ABSTRACT
Surgery is the only curative option for stage I/II pancreatic cancer; nonetheless, most patients will experience a recurrence after surgery and die of their disease.

To identify novel opportunities for management of recurrent pancreatic cancer, we performed whole-exome or targeted sequencing of 10 resected primary cancers and matched intrapancreatic recurrences or distant metastases. We identified that recurrent disease after adjuvant or first-line platinum therapy corresponds to an increased mutational burden. Recurrent disease is enriched for genetic alterations predicted to activate MAPK/ERK and PI3K–AKT signaling and develops from a monophyletic or polyphyletic origin. Treatment-induced genetic bottlenecks lead to a modified genetic landscape and subclonal heterogeneity for driver gene alterations in part due to intermetastatic seeding. In 1 patient what was believed to be recurrent disease was an independent (second) primary tumor. These findings suggest routine post-treatment sampling may have value in the management of recurrent pancreatic cancer.

SIGNIFICANCE: The biological features or clinical vulnerabilities of recurrent pancreatic cancer after pancreaticoduodenectomy are unknown. Using whole-exome sequencing we find that recurrent disease has a distinct genomic landscape, intermetastatic genetic heterogeneity, diverse clonal origins, and higher mutational burden than found for treatment-naïve disease.

See related commentary by Bednar and Pasca di Magliano, p. 762.
INTRODUCTION

Pancreatic ductal adenocarcinoma (PDA) is currently the third leading cause of cancer-related death in the United States and is projected to become the second leading cause of cancer death within 5 years (1). Several reasons account for these statistics, including an inability to diagnose the disease when it is at a curative stage, late presentation, and modest impact of current best available therapies (2). There is a limited understanding of the genetics of recurrent disease which limits targeted therapy opportunities or informed design of clinical trials.

Approximately 10% to 15% of newly diagnosed patients with PDA are diagnosed with early-stage disease (stage I or II). For these patients, surgical resection followed by adjuvant therapy is the only option for cure (2). Although long-term survival following resection of PDA has been reported (3, 4), the majority of patients who undergo resection will experience recurrence locally or at distant sites and die of their disease within 5 years. Several factors have been shown to have predictive or prognostic value for disease-free or overall survival in patients with resected PDA, including a high ratio of involved to total resected lymph nodes, larger tumor size, high tumor grade, the presence of vascular and perineural invasion, or variably positive margins (5, 6). Venous invasion is very common in pancreatic cancer and may contribute to the aggressiveness of this disease (7). Molecular features of PDA have also been attributed to worse outcome after surgery. For example, patients with coincident TP53 and SMAD4 alterations have shorter disease-free survival than patients whose tumors do not have these genetic alterations (8). Alternatively, the presence of a basal expression signature (9), paucity of an immune signature (10), or microbial dysbiosis (11) has also been associated with worse overall survival.

We have previously reported in a small cohort of treatment-naïve patients with stage IV PDA that we found no evidence of driver gene heterogeneity among primary and metastatic sites (12). What heterogeneity was found in those patients corresponded to passenger mutations only. Moreover, the extent of passenger gene heterogeneity was far less than may be seen in spatially distinct cells within normal tissues, indicating at least one clonal sweep occurred within the primary tumor prior to metastatic dissemination. In contrast, the extent to which metastatic PDA that arises following surgical
reection exhibits similar features is unknown. To address this question and to improve our understanding of recurrent PDA following resection, we performed whole-exome and/or targeted sequencing of 10 primary PDAs, matched local (pancreatic resection bed) recurrences, and multiple anatomically diverse metastases. We identified that pancreatic cancer recurrences following surgery have an increased mutational burden and distinct subclonal origins, and in some instances are characterized by somatic mutations with potential implications for clinical management.

RESULTS

We screened a collection of more than 160 PDA research autopsies to identify patients for whom a sample of their original surgical pancreatic resection was available, who underwent adjuvant treatment after surgery, and who had histologically confirmed recurrent disease within the pancreatic remnant and one or more metastases to anatomically distinct sites such as the liver, lungs, or peritoneum (Fig. 1A and B). We identified 9 such patients for study (Supplementary Table S1). An additional patient was included who did not have metastatic disease at autopsy but did have an aggressive local recurrence with multiple geographically distinct samples of this mass available for profiling. One normal tissue sample from each patient was also used to distinguish somatic from germline variants (Supplementary Table S2). Histologic review of each tumor sample indicated that in 2 patients the primary tumor was conventional ductal adenocarcinoma, whereas the recurrent disease had squamous features (PAM39) or squamous differentiation (PAM46). In a third patient, the primary tumor exhibited classic ductal adenocarcinoma, whereas the recurrent disease was anaplastic (PAM37). One patient had a primary small-cell carcinoma of the pancreas (PAM41). The recurrent disease was notable for transition to a large cell phenotype (Supplementary Fig. S1).

All samples were analyzed by whole-exome sequencing (WES) to a median of 330x coverage (range 135x–652x; see Methods; Supplementary Table S3). To supplement the breadth of WES in these samples, we also performed concurrent deep sequencing using a targeted sequencing panel to ensure greater sensitivity for mutations and DNA copy-number alterations in 410 cancer-associated genes (ref. 13; Supplementary Table S4). We identified 4,864 total somatic single-nucleotide variants and small insertions or deletions (indels) with a median of 57 per sample (range 17–203; Fig. 1C; Supplementary Table S5). We found no evidence of microsatellite instability (MSI) in these 10 patients, nor were germline or somatic mutations identified in recognized MSI-related or other DNA damage repair genes (14).

To better understand the patterns of mutation accumulation in primary versus recurrent disease, we performed mutational signatures analysis of all samples that underwent WES (Fig. 1C). Nine signature classes were identified in our cohort: aging (signatures 1 and 5), double-strand break repair (DSBR; signature 3), apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC; signatures 2 and 13), mismatch repair defects (signatures 6, 15, 20, 21, and 26), tobacco (signature 4), somatic hypermutation (signature 9), and the three unknown signatures 8, 17, and 23 (15). A high prevalence of Signature 9 (POLN) was seen in at least one sample in 7 of 8 patients. This signature is associated with somatic hypermutation by polymerase η, a DNA polymerase that plays a role in DNA repair by translesion synthesis (16). As previously reported, we also identified subsets of patients with an abundance of signatures with unknown etiology such as signatures 8 and 17 (10). Patient PAM41 with a primary small-cell carcinoma of the pancreas had a remarkable abundance of mutations characteristic of signature 23, a rare type of signature of unknown etiology (15). Comparisons of the prevalence of each signature class specifically in post-treatment recurrent disease versus treatment-naïve primary carcinomas revealed a significant increase in the DSBR signature only (median 16 somatic mutations per primary tumor vs. 92 in recurrent disease, two-sided χ² test; P < 0.0001; Fig. 1D). Four patients received a platinum agent as part of their adjuvant or first-line therapy for recurrent disease (Supplementary Table S1); thus, we determined the extent to which this signature was enriched in these patients compared with those who received other regimens. Comparison of these two groups revealed that patients who did not receive a platinum agent had a modest yet significant increase in their recurrent disease with respect to the DSBR signature (median 58 somatic mutations per primary vs. 352 in recurrent disease, Mann–Whitney U test; P < 0.002; Fig. 1E). Moreover, as described previously (17, 18) this signature was remarkably more prevalent in the recurrent disease of patients who received one or more platinum agents (median 84.5 somatic mutations per primary vs. 1,379 in recurrent disease, Mann–Whitney U test; P < 0.0001).

We next determined the somatic alterations in known cancer genes in each patient. We identified somatic mutations in known PDA driver genes predicted to have functionally deleterious effects such as those in KRAS, CDKN2A, TP53, SMAD4, and ARID1B (Fig. 2A; Supplementary Tables S6 and S7). Collectively, the genetic features of the resected PDA samples (sample PT1 from each patient) were consistent with previous sequencing studies of these cancers (9). In contrast, the genomic features of recurrent disease (Fig. 2B; Supplementary Table S6) were notable for somatic alterations in additional genes, some of which are predicted to
Origins of Recurrent Pancreatic Cancer

RESEARCH ARTICLE

JUNE 2020  CANCER DISCOVERY | 795

Aging-PAging-RDDR-PDDR-R
APOBEC-P
APOBEC-
MSI-P
MSI-R
POLN-PPOLN-RTobacco-PTobacco-R
Sig17-PSig17-R
Sig23-PSig23-R
Sig8-PSig8-R

Mutation type
- Missense
- Frameshift deletion
- Frameshift insertion
- In-frame deletion
- In frame insertion
- Nonsense
- Nonstop
- Splice site

Location
- Primary
- Local recurrence
- Liver
- Lung
- Abdominal cavity
- Lymph node
- Diaphragm
- Pericardium
- Prostate/perirectal

Radiation
- No
- Yes

Chemotherapy
- None
- Adj/1st line platinum
- Other

Signature
- Aging
- DNA repair
- APOBEC
- MMR/MSI
- POLN
- Tobacco
- Sig 8
- Sig 17
- Sig 23
- Other

Nonsynonymous mutations

Percent of total

Primary
Recurrence
Primary
Recurrence

No platinum
Platinum

P < 0.0001
P < 0.002
P < 0.02
NS

Published OnlineFirst March 19, 2020; DOI: 10.1158/2159-8290.CD-19-1508
A

Legend
- Missense
- Nonsense
- frameshift
- Splice site
- Amplification (≥ 8 copies)
- Deletion
- Loss of heterozygosity

B

MAPK/ERK signaling
- KRAS
- NF1
- PPP6C

PI3K/AKT/MTOR signaling
- AKT1
- PK3CA
- STK11

MYC/MAX gene regulation
- CHD8
- MGA
- MYC
- NOTCH1

Transcriptional regulation by chromatin remodeling
- HIST1H3B
- KMT2B
- KMT2C
- KMT2D

DNA damage
- ATM
- TP53
- TAF12

Innate immunity
- PRKCI
- TRAF3

Nuclear export
- XPO1

C

Primaries

Recurrences

D

Double loss after WGD
CN LOH before WGD and Loss
CN LOH after WGD
CN LOH before WGD
Loss before WGD
Loss after WGD
Loss and gain
Heterozygous loss
Gain/amplification

E

CN LOH before WGD, then Loss
WGD and gain
WGD and amplification

Published Online First March 19, 2020; DOI: 10.1158/2159-8290.CD-19-1508
or likely to activate MAPK/ERK signaling (G12V mutation in KRAS, I679Dfs*21 mutation in NFI, and R111X mutation in PPP6C; refs. 19, 20), PI3K/AKT/MTOR signaling (D323H hotspot mutation in AKT1, E542K hotspot mutation in PIK3CA, and homozygous deletion of STK11; ref. 21), or MYC/MAX-regulated gene expression (Y1312X mutation in CHD8, W1004X mutation in MGA, X863_splice site mutation in NOTCH1, and up to 16-fold amplification of MYC; refs. 22, 23). Somatic alterations in genes predicted to affect chromatin-mediated gene expression (up to 38-fold focal amplification of HIST1H3B, framesshift mutations in KMT2B, KMT2C, and KMT2D, and c.4471-1N>A splicing mutations in TRIP12; refs. 24–26), nuclear export (ES71K hotspot mutation in XPO1; ref. 27), and DNA damage repair (c.497-1N>A splicing mutation in ATM and F134V mutation in TP53) were also found (28, 29). Recurrences from 2 patients contained copy-number alterations of genes implicated in innate immunity signaling (PRKCI 38-fold focal amplification and TRAF3 homozygous deletion; refs. 30, 31). Whole-genome duplication (WGD) was present in one or more samples of recurrent disease in 8 patients (ref. 32; Supplementary Table S7). In only 2 of these 8 patients was WGD detected in the primary tumor, indicating gains in ploidy may accompany PDA progression. Finally, we determined the extent to which alternative mechanisms of increasing mutant KRAS signaling exist in recurrent disease. Allelic imbalance for mutant KRAS was identified in at least one sample of the recurrent disease in all 10 patients including both patients for which KRAS mutations were identified in the recurrent disease (Fig. 2C; Supplementary Table S8). In 5 patients two or more routes to allelic imbalance occurred independently during clonal progression, indicating convergent evolution toward increased KRAS signaling (Fig. 2D and E). Collectively these data identify potential signaling and transcriptional pathways by which recurrent disease develops.

We hypothesized that somatic alterations identified in recurrent PDA samples reflect the clonal expansion of pre-existing populations following selective pressure imposed by adjuvant therapy. Pairwise cancer cell fraction (CCF) plots generated for all patients confirmed the enrichment of one or more subclonal populations in the recurrent disease compared with the primary tumor (Fig. 3A). We specifically focused on the AKT1 D323H in PAM40, PIK3CA E542K in PAM39, and NOTCH1 X863_splice site mutation in PAM41, as they represent functionally relevant alterations and in theory may be clinically actionable (http://oncokb.org). The KRAS p.G12V mutation in PAM40 was of interest given recent reports of subclonal KRAS mutations in PDA (9). Droplet digital PCR confirmed that in all instances these mutations existed in the primary tumor at prevalence rates from 0.2% to 2%, below the level of detection of our WES (Fig. 3B–D).

To understand the evolutionary origins of recurrent disease in each patient, we inferred phylogenies based on high confidence mutations in each patient (Methods). One patient was clinically thought to have an intrapancreatic recurrence of the resected PDA after 18 months, but molecular analysis revealed a second (independent) primary PDA (Fig. 4A). For example, the resected primary tumor (PAM44PT1) had a KRAS p.G12R mutation and 80 private passenger mutations, whereas the “recurrence” samples from this patient (PAM44PT2-PT5) shared 124 mutations not seen in the primary. These latter mutations included a KRAS p.G12D, a TP53 p.F134V, and a 29 base-pair framesshift deletion in KMT2D (Fig. 4B). A CCF plot of the original primary tumor (PAM44PT1) compared with the “recurrence” samples confirmed the mutual exclusivity of the somatic mutations in each lineage (Fig. 4C). Re-review of the histology of the first primary tumor and associated imaging studies did not suggest the presence of a cystic neoplasm. These findings are highly indicative of two independent primary tumors that arose from distinct precursors (33).

In the remaining 9 patients, the sample of resected primary tumor and all samples of recurrent disease arose from a common ancestor of neoplastic cells containing canonical PDA driver mutations. However, there were two distinct evolutionary trajectories by which recurrent disease arose. For 5 patients (PAM37, PAM38, PAM40, PAM42, and PAM46) the resected primary tumor sample formed the outgroup in the phylogeny (Fig. 5A and B; Supplementary Fig. S2A–S2H), indicating that in these patients the recurrent disease developed from a single residual clonal population, that is, a monophyletic origin. In the remaining 4 patients (PAM39, PAM41, PAM43, and PAM45; Fig. 5C and D; Supplementary Fig. S3A–S3F), the recurrent disease was inferred to be seeded by multiple ancestral clones and was polyphyletic in origin. For a more objective metric of relatedness among the primary and recurrent disease in each patient, we calculated pairwise Jaccard similarity coefficients for all samples within a patient. The average Jaccard indices per patient supported the distinction of monophyletic versus polyphyletic origins of recurrent disease (Fig. 5E–G) in that monophyletic recurrences were significantly more distant (divergent) from the primary tumor, whereas polyphyletic recurrences were highly related to the primary tumors. Recognizing the sample numbers are low for robust outcomes analysis, exploratory analysis indicates that patients with monophyletic recurrences had a longer disease-free but not overall survival (Fig. 5H and I). From phylogenetic analysis alone, the timing of dissemination to other organs cannot be readily resolved. However, utilizing mathematical modeling and previously measured metastasis doubling times (34), we found that the minimal time required to grow
from one to a billion cells (roughly 1 cm³) is 1.82 years, or 21.9 months [90% confidence interval (CI), 1.61–2.05 years; Methods]. Because clinical metastases occurred much earlier than the required 1.82 years after surgery in all patients with distant disease (median metastasis-free survival 11.0 months, range 6–18 months; Supplementary Table S1), at least one of the metastases must have been microscopically seeded before surgery. Patient PAM46 typified these dynamics, as they had a grossly positive surgical margin and developed a radiographically evident locoregional recurrence, but no metastases, 17 months after surgery (Supplementary Table S1). Irrespective of origin, in all 9 patients additional subclonal expansions occurred after dissemination, in some cases resulting in spatial heterogeneity for driver gene mutations among the different sites of recurrent disease (Fig. 5; Supplementary Figs. S2 and S3).

We next sought to understand the clonal origins of local recurrences, a major clinical issue for patients who undergo potentially curative resection (35), by inferring the migration patterns of recurrent disease across spatially distinct sites in each patient (Methods).

These analyses indicated that the seeding patterns of recurrent disease were diverse, with multiple patterns coexisting in the same patient. Metastasis-to-metastasis seeding was evident in some patients, typified by PAM37 in whom an omental metastasis seeded three liver metastases (Fig. 6A and B), in PAM38 in whom a liver metastasis seeded a lung metastasis (Fig. 6C and D), and in PAM45 in whom a perirenal metastasis seeded two abdominal wall metastases (Fig. 6E and F). Local recurrences can also be seeded by the primary tumor or by a metastasis. In PAM41 (Figs. 1B, and 7A and B) and PAM43 (Fig. 7C and D), the primary tumor seeded

Figure 3. Somatic alterations in recurrent pancreatic cancer alterations reflect expansion of subclones preexistent in the primary tumor. **A**, Representative density cloud plots of the primary tumor (PT1) and one matched sample of recurrent disease in 3 different patients. In all 3 patients subclones were preexistent at low allele frequency in the primary tumor and one matched sample of recurrent disease for 9 different patients. Subclonal expansion of cells containing a PIK3CA E524K mutation (PAM39), a KRAS G12V and AKT1 D323H mutation (PAM40), a NOTCH1 frameshift mutation (PAM41), a KMT2C frameshift mutation (PAM35), and CDH6 nonsense and missense mutations (PAM46) in the recurrent disease are seen. The CCFs of representative clonal driver genes for all cases (i.e., KRAS, TP53, SMAD4, and/or GNAS) are shown for reference. **B–D**, Droplet digital PCR analysis of mutant allele abundance in the primary tumor and one matched sample of recurrent disease in 3 different patients. In all 3 patients subclones were preexistent at low allele frequencies in the primary tumor. Each dot represents one droplet, and color bars at top right of each plot indicate relative intensity of the VIC-labeled mutant (x-axis) and FAM-labeled wild-type allele (y-axis) fluorescent labels.
Origins of Recurrent Pancreatic Cancer

recurrences, collectively indicating that migration patterns with both monophyletic (PAM37, PAM38, and PAM42; Fig. 6C and D) and PAM42 (Fig. 7E and F) had local recurrences, whereas local recurrences seeded by a metastasis were noted in patients with a positive surgical margin (i.e., PAM38; Fig. 6C and D).

**DISCUSSION**

To date, genomic studies of PDA have primarily relied on resections or biopsies of treatment-naive disease (9, 36, 37). Herein, we demonstrate that study of post-treatment samples may have value in identifying putative therapeutic vulnerabilities in recurrent disease. Of immediate clinical relevance is the finding that recurrent disease is enriched for somatic alterations in genes associated with MAPK/ERK or PI3K–AKT signaling, some of which are potentially actionable (http://oncoka.org). Importantly, these targets are preexistent and selected for during adjuvant treatment itself. Although our sample size is not sufficient for a robust gene discovery in advanced pancreatic cancer, we identified genes not commonly associated with the PDA landscape such as the nuclear exportin XPO1 (21), the serine–threonine protein phosphatase PPP6C (28), or regulators of innate immunity (PRKCI and TRAF3; refs. 30, 31). These findings support the need for prospective and statistically robust efforts to sequence post-treatment PDA to better define the genes or pathways repeatedly targeted by somatic alteration. We caution readers that the evolutionary timing of these potentially actionable events is critical to know. For example, in patient PAM40 all recurrent disease analyzed was the result of expansion of a single subclone with a known deleterious AKTI D323H mutation; in theory all recurrent disease in this patient may have been sensitive to an AKT1 inhibitor. In contrast, in patients such as PAM41 the PIK3CA E542K mutation occurred in a minor subclone of the recurrent disease hence would not be suitable for targeting. Nonetheless, although our sample set is small, it indicates that such instances may occur in PDA; such information could arm a clinician with more information for how to manage recurrent disease.

Another important finding of this study is that it illustrates a taxonomy by which recurrent PDA may be stratified: monophyletic origin, polyphyletic origin, or unique origin (i.e., synchronous/metachronous primaries). In all patients with bona fide recurrent PDA of monophyletic or polyphyletic origin we find that systemic subclinical dissemination had already occurred at the time of surgery, consistent with prior estimates (38). Questions for future investigation thus relate to methods to identify monophyletic versus polyphyletic recurrences in real time, the extent to which subclonal diversity develops within each category, and the clinical significance of this finding in the context of ongoing or planned clinical trials. Finally, although second primary carcinomas of the pancreas have been reported, our finding of a metachronous primary PDA in 1 of 10 otherwise unselected patients with an intrapancreatic mass post-resection suggest that this phenomenon may be more common than previously appreciated (39, 40). Detailed and formal prospective studies of intrapancreatic masses in patients who have
Figure 5. Recurrent pancreatic cancer originates from two distinct evolutionary origins. A, Phylogenetic analysis of the relationships of the primary tumor to the local recurrence and liver metastases in patient PAM40. All recurrent disease is the result of clonal expansion of a single preexistent subclone notable for an AKT1 and KRAS mutation (see Fig. 2B). The primary tumor (PT1) is the outgroup in the tree. A subclonal TRIP12 mutation is also seen in a single liver metastasis. B, Color code of sample origins in PAM40. C, Phylogenetic analysis of the relationships of the primary tumor to the local recurrence and liver metastases in patient PAM39. In this patient the recurrent disease is the result of more than one clonal expansion. The preexistent PIK3CA mutation has expanded in the lineage that gave rise to samples PT2 and PT9. D, Color code of sample origins in PAM39. E and F, Jaccard indices for each pairwise comparison in PAM40 (monophyletic recurrence) versus PAM39 (polyphyletic recurrence). G, Comparison of the average Jaccard index for primary tumors and their matched recurrences in patients with monophyletic recurrences versus those with polyphyletic recurrences. Monophyletic recurrences are significantly different from their matched primary tumor indicating passage through a genetic bottleneck, whereas no difference is found between the primary tumor and matched recurrences in patients with polyphyletic recurrences (comparisons by Student two-sided t test). Monophyletic (“common origin”) recurrences are associated with an improved disease-free survival (H) but not overall survival (I) in this small cohort (comparisons by log-rank test of Kaplan–Meier survival curves).
Figure 6. Examples of intermetastatic seeding in recurrent pancreatic cancer. For each patient the previously resected primary tumor is indicated by PT1 to the left of each patient schematic. +M, positive surgical margin; +LN, positive lymph nodes. In the body maps for all 3 patients (A, C, and E), red lines reflect migration from the primary tumor and blue lines indicate migration from a site of recurrent disease. Solid lines indicate high-confidence migration patterns and dashed lines indicate low-confidence migration patterns inferred by MACHINA (B, D, and F). Each patient had at least one high-confidence migration event from one site of recurrent disease to another. In patient PAM45, two migration patterns were equally likely to have occurred (F) although both predict that the primary tumor seeded the local recurrence (LR) and the perirectal metastasis seeded the abdominal metastases.
undergone prior resection for PDA will be required to more firmly understand the frequency of second primary tumors and the risk factors associated with their development, as has been shown for invasive carcinomas arising in intraductal papillary mucinous neoplasms of the pancreas (41).

Finally, we note that the genomic features of these patients, all of whom presented with stage I/II disease and underwent surgery and therapy, are in stark contrast to those of untreated stage IV PDA (12). In untreated PDA, the genetic features of both the primary tumors and metastases are remarkably uniform, and the genetic heterogeneity seen appears due exclusively to passenger mutations, suggesting at least one clonal sweep occurred prior to metastatic dissemination. In contrast, stage I/II disease is notable for subclonal heterogeneity for driver genes as reported in other tumor types (42, 43). Treatment-induced genetic bottlenecks that sculpt the genomic landscape of PDA, together with intermetastatic seeding, likely contribute to subclonal and intermetastatic heterogeneity for driver gene alterations observed in recurrent disease. Thus, context appears key in the interpretation of heterogeneity in PDA.

In summary, we identify novel genetic features of PDA in the context of recurrent disease after surgical resection and treatment with potential clinical implications for use of targeted therapies in disease management. In the event that therapeutically targetable gene alterations are validated and treatment with potential clinical implications for use of targeted therapies in disease management, as has been shown for invasive carcinomas arising in intraductal papillary mucinous neoplasms of the pancreas (41).

Figure 7. Origins of local recurrences after surgical resection. For each patient the previously resected primary tumor is indicated by PT1 to the left of each patient schematic. −M, negative surgical margin; +LN, positive lymph nodes; −LN, negative lymph nodes. In the body maps for all 3 patients (A, C, and E), red lines reflect migration from the primary tumor and blue lines indicate migration from a site of recurrent disease. Solid lines indicate high-confidence migration patterns and dashed lines indicate low-confidence migration patterns inferred by MACHINA (B, D, and F). In PAM41 (A and B) and PAM43 (C and D) the local recurrence is seeded by the originally resected primary tumor, whereas in PAM42 (E and F) it is seeded by a lung metastasis. In patients PAM41 (A and B) and PAM42 (E and F) three or more migration patterns were equally likely to have occurred.

METHODS

Tissue Samples

Patient samples were generously shared by the Gastrointestinal Cancer Rapid Medical Donation Program resource at The Johns Hopkins Hospital (Baltimore, MD). Sections were cut from formalin-fixed, paraffin embedded (FFPE) or frozen sections available and reviewed to identify those with at least 20% neoplastic cellularity and preserved tissue quality. Samples meeting these criteria were macrodissected from serial unstained sections before extraction of genomic DNA using DNeasy Blood & Tissue Kits for frozen samples or QIAamp DNA FFPE Tissue Kits for FFPE materials (Qiagen) following the manufacturer’s instructions.

WES

DNA quantification, library preparation, and sequencing were performed in the Integrated Genomics Operation and bioinformatics analysis of somatic variants by the Bioinformatics Core at Memorial Sloan Kettering Cancer Center (New York, NY). Libraries were created using AgilentExon_51MB_hg19_v3 as bait and sequenced on an Illumina HiSeq 2500. WES resulted in a median target sequence coverage of 317× and a calculated average tumor cellularity of 35.7% (min–max) with 83% of the coding regions covered at least 100×.

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 ≥1, patient-matched normal variant frequency <0.02, tumor coverage ≥20 reads, and tumor variant allele frequency ≥0.05 in at least one tumor sample. The resulting list of somatic variants was filtered for those present in the coding regions only and subject to further bioinformatic annotation for pathogenicity and germline allele frequencies from healthy populations distributed worldwide using LiFD (ref. 45; Supplementary Table S5). MSI status was inferred from the sequencing data using a clinically validated algorithm (46), with MSI-H defined as an MSIsensor score ≥10. For all mutations, we inferred the CCF from the mutant allele fraction, local copy number, and FACETS estimate of tumor purity according to previously described methods (47, 48). CIs for the CCF therefore reflect the 95% binomial CI for the underlying mutant allele fraction. Whole-gene copy-number alterations and WGD were inferred by FACETs. Genes with copy number of 8-fold or greater inferred by FACETs were considered amplifications.

Phylogenetic Analysis

We applied Treeomics v1.6.0 to reconstruct the phylogenies of recurrent disease using high-quality somatic variants identified by WES. Treeomics uses a Bayesian inference model to account for sequencing errors and low purity and employs Integer Linear Programming to infer a maximum likelihood tree.

Mathematical Modeling

To calculate the minimal required time a metastasis founding cell needs to grow to a detectable lesion of 1 cm³ (∼10⁸ cells), we used the smallest previously measured PDA metastasis doubling time of 27 days (49) leading to an exponential growth rate of r = 0.026 per day. Assuming a PDA cell division time of 2.3 days (50), the expected minimal time a tumor conditioned on survival takes to reach 10⁸ cells is 1.82 years (90% CI, 1.61–2.05 years).

Mutational Signatures

Mutational signatures were derived using the methods described by Alexandrov and colleagues (15). To enrich for the most abundant signatures we merged those with similar putative etiologies into a single group. Signature groups present in at least 20% abundance in at least one sample were included for additional study and statistical analysis.
Droplet Digital PCR

Absolute quantification of mutant alleles was determined using a RainDrop Plus digital PCR system according to the manufacturer’s instructions. Predesigned or custom designed TaqMan assays were obtained for variants of interest (Thermo Fisher Scientific). Approximately 75 ng of gDNA was used per reaction in a 25 μl volume. Each reaction contained 5 × 10⁶ droplets at a target inclusion rate of 10% (5 × 10⁵ target molecules).

Migration Pattern Inferences

PyClone and MACHINA were used to infer seeding patterns associated with metastasis and local recurrence. To alleviate long run times associated with the high sample number context, we applied specific filters to focus on the most informative loci. These were: (i) the locus was sequenced to a depth of at least 60 in at least one sample; (ii) the locus had a copy-number profile consistent with a relatively simple genomic history in all samples (the combination of major allele A and minor allele B at the locus was required to be one of AB, AAB, AAAB, or AABBB in each sample, although not necessarily the same across samples), and (iii) each locus was required to be sequenced to nonzero depth in all remaining samples. Samples that did not contain more than 20 variant loci after applying this filter were excluded. In patient PAM41, a large cluster of highly related liver metastases (PT4, 8, 9, 10, and 11) was simplified by including only sample PT9 to improve runtime. The resulting mutations and associated major and minor copy numbers were clustered with PyClone using default settings. The PyClone consensus cluster files were used to enumerate evolutionary relationships, then the combination of PyClone cluster frequency estimates and enumerated trees together were used to search for the most parsimonious migration patterns consistent with each tree topology. The resulting solution with the lowest overall migration number, and then the lowest comigration number, were determined. No other constraints were applied to the migration plots.

Statistical Analysis

Descriptive data are represented as a mean and SD unless otherwise mentioned. Parametric distributions were compared by a Student t test, whereas nonparametric distributions were compared by a Mann–Whitney U test. Frequency data were compared using a χ² test. All comparisons were two-sided. Survival curves were plotted according to the methods of Kaplan and Meier and compared using a log-rank test.

Disclosure of Potential Conflicts of Interest

E.M. O’Reilly is a DSMB at CytoMx Therapeutics and Rafael, is a consultant at Polaris, Ipsen, Sobi, and Merck, has immediate family members who consult for Celgene-BMS, and reports receiving commercial research grants from Celgene-BMS, MabVax Therapeutics, ActaBiologica, AstraZeneca, and Genentech Roche. R.H. Hruban has ownership interest (including patents) in Thrive Earlier Detection and Precision Lifesciences Group. N.D. Socci has a consultant/advisory board relationship with Solvuu. B.S. Taylor reports receiving a commercial research grant from Genentech, Inc., has received speakers bureau honoraria from Genentech, Inc., and has a consultant/advisory board relationship with Boehringer Ingelheim and Loxo Oncology, a wholly owned subsidiary of Eli Lilly, Inc. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: H. Sakamoto, M.A. Attiyeh, A.P. Makohon-Moore, C.A. Iacobuzio-Donahue

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.A. Attiyeh, A. Hayashi, R. Kappagantula, L. Zhang, E.M. O’Reilly, L.D. Wood, C.A. Iacobuzio-Donahue


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Sakamoto, M.A. Attiyeh, A.P. Makohon-Moore, R. Kappagantula, L. Zhang, J.P. Melchor, C.A. Iacobuzio-Donahue

Study supervision: C.A. Iacobuzio-Donahue

Acknowledgments

This work was supported by NIH/NCI grants R01 CA179991 and R35 CA220508 to C.A. Iacobuzio-Donahue, 2T32 CA160001-06 to A.P. Makohon-Moore, the Daisichi-Sankyo Foundation of Life Science Fellowship to A. Hayashi, the Mochida Memorial Foundation for Medical and Pharmaceutical Research Fellowship to A. Hayashi, R00 CA22991102 to J.G. Reiter, and NIH/NCI P50 CA62924. This work was funded in part by the Marie-Josée and Henry R. Kravis Center for Molecular Oncology and the National Cancer Institute Cancer Center Core Grant No. P30-CA008748.

Received December 19, 2019; revised March 6, 2020; accepted March 16, 2020; published first March 19, 2020.

REFERENCES


