticals, KeyQuest and TD Cowen; and received honoraria for Continuing Medical Education—American Physician Institute, Curio, Integrity and PER. E.M.O. receives institutional research funding from Genentech/Roche, BioNTech, AstraZeneca, Arcus, Elicio Therapeutics, Parker Institute, the NIH/NCI, Digestive Care, Break Through Cancer, Agenus, Amgen and Revolution Medicines; consults (uncompensated) for Arcus, Amgen, AstraZeneca, Ability Pharma, Alligator BioSciences, Pfizer, Agenus, BioNTech, Ipsen, Ikena, Merck, Immuneering, Moma Therapeutics, Novartis, Astellas, Bristol Myers Squibb, Revolution Medicines, Regeneron and Tango Therapeutics; and receives travel reimbursement from BioNTech and Arcus. A.C.W. receives consulting fees from Histosonics for Data and Safety Monitoring Board membership and receives institutional clinical trial funding from Ipsen. All roles are outside of the scope of this manuscript. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02468-9.

Correspondence and requests for materials should be addressed to Benjamin J. Raphael or Christine A. Iacobuzio-Donahue.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Haochen Zhang, Christine Iacobuzio-Donahue

Last updated by author(s): Oct 22, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The custom computational analysis pipeline consists of three modules: (i) A bioinformatics pipeline for single-cell mutation calling and genotyping, available at https://github.com/haochenz96/TapVarCallSmk or 10.5281/zenodo.17309019. (ii) Copy number clone calling pipeline, Tapestri-CN, available at https://github.com/haochenz96/tap_cn_calling or 10.5281/zenodo.17309013. (iii) The fast-ConDoR pipeline for phylogenetic analysis, available at https://github.com/Bolladeen/full-ConDoR or 10.5281/zenodo.17309066.
Data analysis	The downstream analyses to reproduce Figures in the manuscript are available at: https://github.com/haochenz96/Tapestri_main_manuscript_analysis or 10.5281/zenodo.17309023.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Processed H5 files required for replicating relevant analysis using provided scripts (detailed in below section) are published on Zenodo: https://doi.org/10.5281/zenodo.17155593.
Raw BAM files are deposited to European Genomephenome Archive (EGA) under Study ID EGAS50000001351. Readers interested in gaining access through EGA need to contact the data access committee (DAC) of this dataset and start an application. The DAC will try to respond within two weeks but may take longer in special conditions. The MSKCC MDP DAA, to be provided by the DAC and include guidelines and restrictions on data usage, must be signed. Once the application is approved, an EGA account will be provided for data access. The time length of access to the data will be determined by the DAC on a case-by-case basis. Further information about EGA can be found at https://ega-archive.org and "The European Genomephenome Archive of human data consented for biomedical research" (http://www.nature.com/ng/journal/v47/n7/full/ng.3312.html). Other data may be made available upon reasonable request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex breakdown is as follows: male = 14, female = 10. No sex-based analysis was done as that was deemed irrelevant.
Reporting on race, ethnicity, or other socially relevant groupings	Socially-relevant grouping was not applicable in this study. The race breakdown is as follows: white = 16, white spanish/hispanic/latino not otherwise specified (NOS) = 4, black = 1, native American/American Indian/Alaska = 1. 2 patients preferred not to answer. The imbalanced racial composition might confound this study as pancreatic cancers in different populations might be subject to different oncogenic mutational process and thus different genomic features. This could not be well accounted for because we only had access to patients treated at Memorial Sloan Kettering Cancer Center in New York, USA.
Population characteristics	Age at diagnosis statistics of this cohort of N=24 are: mean = 64.74 years, standard deviation = 12.21 years. 6 patients had germline BRCA2 mutations. These patients were intentionally included as their genomic evolution is of high clinical relevance; the proportion in the cohort (25%) is higher than that of the general pancreatic cancer population (~5%, according to Park et al., JAMA 2021). 8 patients' samples in the resected tissue biobank program were treatment naive; 11 patients' samples in the autopsy program were post-treatment; the other 6 patients had longitudinal samples before and after treatment.
Recruitment	Patients treated at Memorial Sloan Kettering Cancer Center were given the option to participate in the relevant research programs (resected tissue biobanking, Last Wish Program and Tracking of Pancreatic Cancer Regression and Resistence) and elected to donate their tissues with written consent.
Ethics oversight	Use of human samples in this study was approved by the institutional review board at Memorial Sloan Kettering Cancer Center (under protocols #06-107, #15-021, #15-149).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The total cohort size (N=24) was as big as we could achieve given the economic cost and logistical complexity of conducting the deep single-cell DNA sequencing analysis on primary human patient samples.
Data exclusions	No patient was excluded from this study.
Replication	Replication was not applicable due to the use of human patient samples. Technical replication was done at each step: single-nucleus DNA sequencing and downstream analyses.
Randomization	Not applicable.

Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*