

RESEARCH ARTICLE

Whole Exome Sequencing of Rapid Autopsy Tumors and Xenograft Models Reveals Possible Driver Mutations Underlying Tumor Progression

Tao Xie^{1*}, Monica Musteanu², Pedro P. Lopez-Casas², David J. Shields¹, Peter Olson¹, Paul A. Rejto¹, Manuel Hidalgo^{2*}

1 Oncology Research, Pfizer Worldwide Research and Development, San Diego, California, United States of America, **2** CNIO (Spanish National Cancer Research Centre), E-28029, Madrid, Madrid, Spain

* tao.xie@pfizer.com (TX); mhidalgo@cniio.es (MH)



OPEN ACCESS

Citation: Xie T, Musteanu M, Lopez-Casas PP, Shields DJ, Olson P, Rejto PA, et al. (2015) Whole Exome Sequencing of Rapid Autopsy Tumors and Xenograft Models Reveals Possible Driver Mutations Underlying Tumor Progression. PLoS ONE 10(11): e0142631. doi:10.1371/journal.pone.0142631

Editor: Ilse Rooman, Garvan Institute of Medical Research, AUSTRALIA

Received: July 24, 2015

Accepted: October 23, 2015

Published: November 10, 2015

Copyright: © 2015 Xie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The WES data from this study has been made fully available through the European Nucleotide Archive (ENA), with accession number: PRJEB9296 (<http://www.ebi.ac.uk/ena/data/view/PRJEB9296>).

Funding: This research was funded by Pfizer Inc. The funder provided support in the form of salaries for authors TX, DJS, PO, and PAR and/or research materials, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Pancreatic Ductal Adenocarcinoma (PDAC) is a highly lethal malignancy due to its propensity to invade and rapidly metastasize and remains very difficult to manage clinically. One major hindrance towards a better understanding of PDAC is the lack of molecular data sets and models representative of end stage disease. Moreover, it remains unclear how molecularly similar patient-derived xenograft (PDX) models are to the primary tumor from which they were derived. To identify potential molecular drivers in metastatic pancreatic cancer progression, we obtained matched primary tumor, metastases and normal (peripheral blood) samples under a rapid autopsy program and performed whole exome sequencing (WES) on tumor as well as normal samples. PDX models were also generated, sequenced and compared to tumors. Across the matched data sets generated for three patients, there were on average approximately 160 single-nucleotide mutations in each sample. The majority of mutations in each patient were shared among the primary and metastatic samples and, importantly, were largely retained in the xenograft models. Based on the mutation prevalence in the primary and metastatic sites, we proposed possible clonal evolution patterns marked by functional mutations affecting cancer genes such as KRAS, TP53 and SMAD4 that may play an important role in tumor initiation, progression and metastasis. These results add to our understanding of pancreatic tumor biology, and demonstrate that PDX models derived from advanced or end-stage likely closely approximate the genetics of the disease in the clinic and thus represent a biologically and clinically relevant pre-clinical platform that may enable the development of effective targeted therapies for PDAC.

Introduction

Pancreatic Ductal Adenocarcinoma (PDAC) is a highly lethal malignancy that represents a major therapeutic challenge[1]. In the United States, PDAC ranks fourth in mortality with a

Competing Interests: The authors have the following interests: The study was funded by Pfizer Inc. Tao Xie, David J. Shields, Peter Olson and Paul A. Rejto are employees of Pfizer. There are no patents, products in development or marketed products to declare. This did not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

median survival of only six to nine months. The overall five-year survival rate is less than 5%, and even for patients who have surgical resection, the rate is still less than ~25% [2]. The poor prognosis for patients with PDAC partially stems from its propensity to rapidly metastasize [3]. Despite decades of basic and translational research, very few new therapies have been approved. To help develop novel targeted therapeutic strategies to treat PDAC patients, it is critical to understand the cellular and molecular mechanisms that facilitate metastasis formation, and to generate models that more faithfully represent the end-stage disease. Previous studies [4–7] found that PDAC tumors have numerous somatic alterations that affect key oncogenic drivers and tumor suppressor genes including KRAS, TP53, CDKN2A and SMAD4, although it is still not clear which mutations lead to the acquisition of metastatic ability.

Our understanding of the mutational landscape in pancreatic cancer has evolved with larger data sets, newer sequencing technologies and more advanced bioinformatics analyses [8]. In addition to the aforementioned mutated pancreatic cancer genes, lower incidence mutations and structural rearrangements have emerged more recently including mutations in genes involved in DNA repair, axon guidance, receptor tyrosine kinases and epigenetic modifiers [6, 7, 9]. While gemcitabine plus Abraxane and FOLFIRINOX have recently been approved for pancreatic cancer, response rates are still at or below 50% which may be due to additional genetic loci that confer sensitivity or resistance to therapy [10, 11]. Mutational heterogeneity has also been observed in pancreatic cancer. Comparing mutational profiles between primary tumors and paired metastases has been shown to afford an opportunity to identify mutations that may underlie the metastatic process [12–15]. Mutational heterogeneity may also at least partially explain the short-lived responses to virtually all drugs as patients may harbor pre-existing, therapy-resistant clones.

PDX (patient-derived xenograft) models have emerged as preclinical models that better reflect human cancer since the heterogeneity of the tumor more closely reflects patient tumors than traditional cell line xenografts [16]. A key gap in the field has been to generate and characterize models across the spectrum of disease stages [17–19]. Since most clinical trials are tested in late-stage disease, PDX models derived from advanced or end-stage likely represent a closer approximation to the clinic compared to resection-derived models which represent the majority of models generated in the field to date [20, 21]

In this study, we sought to comprehensively characterize somatic SNVs (single-nucleotide variations) in end stage pancreatic cancer and performed whole exome sequencing (WES) on paired primary tumors and metastases. Through this analysis, we identified potential molecular cancer drivers and related dysregulated pathways that might underlie the progression of PDAC. PDX models derived from the same primary and metastatic tumor samples were also sequenced to assess differences in the mutation profiles between primary tumor samples and xenograft models.

Results

Reads data generated by whole exome sequencing

Three PDAC patients (patient ID: P042, P047 and P059) were characterized in this study by DNA sequencing of tumors obtained at autopsy, as well as derived xenograft models. We conducted whole-exome sequencing (WES) for 13 samples, including ten tumor samples plus three normal DNA samples from each patient's peripheral blood (Table 1) and generated about 40–50 millions of sequencing read pairs (100bp, paired end) for each sample. After raw reads QC and removal of mouse reads from the PDX samples, we got on average 37.8 million read pairs per sample and mapped the majority (close to 98%) to the human reference genome

Table 1. Summary statistics of WES data.

| Sample Type | Short Name | %mouse | Read Pairs (in mil.) * | Mean depth |
|-----------------------|--------------|--------|------------------------|------------|
| P042 | | | | |
| Normal | P042_NOR | - | 43.7 | 141.0 |
| Primary | P042_PRI | - | 57.1 | 184.4 |
| Primary PDX | P042_PRI_PDX | 6.0% | 37.3 | 120.4 |
| Liver metastasis | P042_LIV | - | 30.0 | 96.9 |
| Liver metastasis PDX | P042_LIV_PDX | 17.9% | 35.0 | 112.9 |
| P047 | | | | |
| Normal | P047_NOR | - | 34.6 | 111.6 |
| Primary | P047_PRI | - | 38.0 | 122.8 |
| Liver metastasis | P047_LIV | - | 34.9 | 112.5 |
| Liver metastasis PDX | P047_LIV_PDX | 10.5% | 35.7 | 115.2 |
| Peritoneum metastasis | P047_PER | - | 44.7 | 144.4 |
| P059 | | | | |
| Normal | P059_NOR | - | 32.6 | 105.4 |
| Primary PDX | P059_PRI_PDX | 7.1% | 34.9 | 112.8 |
| Liver metastasis PDX | P059_LIV_PDX | 2.4% | 32.7 | 105.6 |

*For PDX samples, predicted mouse reads were excluded.

doi:10.1371/journal.pone.0142631.t001

(human NCBI build 37), which gives an average sequencing mean depth of greater than 120x on covered exonic regions.

Overview of somatic SNVs

By comparing to matched normal DNA, we initially identified an average of 160 somatic SNVs (single nucleotide variants or point mutations) per cancer sample, ranging from as few as 80 (P047_PER) to as many as 312 (P059_PRI_PDX). There were significantly more C>T variations (transitions) in every sample included in this study (Fig 1), similar to what has been reported elsewhere (for example: [22]). The paired primary and metastatic tumors as well as matched tumor and xenograft samples had similar mutation class distributions in most cases: e.g. P047_PRI has a very similar mutation pattern as the two metastatic samples (P047_PER, peritoneum and P047_LIV, liver) from the same patient; for patient P059, though no tumor samples were available for WES analysis, the mutation patterns of the two xenografts (P059_PRI_PDX and P059_LIV_PDX) are more similar to each other than to any sample from different patients. These results suggest that the underlying mutational processes remain the same in primary and metastatic lesions, as well as in their xenograft models. PDX samples tend to have a slightly lower proportion of C>T transitions than their originating tumors. The only exception is P042_LIV which shows substantially more T->G variation (and thus proportionally much less C>T transitions) than any other sample in this study.

We classified mutations into several different categories based on their location (Table 2). In general, the majority of somatic mutations are in exonic regions, including UTRs and CDS regions where a subset can alter encoded protein sequences (including missense and non-sense mutations, predicted by the Sift tool [23]), and are likely to have a functional impact. We found on average 55 somatic protein-altering SNVs per sample, ranging from as few as 35 (P047_PER/P047_PRI) to as many as 99 (P059_PRI_PDX). Detailed information can be found in S1 Table.

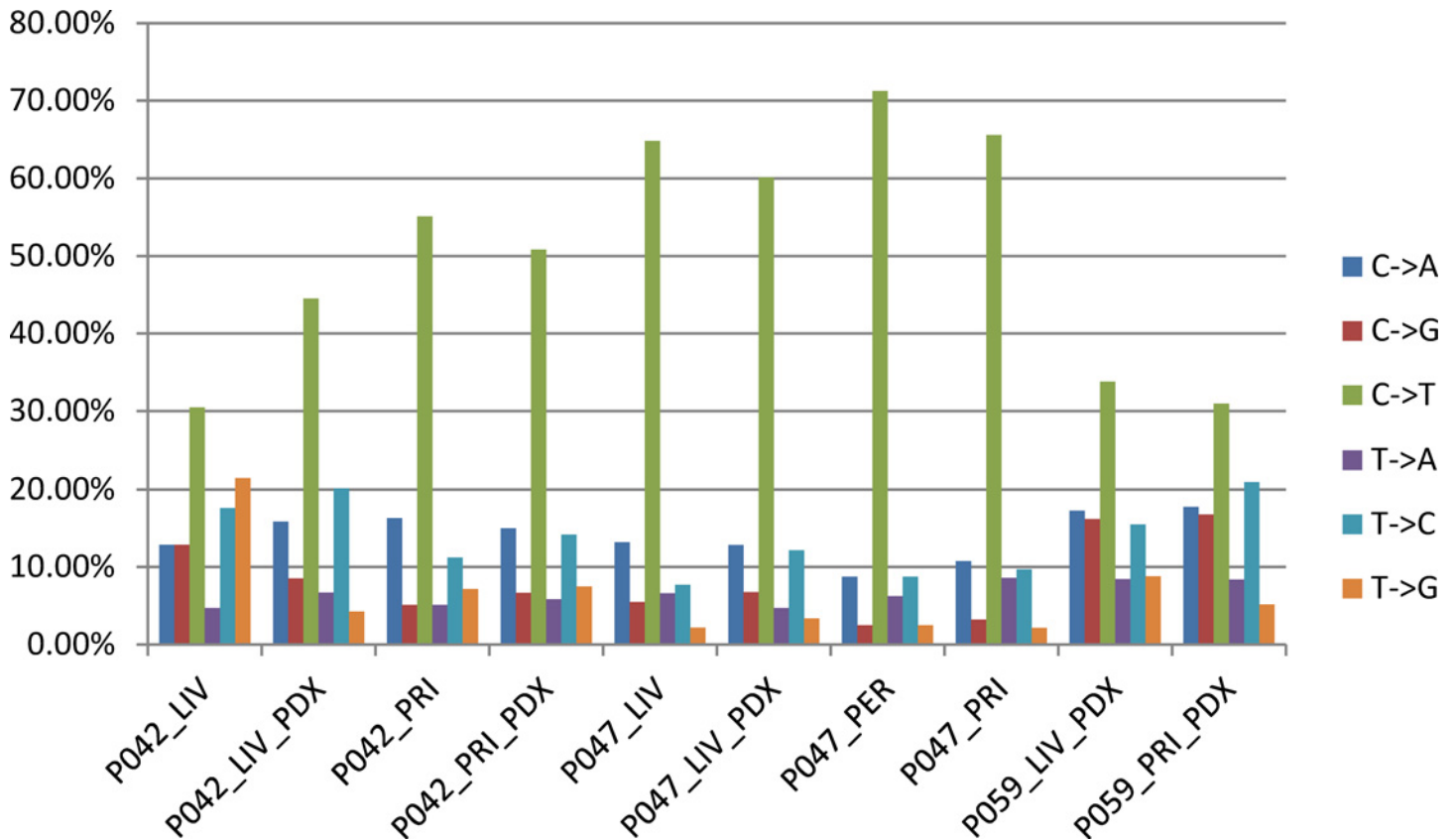


Fig 1. Somatic mutation patterns of primary, metastatic and PDX tumors. Proportion of somatic SNVs by class (C->A, C->G, C->T, T->A, T->C and T->G) in the primary, metastatic and PDX tumor samples are shown for the ten cancer samples included in this study.

doi:10.1371/journal.pone.0142631.g001

Common and distinct somatic SNVs between primary, metastatic tumors and derived xenografts

Characterization of common and distinct genetic changes between primary and metastatic tumors is critical for understanding tumor biology and developing targeted therapies. In general, the majority of mutations from the primary tumors are retained in their paired metastases, with further enrichment of common somatic SNVs within the protein-altering mutation subset (Table 3). For example, 78.6% of somatic SNVs found in P042_PRI were also reported in P042_LIV (77 out of in total 98 somatic SNVs found in P042_PRI) while for protein-altering mutations, the number increases to 88.9% (33 out of in total 36 protein-altering mutation found in P042_PRI), suggesting positive selection pressure on functional alterations. When comparing xenografts to their primary tumors, PDX models (P042_PRI_PDX, P042_LIV_PDX and P047_LIV_PDX) retained the majority of the mutations and also proportionally more of the functional mutations (on average 66.7% and 86.1%, respectively). Both paired primary/metastatic tumors as well as matched patient and xenograft samples have high concordance for somatic SNVs and an even higher concordance for functional alterations.

We extended the comparison to samples from different patients and measured sample similarity using the fraction of common functional SNVs from any two samples (Table 3 and S1 Table). While no common mutations were observed in samples from all three patients, P042 and P047 share two common mutations (Table 3). Both mutations are known to be functional: TP53 (R282W) and KRAS (G12V). These findings are not surprising since tumor suppressors

Table 2. Summary of locations of somatic SNVs.

| Patient | P042 | | | | P047 | | | | P059 | | |
|---------------|------------|------------|-----------|------------|-----------|------------|-----------|-----------|------------|------------|--|
| | LIV | LIV_PDX | PRI | PRI_PDX | LIV | LIV_PDX | PER | PRI | LIV_PDX | PRI_PDX | |
| 3' UTR | 3 | 2 | 5 | 2 | 1 | 0 | 0 | 0 | 7 | 7 | |
| 5' UTR | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | |
| CDS | 68 | 74 | 44 | 53 | 48 | 62 | 39 | 43 | 102 | 131 | |
| Downstream | 12 | 3 | 2 | 3 | 3 | 4 | 3 | 2 | 9 | 8 | |
| Intergenic | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 6 | 14 | |
| Intronic | 97 | 42 | 26 | 32 | 21 | 50 | 20 | 28 | 90 | 98 | |
| Splice_site | 6 | 6 | 6 | 4 | 3 | 6 | 5 | 4 | 9 | 8 | |
| Upstream | 11 | 6 | 4 | 6 | 3 | 6 | 1 | 3 | 17 | 12 | |
| Other types | 33 | 30 | 10 | 19 | 11 | 18 | 11 | 10 | 44 | 31 | |
| Total: | 233 | 164 | 98 | 120 | 91 | 148 | 80 | 93 | 287 | 312 | |

Note: UTR, untranslated region; CDS, coding sequence.

doi:10.1371/journal.pone.0142631.t002

such as TP53, SMAD4 and CDKN2A (P16) and oncogenes such as KRAS are frequently mutated in PDAC [4–6]. Interestingly, several mutations were identified in common across distinct PDX models, but were not detectable in the originating patient material. For example (Table 3), P047_LIV_PDX has 15 more common mutations with P042_LIV_PDX (17 in total) than with P042_LIV (2 in total) and 10 more common mutations with P042_PRI_PDX (12 in total) than with P042_PRI (2 in total). Though a few of these xenograft exclusive calls might be real, most are likely to be false positives since some mouse reads can map to regions such as ultraconserved elements in the human genome [24, 25] that are almost 100% identical to their mouse orthologous regions and thus could not be fully detected and removed completely before mutation calling.

Besides the number of somatic SNVs, we also compared the distribution of mutant allele frequency (MAF) across all the cancer samples (Fig 2). The wide range of MAFs revealed considerable genetic heterogeneity in the samples. While xenografts normally have a median MAF close to 0.5, tumor samples present much lower median MAFs (0.15–0.3), most likely

Table 3. Pairwise comparison matrix for common somatic SNVs across all samples.

| Patient | P042 | | | | P047 | | | | P059 | | |
|--------------|---------------|---------------|--------------|---------------|--------------|---------------|--------------|--------------|---------------|---------------|--|
| | LIV | LIV_PDX | PRI | PRI_PDX | LIV | LIV_PDX | PER | PRI | LIV_PDX | PRI_PDX | |
| P042_LIV | 233/64 | 95/42 | 77/33 | 81/34 | 2/2 | 2/2 | 2/2 | 2/2 | 1/0 | 3/0 | |
| P042_LIV_PDX | | 164/58 | 74/31 | 95/36 | 2/2 | 17/4 | 2/2 | 2/2 | 3/0 | 14/1 | |
| P042_PRI | | | 98/36 | 71/31 | 2/2 | 2/2 | 2/2 | 2/2 | 0/0 | 0/0 | |
| P042_PRI_PDX | | | | 120/42 | 2/2 | 12/2 | 2/2 | 2/2 | 1/0 | 6/0 | |
| P047_LIV | | | | | 91/44 | 68/31 | 60/29 | 65/31 | 1/0 | 0/0 | |
| P047_LIV_PDX | | | | | | 148/52 | 67/32 | 74/31 | 0/0 | 7/1 | |
| P047_PER | | | | | | | 80/35 | 66/29 | 0/0 | 0/0 | |
| P047_PRI | | | | | | | | 93/35 | 0/0 | 0/0 | |
| P059_LIV_PDX | | | | | | | | | 287/87 | 158/58 | |
| P059_PRI_PDX | | | | | | | | | | 312/99 | |

In the table, the first number is the total number of common somatic SNVs while the second is the number of common protein-altering ones. The original numbers of somatic mutations and protein-altering ones for each sample are boxed (diagonal).

doi:10.1371/journal.pone.0142631.t003

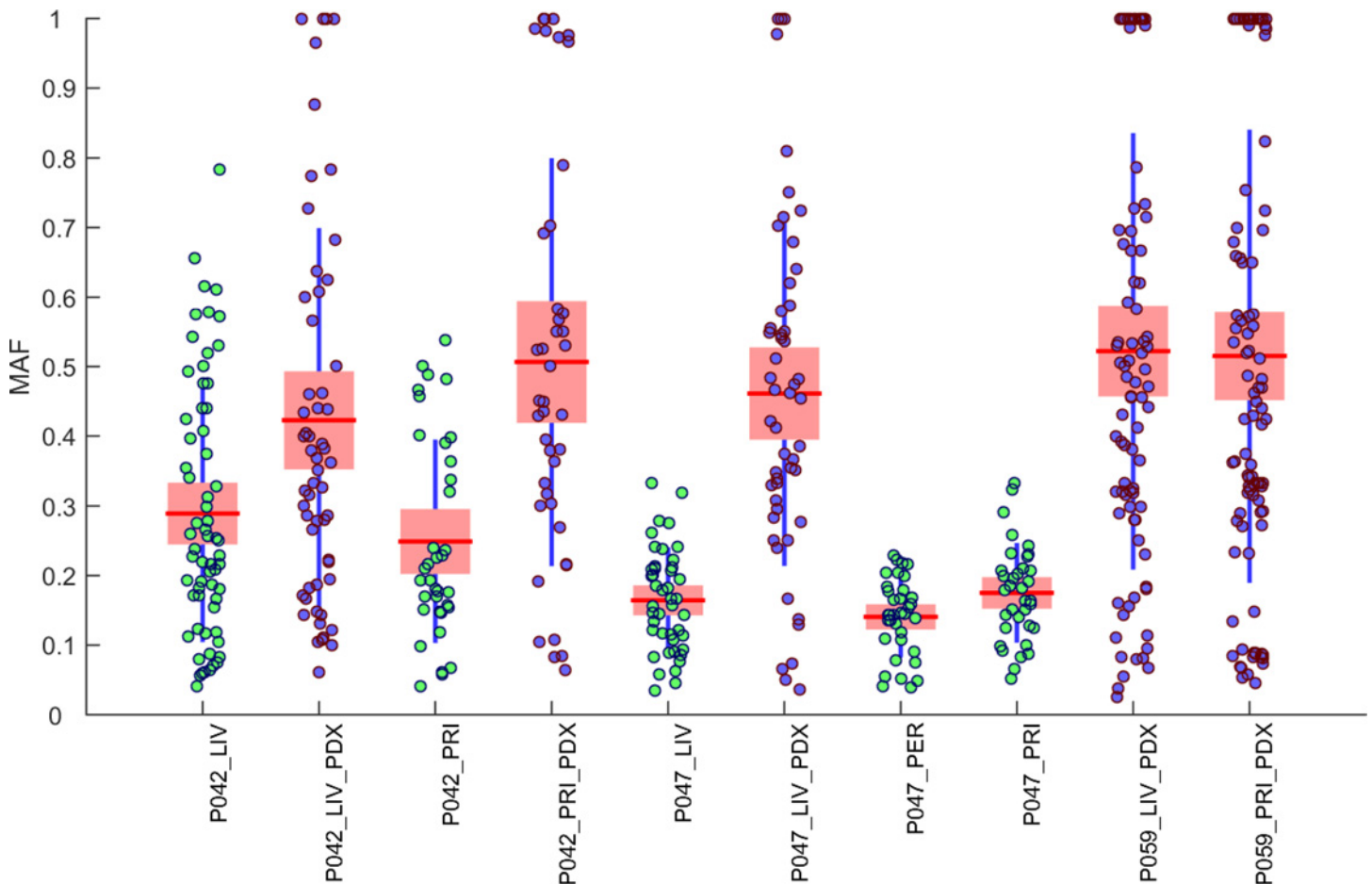


Fig 2. Distribution of MAFs of protein-altering mutations identified in cancer samples. Each boxplot shows the MAFs (from 0 to 1) of somatic SNVs found in an individual sample along with the mean (the red horizontal line), SEM (standard error of the mean, shown as a blue box) and SD (standard deviation, shown as a blue vertical line). PDX samples (MAFs in blue dots) in general show higher mean MAF values (close to 0.5) than those from tumor samples (MAFs in green dots), which indicates that the PDX samples were of much higher tumor purity since reads from mouse stromal cells were effectively removed from the original sequencing data.

doi:10.1371/journal.pone.0142631.g002

due to different level of stromal contamination. For xenografts, a higher median MAF means most sequencing reads from mouse stroma were effectively excluded prior to alignment, so they appear much purer (without reads from mouse stromal cells) than tumor samples (with reads from human stromal cells). This can also explain why in general there are more detectable mutations in PDX samples than in tumor samples (Table 3).

To further pinpoint key mutations that might underlie tumor initiation, progression and metastasis, we focused on the predicted functional mutations affecting known cancer drivers and tumor suppressors. Table 4 lists point mutations affecting TP53, KRAS and SMAD4 while the complete list of protein-altering mutations can be found in S1 Table. As previously described, the R282W mutation in TP53 and the G12V mutation in KRAS are the only two common mutations found in samples from both P042 and P047, while two other TP53 mutations were found in samples from P059.

Besides well-known oncogenic driver mutations, we also found several additional mutations that might be interesting. In samples from P042, a non-synonymous mutation (M9930I) was called for MUC16 in the liver metastatic sample (P042_LIV) but not in the primary sample (P042_PRI). By further visual analysis of available sequencing reads mapping to the MUC16

Table 4. Somatic SNVs affecting key cancer genes.

| Sample | Gene | PRI | LIV | PRI_PDX | LIV_PDX | AA_change |
|--------|-------|------|------|-------------------|---------|--------------|
| P042 | TP53 | 0.50 | 0.50 | 1.00 | 1.00 | R282W |
| | KRAS | 0.40 | 0.61 | 0.79 | 0.77 | G12V |
| | SMAD4 | 0.24 | 0.42 | 0.97 | 1.00 | W524C |
| P047 | TP53 | 0.14 | 0.18 | 0.17 [§] | 1.0 | R282W |
| | KRAS | 0.23 | 0.21 | 0.14 [§] | 0.98 | G12V |
| P059 | TP53 | n.a. | n.a. | 0.48 | 1.00 | P152R |
| | TP53 | n.a. | n.a. | 0.51 | 0.00 | C182* |

§: from P047_PER

*: stop codon.

doi:10.1371/journal.pone.0142631.t004

locus in P042_PRI, we successfully identified two sequencing reads carrying the mutant allele out of a total of 32 reads covering the mutant position, which explained why this SNV was not called in P042_PRI (due to the very low incidences). When checking PDX samples from the same patient, the mutation was also called only in P042_LIV_PDX but not in P042_PRI_PDX. Similarly, by looking into raw sequencing reads from P042_PRI_PDX, we also located two sequencing reads carrying the mutant allele out of 20 reads covering the mutant position. MUC16, also known as CA125 (cancer antigen125), is a mucin, a family of high molecular weight, heavily glycosylated proteins that are known to play important roles in pancreatic cancer pathogenesis [26, 27]. MUC16 is significantly upregulated in pancreatic cancer and co-overexpressed with MSLN (Mesothelin) at the invading edge [28, 29]. It has also been implicated in cell motility and invasion via matrix metalloproteinase 7 (MMP7), and may stimulate cancer metastasis [30]. On the DNA level, MUC16 is frequently mutated in many cancer types, though more likely due to its very large gene size [31]. The mutation M9930I is located in the amino-terminal domain and its functional impact needs to be further studied to see if it plays a role in facilitating pancreatic cancer progression and metastasis. Besides MUC16, there is a deleterious mutation (P1598S, COSMIC Mutation Id[32]: COSM922860) affecting another mucin family gene, MUC2, with elevated MAFs in both metastatic samples (P042_LIV and P042_LIV_PDX) compared to the primary samples (P042_PRI and P042_PRI_PDX). MUC2 has also been implicated in cancer metastasis previously through a tumor-associated macrophage (TAM)-dependent mechanism [33, 34]. Besides P1598S, we also pinpointed two additional protein-altering mutations affecting MUC2 in two metastatic samples from P047 (T1683M in P047_LIV and T1722I in P047_PER, S1 Table). While the functional impact of these mutations in mucin proteins remains to be further validated experimentally, the present data suggest that mucin may play a role in helping establish the metastatic founder clone.

In addition to mutations shared between the primary and metastatic tumors, there were predicted functional mutations present in only one or the other tumor. We listed point mutations that appeared to arise *de novo* in the metastatic lesions in S1 Table. One interesting gene, ODZ1 is only found to be mutated in metastatic samples (P042_LIV and P042_LIV_PDX). It encodes Tenascin, an oligomeric glycoprotein of the extracellular matrix that is a putative tumor suppressor involved in morphogenetic movement, tissue patterning, repair, and tumor invasion, and was reported frequently deleted in malignant gastroenteropancreatic endocrine carcinomas [35, 36]. The mutation N570K, predicted as “Deleterious” by the Sift tool [23], is located in a EGF-like domain and thus might play a role in the development of metastasis by disrupting the gene’s tumor suppressor function. It is possible that, after the physical translocation of cancer cells from primary tumor to distant organ (liver in this case), a subsequent

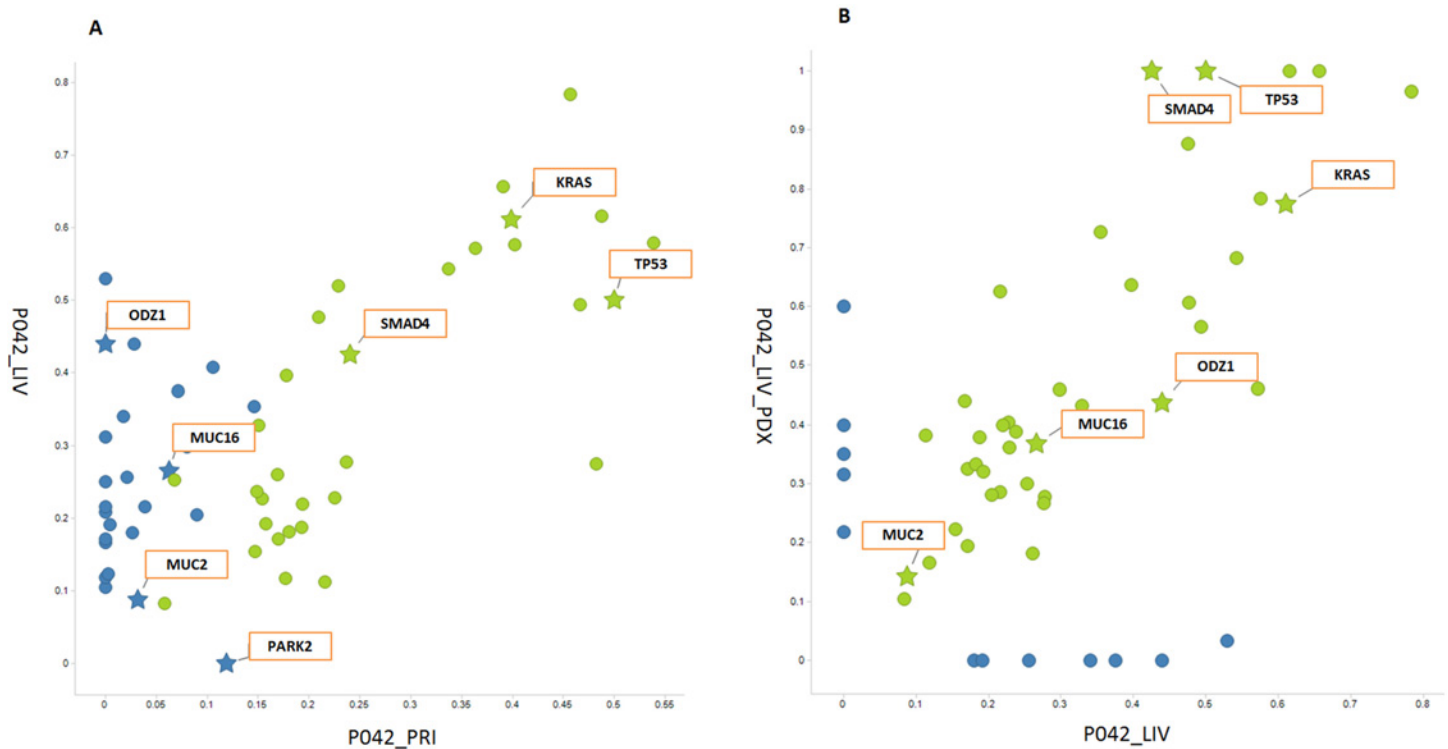


Fig 3. Scatter plots of non-synonymous and non-sense mutations for (A) P042_PRI vs. P042_LIV and (B) P042_LIV vs. P042_LIV_PDX. The plots are based on the mutant allele frequencies (MAFs, from 0 to 1) of non-synonymous and non-sense mutations from the paired samples. The mutations that are called in both samples are marked by green dots, while ones called in only one sample are colored blue. Gene names are shown for those key mutations discussed in the text.

doi:10.1371/journal.pone.0142631.g003

hit on ODZ1 (or other rate limiting events[37]) might have been selected in a subpopulation of tumor cells leading to the formation of a dominant clone at the metastatic site.

Comparison of MAF distributions between the primary tumors and their paired liver metastases can help reveal clonal or subclonal mutation patterns underlying PDAC initiation, progression and metastasis. Focusing on protein-altering somatic SNVs, we compared a liver metastasis (P042_LIV) to its xenograft model (P042_LIV_PDX) (Fig 3A). The clonal mutations shared between the primary and metastatic tumors (green dots) have a correlated pattern of MAFs, while some subclonal mutations in the primary site have significantly elevated MAFs in the metastatic site, possibly composing the dominant clone in the metastatic tumor. For example, the MUC16:M9903I mutation mentioned previously shows a significant MAF increase from the primary site to the metastatic site in both tumor samples (from 6.3% in P042_PRI to 27% in P042_LIV) and in xenografts (from 10% in P042_PRI_PDX to 37% in P042_LIV_PDX). On the other hand metastasis-specific mutations (the blue dots with MAF (P042_PRI) = 0 in Fig 3A) reflect ongoing clonal dynamics in the liver site after metastasis. For primary-specific SNVs, we found only one protein-altering mutation in the gene PARK2 (Parkin RBR E3 Ubiquitin Protein Ligase). It possibly marks a subclone(s) in the primary site that did not migrate to the metastatic site.

Based on the clonal/subclonal mutation information, a parsimonious evolutionary history for P042 can be proposed: first, clonal driver mutations (TP53 and KRAS mutations) appeared early in the primary site (i.e. tumor initiation) followed by subsequent genetic alterations (concurrently or separately) that formed a number of subclones (tumor progression); then a single

cell from a minor subclone that acquired an invasive phenotype (marked by mutations such as MUC16:M9930I), migrated to a distant organ (liver in this case) and successfully established a metastasis. We also compared the MAF ([S1 Table](#)) of the primary tumor samples to their paired metastases from the other two patients (P047 and P059) and found similar patterns of clonal and subclonal mutations. Particularly, in P059, there appeared to be two subclones in the primary site (P059_PRI_PDX) carrying either a missense (P152R, MAF = 48%) or a non-sense (C182*, MAF = 51%) mutation affecting TP53 ([Table 4](#)). We examined the sequencing reads spanning the two sites and found that the two somatic SNVs are mutually exclusive (data not shown). More interestingly, in the metastatic site (P059_LIV_PDX), it appeared that only the subclone carrying the missense mutation (P152R, MAF = 100%) rather than the non-sense mutation (C182*, MAF = 0%) successfully established the liver metastasis. The apparent hierarchical structure of clones in the primary and metastatic tumors is similar to what had been reported elsewhere [[5](#), [13](#), [38–40](#)].

We also compared MAF distributions between the tumor samples and their derived xenograft models. [Fig 3B](#) shows the scatter plot of MAF values for protein-altering mutations in P042_LIV and P042_LIV_PDX. We found that the majority of mutations detected in P042_LIV, including all the key mutations affecting cancer genes such as TP53, SMAD4, KRAS, MUC16 and ODZ1, are retained in the PDX model (green dots in [Fig 3B](#)). We also noticed a small subset of protein-altering mutations that were either found only in the xenograft or only in the tumor sample. Overall the MAF data demonstrated that the clonal and subclonal mutations from the original tumor are largely retained in xenograft models and this high concordance of somatic alterations indicates that the genomic mutational pattern remains broadly stable in xenografts, consistent with previous reports [[17–19](#), [41](#)].

Discussion

PDAC patients suffer from rapid disease progression and metastasis, which motivated the search for somatic alterations driving metastasis. In this study, we conducted WES analysis on a collection of rapid autopsy (RA) PDAC primary and matched metastatic tumors and xenografts, and generated a comprehensive portrait of somatic SNVs. Mutations identified from our study in genes such as MUC16, MUC2 and ODZ1 could shed new insights into metastasis-promoting pathways in tumor cells. Although the functional impact of those highlighted somatic SNVs remains to be further tested experimentally and their incidence needs to be determined in a larger cohort, they might serve as a starting point to facilitate our understanding of the molecular mechanisms that underlie metastasis formation in PDAC.

The somatic variants in the PDX models substantially recapitulated their tissues of origin and represent a promising platform to identify and validate potential new pancreatic cancer targets. Based on the somatic SNVs data from this study, we can see that all the key mutations including those on cancer driver genes and tumor suppressors are retained in the xenograft models. Moreover, experimental therapeutic trials in these end-stage derived models should represent a higher bar in which meaningful responses are likely more difficult to achieve but when seen, would generate more confidence to move into the clinic with a new therapy. Overall, these results provide insight into the genetic basis of PDAC metastasis and can help advance the development of targeted therapeutics aimed at controlling progression of pancreatic cancer.

Materials and Methods

Patient information

Detailed clinical information can be found in [Table 5](#). The rapid autopsy program was established at CNIO for the purpose of harvesting pancreatic cancer patients' tissue samples within

Table 5. Patient information.

| Patient ID | Age (years) | Gender | T/N Stage | Treatment | Met. Site(s) |
|------------|-------------|--------|-----------|--|-------------------|
| P042 | 70 | Male | T4N1M1 | 1st line: Gemcitabine + Abraxane; 2nd line: Ci.Tr. PM 1183-B-001-1; 3rd line: Xeloda | liver |
| P047 | 75 | Female | T3-4N1M1 | Gemcitabine + DLL4 Inhibitor | liver, peritoneum |
| P059 | 69 | Male | T3-4N1M2 | Gemcitabine + Abraxane (4 cycles) | liver |

doi:10.1371/journal.pone.0142631.t005

1–3 hours after death. All samples were obtained with written informed consent. The consent and this study were reviewed and approved by the IRB of Hospital de Madrid, in accordance with the tenets of the Declaration of Helsinki.

Sample preparation and whole exome sequencing

Frozen normal, primary tumor and matched metastasis samples were collected and five xenografts were successfully generated from primary tumor or metastases. For P047, the primary tumor xenograft didn't grow successfully and for P059, not enough materials were found for the primary tumor and liver metastasis. Exome capture was performed using the Illumina TrueSeq exome enrichment kit after DNA sample preparation. The kit targeted for more than 200,000 exons spanning over 62 Mb of human genome (http://images.illumina.com/products/truseq_exome_enrichment_kit.ilmn). Captured DNA from the samples was fragmented and sequenced using HiSeq 2000 sequencer to generate paired end, 100-bp reads. On average, approximately 35–40 million reads were generated and assessed by their quality scores. Low quality reads/bases (Phred score < 20) were removed or trimmed before mapping. For PDX samples mouse reads from the mixture of both human and mouse tissue were removed prior to mapping and mutation analysis using a dictionary based method [42]. This approach identified between 2.4% and 17.9% mouse contaminations in the five sequenced xenograft samples. After QC and cleaning of the raw reads, clipped reads were mapped to the hg19 human reference genome using the Burrows-Wheeler Alignment tool (BWA) [43] with default parameters. The mapping quality of the resulting BAM files was inspected, and those with zero quality were filtered out to reduce the false positive mapped reads. PCR duplicates were identified and removed using Samtools [44]. For generating somatic SNV calls, we used Strelka [45] to compare tumor and matched normal samples with a standard Q-score (a measure of the confidence of the somatic calls [45]) to be equal or greater than 15. Then the somatic SNVs were annotated based on the ENSEMBL [46] and their functional impact was predicted by the SIFT tool [23]. The raw data from this study has been made available through the European Nucleotide Archive (ENA), with accession number: PRJEB9296. (<http://www.ebi.ac.uk/ena/data/view/PRJEB9296>).

Supporting Information

S1 Table.

(XLSX)

Author Contributions

Conceived and designed the experiments: TX DJS PO PAR MH. Performed the experiments: MM PPL. Analyzed the data: TX DJS PO. Contributed reagents/materials/analysis tools: TX MM PPL DJS PO. Wrote the paper: TX MM PPL DJS PO PAR MH.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer Statistics, 2008. CA: A Cancer Journal for Clinicians. 2008; 58(2):71–96. doi: [10.3322/ca.2007.0010](https://doi.org/10.3322/ca.2007.0010)
2. Winter JM, Cameron JL, Campbell KA, Arnold MA, Chang DC, Coleman J, et al. 1423 pancreaticoduodenectomies for pancreatic cancer: A single-institution experience. J Gastrointest Surg. 2006; 10(9):1199–210; discussion 210–1. Epub 2006/11/23. doi: [10.1016/j.gassur.2006.08.018](https://doi.org/10.1016/j.gassur.2006.08.018) PMID: [17114007](https://pubmed.ncbi.nlm.nih.gov/17114007/).
3. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev. 2006; 20(10):1218–49. Epub 2006/05/17. doi: [10.1101/gad.1415606](https://doi.org/10.1101/gad.1415606) PMID: [16702400](https://pubmed.ncbi.nlm.nih.gov/16702400/).
4. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010; 467(7319):1114–7. Epub 2010/10/29. doi: [10.1038/nature09515](https://doi.org/10.1038/nature09515) PMID: [20981102](https://pubmed.ncbi.nlm.nih.gov/20981102/); PubMed Central PMCID: [PMC3148940](https://pubmed.ncbi.nlm.nih.gov/PMC3148940/).
5. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010; 467(7319):1109–13. Epub 2010/10/29. doi: [10.1038/nature09460](https://doi.org/10.1038/nature09460) PMID: [20981101](https://pubmed.ncbi.nlm.nih.gov/20981101/); PubMed Central PMCID: [PMC3137369](https://pubmed.ncbi.nlm.nih.gov/PMC3137369/).
6. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science. 2008; 321(5897):1801–6. Epub 2008/09/06. doi: [10.1126/science.1164368](https://doi.org/10.1126/science.1164368) PMID: [18772397](https://pubmed.ncbi.nlm.nih.gov/18772397/); PubMed Central PMCID: [PMC2848990](https://pubmed.ncbi.nlm.nih.gov/PMC2848990/).
7. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012; 491(7424):399–405. doi: [10.1038/nature11547](https://doi.org/10.1038/nature11547) PMID: [23103869](https://pubmed.ncbi.nlm.nih.gov/23103869/); PubMed Central PMCID: [PMC3530898](https://pubmed.ncbi.nlm.nih.gov/PMC3530898/).
8. Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. Nature reviews Gastroenterology & hepatology. 2012; 9(8):477–86. doi: [10.1038/nrgastro.2012.126](https://doi.org/10.1038/nrgastro.2012.126) PMID: [22751458](https://pubmed.ncbi.nlm.nih.gov/22751458/).
9. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015; 518(7540):495–501. doi: [10.1038/nature14169](https://doi.org/10.1038/nature14169) PMID: [25719666](https://pubmed.ncbi.nlm.nih.gov/25719666/).
10. Von Hoff DD, Ramanathan RK, Borad MJ, Laheru DA, Smith LS, Wood TE, et al. Gemcitabine plus nab-paclitaxel is an active regimen in patients with advanced pancreatic cancer: a phase I/II trial. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2011; 29(34):4548–54. doi: [10.1200/JCO.2011.36.5742](https://doi.org/10.1200/JCO.2011.36.5742) PMID: [21969517](https://pubmed.ncbi.nlm.nih.gov/21969517/); PubMed Central PMCID: [PMC3565012](https://pubmed.ncbi.nlm.nih.gov/PMC3565012/).
11. Conroy T, Desseigne F, Ychou M, Bouche O, Guimbaud R, Becouarn Y, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. The New England journal of medicine. 2011; 364(19):1817–25. doi: [10.1056/NEJMoa1011923](https://doi.org/10.1056/NEJMoa1011923) PMID: [21561347](https://pubmed.ncbi.nlm.nih.gov/21561347/).
12. Christofori G. New signals from the invasive front. Nature. 2006; 441(7092):444–50. Epub 2006/05/26. doi: [10.1038/nature04872](https://doi.org/10.1038/nature04872) PMID: [16724056](https://pubmed.ncbi.nlm.nih.gov/16724056/).
13. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science. 2011; 331(6024):1559–64. Epub 2011/03/26. doi: [10.1126/science.1203543](https://doi.org/10.1126/science.1203543) PMID: [21436443](https://pubmed.ncbi.nlm.nih.gov/21436443/).
14. Nguyen DX, Massague J. Genetic determinants of cancer metastasis. Nat Rev Genet. 2007; 8(5):341–52. Epub 2007/04/19. doi: [10.1038/nrg2101](https://doi.org/10.1038/nrg2101) PMID: [17440531](https://pubmed.ncbi.nlm.nih.gov/17440531/).
15. Hirbe A, Dahiya SM, Miller C, Li T, Fulton R, Zhang X, et al. Whole exome sequencing reveals the order of genetic changes during malignant transformation and metastasis in a single patient with NF1-plexiform neurofibroma. Clinical cancer research: an official journal of the American Association for Cancer Research. 2015. doi: [10.1158/1078-0432.CCR-14-3049](https://doi.org/10.1158/1078-0432.CCR-14-3049) PMID: [25925892](https://pubmed.ncbi.nlm.nih.gov/25925892/).
16. Siolas D, Hannon GJ. Patient-derived tumor xenografts: transforming clinical samples into mouse models. Cancer Res. 2013; 73(17):5315–9. Epub 2013/06/05. doi: [0008-5472.CAN-13-1069](https://doi.org/10.1158/0008-5472.CAN-13-1069) [pii]doi: [10.1158/0008-5472.CAN-13-1069](https://doi.org/10.1158/0008-5472.CAN-13-1069) PMID: [23733750](https://pubmed.ncbi.nlm.nih.gov/23733750/); PubMed Central PMCID: [PMC3766500](https://pubmed.ncbi.nlm.nih.gov/PMC3766500/).
17. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464(7291):999–1005. Epub 2010/04/16. doi: [10.1038/nature08989](https://doi.org/10.1038/nature08989) PMID: [20393555](https://pubmed.ncbi.nlm.nih.gov/20393555/); PubMed Central PMCID: [PMC2872544](https://pubmed.ncbi.nlm.nih.gov/PMC2872544/).
18. Rossello FJ, Tothill RW, Britt K, Marini KD, Falzon J, Thomas DM, et al. Next-generation sequence analysis of cancer xenograft models. PloS one. 2013; 8(9):e74432. doi: [10.1371/journal.pone.0074432](https://doi.org/10.1371/journal.pone.0074432) PMID: [24086345](https://pubmed.ncbi.nlm.nih.gov/24086345/); PubMed Central PMCID: [PMC3784448](https://pubmed.ncbi.nlm.nih.gov/PMC3784448/).
19. Gu Q, Zhang B, Sun H, Xu Q, Tan Y, Wang G, et al. Genomic characterization of a large panel of patient-derived hepatocellular carcinoma xenograft tumor models for preclinical development. Oncotarget. 2015. PMID: [26062443](https://pubmed.ncbi.nlm.nih.gov/26062443/).

20. Aparicio S, Hidalgo M, Kung AL. Examining the utility of patient-derived xenograft mouse models. *Nature reviews Cancer*. 2015; 15(5):311–6. doi: [10.1038/nrc3944](https://doi.org/10.1038/nrc3944) PMID: [25907221](https://pubmed.ncbi.nlm.nih.gov/25907221/).
21. Majumder B, Baraneedharan U, Thiyagarajan S, Radhakrishnan P, Narasimhan H, Dhandapani M, et al. Predicting clinical response to anticancer drugs using an ex vivo platform that captures tumour heterogeneity. *Nature communications*. 2015; 6:6169. doi: [10.1038/ncomms7169](https://doi.org/10.1038/ncomms7169) PMID: [25721094](https://pubmed.ncbi.nlm.nih.gov/25721094/); PubMed Central PMCID: PMC4351621.
22. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007; 446(7132):153–8. Epub 2007/03/09. doi: [10.1038/nature05610](https://doi.org/10.1038/nature05610) PMID: [17344846](https://pubmed.ncbi.nlm.nih.gov/17344846/); PubMed Central PMCID: PMC2712719.
23. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4(7):1073–81. Epub 2009/06/30. doi: [10.1038/nprot.2009.86](https://doi.org/10.1038/nprot.2009.86) PMID: [19561590](https://pubmed.ncbi.nlm.nih.gov/19561590/).
24. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, et al. Ultraconserved elements in the human genome. *Science*. 2004; 304(5675):1321–5. doi: [10.1126/science.1098119](https://doi.org/10.1126/science.1098119) PMID: [15131266](https://pubmed.ncbi.nlm.nih.gov/15131266/).
25. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic biology*. 2014; 63(1):83–95. doi: [10.1093/sysbio/syt061](https://doi.org/10.1093/sysbio/syt061) PMID: [24021724](https://pubmed.ncbi.nlm.nih.gov/24021724/).
26. Kaur S, Kumar S, Momi N, Sasson AR, Batra SK. Mucins in pancreatic cancer and its microenvironment. *Nature reviews Gastroenterology & hepatology*. 2013; 10(10):607–20. doi: [10.1038/nrgastro.2013.120](https://doi.org/10.1038/nrgastro.2013.120) PMID: [23856888](https://pubmed.ncbi.nlm.nih.gov/23856888/); PubMed Central PMCID: PMC3934431.
27. Moniaux N, Andrianifahanana M, Brand RE, Batra SK. Multiple roles of mucins in pancreatic cancer, a lethal and challenging malignancy. *British journal of cancer*. 2004; 91(9):1633–8. doi: [10.1038/sj.bjc.6602163](https://doi.org/10.1038/sj.bjc.6602163) PMID: [15494719](https://pubmed.ncbi.nlm.nih.gov/15494719/); PubMed Central PMCID: PMC2409950.
28. Haridas D, Chakraborty S, Ponnusamy MP, Lakshmanan I, Rachagani S, Cruz E, et al. Pathobiological implications of MUC16 expression in pancreatic cancer. *PloS one*. 2011; 6(10):e26839. doi: [10.1371/journal.pone.0026839](https://doi.org/10.1371/journal.pone.0026839) PMID: [22066010](https://pubmed.ncbi.nlm.nih.gov/22066010/); PubMed Central PMCID: PMC3204976.
29. Shimizu A, Hirano S, Tani M, Kawai M, Okada K, Miyazawa M, et al. Coexpression of MUC16 and mesothelin is related to the invasion process in pancreatic ductal adenocarcinoma. *Cancer science*. 2012; 103(4):739–46. doi: [10.1111/j.1349-7006.2012.02214.x](https://doi.org/10.1111/j.1349-7006.2012.02214.x) PMID: [22320398](https://pubmed.ncbi.nlm.nih.gov/22320398/).
30. Chen SH, Hung WC, Wang P, Paul C, Konstantopoulos K. Mesothelin binding to CA125/MUC16 promotes pancreatic cancer cell motility and invasion via MMP-7 activation. *Sci Rep*. 2013; 3:1870. Epub 2013/05/23. doi: [10.1038/srep01870](https://doi.org/10.1038/srep01870) PMID: [23694968](https://pubmed.ncbi.nlm.nih.gov/23694968/); PubMed Central PMCID: PMC3660778.
31. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499(7457):214–8. doi: [10.1038/nature12213](https://doi.org/10.1038/nature12213) PMID: [23770567](https://pubmed.ncbi.nlm.nih.gov/23770567/); PubMed Central PMCID: PMC3919509.
32. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 2011; 39(Database issue):D945–50. Epub 2010/10/19. doi: [10.1093/nar/gkq929](https://doi.org/10.1093/nar/gkq929) PMID: [20952405](https://pubmed.ncbi.nlm.nih.gov/20952405/); PubMed Central PMCID: PMC3013785.
33. Inaba T, Sano H, Kawahito Y, Hla T, Akita K, Toda M, et al. Induction of cyclooxygenase-2 in monocyte/macrophage by mucins secreted from colon cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(5):2736–41. doi: [10.1073/pnas.0435410100](https://doi.org/10.1073/pnas.0435410100) PMID: [12598658](https://pubmed.ncbi.nlm.nih.gov/12598658/); PubMed Central PMCID: PMC151410.
34. He YF, Zhang MY, Wu X, Sun XJ, Xu T, He QZ, et al. High MUC2 expression in ovarian cancer is inversely associated with the M1/M2 ratio of tumor-associated macrophages and patient survival time. *PloS one*. 2013; 8(12):e79769. doi: [10.1371/journal.pone.0079769](https://doi.org/10.1371/journal.pone.0079769) PMID: [24324582](https://pubmed.ncbi.nlm.nih.gov/24324582/); PubMed Central PMCID: PMC3855613.
35. Azzoni C, Bottarelli L, Pizzi S, D'Adda T, Rindi G, Bordi C. Xq25 and Xq26 identify the common minimal deletion region in malignant gastroenteropancreatic endocrine carcinomas. *Virchows Archiv: an international journal of pathology*. 2006; 448(2):119–26. doi: [10.1007/s00428-005-0058-4](https://doi.org/10.1007/s00428-005-0058-4) PMID: [16244870](https://pubmed.ncbi.nlm.nih.gov/16244870/).
36. Jones FS, Jones PL. The tenascin family of ECM glycoproteins: structure, function, and regulation during embryonic development and tissue remodeling. *Developmental dynamics: an official publication of the American Association of Anatomists*. 2000; 218(2):235–59. doi: [10.1002/\(SICI\)1097-0177\(200006\)218:2<235::AID-DVDY2>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0177(200006)218:2<235::AID-DVDY2>3.0.CO;2-G) PMID: [10842355](https://pubmed.ncbi.nlm.nih.gov/10842355/).
37. Setlur SR, Lee C. Tumor archaeology reveals that mutations love company. *Cell*. 2012; 149(5):959–61. doi: [10.1016/j.cell.2012.05.010](https://doi.org/10.1016/j.cell.2012.05.010) PMID: [22632962](https://pubmed.ncbi.nlm.nih.gov/22632962/).

38. Hidalgo M, Von Hoff DD. Translational therapeutic opportunities in ductal adenocarcinoma of the pancreas. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2012; 18(16):4249–56. Epub 2012/08/17. doi: 18/16/4249 [pii]doi: [10.1158/1078-0432.CCR-12-1327](https://doi.org/10.1158/1078-0432.CCR-12-1327) PMID: [22896691](https://pubmed.ncbi.nlm.nih.gov/22896691/).
39. Xie T, Cho YB, Wang K, Huang D, Hong HK, Choi YL, et al. Patterns of somatic alterations between matched primary and metastatic colorectal tumors characterized by whole-genome sequencing. *Genomics*. 2014; 104(4):234–41. Epub 2014/07/30. doi: S0888-7543(14)00127-X [pii]doi: [10.1016/j.ygeno.2014.07.012](https://doi.org/10.1016/j.ygeno.2014.07.012) PMID: [25066378](https://pubmed.ncbi.nlm.nih.gov/25066378/).
40. Vignot S, Frampton GM, Soria JC, Yelensky R, Commo F, Brambilla C, et al. Next-generation sequencing reveals high concordance of recurrent somatic alterations between primary tumor and metastases from patients with non-small-cell lung cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2013; 31(17):2167–72. doi: [10.1200/JCO.2012.47.7737](https://doi.org/10.1200/JCO.2012.47.7737) PMID: [23630207](https://pubmed.ncbi.nlm.nih.gov/23630207/).
41. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*. 2015; 518(7539):422–6. doi: [10.1038/nature13952](https://doi.org/10.1038/nature13952) PMID: [25470049](https://pubmed.ncbi.nlm.nih.gov/25470049/).
42. Conway T, Wazny J, Bromage A, Tymms M, Sooraj D, Williams ED, et al. Xenome—a tool for classifying reads from xenograft samples. *Bioinformatics*. 2012; 28(12):i172–8. Epub 2012/06/13. doi: bts236 [pii]doi: [10.1093/bioinformatics/bts236](https://doi.org/10.1093/bioinformatics/bts236) PMID: [22689758](https://pubmed.ncbi.nlm.nih.gov/22689758/); PubMed Central PMCID: PMC3371868.
43. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. Epub 2009/05/20. doi: btp324 [pii]doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/); PubMed Central PMCID: PMC2705234.
44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. Epub 2009/06/10. doi: btp352 [pii]doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/); PubMed Central PMCID: PMC2723002.
45. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012; 28(14):1811–7. Epub 2012/05/15. doi: bts271 [pii]doi: [10.1093/bioinformatics/bts271](https://doi.org/10.1093/bioinformatics/bts271) PMID: [22581179](https://pubmed.ncbi.nlm.nih.gov/22581179/).
46. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. *Nucleic Acids Res*. 2012; 40(Database issue):D84–90. Epub 2011/11/17. doi: gkr991 [pii]doi: [10.1093/nar/gkr991](https://doi.org/10.1093/nar/gkr991) PMID: [22086963](https://pubmed.ncbi.nlm.nih.gov/22086963/); PubMed Central PMCID: PMC3245178.