

The Evolutionary Forest of Pancreatic Cancer



Katelyn M. Mullen¹, Jungeui Hong¹, Marc A. Attiyeh¹, Akimasa Hayashi¹, Hitomi Sakamoto¹, Zachary A. Kohutek¹, Caitlin A. McIntyre¹, Haochen Zhang¹, Alvin P. Makohon-Moore¹, Amanda Zucker¹, Laura D. Wood², Matthew A. Myers³, Brian J. Arnold³, Simone Zaccaria³, Joanne F. Chou⁴, Marinela Capanu⁴, Nicholas D. Socci⁵, Benjamin J. Raphael³, and Christine A. Iacobuzio-Donahue^{1,6,7}



ABSTRACT

The genomic features of pancreatic ductal adenocarcinoma (PDAC) have been well described, yet the evolutionary contexts within which these features occur remain unexplored. We studied genome landscapes, phylogenies, and clonal compositions of 91 PDACs in relation to clinicopathologic features. There was no difference in the number of driver mutations or evolutionary timing when each mutation occurred. High truncal density, a metric of the accumulation of somatic mutations in the lineage that gave rise to each PDAC, was significantly associated with worse overall survival. Polyclonal, monoclonal, or mixed polyclonal/monoclonal metastases were identified across the cohort, highlighting multiple forms of intertumoral heterogeneity. Advanced stage and treated PDACs had higher odds of being polyclonal, whereas oligometastatic PDACs had fewer driver alterations, a lower fractional allelic loss, and increased likelihood of being monoclonal. In sum, our findings reveal novel insights into the dynamic nature of the PDAC genome beyond established genetic paradigms.

SIGNIFICANCE: Although the pancreatic cancer genome has been described, it has not been explored with respect to stages of diagnosis or treatment bottlenecks. We now describe and quantify the genomic features of PDAC in the context of evolutionary metrics and in doing so have identified a novel prognostic biomarker.

INTRODUCTION

The genetic basis of pancreatic ductal adenocarcinoma (PDAC) has been well documented throughout the era of next-generation sequencing. Through the International Cancer Genome Consortium (ICGC; refs. 1–3), The Cancer Genome Atlas (TCGA; ref. 4), Pan-Cancer Analysis of Whole Genomes (5, 6), and others (7, 8), whole-genome sequencing (WGS) and whole-exome sequencing (WES) have been performed on 100 samples to reveal the complex mutational landscape of primary tumors. These studies have revealed both common and infrequent drivers associated with PDAC and clarified the genetic basis of responsiveness to different

standard-of-care therapies (1, 4). Notably, most samples used in these studies were obtained from surgical resections. This is rational given the amount of material needed for sequencing and that surgical resection is the most common mode of obtaining PDAC tissue for research purposes. However, patients with resectable disease comprise only 12% of newly diagnosed PDAC cases (seer.cancer.gov). Furthermore, of those who undergo resection followed by adjuvant therapy, more than 80% relapse and ultimately die of their disease (9–11). These statistics highlight the need to better understand the more common clinical contexts of PDAC such as patients with locally advanced or metastatic disease.

Nonetheless, collecting tumor tissue from patients with stage III or IV disease has proven to be challenging in PDAC. Unlike patients with early-stage disease, patients with late-stage disease do not undergo surgical resection as part of their disease management. As a result, samples are typically collected using either a small tumor core biopsy or fine-needle aspiration, often yielding samples with relatively lower overall cellularity than resections, which makes comprehensive genomic assessment more challenging (12). Owing to these technical challenges, PDAC has been underrepresented in recent studies of metastatic cancer genomes across tumor types (13, 14). Conversely, the MSK-MET pan-cancer cohort contains nearly 1,800 PDAC samples, the largest study of metastatic PDAC to date (15). Although this cohort is sizable, genomic characterization is limited due to the use of a targeted sequencing approach. Furthermore, only a single metastatic sample was studied for most patients. Single-sample analyses can underestimate intratumoral heterogeneity because variants identified as clonal in one sample may be subclonal or even absent in another, giving rise to the “illusion of clonality” (16).

To circumvent some of these obstacles, sampling can be conducted postmortem via research autopsies to enable more extensive sampling than otherwise possible in a living

¹Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York. ²Division of Gastrointestinal Pathology, Department of Pathology, Johns Hopkins Hospital, Baltimore, Maryland. ³Department of Computer Science, Princeton University, Princeton, New Jersey. ⁴Bioinformatics and Epidemiology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York. ⁵Bioinformatics Core, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York. ⁶The David M. Rubenstein Center for Pancreatic Cancer Research, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York. ⁷Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York.

Corresponding Authors: Christine A. Iacobuzio-Donahue, The David Rubenstein Pancreatic Cancer Research Center, Memorial Sloan Kettering Cancer Center, Zuckerman Research Building, 417 East 68th Street, Room 761, New York, NY 10065. E-mail: iacobuzc@mskcc.org; and Benjamin J. Raphael, Department of Computer Science, Princeton University, 35 Olden Street, Room 309 Computer Science, Princeton, NJ 08540. E-mail: braphael@cs.princeton.edu

Cancer Discov 2025;15:329–45

doi: 10.1158/2159-8290.CD-23-1541

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2024 The Authors; Published by the American Association for Cancer Research

patient (17). Although it remains unclear how many samples are required to conclusively determine the composition and clonality of drivers present in a patient's tumor, a multiregion study of clear-cell renal cell carcinoma suggests that, for larger tumors, four to eight samples are sufficient to capture the majority of events (18). Thus far, a handful of multiregion studies have been published addressing specific clinical contexts of PDAC, including treatment-naïve patients with stage IV disease (19), patients with recurrent disease after resection and adjuvant therapy (20), and treated patients with late-stage disease (21). Additionally, the genetic correlates of transcriptional phenotypes in metastatic patients have been explored (21, 22). Cumulatively, these cohorts were relatively small (21, 22) and focused on transcriptional and genomic features during metastatic progression, thus limiting the statistical power for a broader investigation of the evolutionary histories of PDAC. Consequently, we believe that a comprehensive analysis of PDAC, including all stages of diagnosis and standard-of-care treatment paradigms, is lacking. To this end, we aim to quantify the evolutionary features of PDAC that are stage- and/or context-dependent using a multiregion sampling approach on a scale that has not yet been attempted.

RESULTS

Overview of Cohort

We identified 68 PDAC research autopsies from two institutions for inclusion in this study (Fig. 1A). Inclusion and exclusion criteria for patients and samples used are detailed in the "Methods" section. Forty-nine of the 68 research autopsies were previously sequenced (19–21) and 19 were newly sequenced for the purposes of this study (Fig. 1B; Supplementary Table S1). To supplement the multiregion sampled autopsy cohort, we also included 23 multiregion sampled PDACs from surgical resection specimens (Fig. 1A). All clinicopathologic details of the final cohort of 91 patients are shown in Supplementary Table S1.

Collectively, our cohort included 660 distinct samples corresponding to 262 distinct primary tumor samples, 307 metastatic samples, and one normal tissue sample from each of the 91 patients (Fig. 1A; Supplementary Table S2). The median number of tumor samples per resected primary pancreatic cancer was three (range, 1–6). The median number of primary tumor samples per research autopsy was three (range, 1–9), and the median number of metastases per research autopsy was four (range, 1–15). All resected primary tumor samples were treatment-naïve, whereas research autopsy samples were taken from both untreated and treated patients. Metastases encompassed a broad spatial representation of 11 distinct metastatic sites that included locoregional recurrences within the pancreas (Fig. 1C). Fourteen patients were oligometastatic, which we defined as having zero to five metastatic lesions cumulatively from diagnosis to death, as detected by serial imaging results and comprehensive sampling at autopsy (Fig. 1D; Supplementary Table S1; refs. 23, 24). Patients with oligometastatic disease at autopsy were significantly more likely to have been diagnosed with stage III disease ($P < 0.005$; two-sided χ^2 test) than patients with high metastatic burden (Fig. 1E).

All stages of diagnosis were represented, with 36 (40%), 25 (27%), and 30 (33%) patients initially presenting with stage I/II, III, and IV disease, respectively. The median overall survival (OS) of the patients diagnosed with stage I/II (resectable) disease was 29 months (range, 4–94 months), with stage III (locally advanced, nonmetastatic) disease was 17 months (range, 3–62 months), and with stage IV (metastatic) disease was 6 months (range, 1–41 months; Fig. 1F; Supplementary Table S1; ref. 25), consistent with expected outcomes. Male and female patients were nearly equally represented, with the average age at diagnosis for females being 66 ± 13 years and 64 ± 12 years for males, respectively. Forty-two patients (46% of cohort) reported a former or current history of smoking and 25 patients (27% of cohort) had a history of hyperglycemia or type 2 diabetes mellitus, both of which are known risk factors for PDAC (26, 27). Sixteen patients (18% of cohort) had histologic and immunohistochemical features indicative of basal-like phenotypes ranging from focal squamous features present in the recurrent disease to frank squamous differentiation in one or more spatially distinct sites of the primary tumor and matched metastases (Supplementary Table S1). Oligometastatic PDACs were less likely to have basal-like features than non-oligometastatic PDACs, although this difference was not statistically significant (2/14 oligometastatic vs. 14/54 non-oligometastatic; $P = 0.49$; Fisher exact test).

Annotation of Driver Events

A major goal of next-generation sequencing analyses of human tumor tissue is to identify driver gene alterations, i.e., mutations that are predicted to confer a selective survival advantage for the neoplastic cells they occur in. Although *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* are the most common PDAC drivers, other genes may acquire deleterious somatic alterations that also drive cancer cell survival, albeit at a lower frequency (1, 4). Thus, when selecting a next-generation sequencing assay, there is a trade-off between depth and breadth; a high depth is required to accurately identify driver gene mutations, whereas genome-wide detection of passenger mutations helps identify distinct clones and copy-number aberrations (28). We therefore sequenced samples using at least two different methods (Fig. 1B; Supplementary Table S3) and identified 77,741 unique somatic variants (Supplementary Table S4). These variants were screened using a multitool bioinformatics annotation approach ("Methods" section), resulting in the identification of 2,120 distinct driver gene variants across 121 genes, in which the most frequently observed PDAC drivers (i.e., *KRAS* and *TP53*) generally yielded the highest support values (Fig. 2A; Supplementary Fig. S1; Supplementary Table S5). Fourteen percent of drivers were identified exclusively in the targeted sequencing dataset, including major drivers such as *KRAS*, *TP53*, *CDKN2A*, and *SMAD4*, highlighting the importance of deep sequencing to detect driver mutations in lower-purity samples (Supplementary Table S5). The median number of coding driver mutations identified per patient was three (range, 1–15), in line with other recent reports (29, 30). Of the five patients with eight or more driver mutations (PAM10, PAM40, MPAM08, MPAM18, and MPAM32), four harbored somatic

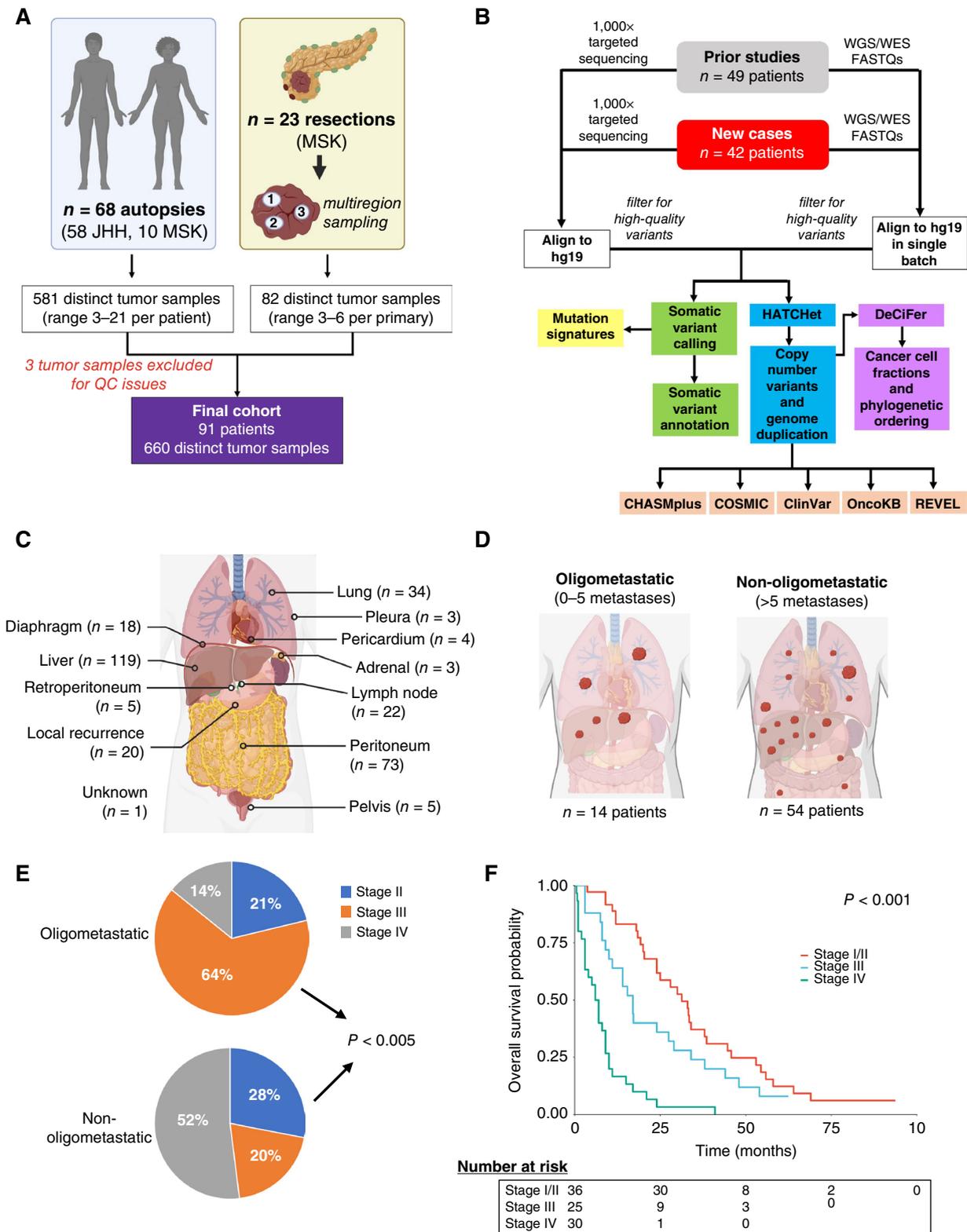


Figure 1. Overview of study and clinical features of cohort. **A**, Schematic illustrating the source of samples from each institution. **B**, Overview of analysis pipeline and sequencing strategy for identification of somatic alterations in tumor tissues from 91 PDAC cases. Mutations were annotated using five independent tools available from OpenCRAVAT. **C**, Locations of all metastatic samples in the study. **D**, Proportion of patients with oligometastatic disease in the study. Oligometastatic disease was considered to be any patient with 0–5 metastases cumulatively from diagnosis to death. **E**, Proportion of patients with oligometastatic disease stratified by stage at diagnosis. **F**, Kaplan–Meier survival curve of all 91 patients stratified by the stage at diagnosis. JHH, Johns Hopkins School of Medicine; MSK, Memorial Sloan Kettering Cancer Center. (Created with BioRender.com.)

Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/15/2/2329/3538220/od-23-1541.pdf> by guest on 25 February 2025

mutations in mismatch repair genes, including *MSH6* and *MSH2*, of which 50% concurrently demonstrated the loss of heterozygosity (LOH). An additional patient harbored a somatic mutation in *ATM*, indicating a potential defect in homologous recombination. The median number of driver mutations identified per sample was three (range, 1–7), indicating a degree of driver gene heterogeneity within the cohort.

To address whether the genomic landscape of end-stage PDAC differs from that reported for early-stage PDAC, we compared our findings to other published datasets that contain a predominance of resectable PDAC (1, 4). Common driver genes, including *KRAS*, *TP53*, *CDKN2A*, and *SMAD4*, among others, were mutated at similar frequencies across the ICGC, TCGA, and our multiregion cohort (Fig. 2B; Supplementary Table S6). We next compared the extent that driver genes with recurrent nonsilent coding mutations (defined as two or more patients) were identified in each study (Fig. 2C; Supplementary Table S7). By this criteria, we find that five driver genes are common to this current cohort, the ICGC and TCGA (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*, and *GNAS*). Six recurrently mutated driver genes were identified in the current cohort and the ICGC but not in the TCGA (*ARID1A*, *ATM*, *PIK3CA*, *RNF43*, *TTN*, and *SMARCA4*), whereas a single recurrently mutated driver gene (*U2AF1*) was identified in the current cohort and the TCGA but not in the ICGC. Recurrent mutations in 13 driver genes were unique to our cohort, including *SMAD2*, *SMAD3*, and *SF3B1*. Although *KRAS* mutations were identified in 84/91 (92%) patients, we identified a broader spectrum of mutant alleles, including E31K, G13P, and G12L, than that reported in other studies (Fig. 2D; refs. 1, 4). Our cohort contained seven patients with wild-type (WT) *KRAS* who harbored driver alterations in *BRAF*, *ATM*, *P TEN*, *CDKN2A*, *TP53*, and *RBM10* (Fig. 2E). One patient had a known germline *BRCA2* mutation and an inactivating somatic alteration of the second allele. Two patients with WT *KRAS* had deletions of *CTNNA2*; in one of the patients (PAM30), this was the only driver alteration identified despite multiple independent sequencing efforts and adequate tumor cellularity (31). We also identified 10 patients with *KRAS* mutations who were found to have additional mutations in genes involved in ERK signaling (Fig. 2F). These included *NF1*, *ERBB2*, *ERBB3*, *ERBB4*, and *EGFR*, among others. For seven of these patients, these mutations occurred in a subclonal manner (Fig. 2G–I). These mutations occurred in both the primary and secondary sites and in both treated and untreated patients.

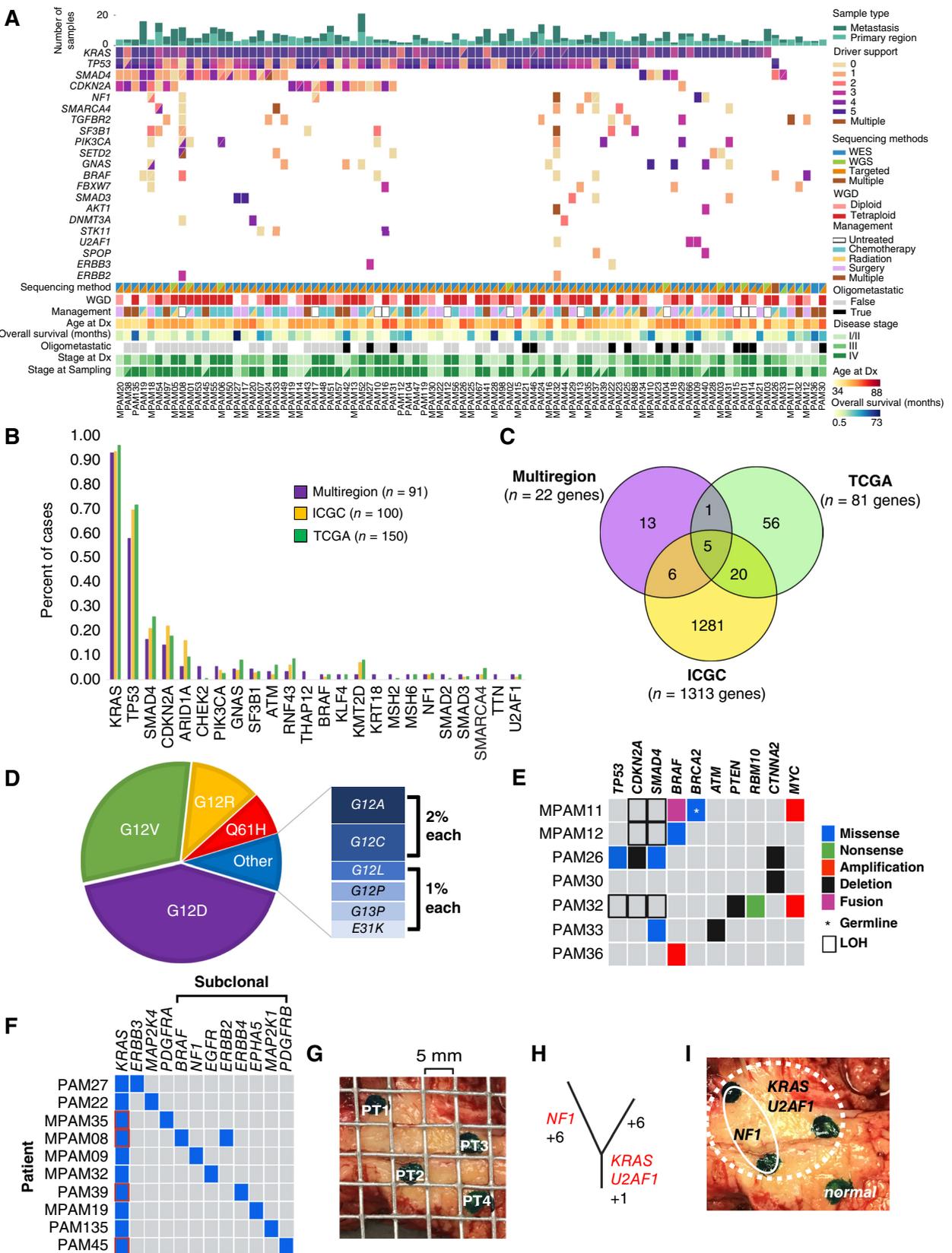
Longitudinal data were available for nine patients in our cohort, with one sample collected at surgical resection and the rest collected after disease recurrence at autopsy. Seven patients acquired additional driver mutations after treatment, including alterations in *PIK3CA*, *PDGFRB*, *HLA-B*, and *MSH2*. In the absence of available longitudinal samples for most patients, we compared drivers identified among patients diagnosed at different stages of disease, i.e., in resectable PDACs to unresectable PDACs. Resectable PDACs were all treatment-naïve, whereas unresectable PDACs were obtained from both treated and untreated patients, allowing us to compare the extent to which drivers were stage specific (only in stage I/II vs. stages III/IV) or emerged after a treatment-induced

genetic bottleneck. Apart from known high-frequency driver genes, the vast majority of drivers identified were observed in a single patient. However, recurrent mutations in *SMAD3* were found in three PDACs (MPAM17, MPAM27, and MPAM29), all of which corresponded to treatment-naïve resection specimens. We also identified recurrent mutations in *THAP12* in the metastases of three additional PDACs (MPAM07, MPAM32, and PAM03), all of which were conservative p.Ala81Val missense variants. In two of the three patients, these mutations were private to a single metastatic lesion. Finally, we did not identify a significant difference in driver counts between the 35 treatment-naïve and 56 treated patients (Wilcoxon rank-sum test, $P = 0.55$).

Given the range of driver mutations identified, we also explored the extent to which there was a relationship between the number of altered driver genes identified and tissue site (primary vs. metastasis). To account for the variable number of primary or metastatic samples per patient, we used a generalized linear mixed model with random intercept and covariance structure to model the number of distinct driver mutations with respect to sample type (primary vs. metastasis). In a univariate analysis, we observed a significant increase in the mean driver mutation count in metastatic samples compared with primary samples [$\beta = 0.178$; 95% confidence interval (CI), 0.027–0.32; $P = 0.02$]. Given this observation, we further investigated whether there was any relationship between the driver count and different metastatic routes (lymphatic, hematogenous, and directly seeded; ref. 32). Using the model outlined above, we found no significant association between the route of metastasis and number of driver genes ($P = 0.60$). Although multiple driver genes were identified across distinct metastatic sites or routes, some were only observed in the context of specific sites or routes. However, these mutations were rare events, and we did not find any significant associations by gene enrichment of different core signaling pathways and routes of metastasis. The lack of association between any specific gene or pathway and a particular metastatic route is in keeping with early accumulation of driver gene alterations that establish metastatic propensity, followed by alternative mechanisms that promote organ-specific colonization, as previously reported (33).

Somatic Copy-Number Alterations

Previous genomic analyses of PDAC have revealed numerous somatic copy-number alterations (CNA) affecting key oncogenes and tumor suppressor genes, including *KRAS*, *TP53*, *SMAD4*, and *CDKN2A*, among others (1, 2, 4, 7). To this end, we used Holistic Allele-specific Tumor Copy-number Heterogeneity (HATCHet), an algorithm developed for analysis of bulk sequencing of multiregion sampled tumors, to infer both allele- and clone-specific CNAs and their relative proportions across multiple samples from a subset of 70 patients in which these metrics could be reliably derived (34, 35). Our analysis revealed a notably high frequency of both clonal and subclonal gains in *KRAS* and *MYC* compared with the rest of the genome (Fig. 3A and B). A high frequency of clonal and subclonal LOHs of *CDKN2A*, *TP53*, *SMAD4*, *TGFBR2*, and *ARID1A* was also observed. Unlike gains of *KRAS* and *MYC*



Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/15/2/2329/3538220/od-23-1541> pdf by guest on 25 February 2025

Figure 2. Genetic alterations within multiregion cohort. **A**, Overview of somatic alterations detected in tumor samples of 90 PDAC cases. One case (PAM30) was not included due to lack of driver mutations identified by three sequencing modalities. Only those genes with mutations in at least two patients, one of which had a driver score of 3 or higher, are shown. The complete list of annotated driver mutations for all patients (*continued on following page*)

that were predominantly subclonal in nature, LOH of PDAC tumor suppressor genes were more likely to be clonal (Fig. 3B). Additionally, homozygous deletions in these known PDAC tumor suppressor genes were identified in 3% to 9% of patients, the majority of which were subclonal and co-occurred with LOH events in the same gene. LOH of 8p was also notable among these 70 patients, although the multitude of genes present in the region precludes identification of any candidate driver genes. Of the four patients with WT *KRAS*, we were able to obtain CNA calls for MPAM11, MPAM12, PAM26, and PAM32; two exhibited gains in *KRAS* (PAM26 and PAM32). All four patients demonstrated LOH of *SMAD4* and *CDKN2A*, whereas MPAM11, PAM26, and PAM32 also exhibited LOH of *TP53*.

Overall, both genome-wide clonal (Wilcoxon rank-sum test, $P = 2e-11$) and subclonal (Wilcoxon rank-sum test, $P = 0.0001$) LOH events were significantly more common than gains in our cohort (Supplementary Tables S8 and S9). An LOH event at any location was detected in all patients (Supplementary Table S8); however, PAM25 was the sole patient without any copy-number gains (Supplementary Table S10). Furthermore, both subclonal LOH events (Wilcoxon rank-sum test, $P = 7.3e-12$) and gains (Wilcoxon rank-sum test, $P = 4.7e-14$) were significantly more common than clonal copy-number events (Fig. 3C and D), suggesting that the majority of CNA events occurred relatively later in tumor evolution, as reported previously (3), and that they may play a crucial role in driving intratumoral heterogeneity and tumor progression.

Whole-genome duplication (WGD) was identified in nearly two-thirds (45/70) of patients (Fig. 3B). This exceeds previously reported rates of WGD in metastatic PDAC (5, 13), potentially due to more comprehensive sampling per patient. No significant association was found between the number of driver mutations and tetraploid status [odds ratio (OR): 1.11; 95% CI, 0.88–1.51; $P = 0.4$]. We observed that tetraploid patients had a significantly higher number of clonal (Wilcoxon rank-sum test, $P = 0.013$) and subclonal (Wilcoxon rank-sum test, $P = 2.99e-6$) LOH events (i.e., fractional allelic loss) than diploid patients, likely due to increased genomic instability (Fig. 3E and F). Neither the proportion of clonal nor subclonal gain events differed significantly with respect to WGD status (Fig. 3F). Furthermore, we found that WGD did not occur more frequently in treated than in treatment-naive patients (Fisher exact test; $P = 0.23$). However, WGD was more frequent in PDACs with basal-like features (13/14, 93%) than in classical-type PDACs (27/56, 48%; Fisher exact test; $P < 0.003$).

Timing of Somatic Events

We quantified intratumoral heterogeneity and evolutionary timing of somatic events in our cohort of 70 patients by computing both cancer cell fractions (CCF) and descendent

cell fractions of single nucleotide variants (SNVs) in regions of copy-number variation using the DeCiFer algorithm (“Methods” section; Fig. 4A; ref. 36). SNVs were classified as truncal if they were inferred to occur before the most recent common ancestor (MRCA) and subtruncal if otherwise. Of the 63% of putative drivers for which a truncal status could be determined (“Methods” section), 81% were classified as truncal and the other 19% as subtruncal (Fig. 4B). PDACs with WGD did not have a significantly different number of truncal or subtruncal driver mutations compared with diploid patients. To compare our results to traditional CCF methods, we ran DeCiFer in CCF mode to determine whether truncal mutations were consistently clonal. Overall, we found that most truncal driver mutations were considered clonal based on CCF calculations. Notably, we identified two *KRAS* mutations and one *TP53* mutation that were all found to be truncal yet had variant allele frequencies that were consistent with subclonality. Upon further investigation, we found that subclonal deletions and gains likely contributed to these truncal mutations having subclonal mutation frequencies, leading to erroneous conclusions regarding the evolutionary timing of somatic mutations from CCF estimates alone. Notably, *KRAS* driver mutations were not universally determined to be truncal events. Of the two patients harboring subtruncal *KRAS* mutations, one had alternative truncal driver events in *AKT1* and *GNAS* (PAM40). The second patient, PAM46, underwent surgical resection followed by adjuvant chemoradiation before passing away from locally recurrent disease. Both the original surgical resection and all samples of recurrent disease harbored a G12R mutation, which was determined to be truncal, whereas a subtruncal G12D mutation was identified in two of the eight samples of locoregional recurrence. Finally, in patient MPAM26, two distinct truncal *KRAS* mutations were found (G12D and E31K); however, manual review of sequencing reads indicated that they occurred on different alleles. Given the spectrum of clinical contexts represented in our dataset, we studied whether the timing of accumulation of driver mutations, i.e., before or after the MRCA (Fig. 4A), differed across these contexts. First, we found no differences in the number of truncal or subtruncal drivers between the 15 treatment-naive stage I/II PDACs versus the 10 treatment-naive stage III/IV PDACs in our cohort (Fig. 4C); when clonal and subclonal amplifications or homozygous deletions of driver genes were included in this analysis (Supplementary Table S7), there remained no statistically significant difference in the number of driver events in untreated early-stage versus untreated late-stage disease. We next focused on late-stage PDACs specifically by comparing the number of truncal or subtruncal drivers in the same 10 treatment-naive stage III/IV PDACs versus the 28 stage III/IV PDACs treated by standards of care; for the purposes of this study, we consider treatments of any kind to indicate PDACs were exposed to one or more treatment bottlenecks

Figure 2. (Continued) can be found in Supplementary Table S4. **B**, Frequency of somatic alteration for driver gene mutations identified in two or more patients in the current study compared with the ICGC and TCGA. **C**, Overlap of driver genes with two or more nonsynonymous mutations in the current study, ICGC (1), and TCGA (4) PDAC cohorts. The list of genes corresponding to the Venn diagram are in Supplementary Table S6. **D**, Frequency of *KRAS* alleles identified. **E**, OncoPrint illustrating *KRAS*-mutant PDACs with mutations in genes also associated with increased ERK signaling. **F**, OncoPrint illustrating driver gene alterations found in *KRAS* WT PDACs. **G**, Image of resection specimen MPAM09 sampled at four independent regions. **H**, Phylogeny of MPAM09 indicating a truncal *KRAS* mutation and a subtruncal *NFI* inactivating mutation. **I**, Overlay of phylogeny onto sample site from MPAM09. Sample PT4 was found to have no tumor cells. (Created with BioRender.com.)

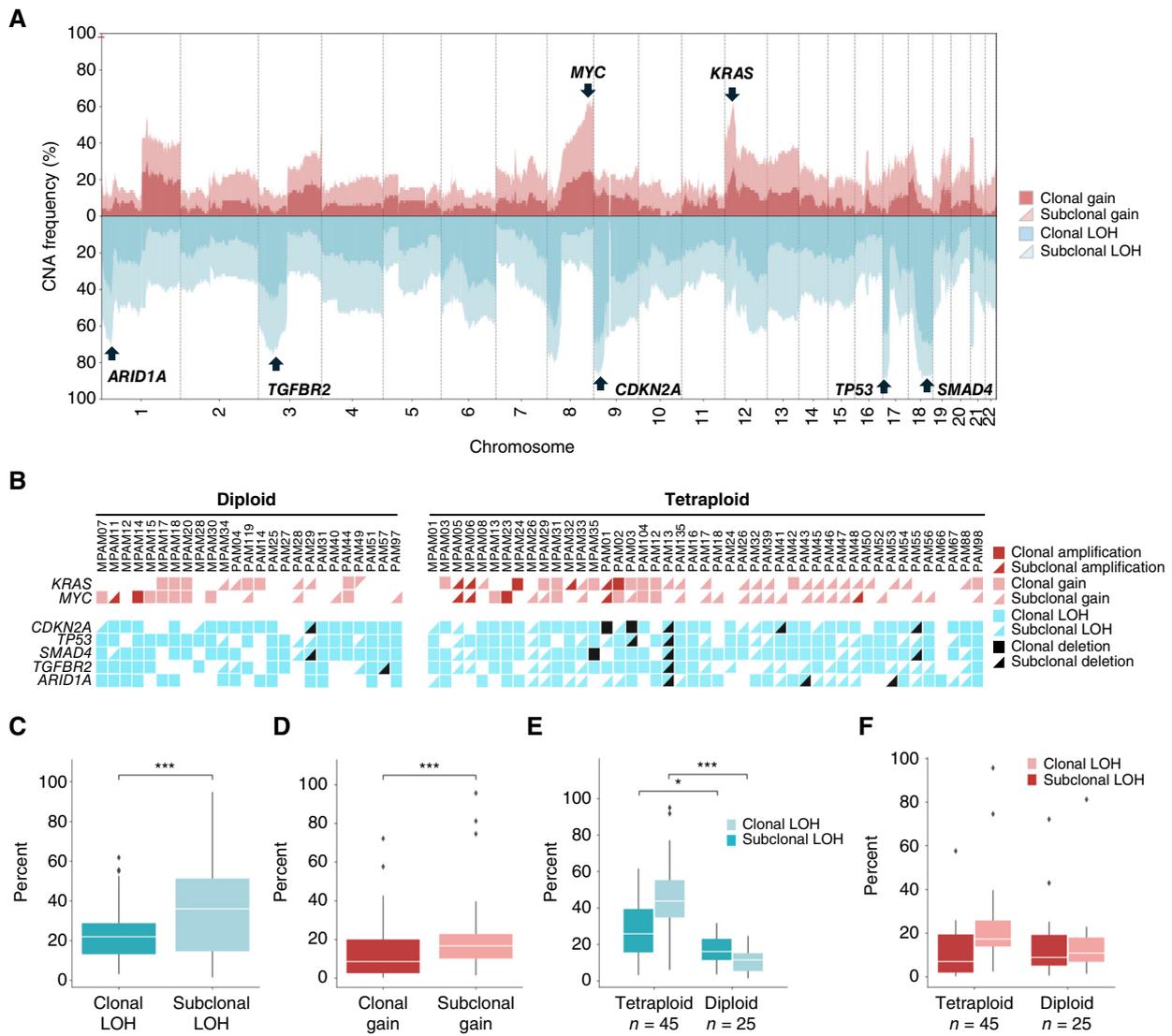


Figure 3. CNAs within multiregion cohort as determined by HATCHet. **A**, Genome wide frequency of CNAs in the cohort. Copy-number gains and losses are indicated in red and blue respectively. Clonal CNAs are shown in darker and subclonal CNAs in lighter shades of their respective colors. Genomic regions containing known driver mutations are indicated. **B**, CNAs for seven major driver genes in PDAC, stratified by ploidy status. **C**, Percent of genome wide copy-number loss events that are clonal vs. subclonal in origin. **D**, Percent of genome wide copy-number gain events that are clonal vs. subclonal in origin. **E**, Relationship of clonal/subclonal loss events to ploidy. **F**, Relationship of clonal/subclonal gain events to ploidy. *, $P < 0.05$; ***, $P < 0.001$. (Created with BioRender.com.)

(Fig. 4C; Supplementary Table S1). Again, no differences were found including when clonal and subclonal amplifications and deletions were included. The median number of truncal drivers per PDAC was 2 (Supplementary Table S11). Using this cutoff, we did not observe any difference in OS between those with more than versus less than two truncal drivers (log-rank test; $P = 0.2$; Fig. 4D). We did find that a one-unit increase in the number of truncal drivers increased all-cause mortality by 20%, although this association did not reach statistical significance (HR: 1.20; 95% CI, 0.95–1.53; $P = 0.14$). There was no difference in the number of truncal or subtruncal drivers among classical PDACs when compared with those with basal-like features.

Given how few subtruncal drivers were identified, it is not possible to know if truncal alterations in specific genes were associated with an increased or decreased likelihood

of subsequent subtruncal alterations (37). In light of these observations, we expanded our scope to assess the timing of all SNV/insertion–deletion events. To do this, we calculated truncal and subtruncal densities for each patient (Fig. 4E). Across the cohort, we observed that subtruncal densities were significantly larger than truncal densities (Wilcoxon rank-sum test, $P = 2.17 \times 10^{-9}$; Supplementary Table S11). No relationship was observed between subtruncal density and the number of samples analyzed per patient, confirming that increased sample number did not introduce bias (Fig. 4F). The majority of patient outliers harbored mutations in genes associated with DNA damage response, including *MSH2*, *MSH3*, *ATM*, and *POLQ*. When comparing treatment-naïve patients with early-stage versus late-stage disease, we found that patients with late-stage disease had larger truncal densities than patients

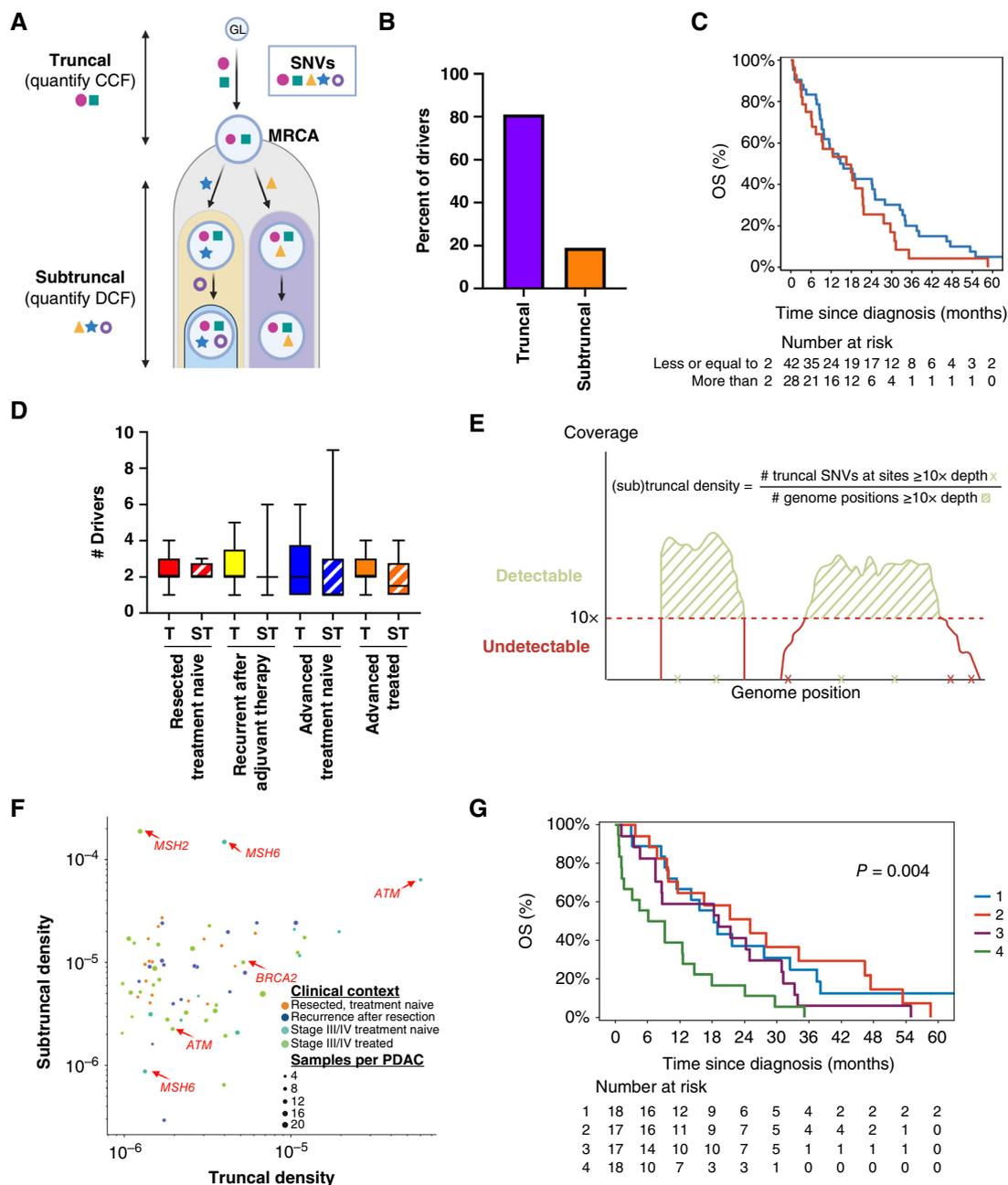
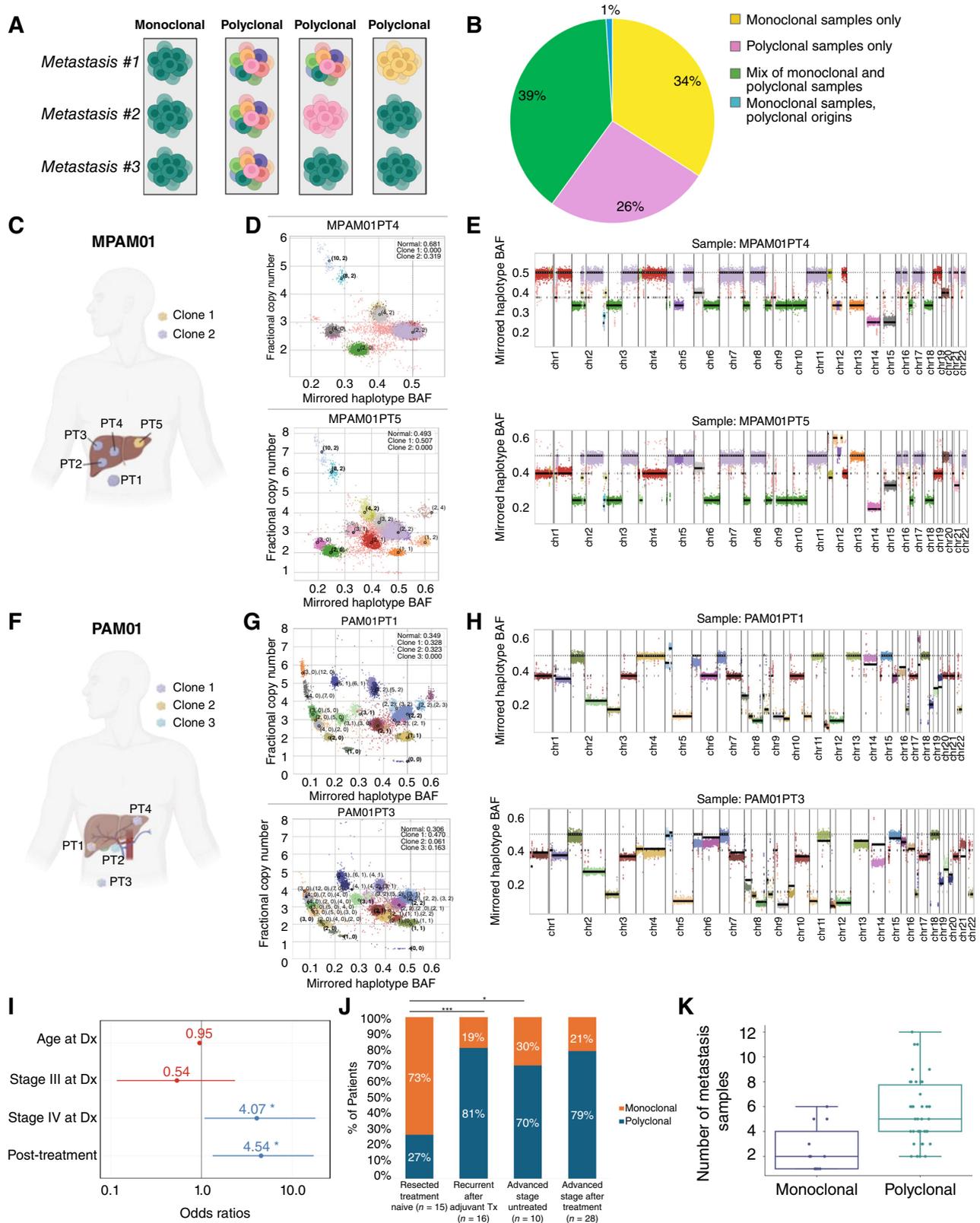


Figure 4. Timing of somatic alterations in PDAC as determined by DeCiFer. **A**, Schematic illustrating timing of occurrence of truncal vs. subtruncal somatic nucleotide variants (SNV) and their role in determining CCF of truncal mutations (occur prior to MRCA) vs. descendant cell fraction (DCF) of subtruncal mutations (occur after MRCA). **B**, Proportion of SNVs determined to be truncal vs. subtruncal in PDACs in current cohort. **C**, Kaplan-Meier survival curve illustrating no OS difference in PDACs with 0–2 vs. >2 truncal driver gene alterations. **D**, Box plot illustrating the median number of truncal and subtruncal drivers in the four clinical categories represented in the cohort. No differences in the number of truncal or subtruncal drivers were found. **E**, Schematic illustrating the method for how (sub)truncal densities are derived from sequencing data. **F**, Scatterplot illustrating the relationship of truncal density to subtruncal density for each patient in which it could be derived. Subtruncal densities are larger than truncal densities. No correlation between the number of samples sequenced per patient and (sub)truncal densities were found nor were differences found between the four clinical categories represented within the dataset. Red arrows indicate PDACs with mutations in DNA damage or mismatch repair genes. See also Supplementary Table S11. **G**, Kaplan-Meier survival curve based on quartiles of truncal density. The highest quartile of truncal density (#4, green line) is significantly correlated with worse OS. GL, germline. (Created with BioRender.com.)

with early-stage disease (Wilcoxon rank-sum test, $P = 0.02$); however, no significant differences were observed with respect to subtruncal densities between these two groups (Wilcoxon rank-sum test, $P = 0.74$). There were no significant differences

in the truncal or subtruncal densities between patients before versus after surgical resection or in those with or without basal-like features. Although there were no significant differences between the subtruncal densities of late-stage treatment-naive



Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/15/2/329/3538220/ced-23-1541> pdf by guest on 25 February 2025

Figure 5. Subclonal copy-number events define PDAC subclones. **A**, Schematic of possible clonal compositions of PDAC progressive disease. Unlike monoclonal PDACs, polyclonal PDACs may be mixtures of clones within the same site or different sites. **B**, Proportion of patients with polyclonal or monoclonal metastases. **C**, Anatomical locations of discrete metastases used for bulk DNA sequencing from patient MPAM01. **D**, Inferred copy-number states for clones identified in MPAM01. Each point represents a genomic bin whose position corresponds to its inferred mirrored (continued on following page)

versus late-stage treated groups (Wilcoxon rank-sum test, $P = 0.53$), treatment-naïve patients had significantly larger truncal densities (Wilcoxon rank-sum test, $P = 0.03$). Moreover, we did not observe a relationship between truncal density and the number of truncal drivers identified per patient.

Given these findings, we determined the relationship between truncal and subtruncal density and survival (Fig. 4G). We found a significant association between truncal densities (per one unit increase in log-scale) and OS (HR: 1.67; 95% CI, 1.16–2.42; $P = 0.006$). Upon categorizing truncal densities into quartiles and adjusting for stage at diagnosis, age at diagnosis, and smoking history, we continued to observe worse survival for patients with higher truncal densities (Q4) than the reference group (Q1) with the lowest truncal densities (HR: 2.91; 95% CI, 1.42–5.99; $P = 0.004$). Conversely, we did not observe any association between subtruncal density and OS (HR: 1.07; 95% CI, 0.81–1.42; $P = 0.64$). Cumulatively, these data suggest that the extent of accumulation of somatic alterations in the cell lineage that gives rise to the MRCA of PDAC is a prognostic marker.

Smoking is a common risk factor for PDAC, contributing to the development of up to 25% of cases (38, 39). We did not observe a significant association between smoking status and truncal density (Fisher exact test; $P = 0.2$), thus we investigated the relationship between smoking and mutational signatures prevalent within each sample. *De novo* extraction of mutational signatures from 134,772 somatic alterations identified seven double-base substitution (DBS), two insertion-deletion (ID), and six single-base substitution (SBS) signatures. SBS *de novo* signature-4, representing 21.0% of all SBSs, was found to be most similar to Catalogue of Somatic Mutations in Cancer (COSMIC) signature SBS29, whose etiology is tobacco related (40), although it also showed similarity to COSMIC SBS4 (Supplementary Fig. S2A and S2B; “Methods” section). DBS *de novo* signature-7 matched DBS2, also a tobacco-related signature. However, as SBSs accounted for 92.3% of all somatic mutations identified, we focused solely on this subset for evaluating the relationship of smoking-related mutational signatures to PDAC evolutionary features. To account for correlations within samples, we used a generalized linear mixed model with random intercept and covariance structure to model the number of SBS *de novo* signature-4 mutations with respect to smoking history status. In a univariate analysis, we did not observe a relationship between smoking history and the number of mutations attributed to the *de novo* SBS4 signature ($P = 0.72$). Similar to previous reports (37), some patients with smoking histories (current or former) contained samples that did not harbor any smoking-related mutations. Conversely, several never-smokers harbored mutations attributed to SBS *de novo* signature-4. These findings suggest that, despite substantial tobacco exposure in some patients, PDAC initiation may be independent of smoking-mediated mutagenesis.

Quantification of Subclones and Clinical Correlates

To date, the clonal composition of PDAC remains poorly understood. To this end, we used HATCHet (34, 35) to infer clonal populations and their relative proportions jointly across multiple samples from the same patient (Supplementary Table S8). The number of subclones identified per patient ranged from one to five, with 34% of patients being classified as monoclonal (Fig. 5A and B). Patients with polyclonal disease demonstrated varying degrees of clonal mixing, with some patients exclusively comprising polyclonal samples (26%) and others harboring a mix of monoclonal and polyclonal samples (39%).

We identified a single polyclonal patient who did not exhibit any clonal mixing, with two clones identified across five monoclonal samples (Fig. 5C). Notably, all samples collected from the right liver and abdominal wall metastasis were composed of one clone, and metastasis from the left liver (MPAM01PT5) comprised exclusively of a different clone. Although all five of the driver mutations identified in MPAM01 were truncal and present in every sample, we observed mirrored subclonal CNAs, or differential gains or losses of the maternal and paternal chromosomes in distinct tumor clones (37). We found that clone 1 had a copy state of (1, 2) or of (2, 4) across regions of chromosome 12 totaling 46.3 Mb, excluding *KRAS* (Fig. 5D and E). Within these same regions, clone 2 demonstrated LOH and amplifications of the opposite allele with copy states of (2, 0) and of (4, 2). Numerous additional subclonal CNA events were also observed on chromosomes 1, 4, 5, 13, 15, 19, 20, and 21 (Fig. 5D). Similarly, PAM01 demonstrated an abundance of subclonal CNAs spanning across all chromosomes and totaling 1.7 Gb (Fig. 5F–H). These events varied in size, ranging from relatively focal events (8q, olive green) to entire chromosomes (chromosome 4, goldenrod; Fig. 5G). The copy-number state of *KRAS* differed in each of the three identified clones (Fig. 5E); however, all clones exhibited LOH of the B-allele (clone 1: 3, 0; clone 2: 12, 0; clone 3: 5, 0; Supplementary Table S8). Notably, mirrored subclonal CNAs were observed on different chromosomes compared with MPAM01, including 5p and 18p (Fig. 5G and H). Overall, 51% (36/70) of patients harbored mirrored subclonal CNAs, and the average frequency of any genomic bin harboring such an event was 5.4%. This phenomenon was observed on every chromosome, with chromosomes 3, 4, and 7 being the most commonly altered across patients. Notable genes exhibiting the highest frequencies of mirrored subclonal CNAs included *TGFBR2*, *MLH1*, and *SETD2*, all of which localize to chromosome 3p (41).

We found that treated PDACs had increased odds of being polyclonal compared with treatment-naïve cancers in a multivariable adjusted analysis (OR: 4.54; 95% CI, 1.33–17.3; $P = 0.019$; Fig. 5I). No significant association was found

Figure 5. (Continued) haplotype B-allele frequency (BAF; x-axis) and fractional copy number (y-axis) seen in **E, F**, Anatomical locations of discrete metastases used for bulk DNA sequencing from patient PAM01. **G**, Inferred copy-number states for clones identified in PAM01. Each point represents a genomic bin whose position corresponds to its inferred mirrored haplotype BAF (x-axis) and fractional copy number (y-axis) seen in **H**. In **H**, points labeled a, b denote the corresponding haplotype-specific copy-number state with “a” indicating copies of the major haplotype and “b” indicating copies of the minor haplotype. **I**, Multivariate analysis illustrating the odds of having polyclonal disease in relation to age, stage, and prior treatment. **J**, Proportion of patients with monoclonal vs. polyclonal disease, categorized by four clinical scenarios of management of patients with PDAC. *, $P < 0.05$; ***, $P < 0.001$. **K**, Number of metastatic samples found to have monoclonal vs. polyclonal compositions. Dx, diagnosis (Created with BioRender.com.)

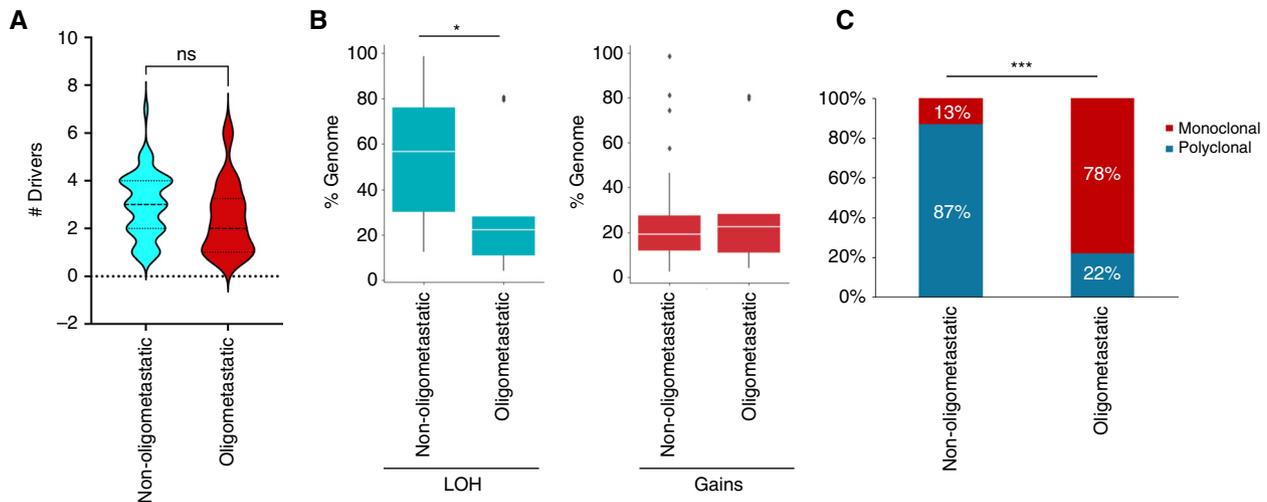


Figure 6. Genetic features of patients with oligometastatic disease. Shown are the comparisons of the total number of driver gene mutations (A), the prevalence of LOH and copy-number gain events (B) and the proportions of mono- vs. polyclonal disease (C) identified in oligometastatic PDACs compared with non-oligometastatic PDACs. *, $P < 0.05$; ***, $P < 0.001$; ns, nonsignificant.

between the number of driver genes and polyclonal status (OR: 1.08; 95% CI, 0.86–1.45; $P = 0.6$). Treatment-naïve resection samples were more often monoclonal compared with samples collected from recurrent disease after adjuvant therapy (Fisher exact test, $P = 0.004$) and treatment-naïve late-stage PDACs (Fisher exact test, $P = 0.049$; Fig. 5J). Among patients with late-stage disease, we did not find a significant difference in the proportion of patients with polyclonal disease (Fisher exact test, $P = 0.7$) who were or were not treated. Cumulatively, our findings indicate that a polyclonal composition of PDAC is significantly associated with disease recurrence after adjuvant therapy and advanced stage at diagnosis. However, among late-stage PDACs, specifically, genetic bottlenecks imposed by treatment do not lead to an increase in proportion of polyclonality.

Expanding upon these observations, we considered the diversity of metastatic sites represented in our cohort and investigated the prevalence of polyclonality with respect to different tissues. Although we found that metastases were frequently polyclonal with respect to different sites and routes of metastatic dissemination (Supplementary Table S2), none of these observations reached statistical significance. However, the number of metastatic samples sequenced was significantly higher in polyclonal patients than in monoclonal patients (Wilcoxon rank-sum test, $P = 0.00096$; Fig. 5K), likely due to polyclonal PDACs being associated with more aggressive features. Additionally, we observed a higher proportion of polyclonal disease in tetraploid patients (33/45) than in diploid patients (13/25); however, this did not reach statistical significance (Fisher exact test, $P = 0.11$).

Features of Patients with Oligometastatic Disease

Although chemotherapy is the standard of care for patients with metastatic disease, clinical management strategies remain poorly defined for patients with oligometastatic disease (42, 43). To this end, we determined the extent to which

patients with oligometastatic disease harbored genetic differences compared with patients with widespread metastatic disease. We found that patients with oligometastatic disease have a median of one fewer drivers than patients with metastatic disease, but this did not reach statistical significance (Wilcoxon rank-sum test, $P = 0.081$; Fig. 6A). No significant associations between gene enrichment of various core signaling pathways and oligometastatic status were found. Furthermore, we found no differences in the number of truncal or subtruncal drivers between patients with oligometastatic versus metastatic disease nor did we find a difference with respect to truncal or subtruncal densities.

With respect to CNAs, patients with oligometastatic disease contained significantly fewer LOH events than patients with metastatic disease (Wilcoxon rank-sum test, $P = 0.02$), although no differences were observed with respect to gains (Fig. 6B). Notably, LOH in both *TP53* and *SMAD4* was significantly less common in patients with oligometastatic disease (Fisher exact test, $P = 0.049$; Supplementary Table S8); however, no difference was observed in other common tumor suppressor genes. Although *MYC* gains were less prevalent in patients with oligometastatic disease, this did not reach statistical significance (Fisher exact test, $P = 0.16$). Extending upon these findings, we found that patients with oligometastatic disease were more often monoclonal than patients with widespread metastatic disease (Fisher exact test, $P < 0.001$; Fig. 6C). Cumulatively, these data suggest that patients with oligometastatic disease have more genomically stable tumors, which in turn may restrain metastatic efficiency (44).

DISCUSSION

This study both corroborates and enriches existing knowledge of the PDAC genome (1, 4). Collectively, our multiregion sampling approach (17) and broadened spectrum of disease

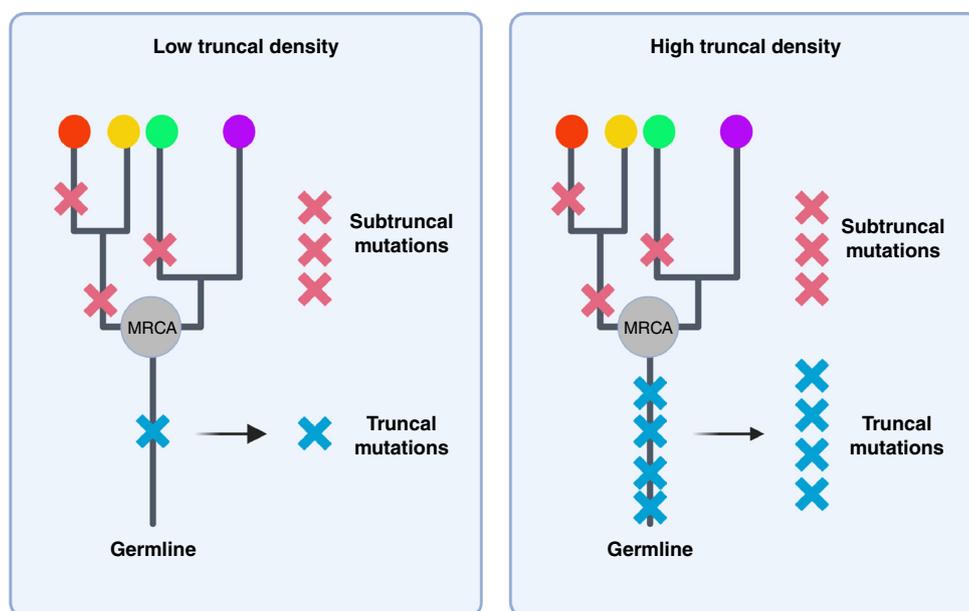


Figure 7. Schematic illustration of truncal density, a novel biomarker of PDAC prognosis. Two identical phylogenies are shown with colored circles indicating distinct clonal populations defined by copy-number states. The location of the MRCA within the evolutionary life history of each PDAC is also shown. Truncal mutations (blue x) are those that occur at any time in the cell lineage spanning oocyte fertilization to the MRCA of the neoplasm, whereas subtruncal mutations (red x) occur on shared or private branches corresponding to distinct subclones. Patients with high truncal densities (fourth quartile) were found to have poor prognosis in a multivariate analysis after controlling for age, stage at diagnosis, and smoking history. (Created with BioRender.com.)

presentations enabled us to quantify the diversity of evolutionary features in this tumor type and correlate them with clinical attributes. A notable insight from this study was derived from our introduction of the concept of truncal density (Fig. 7). This metric quantifies the accumulation of somatic alterations in lineages giving rise to infiltrating carcinoma and was found to be an independent prognostic variable, regardless of disease stage at diagnosis, patient age, or smoking history. Truncal density can be influenced by various endogenous factors such as age-related clock-like mutagenesis (45), chronic genotoxic stress induced by reactive oxygen species (46), inflammation and increased cellular turnover (47), activation of mutational stress responses (47), and exogenous factors such as carcinogens from tobacco smoke (48). Conversely, there are mechanisms that could potentially decrease truncal density during an individual's lifetime, including inherent efficiency of DNA repair (49, 50), immunoediting (51), or genetic drift (52). Notably, smoking history was neither related to truncal density nor was it correlated with the presence of a tobacco-related mutational signature. We stress that this does not imply that smoking is not a risk factor for PDAC, but only that its contribution to PDAC incidence may extend beyond accumulation of mutations. This finding aligns with recent research on smoking and lung carcinogenesis (37), presenting significant implications for both early detection and prevention of PDAC. Furthermore, truncal density did not demonstrate any relationship with the number or type of driver mutations. This observation could reflect a more generalized feature of the PDAC lineage, such as the extent of epigenetic memory after the resolution of inflammation or other injuries (53).

We confirmed that the majority of driver mutations were truncal, and subtruncal drivers were relatively uncommon (54), irrespective of disease stage and treatment. This observation lends optimism to the potential application of targeted therapies against prevalent driver genes in this disease, including *KRAS* (55, 56), as driver events are common to virtually all cells in most PDACs. We also found that patients with oligometastatic disease had a median of one fewer driver mutations than those with widespread metastasis. This observation might be indicative of differences in the number of clonal expansions between these two groups in which one such expansion in the widespread metastasis group may have encompassed an additional driver event, for example, deletion of *SMAD4*. These clonal expansions likely occurred prior to diagnosis given that nearly half of the patients with oligometastatic disease did not experience disease progression between diagnosis and death. Moreover, these expansions may have occurred in association with microenvironmental cues or cell intrinsic features that provided a survival advantage (57). Validation of this theory may guide surgical management in the setting of oligometastatic disease, which remains a controversial topic (42, 43). Subclonal LOH and gains were significantly more common than clonal copy-number events, suggesting that the majority of CNAs occurred relatively later in tumor evolution, as reported previously (3), and that they may play a crucial role in driving intratumoral heterogeneity and tumor progression. Subclonal CNAs, ranging in size from focal events to entire chromosomes, defined distinct populations in polyclonal patients. Notably, some of these events had different numbers of copies of both parental haplotypes in different tumor clones, which we refer to as mirrored

subclonal CNAs (37). Although mirrored subclonal CNAs were identified genome-wide, their biological significance remains unknown. Chromosome 3p had the highest rate of mirrored subclonal CNAs, which may indicate that genes in this region are particularly impacted by convergent evolution (41).

Finally, our findings underscore the necessity of analyzing evolutionary features within the context of different clinical scenarios. In doing so, realization of the features of the “evolutionary forest” of each tumor type will indicate how to best use evolutionary metrics in clinical management and adaptive clinical trials (58, 59). For example, the evolutionary histories exhibited in our study display profound differences compared with other solid tumors such as lung or renal cell cancer, in which subtruncal drivers are more prevalent (54), and current investigations are examining their role in shaping clinical management strategies (52). Specifically in PDAC, prior studies have shown a correlation between the number of driver gene alterations and outcome in resectable PDAC (60). In this instance, it is conceivable that determinations of truncal density, combined with the number and characteristics of driver gene alterations in treatment-naïve PDACs, may aid clinical management by redefining the optimal candidates for surgical resection, radiation, or targeted therapies. Elucidating these patterns in large-scale datasets has the potential to unravel unique, disease-specific, therapeutic approaches, fostering an era of more personalized and effective cancer treatment strategies.

METHODS

Ethics Statement

Written informed consent was obtained from all patients whose tissues were used. The study was conducted in accordance with recognized ethical guidelines (e.g., Declaration of Helsinki, CIOMS, Belmont Report, U.S. Common Rule) and approved by the Institutional Review Boards of the Johns Hopkins School of Medicine and Memorial Sloan Kettering Cancer Center.

Tissue Sample Collection and Processing

Tumor and matched normal tissues were collected through the Gastrointestinal Cancer Rapid Medical Donation Program at the Johns Hopkins Hospital (samples prefixed with PAM) and the Last Wish Program at Memorial Sloan Kettering Cancer Center (samples prefixed with MPAM). Premortem informed consent was obtained from all subjects. Subsequent to their demise, a research autopsy was conducted, and samples from the primary tumor (if not already resected), local recurrence, and metastasis were harvested. All samples were split into equal halves for snap-freezing in liquid nitrogen and formalin fixing respectively such that the fresh frozen sample was a mirror image of the formalin-fixed, paraffin-embedded (FFPE) sample. Hematoxylin and eosin sections were prepared from either frozen or FFPE tissues and reviewed by a gastrointestinal pathologist (A. Hayashi and C.A. Iacobuzio-Donahue) (i) to confirm the diagnosis of PDAC and (ii) to identify tumor-rich regions (>30% cellularity) for DNA purification. For frozen tissues, serial 20- μ m sections were cut from optimal cutting temperature embedded tissue and the area of interest scraped from the slide using a blade. For FFPE tissues, a core was directly punched from the tissue block.

DNA Sequencing

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer’s protocol. DNA quantification, library preparation, and sequencing were performed in the

Integrated Genomics Operation, and preliminary bioinformatics analysis was performed by the Bioinformatics Core at the Memorial Sloan Kettering Cancer Center. Briefly, the Illumina HiSeq 2000, HiSeq 2500, HiSeq 4000, or NovaSeq 6000 platform was used. The majority of samples ($n = 545$) had WES performed at 250 \times coverage; 34 samples had WES performed at 150 \times coverage and 77 samples had WGS performed at 60 \times coverage. The resulting sequencing reads were analyzed *in silico* to assess quality and overall coverage, and alignment to the human reference genome hg19 was performed using BWA v0.7.17 (61). Read deduplication, base quality recalibration, and multiple sequence realignment were performed using the Picard suite and GATK v.3.1 (62). Somatic single-nucleotide variants and ID mutations were detected using Mutect2 (v4.1.2.0) and HaplotypeCaller v.2.4 (63). To validate the mutations found from WES and WGS datasets, one of two different targeted sequencing approaches were used. The majority of samples was sequenced with one of multiple versions of the MSK-IMPACT panel (IMPACT 410, $n = 68$; IMPACT 468, $n = 105$; and IMPACT 505, $n = 174$) with a mean coverage of 500 to 1,000 \times . Another 285 samples were sequenced using a custom targeted panel described elsewhere (Supplementary Table S3). BAM files and associated metadata have been uploaded to the European Genome-Phenome Archive (64).

Filtering and Annotation of Variants

For each patient, somatic variants were filtered using the following criteria: patient-matched normal coverage ≥ 10 reads, variant count in patient-matched normal < 2 reads, patient-matched normal variant frequency < 0.02 , tumor coverage ≥ 20 reads, and tumor variant allele frequency (VAF) ≥ 0.05 in at least one tumor sample. Variants were further filtered to include those present in coding regions only. Mutations located in blacklisted regions defined by ENCODE and RepeatMasker (<https://github.com/mskcc/ngs-filters/blob/master/data/source.txt>) were ignored. In FFPE samples, if a mutation exhibited VAF < 0.1 and was identified as a C>T substitution, it was considered an FFPE artifact and thus excluded from subsequent analyses. Filtered variants were then annotated by OpenCRAVAT v2.2.7 (65) to identify likely functional driver mutations. When available, criteria for determining whether a mutation should be considered a driver by a given module were informed by recommendations made for interpreting results from the module in the OpenCRAVAT store. For CHASMPplus (66), the adjusted P -value had to be < 0.05 . For COSMIC (67), variants had to be present at least four times in the database to be considered a driver. For ClinVar (68), the clinical significance value had to be “Pathogenic,” “Pathogenic/Likely pathogenic,” “Likely pathogenic,” “Pathogenic, drug response, other,” or “drug response”, and the Review Status had to be “criteria provided, multiple submitters, no conflicts” or “reviewed by expert panel.” For OncoKB (69), a mutation had to be labeled as either “Oncogenic” or “Likely Oncogenic.” For REVEL, the score had to be > 0.7 (70, 71). These modules were selected on the basis of a survey of the most commonly used variant annotators in scientific literature. A final driver score for each mutation was calculated by tallying how many annotators classified the mutation as a driver event (maximum score = 5). Mutations with a score of 2 or higher were considered driver mutations and those with a score of 1 or lower were not.

Copy-Number Analysis and WGD Prediction

WGD and allele-specific CNAs were inferred for WES and WGS datasets using HATCHet v1.2.0 (34, 35). Only autosomes were used for copy-number analysis, and phasing was performed with SHAPEIT v2.r90 (72). Upon manual review of computed read-depth ratios, B-allele frequencies, and clusters, parameters for clustering refinement were reviewed for consensus by four of the authors (K.M. Mullen, B.J. Arnold, M.A. Myers, and B.J. Raphael). Copy-number calls of sufficient quality could not be obtained for 21 patients.

Evolutionary Analysis of Driver Mutations

To identify clusters of SNVs which occur in the same phylogenetic branch of tumor evolution, we used DeCiFer v2.1.3 (36). HATCHet output was used as the copy-number input for this analysis. Therefore, only the subset of 70 patients with HATCHet results underwent analysis with DeCiFer. To ensure timely run completion, we required a minimum VAF of 0.05 for WES datasets and 0.1 for WGS datasets. To generate custom state trees, a maximum copy number of 6 was used for each patient. When the total copy-number profile of a given mutation was >6 or read depths were not >1 across all samples, mutations could not be analyzed by DeCiFer. Furthermore, the timing of driver mutations identified exclusively in targeted sequencing datasets could not be analyzed with DeCiFer because HATCHet could not be performed on these datasets. Truncal and subtruncal densities were calculated by dividing the number of truncal or subtruncal SNVs at sites with >10 \times depth by the number of genomic positions using the same depth threshold.

De Novo Mutational Signature Analysis

De novo mutational signatures including SBS, DBS, and ID were evaluated in comparison with COSMIC Mutational Signatures version 3 (<https://cancer.sanger.ac.uk/signatures/>) using the R package Palimpsest (73). Only somatic variants that met the filtering criteria outlined in the “Filtering and Annotation of Variants” section, and those located within 5' or 3' UTR regions, were included in this analysis. Additionally, FFPE tumor samples were excluded as they induce massively unique signatures that are not relevant with cancer progression. Significant *de novo* mutational signatures were extracted using the nonnegative matrix factorization algorithm (74). The maximum number of nonnegative matrix factorization runs and *de novo* signatures were set to 30 and 20, respectively. Finally, we assigned the most representative *de novo* signature of SBS, DBS, and ID types among all those extracted to each variant based on probability scores generated by Palimpsest.

Data Visualizations

The oncoPrint was created with CoMut (75). Anatomic diagrams and schematics and other annotations were created using BioRender (<https://biorender.com/>).

Data Availability

All BAM files and associated metadata for each patient in this study are publicly available through the European Genome-Phenome Archive (EGA; <http://www.ebi.ac.uk/ega>) under the accession number EGAS00001007379 (64).

Authors' Disclosures

H. Zhang reports other support from Valar Labs outside the submitted work. A.P. Makohon-Moore reports grants from the NCI, the American Cancer Society, the Scott Mackenzie Foundation, the American Association for Cancer Research, and the New Jersey Health Foundation during the conduct of the study. A. Zucker reports grants from Ruth L Kirschstein T32 MD/PG Training Grant outside the submitted work. N.D. Socci reports grants from NIH during the conduct of the study. C.A. Iacobuzio-Donahue reports other support from Bristol Myers Squibb outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

K.M. Mullen: Conceptualization, resources, data curation, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **J. Hong:**

Resources, formal analysis, visualization, methodology. **M.A. Attiyeh:** Resources, software. **A. Hayashi:** Resources, formal analysis. **H. Sakamoto:** Resources, formal analysis. **Z.A. Kohutek:** Resources, formal analysis. **C.A. McIntyre:** Conceptualization, resources, formal analysis, methodology. **H. Zhang:** Resources, formal analysis, methodology. **A.P. Makohon-Moore:** Conceptualization, formal analysis, supervision, methodology. **A. Zucker:** Resources, methodology. **L.D. Wood:** Resources. **M.A. Myers:** Resources, software, formal analysis, visualization, methodology. **B.J. Arnold:** Conceptualization, resources, data curation, software, formal analysis, visualization. **S. Zaccaria:** Resources, software, visualization. **J.F. Chou:** Formal analysis, investigation, writing—review and editing. **M. Capanu:** Resources, formal analysis, supervision, investigation, writing—review and editing. **N.D. Socci:** Resources, software, supervision, methodology. **B.J. Raphael:** Conceptualization, resources, software, formal analysis, supervision, validation, investigation, visualization, methodology, project administration, writing—review and editing. **C.A. Iacobuzio-Donahue:** Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, visualization, methodology, writing—original draft, project administration, writing—review and editing.

Acknowledgments

This work was funded by grants from NIH/NCI R35 CA220508-03 and U2C CA233284-03 (to C.A. Iacobuzio-Donahue) and P50 CA25788 (to C.A. Iacobuzio-Donahue); the Cycle for Survival Foundation, the David M. Rubenstein Center for Pancreatic Cancer Research, the Daiichi-Sankyo Foundation of Life Science Fellowship, and the Mochida Memorial Foundation for Medical and Pharmaceutical Research Fellowship (to A. Hayashi); NIH/NCI K99/R00 CA229979 (to A.P. Makohon-Moore); and F31 CA260796-01 (to K.M. Mullen). We are grateful to the Integrated Genomics Operations and the Bioinformatics Core at Memorial Sloan Kettering Cancer Center for their technical support, as well as to Rajya Kappagantula and Elias-Ramzey Karnoub for technical and computational assistance during the resubmission phase of this article.

Note

Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

Received January 18, 2024; revised August 6, 2024; accepted October 4, 2024; published first October 8, 2024.

REFERENCES

- Waddell N, Pajic M, Patch A-M, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015;518:495–501.
- Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;531:47–52.
- Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 2016;538:378–82.
- Raphael BJ, Hruban RH, Aguirre AJ, Moffitt RA, Yeh JJ, Stewart C, et al. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* 2017;32:185–203.e13.
- Gerstung M, Jolly C, Leshchiner I, Dentre SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. *Nature* 2020;578:122–8.
- Dentre SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* 2021;184:2239–54.e39.

7. Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008;321:1801–6.
8. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin W-C, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun* 2015;6:6744.
9. Groot VP, Gemenetzi G, Blair AB, Ding D, Javed AA, Burkhart RA, et al. Implications of the pattern of disease recurrence on survival following pancreatectomy for pancreatic ductal adenocarcinoma. *Ann Surg Oncol* 2018;25:2475–83.
10. Jones RP, Psarelli E-E, Jackson R, Ghaneh P, Halloran CM, Palmer DH, et al. Patterns of recurrence after resection of pancreatic ductal adenocarcinoma. *JAMA Surg* 2019;154:1038.
11. Daamen LA, Dorland G, Brada LJH, Groot VP, van Oosten AF, Besselink MG, et al. Preoperative predictors for early and very early disease recurrence in patients undergoing resection of pancreatic ductal adenocarcinoma. *HPB (Oxford)* 2022;24:535–46.
12. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. *Cell* 2017;168:584–99.
13. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* 2019;575:210–16.
14. van de Haar J, Hoes LR, Roepman P, Lolkema MP, Verheul HMW, Gelderblom H, et al. Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. *Nat Med* 2021;27:1553–63.
15. Nguyen B, Fong C, Luthra A, Smith SA, DiNatale RG, Nandakumar S, et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* 2022;185:563–75.e11.
16. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014;346:251–6.
17. Iacobuzio-Donahue CA, Michael C, Baez P, Kappagantula R, Hooper JE, Hollman TJ. Cancer biology as revealed by the research autopsy. *Nat Rev Cancer* 2019;19:686–97.
18. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* 2018;173:595–610.
19. Makohon-Moore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet* 2017;49:358–66.
20. Sakamoto H, Attiye MA, Gerold JM, Makohon-Moore AP, Hayashi A, Hong J, et al. The evolutionary origins of recurrent pancreatic cancer. *Cancer Discov* 2020;10:792–805.
21. Hayashi A, Fan J, Chen R, Ho Y-J, Makohon-Moore AP, Lecomte N, et al. A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. *Nat Cancer* 2020;1:59–74.
22. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O’Kane GM, et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat Genet* 2020;52:231–40.
23. Hellman S, Weichselbaum RR. Oligometastases. *J Clin Oncol* 1995;13:8–10.
24. Kaneda H, Saito Y. Oligometastases: defined by prognosis and evaluated by cure. *Cancer Treat Commun* 2015;3:1–6.
25. Wang H, Liu J, Xia G, Lei S, Huang X, Huang X. Survival of pancreatic cancer patients is negatively correlated with age at diagnosis: a population-based retrospective study. *Sci Rep* 2020;10:7048.
26. Lynch SM, Vrieling A, Lubin JH, Kraft P, Mendelsohn JB, Hartge P, et al. Cigarette smoking and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Am J Epidemiol* 2009;170:403–13.
27. Huxley R, Ansary-Moghaddam A, Berrington de González A, Barzi F, Woodward M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br J Cancer* 2005;92:2076–83.
28. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. *Nat Rev Genet* 2019;20:404–16.
29. Martínez-Jiménez F, Movasati A, Brunner SR, Nguyen L, Priestley P, Cuppen E, et al. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* 2023;618:333–41.
30. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018;174:1034–35.
31. Kohutek ZA, Rosati LM, Hong J, Poling J, Attiye MA, Makohon-Moore AP, et al. An unusual genomic variant of pancreatic ductal adenocarcinoma with an indolent clinical course. *Cold Spring Harb Mol Case Stud* 2017;3:a001701.
32. Cortés-Guiral D, Hübner M, Alyami M, Bhatt A, Ceelen W, Glehen O, et al. Primary and metastatic peritoneal surface malignancies. *Nat Rev Dis Primers* 2021;7:91.
33. Reichert M, Bakir B, Moreira L, Pitarresi JR, Feldmann K, Simon L, et al. Regulation of epithelial plasticity determines metastatic organotropism in pancreatic cancer. *Dev Cell* 2018;45:696–711.e8.
34. Zaccaria S, Raphael BJ. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat Commun* 2020;11:4301.
35. Myers MA, Arnold BJ, Bansal V, Balaban M, Mullen KM, Zaccaria S, et al. HATCHet2: clone- and haplotype-specific copy number inference from bulk tumor sequencing data. *Genome Biol* 2024;25:130.
36. Satas G, Zaccaria S, El-Kebir M, Raphael BJ. DeCifering the elusive cancer cell fraction in tumor heterogeneity and evolution. *Cell Syst* 2021;12:1004–18.e10.
37. Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* 2023;616:525–33.
38. Fuchs CS, Colditz GA, Stampfer MJ, Giovannucci EL, Hunter DJ, Rimm EB, et al. A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med* 1996;156:2255–60.
39. Momi N, Kaur S, Ponnusamy MP, Kumar S, Wittel UA, Batra SK. Interplay between smoking-induced genotoxicity and altered signaling in pancreatic carcinogenesis. *Carcinogenesis* 2012;33:1617–28.
40. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature* 2020;578:94–101.
41. Zbarovsky ER, Lerman MI, Minna JD. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* 2002;21:6915–35.
42. Ghidini M, Petrillo A, Salati M, Khakoo S, Varricchio A, Tomasello G, et al. Surgery or locoregional approaches for Hepatic oligometastatic pancreatic cancer: myth, hope, or reality? *Cancers* 2019;11:1095.
43. De Simoni O, Scarpa M, Tonello M, Pilati P, Tolin F, Spolverato Y, et al. Oligometastatic pancreatic cancer to the liver in the era of neoadjuvant chemotherapy: which role for conversion surgery? A systematic review and meta-analysis. *Cancers (Basel)* 2020;12:3402.
44. Zhong Y, Macgregor-Das A, Saunders T, Whittle MC, Makohon-Moore A, Kohutek ZA, et al. Mutant p53 together with TGFβ signaling influence organ-specific hematogenous colonization patterns of pancreatic cancer. *Clin Cancer Res* 2017;23:1607–20.
45. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. *Nat Genet* 2015;47:1402–7.
46. Srinivas US, Tan BWQ, Vellayappan BA, Jeyasekharan AD. ROS and the DNA damage response in cancer. *Redox Biol* 2019;25:101084.
47. Greten FR, Grivennikov SI. Inflammation and cancer: triggers, mechanisms, and consequences. *Immunity* 2019;51:27–41.
48. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* 2016;354:618–22.
49. Hoeijmakers JHJ. DNA damage, aging, and cancer. *N Engl J Med* 2009;361:1475–85.
50. Roos WP, Thomas AD, Kaina B. DNA damage and the balance between survival and death in cancer biology. *Nat Rev Cancer* 2016;16:20–33.
51. Luksza M, Sethna ZM, Rojas LA, Lihm J, Bravi B, Elhanati Y, et al. Neoantigen quality predicts immunoeediting in survivors of pancreatic cancer. *Nature* 2022;606:389–95.

52. Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer evolution and the limits of predictability in precision cancer medicine. *Trends Cancer* 2016;2:49–63.
53. Del Poggetto E, Ho I-L, Balestrieri C, Yen E-Y, Zhang S, Citron F, et al. Epithelial memory of inflammation limits tissue damage while promoting pancreatic tumorigenesis. *Science* 2021;373:0486.
54. Iacobuzio-Donahue CA, Litchfield K, Swanton C. Intratumor heterogeneity reflects clinical disease course. *Nat Cancer* 2020;1:3–6.
55. Bekaii-Saab TS, Spira AI, Yaeger R, Buchsacher GL, McRee AJ, Sabari JK, et al. KRYSTAL-1: updated activity and safety of adagrasib (MRTX849) in patients (Pts) with unresectable or metastatic pancreatic cancer (PDAC) and other gastrointestinal (GI) tumors harboring a KRAS^{G12C} mutation. *J Clin Oncol* 2022;40(4 Suppl):519.
56. Hallin J, Bowcut V, Calinisan A, Briere DM, Hargis L, Engstrom LD, et al. Anti-tumor efficacy of a potent and selective non-covalent KRAS^{G12D} inhibitor. *Nat Med* 2022;28:2171–82.
57. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* 2015;17:816–26.
58. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. *Cancer Res* 2009;69:4894–903.
59. Fittall MW, Van Loo P. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Med* 2019;11:20.
60. Qian ZR, Rubinson DA, Nowak JA, Morales-Oyarvide V, Dunne RF, Kozak MM, et al. Association of alterations in main driver genes with outcomes of patients with resected pancreatic ductal adenocarcinoma. *JAMA Oncol* 2018;4:e173420.
61. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
62. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
63. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–19.
64. Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, Ur-Rehman S, et al. The European Genome-Phenome Archive of human data consented for biomedical research. *Nat Genet* 2015;47:692–5.
65. Pagel KA, Kim R, Moad K, Busby B, Zheng L, Tokheim C, et al. Integrated informatics analysis of cancer-related variants. *JCO Clin Cancer Inform* 2020;4:310–17.
66. Tokheim C, Karchin R. CHASMPplus reveals the scope of somatic missense mutations driving human cancers. *Cell Syst* 2019;9:9–23.e8.
67. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019;47:941–7.
68. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:1062–7.
69. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol* 2017;1:1–16.
70. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet* 2016;99:877–85.
71. Tian Y, Pesaran T, Chamberlin A, Fenwick RB, Li S, Gau C-L, et al. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci Rep* 2019;9:12752.
72. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* 2014;10:1004234.
73. Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics* 2018;34:3380–1.
74. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 2004;101:4164–9.
75. Crowdis J, He MX, Reardon B, Van Allen E.M. CoMut: visualizing integrated molecular information with comutation plots. *Bioinformatics* 2020;36:4348–9.