The Genetic Basis of Transcriptional and Spatial Heterogeneity of Squamous Features in Pancreatic Ductal Adenocarcinoma

Akimasa Hayashi[1,2], Jun Fan[1,2], Ruoyao Chen[3], Yu-jui Ho[4], Alvin P. Makohon-Moore[1,2], Yi Zhong[1,2], Jungeui Hong[1,2], Hitomi Sakamoto[1,2], Marc A. Attiyeh[1,3], Zachary A. Kohutek[1,2], Lance Zhang[1,2], Jinlong Huang[1,2], Aida Boumiza[1,2], Rajya Kappagantula[1,2], Priscilla Baez[1,2], Laura D. Wood[5], Ralph H. Hruban[5], Lisi Marta[7], Kalyani Chadalavada[7], Gouri J. Nanjangud[7], Olca Basturk[6], David S. Klimstra[1, 6], Michael Overholtzer[3], Christine A. Iacobuzio-Donahue[1,2,6*]

Affiliations:

[1] The David M. Rubenstein Center for Pancreatic Cancer Research, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[2] Human Oncology and Pathogenesis Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY USA.

[3] Cell Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[4] Cancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[5] Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[6] Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[7] Molecular Cytogenetics Core, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

* Correspondence should be addressed to iacobuzc@mskcc.org.

Summary (132 words)

Recent studies indicate that pancreatic cancer expression profiles are variable and largely reflect a classical or basal-type phenotype. We performed genetic sequencing, RNA-seq, and histologic review of multiregion sampled pancreatic cancers and found that squamous and squamoid features, indicators of poor prognosis, correlate with a “basal -like” expressional type. Cancers with squamous features were more likely to have truncal mutations in chromatin modifier genes and intercellular heterogeneity for *MYC* amplification that was associated with entosis. In most patients the basal phenotype coexisted with a glandular component, and phylogenetic studies indicated that it arose from a subclonal population in the tumor. These data provide a unifying paradigm for understanding the interrelationship of basal-type features, squamous histology, and somatic mutations in chromatin modifier genes in the context of the clonal evolution of pancreatic cancer.

Text:

Despite advances in diagnostic tools, surgery, chemotherapy and radiation therapy, pancreatic ductal adenocarcinoma (PDAC) remains one of the most lethal tumor types [1]. The five-year survival is less than 10% with 56,770 new cases diagnosed and 45,750 deaths estimated for 2019 in the United States alone [2,3]. Large scale sequencing studies have revealed the recurrent genomic features of this disease that target a defined number of core pathways [4-8]. In some patients a genome instability signature is also seen based on either microsatellite instability or on a high number of structural rearrangements [5,9]. Transcriptional studies have revealed that PDAC can be segregated into two major subtypes termed "classical" and "basal-like" [6,7,10,11]. Moreover, these pivotal studies have revealed that PDAs with squamous differentiation specifically correspond to the "basal-like" subtype [11], alternatively designated the "squamous" subtype by some [6]. This overlap and strong relationship of the adenosquamous phenotype with the "basal-like/squamous" signature was confirmed in TCGA re-analysis [7].

We have previously reported that, compared to resected PDACs, advanced stage neoplasms have a higher incidence of variant histomorphologic features that include adenosquamous morphology [12]. Adenosquamous carcinoma (ASC) is a morphologic subtype of PDAC defined by its having >30% squamous morphology. ASCs account for 0.9% of all PDACs and are associated with a poorer prognosis than carcinomas with conventional glandular morphologies [13-15]. In light of this observation, and transcriptional studies indicating PDAC is represented by two transcriptional subtypes, we posited that an integrated analysis of the histologic, genomic and transcriptional features of PDAC would provide insight into the dynamics of transcriptional alterations and adenosquamous features during clonal evolution and development of metastatic disease.

We reviewed hematoxylin and eosin stained sections prepared from more than 7,000 unique formalin-fixed paraffin-embedded tissues from 156 research autopsy participants spanning two institutions, all of whom had been clinically or pathologically diagnosed with PDAC premortem. After histologic review 33 cases were excluded (see Extended Data Figure 1 and Supplementary Data Table 1) leaving 2944 individual sections from 123 cases (median 17 tumor sections per case) that fulfilled our criteria for further study. Histologic review in combination with immunohistochemical labeling of representative blocks for the common squamous differentiation markers CK5/6 and p63 [16,17] was performed so that each individual formalin-fixed section was categorized as having a conventional glandular pattern of growth (GL), squamoid features (SF), or squamous differentiation (SD) (Figure 1a,1b). Squamoid features were defined as the presence of an alveolar or solid morphologic component with CK5/6 or p63 positivity, but not conventional squamous morphology, as recently described [18]. By contrast, squamous differentiation was defined as the presence of a solid growth pattern of neoplastic cells with abundant pink cytoplasm, prominent cell borders, and intercellular junctions in addition to having both CK5/6 and p63 positivity. Of 2944 blocks, 490 (16.6%) showed squamoid features (SF) or squamous differentiation (SD) (Figure 1c). When these sections were stratified by patient we determined that seven PDACs (5.7%) met criteria for adenosquamous carcinoma (ASC), six PDACs (4.9%) had focal (<30%) squamous differentiation and two PDACs (1.6%) had focal squamoid features (Figure 1d; Extended Data Figure 2). Three (PAM16, PAM54 and PAM73) of the seven ASCs were notable for having admixing of squamous and glandular morphologies throughout the neoplasm, with an estimated >70% of these neoplasms having squamous differentiation overall. In the remaining four ASCs (PAM32, PAM46, PAM54 and PAM55) the squamous morphology formed a large discrete focus in a neoplasm with regions of otherwise conventional glandular

morphology. By univariate analysis patients with ASCs or PDACs with SF/SD had a poorer survival than did patients with PDACs without SF/SD (Figure 1e), compatible with previous findings [14]. The prevalence of ASC in this cohort of 129 patients with end stage disease is higher than that reported for surgically resected tumors (5.7% of our cohort vs 0.9% in Boyd et al. $P < 0.0001$, Chi-square test) [14].

Based on histologic review of all 2944 sections we also noted that ASCs and PDACs with SF/SD exhibited entosis, a distinct form of cell death in which one cancer cell engulfs another (Figure 1f) [19]. Entosis was recognized as distinct from epithelial pearls that may also be seen in squamous carcinomas [20] as they were not associated with concentric layers of squamous cells, with aberrant keratinization and were not located centrally within nests of squamous carcinoma. Entosis was also identified in regions of conventional glandular morphology whereas squamous pearls would not be. To more rigorously determine the relationship of SF/SD to entosis we therefore adopted strict criteria to count entotic cell-in-cell structures (CIC) (Methods) [21]. The number of entotic CIC was higher in PDACs with SF/SD or ASCs compared to PDACs without SF/SD in our cohort (mean ± standard deviation per 10HPF is 1.095±1.240 versus 0.365±0.437 respectively, $P < 0.0001$, Mann–Whitney U test). To determine if entosis is more reflective of stage of disease versus morphology, we reviewed an independent cohort of 30 resected PDACs that included eight ASCs. Similar to the findings in the autopsy cohort, resected ASCs had more entotic CIC than conventional PDACs (0.441±0.213 versus 0.226±0.219, respectively, $P = 0.007$, Mann–Whitney U test) (Figure 1g). Autopsy PDACs with SF/SD or ASCs had more entotic CIC than surgically resected ASCs, although the difference was not statistically significant. Collectively, we conclude that squamous morphologic features, characterized in part by entosis, becomes more prominent with PDAC progression.

To determine the extent that the classical and basal-type transcriptional signatures correspond to our morphologic findings, we extracted total RNA from 480 frozen samples in triplicate; in all cases the frozen tissue was matched to the formalin-fixed sections used for morphologic and immunohistochemical analyses. A total of 214 frozen samples from 27 patients (median 6 samples, range 1 to 26 samples per patient) meeting quality criteria (Methods) were used for RNA sequencing (Supplementary Data Table 2). These 27 cases included five ASCs (three with focal glandular dominant areas) and five PDAC with focal SF/SD. When possible both the conventional glandular (GL) and squamoid/squamous morphologies (SF/SD) from the same carcinoma were separately analyzed. Normalized mRNA expression levels of *TP63*, *KRT5* and *KRT6A* confirmed that GL-PDAC samples had the lowest expression of all three markers whereas SD-PDAC samples had the highest expression levels of all three markers. SF-PDACs had an intermediate expression pattern between GL-PDACs and SD-PDACs (Extended Data Figure 3). Consistent with this finding, network analysis highlights KRT5 and KRT6A as "hub" genes in samples with SF/SD morphology, and SF/SD morphology shows more complex co-expression patterns in keratin filament & keratinization pathway than in samples with GL morphology. We next classified our samples into "classical" and "basal-like" PDAC subtypes using the 50 pancreas cancer gene set reported by Moffitt et al. [11]. This revealed an almost perfect concordance of morphologic features with transcriptional subtype, as most SF-PDAC and all SD-PDAC samples corresponded to the "basal-like" expression pattern, whereas most GL-PDACs corresponded to the "classical" type pattern (Figure 2a). Principal component analysis using this same gene set revealed a similar distribution based on morphologic features or expression subtype, whereas no relationship was found for site of harvesting of each sample (primary or metastasis) (Figure 2b). Finally, we compared the morphologic features to the transcriptional subtypes of each sample analyzed for 23

of these 27 patients for which two or more samples were analyzed. In 15 patients all samples studied were homogeneous with respect to both their transcriptional subtype and morphologic pattern (Figure 2c). The majority of these cases had a glandular morphology (PDAC-GL) and a classical-type expression signature, although in two patients (PAM16, PAM54) prominent squamous differentiation was identified in entire neoplasm and had a basal-type expression signature. In a separate set of three patients (PAM28, PAM39, PAM53) all samples analyzed were homogenous for their transcriptional subtype despite a degree of morphologic heterogeneity (Figure 2d). These included a basal-type transcriptional signature but glandular morphology in the metastases of PAM28 and PAM53, and a classical expression signature in a metastasis with squamoid features in PAM39. Finally, in five patients (PAM02, PAM22, PAM46, PAM55, MPAM6) we observed that both the classical and basal-like subtypes co-existed within a single patient. With two exceptions (one primary tumor sample each in PAM02 and PAM55) the transcriptional signatures correlated with the histologic features of the sample.

We mined our RNAseq dataset to determine the transcriptional differences between samples with GL morphology and SF/SD morphology in an unbiased manner. Gene set enrichment analysis (GSEA) using Hallmark genesets and transcription factor target genesets (Methods, Supplementary Data Table 4) revealed MYC target gene expression as significantly enriched in samples with SF/SD compared to GL morphology (Figure 3a, Supplementary Data Table 5, 6), a finding similar to that reported by Bailey et al. [6]. To further determine the significance of this observation we reviewed our RNAseq data specifically for MYC gene expression. This revealed a significantly higher MYC transcript abundance in samples with SF/SD morphology compared to those with GL morphology (Figure 3b). As *MYC*, located in chromosome 8q 24.21, is a known target of amplification in PDAC [22,23] we performed fluorescent *in situ* hybridization (FISH)

analysis to see if increased MYC expression was due to increased *MYC* copy number. We found eight cases wherein both GL and SF/SD morphologies were present within the same tumor/section and in all, *MYC* copy number was significantly higher in regions with SF/SD morphology compared to regions with GL morphology (Figure 3c-d). To further assess the relationship of *MYC* amplification to PDAC progression, we correlated *MYC* copy number with morphologic features in 44 patients in our cohort (32 GL-PDAC, six SF/SD-PDAC, six ASC) (Figure 4, Extended Data Figure 1, Supplementary Data Table 4). *MYC* amplification (≥6-fold by FACETS or FISH) was significantly associated with higher tumor grade and with squamous subtype (amplification observed in two of 13 Grade 2 PDACs, nine of 19 Grade 3 PDACs, 10 of 12 PDACs with SF/SD or ASCs, $P = 0.003$, two-sided Fisher Exact Test). *MYC* amplification also correlated significantly with poor outcome (Figure 3e) and high number of entotic CICs ($P = 0.0027$, Chi-square test with Yates continuity correction) (Figure 4). *MYC* amplification (≥6-fold) was more prevalent in this cohort compared to the reported prevalence in resectable disease (21/44 in our cohort vs 5/149 in TCGA, $P < 0.0001$, Chi-square) [5,7]. In addition, knockdown of MYC in the adenosquamous cell line L3.3 led to a decrease in ΔN-p63 transcript abundance (Extended Data Figure 4) in keeping with the findings of Andricovich et al.[18]. Overall these findings indicate that gains of *MYC* copy number are common during PDAC progression.

In light of the correlation of both *MYC* amplification and entosis with tumor progression in PDAC, we more closely determined the relationship if any between these two observations. In nine cases with *MYC* amplification and entotic CICs, FISH revealed remarkable intercellular heterogeneity for *MYC* copy number. We therefore determined *MYC* copy number in a minimum of five entotic CICs in four different neoplasms, i.e. in matched winner cells (eating) and loser cells (the cell being eaten) (Figure 3f). The winner cells had 8.3 ± 10.5 copies of *MYC* compared

to only 2.4 ± 3.2 copies per loser cell ($P < 0.0001$, Mann-Whitney U-test). After normalization for chromosome 8, the winner cells had still higher copy number (2.1 ± 1.4 copies per winner cell compared to 1.7 ± 0.9 copies per loser cell), but was not statistically significant ($P = 0.103$, Mann-Whitney U-test). Collectively this suggests that gain of *MYC* copy number is selected for in the context of gains in ploidy[24], and illustrates that intercellular heterogeneity for *MYC* amplification specifically is correlated with entosis.

We next determined the relationship of the coding genomic landscape to the development of squamous morphology by performing multiregion whole exome or whole genome sequencing on DNA extracted from frozen samples matched to the formalin-fixed sections in all 44 patients. High quality single nucleotide variants and small insertions/deletions were identified for each sample and used to recreate the phylogenetic relationships among the spatially distinct samples within each patient (Methods). Overall the genetic features of this cohort were consistent with the PDAC genomic landscape (Figure 4, Supplementary Data Table 4)[4-8]. We did not identify mutations of *UPF1* that have previously been reported in ASC [25]. However, cancers from two of 44 patients had a *KDM6A* mutation [6,18], both of whom were female and one with an ASC, leading us to more closely evaluate the chromatin modifier gene mutations in all 44 patients. The most common chromatin modifier gene with a deleterious mutation was *KMT2C* (seven of 44 cases, 16%), followed by *ARID1A* (four cases, 9%), *KMT2D* (three cases, 7%), *ARID2* and *KDM6A* (two cases each, 5%). With two exceptions the mutations in chromatin modifier genes were mutually exclusive (Figure 3e). Seven of 12 patients (58%) with a PDAC with SF/SD morphology or ASC had a mutation in a chromatin modifier gene compared to 10 of 32 patients (31%) with a PDAC with GL morphology, a difference that did not reach statistical significance ($P = 0.195$, Chi-square with Yates' continuity correction).

While there was no difference in the prevalence of mutations in chromatin modifier genes in cancers with or without SF/SD, phylogenetic trees indicated an influential relationship among the evolutionary timing that a mutation in a chromatin modifier gene arose and the extent of squamous morphology. For example, seven of seven chromatin modifier gene mutations identified in PDACs with SF/SD morphology or ASCs were truncal in origin (Figures 4 and 5, Extended Data Figures 5 and 6), compared to only five of the eleven mutations in PDACs with GL morphology (P = 0.038, two-sided Fisher Exact Test). In the remaining PDACs with GL morphology, mutations in chromatin modifier genes were assigned to a branch or were private to a single sample in that patient. Curiously, we noted that two PDACs with SF/SD and wild type chromatin modifier genes (PAM28, MPAM6) had deleterious truncal mutations of *RB1* (Figure 5b, Extended Data Figure 7).

We also evaluated the approximate evolutionary timing of *MYC* copy number gain during clonal evolution based on FACETS copy number and ploidy estimations generated for each sequenced sample (Methods). *MYC* amplification was identified in five cases in a subclonal manner; in four of these cases *MYC* amplification followed whole genome duplication (Figure 5c, Extended Data Figures 6 and 8). Three of five cases with *MYC* amplification had a truncal mutation in *ARID1A* (Extended Data Figure 6). The timing of or subclonality of *MYC* amplification did not correlate with SF/SD morphology. Taken together, we conclude that *MYC* amplification follows truncal mutations in chromatin modifier genes and is in part selected for during whole genome duplication. However, the imperfect correlation of *MYC* amplification with development of squamous morphology suggests that it is a marker of aggressive tumor features in general but not progression to SF/SD or ASC per se.

We further evaluated heterogeneous morphologic or transcriptional features seen in the cancers of 10 patients (six PDACs with SF/SD and four ASCs) in the context of both phylogenetic relationships and their anatomic distribution. In eight of 10 patients the squamous component was a clonal population, i.e. all samples with squamous morphology were phylogenetically more closely related to each other than to the sample(s) with GL morphology in the same patient (Figure 5b,5c and Extended Data Figures 5a, 6a, 6c, 7, 8a). These phylogenetic relationships did not imply a shared anatomic location, as genetic, morphologically and transcriptionally similar samples could be found in both the primary tumor and in metastatic sites (Figure 6, Extended Data Figures 5-19). In the remaining two patients the SF/SD was exclusive to a single sample analyzed (PAM39, Extended Data Figure 6b and PAM22, Extended Data Figure 8b). Integration of phylogenetic trees with morphologic features also suggested that SF/SD can develop independently in the same neoplasm, for example PAM55 (Figure 5a) in which samples PT8, PT9 and samples PT2-PT6 were contained within three different clades respectively. Thus, beyond genetic alterations such as those in chromatin modifier genes, subclonal populations with SF/SD may further be defined by a combination of epigenetic and/or microenvironmental cues [26,27].

These patterns reveal several novel features of PDAC. While mutations in *KMT2C*, *ARID1A* and related genes have consistently been identified in large scale screens of the PDAC genome [6,7], their significance for the natural history of PDAC has remained unclear. We now show that the evolutionary context in which these mutations occur is related to the likelihood the PDAC will develop squamous morphology. This likelihood is not absolute, as evidenced by the deceased patients in our cohort with poorly differentiated PDACs with truncal mutations in chromatin modifier genes. While our findings are consistent with reports that ASCs are associated with a worse outcome [14], they contradict those that report an improved outcome in PDACs with mutations

in *KMT2C*, *ARID1A* and related chromatin modifier genes [28,29]. Future efforts that consider somatic mutations in these genes specifically in the context of whole genome duplication, *MYC* copy number and morphologic features may resolve this discrepancy.

Our data also illustrate that glandular and SF/SD morphologies, and by extension the classical-type and basal-type expression signatures, coexist in the same PDAC. While prior studies of ASC have also reported this phenomenon [17,20,30], we now show that the SF/SD component arises from a subclonal population. This raises two possibilities for understanding SF/SD. First, SF/SD may develop from classical-type gland forming pancreatic cancer. The paucity of data reporting small, early stage ASCs and that SF/SD are commonly found in association with conventional glandular features are consistent with this possibility [30]. Moreover, whereas we found that SF/SD may arise during the clonal evolution of a PDAC we did not observe the converse scenario by phylogenetic analysis, i.e. a subclonal glandular component arising in a predominant SF/SD neoplasm. While we believe the former is the most parsimonious explanation, we acknowledge a second possibility where a common phenotypic intermediate cell type gives rise to both classical-type and basal-type phenotypes. Our study relied on bulk and macrodissected tissues thus we did not reach the level of resolution required to answer this question definitively.

These data also contextualize the significance of *MYC* copy number gains in PDAC by illustrating that it is selected for during tumor progression and following whole genome duplication. Furthermore, we identify a novel feature of *MYC* in PDAC, intercellular heterogeneity for copy number that is associated with entosis. Entosis, a process in which a cancer cell engulfs its neighbor, represents a form of cell competition that stimulated by low glucose environments[19,31]. Intriguingly, MYC expression has also been shown to promote competition between normal cells in both fly and mammalian tissues during development [32,33] suggesting a new

potential mechanistic parallel between intercellular heterogeneity for *MYC* copy number and stimulation of cell competition. In PDAC specifically, these observations provide clues to the microenvironmental changes that contribute to disease progression [34].

In summary, we describe a multimodal approach to multiregion sampled PDACs that provide a unifying paradigm for transcriptional subtypes, squamous morphology and somatic mutations in chromatin modifier genes that is rooted in phylogenetic analyses. These insights provide optimism for clinical management, as they now provide the context in which to understand the significance of these molecular events for more rigorous stratification of patients for personalized medicine approaches. We expect that our findings will have implications for understanding other solid tumor types as well in which these mutations occur and/or that develop squamous features over time. Ultimately, our hope is that comprehensive studies such as this pave the way for identifying novel therapeutic vulnerabilities or re-evaluation of the utility of currently available therapies based on the genotypes and phenotypes assessed.

Methods

Ethics Statement

This study was approved by the Review Boards of Johns Hopkins School of Medicine and Memorial Sloan Kettering Cancer Center.

Patient selection

A cohort of 156 cases from the Gastrointestinal Cancer Rapid Medical Donation Program at Johns Hopkins Hospital and six cases from the Medical Donation Program at Memorial Sloan Kettering Cancer Center were used. All patients had a premortem diagnosis of PDAC based on pathologic review of resected or biopsy material and/or radiographic and biomarker studies. In addition, hematoxylin and eosin (H&E) stained sections of 30 resected PDACs were used for histologic review.

Histology and Immunohistochemistry

H&E slides cut from all formalin-fixed and paraffin-embedded (FFPE) blocks of each autopsy were reviewed by two gastrointestinal pathologists (A.H. and C.I.D). Based on review and joint discussion a consensus diagnosis was rendered. Immunolabeling was performed on unstained serial sections cut from a subset of FFPE blocks per patient with antibodies against p63 (Ventana, clone 4A4) and CK5/6 (Ventana, clone D5/16B4) according to optimized protocol on a Ventana Benchmark XT autostainer (Ventana Medical Systems Inc.). Appropriate positive and negative controls were included in each run.

Histologic Review for Entosis

All H&E sections of each case were reviewed for entotic cell-in-cell structures (CIC) using the criteria proposed by MacKay [21]: cytoplasm of the host cell (winner or engulfing cell), nucleus of the host cell (typically crescent-shaped, binucleate, or multilobular and pushed against the cytoplasmic wall), an intervening vacuolar space completely surrounding the internalized cell (loser), cytoplasm of internalized cell, and nucleus of internalized cell (often round in shape and located centrally or acentrically). If internalized and/or engulfing cells were undergoing mitosis they were excluded from analysis. Any cases in which we were unable to count 50 high power fields and/or had less than five slides for review were excluded from this analysis. Representative entotic CICs were validated by immunofluorescence labeling for e-cadherin in combination with DAPI to highlight cell nuclei in the Molecular Cytogenetics Core at MSKCC (see MYC Immuno-FISH analysis section below for details).

RNA Sequencing

Frozen sections were cut from samples for histologic review and regions of interest were macrodissected for extracting total RNA using TRIzol (Life Technologies) followed by Rneasy Plus Mini Kit (Qiagen). Each RNA sample was initially quantified by Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Samples were additionally quantified by RiboGreen and assessed for quality control using an Agilent BioAnalyzer in the Integrated Genomics Core at MSKCC, and 513ng-1µg of total RNA with an RNA integrity number ranging from 1.3 to 8.3 underwent ribosomal depletion and library preparation using the TruSeq Stranded Total RNA LT Kit (Illumina catalog # RS-122-1202) according to instructions provided by the manufacturer with 8 cycles of PCR. Samples were barcoded and run on a HiSeq 4000 in a 100bp/100bp or 125/125bp

paired end run, using the HiSeq 3000/4000 SBS Kit (Illumina). On average, 94 million paired reads were generated per sample and 26% of the data mapped to the transcriptome.

RNA sequencing data alignment and analysis

Output data (FASTQ files) were mapped to the target genome using the rnaStar aligner [35] that maps reads genomically and resolves reads across splice junctions. We used the 2 pass mapping method outlined by Engstrom [36] in which the reads were mapped twice, the first mapping using a list of known annotated junctions from Ensembl and the second mapping based on known and novel junctions. Postprocessing of the output SAM files was performed using PICARD tools to add read groups and covert it to a compressed BAM format. The expression count matrix from the mapped reads was determined using HTSeq (www-huber.embl.de/users/anders/HTSeq) and the raw count matrix generated by HTSeq was processed using the R/Bioconductor package DESeq (www-huber.embl.de/users/anders/DESeq) to normalize the entire dataset between sample groups. Normalized log2 expression were used for downstream analyses (Supplementary Data Table 3, ***released after publication***).

Expression type classification and PCA analysis

A 50 pancreatic cancer related gene set identified by Moffitt et al. was used to classify all samples into “classical” and “basal” types [11]. Clustering analysis and heatmaps were displayed using the R package ‘pheatmap’ using spearman's rank correlation. These 50 gene signatures were also used for generating the Primary Component Analysis (PCA) plot using DESeq2 package (https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html).

Gene set enrichment analysis

Gene set enrichment analysis was performed based on methods described [37]. Both gene sets and transcription factor target gene sets (Supplementary Data Table 5) based on ChIP-seq data downloaded from ChIP-Atlas (http://chip-atlas.org/) [38] were used for analysis. Only TOP 500 ChIP peaks located within 1000bp from the TSS with scores over 50 were used.

Network analysis and Cytoscape visualization

Co-expression networks were constructed by first identifying the best predicted soft threshold for transforming the data. Pearson correlation between any two genes across samples was next used as the weight between nodes. Subset of Keratins (KRT) family genes were used to construct the weighted gene-gene network, and the network structure was visualized using Cytoscape [39]. We adjusted the width of edges connecting nodes based on the weights, and weights that are less than 0.05 were removed from the network.

DNA sequencing

Genomic DNA was extracted from each tissue using QIAamp DNA Mini Kits (Qiagen). Whole genome sequencing (WGS) and whole exome sequencing (WES) and alignment performed as previously described [40,41]. Briefly, an Illumina HiSeq 2000 platform was used to target a coverage of 60X for WGS samples and 150X for WES samples. The resulting sequencing reads were analyzed *in silico* to assess quality, coverage, as well as alignment to the human reference genome (hg19) using BWA[42]. After read de-duplication, base quality recalibration, and multiple sequence realignment were completed with the Picard Suite and GATK version 3.1[43,44], somatic SNVs/INDELs were detected using Mutect version 1.1.6 and HaplotypeCaller version 2.4[43,45]. We

excluded low-quality or poorly aligned reads from phylogenetic analysis. Filtering of called somatic mutations required each mutant to be observed in at least one neoplastic sample per patient with at least 5% variant allele frequency and with at least 20x coverage; correspondingly, each mutant must have been observed in less than 2% of the reads (or less than 2 reads total) of the matched normal sample with at least 10x coverage. Copy number analyses were performed using FACETS as previously described [46]. Regarding PAM02, we used the data previously reported [40].

Driver gene annotations

All somatic variants causing a frameshift deletion, frameshift insertion, in-frame deletion, in-frame insertion, non-synonymous missense, nonsense, nonstop, splice site/region, or a translation start site change were considered. Variants were called driver mutations if they passed at least three of the following methods: 20/20+ [47], 20/20+ PDAC[47], TUSON[48] and MutSigCV[49]. For frameshift deletions, frameshift insertions and nonsense mutations specifically, passing only two of these four methods were required. Additionally, we required a CHASM p-value of ≤ 0.05 and an FDR of ≤ 0.25 for the 20/20+ and 20/20+ PDAC methods. Additionally, we also considered genes significantly mutated in large PDAC sequencing studies[4,5,7].

Whole Genome Duplication

Whole genome duplication (WGD) was performed in combination of computational analysis and manually reviewed following Bielski et al.[24], called if MCN ≥ 2 and ploidy ≥ 2.5, and >50% of the autosomal genome was affected. Three low tumor purity samples (PAM22PT5, PAM25PT2 and PAM32PT4) which didn't match these criteria were judged in consideration of expecting WGD occurrent point in phylogenetic trees.

Evolutionary analysis

We derived phylogenies for each set of samples by using Treeomics 1.7.9 [50]. Each phylogeny was rooted at the matched patient's normal sample and the leaves represented tumor samples. Treeomics employs a Bayesian inference model to account for error-prone sequencing and varying neoplastic cell content to calculate the probability that a specific variant is present or absent. The global optimal tree is based on Mixed Integer Linear programming. All evolutionary analyses were performed based on WES data with exception of PAM02 (WGS and additional target sequencing)[40] and MPAM06 (WGS). Somatic alterations present in all analyzed samples of a PDAC were considered truncal, in a subset of samples considered branched, and in a single sample considered private.

MYC Immuno-FISH analysis

Immuno-FISH was performed on paraffin sections according to procedures optimized at the Molecular Cytogenetics Core Facility. The primary (E-Cadherin [24E10] Rabbit mAB) and secondary (Goat anti-Rabbit Alexa 488) antibody was purchased form Cell Signaling Technology and Invitrogen (Thermo Fisher Scientific) respectively. The 2-color *MYC*/Cen8 probe was prepared in-house and consisted of BAC clones containing the full length *MYC* gene (clones RPI-80K22, RP11-1136L8, and CTD-2267H22; labeled with Red dUTP) and a centromeric repeat plasmid for chromosome 8 served as the control (pJM128; labeled with Green dUTP). Briefly, de-waxed paraffin sections were microwaved in 10mM sodium citrate, pretreated with 10% pepsin for 10 minutes at 37°C, rinsed in 2XSSC, dehydrated in ethanol series (70%, 90% and 100%), co-denatured at 80°C for 4 minutes with 5-20uL of *MYC*/Cen8 DNA-FISH probe, and hybridized for

72 hours at 37°C. Following hybridization, sections were washed with wash buffer (0.01% Tween 20 in 2XSSC), fixed in 4% formaldehyde for 15-20 minutes at RT, rinsed in 1XPBS, blocked at RT for 1 hour (blocking buffer: 5% FBS and 0.01% Tween 20 in 1XPBS), and incubated overnight at 4°C with primary antibody (1:100)(dilution buffer: 1% FBS and 0.01% Tween 20 in 1XPBS). Following overnight incubation, sections were washed with wash buffer, rinsed in 1XPBS, incubated with secondary antibody (1:500) for 1 hour at 37°C, rinsed in 1XPBS, stained with DAPI and mounted in antifade (Vectashield, Vector Laboratories). Slides were scanned using a Zeiss Axioplan 2i epifluorescence microscope equipped with Isis 5.5.9 imaging software (MetaSystems Group Inc, Waltham, MA). Metafer and VSlide modules within the software were used to generate virtual image of H&E and DAPI-stained sections. In all, corresponding H&E sections assisted in localizing tumor region and histology (GL, SF or SD). The entire section was systematically scanned under 63 × objectives to assess MYC/Cen8 copy number across different histologies and to identify entotic cell-in-cell structures (CIC). All observed entotic cells and representative regions within a case were imaged through the depth of the tissue (merged stack of 16 z-section images taken at 0.5 micron intervals) and signal counts performed on captured images. For correlation of *MYC/*Cen8 copy number with histology, for each case, a minimum of 50 discrete nuclei were scored (range 50-150). Within a given histology (GL, SF or SD), when *MYC*/Cen8 copy number was heterogenous and topographically distinct, a minimum of 50 discrete nuclei were scored for each distinct region whenever possible. For correlation of *MYC/*Cen8 copy number with entosis, only CICs meeting the selection criteria previously described were scored. For each CIC, *MYC*/Cen8 copy number was recorded separately for the "winner" and "loser". Presence of E-Cadherin staining (which highlights the cell perimeter) and nuclear morphology helped distinguish the "loser" (internalized cell with uniformly round nucleus) from "winner" (host cell

with crescent-shaped, binucleate, or multilobulated nucleus and often pushed against the cytoplasmic wall). To minimize truncation artifacts, only nuclei with at least 1 signal for *MYC* and Cen8 were selected. *MYC* amplification was defined as: ≥2 MYC/Cen8 ratio, ≥6 copies of MYC (discrete signal) or presence of at least one *MYC* cluster (≥4 copies; tandem duplications). 3~5 copies of MYC/Cen8 were regarded as copy number gain (polysomy).

Cell Culture and MYC RNAi

L3.3 cells (gift from laboratory of Dr. Scott Lowe) were cultured in DMEM supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin-streptomycin. L3.3 scramble and MYC knockdown cells were prepared by retroviral transduction with pRFP-C-RS scramble (TF30015, Origene) and pRFP-C-RS *c-myc* shRNA constructs (TF311323, Origene), respectively. Cells were selected with puromycin (2 μg/mL) after 48 hr of transduction.

Western Blotting

Cells were scraped with lysis buffer (50mM Tris-HCl at pH 6.8, 2% SDS, 10% glycerol), vortexed, then boiled at 100 °C for 5 min. Protein concentration was measured by BCA assay (Pierce), followed by western blotting as described previously (Florey et al., 2011). The antibodies used include: anti-c-myc (sc-40, Santa Cruz), anti-GAPDH (sc-47724, Santa Cruz), and anti-p63 (619002, BioLegend).

Statistics

All statistics was performed using XLSTAT (version 2018.2).

Acknowledgements:

Supported by NIH grants R01 CA179991 and R35 CA220508 to C.I.D., F31 CA180682 and 2T32 CA160001-06 to A.M.M., CA62924 to R.H.H., the NCI Cancer Center Support Grant P30 CA08748 to MSKCC, the Daiichi-Sankyo Foundation of Life Science Fellowship to A.H, the Mochida Memorial Foundation for Medical and Pharmaceutical Research Fellowship to A.H, Cycle for Survival and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. We are grateful to Gokce Askan and Jacklynn V. Egger for assistance in identifying resected adenosquamous samples for use in this study, and to Shinya Oki for technical support.

Author Contributions:

A.H., J.F. and C.I.-D. designed the study; A.H., J.F., A.P.M.-M, H.S., M.A.A., A.B., R.K., P.B., L.D.W., R.H.H., C.I.-D. collected autopsy samples; A.H. and C.I.-D. reviewed histology of autopsy samples and selected cases; O.B., D.K., A.H. and C.I.-D. reviewed pathology of surgical cases; A.H., R.C., M.O., G.J.N. and C.I.-D. reviewed entosis of Immuno-FISH slides; A.H. and J.F. prepared RNA samples; A.P.M.-M., J.H., H.S., Z.K. and A.H., prepared the DNA samples; A.H., Y.Z. and C.I.-D. performed RNA sequencing; Y.H., A.H., L.Z. and J.H. analyzed RNA sequencing results; A.P.M.-M., J.H., Z.K., H.S. M.A.A., A.H., and C.I.-D. performed DNA sequencing; A.P.M.-M., M.A.A., J.H., A.H. and C.I.-D. analyzed DNA sequencing results and derived the phylogenies; L.M., K.C. and G.J.N. performed Immuno-FISH; R.C. performed knockdown experiments; A.H., R.C., Y.H., M.O. and C.I.-D. wrote the manuscript; all authors reviewed and edited the final manuscript.

Competing Interests:

The authors declare no competing interests. D.K. is a consultant and equity holder to Paige.AI, consultant to Merck Pharmaceuticals, and receives royalties from UpToDate and the American Registry of Pathology.

References

1 Kamisawa, T., Wood, L. D., Itoi, T. & Takaori, K. Pancreatic cancer. *Lancet* **388**, 73-85, doi:10.1016/s0140-6736(16)00141-0 (2016).

2 Gillen, S., Schuster, T., Meyer Zum Buschenfelde, C., Friess, H. & Kleeff, J. Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages. *PLoS Med* **7**, e1000267, doi:10.1371/journal.pmed.1000267 (2010).

3 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, doi:10.3322/caac.21551 (2019).

4 Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405, doi:10.1038/nature11547 (2012).

5 Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501, doi:10.1038/nature14169 (2015).

6 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52, doi:10.1038/nature16965 (2016).

7 Cancer_Genome_Atlas_Research_Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203 e113, doi:10.1016/j.ccell.2017.07.007 (2017).

8 Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature communications* **6**, 6744, doi:10.1038/ncomms7744 (2015).

9 Wilentz, R. E. *et al.* Genetic, immunohistochemical, and clinical features of medullary carcinoma of the pancreas: A newly described and characterized entity. *The American journal of pathology* **156**, 1641-1651, doi:10.1016/s0002-9440(10)65035-3 (2000).

10 Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine* **17**, 500-503, doi:10.1038/nm.2344 (2011).

11 Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature genetics* **47**, 1168-1178, doi:10.1038/ng.3398 (2015).

12 Iacobuzio-Donahue, C. A. *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27**, 1806-1813, doi:10.1200/jco.2008.17.7188 (2009).

13 Fukushima, N. *et al.* Ductal adnocarcinoma variants and mixed neoplasms of the pancreas. *World Health Organization Classification of Tumors* **4th Edition**, 292-296 (2010).

14 Boyd, C. A., Benarroch-Gampel, J., Sheffield, K. M., Cooksley, C. D. & Riall, T. S. 415 patients with adenosquamous carcinoma of the pancreas: a population-based analysis of prognosis and survival. *J Surg Res* **174**, 12-19, doi:10.1016/j.jss.2011.06.015 (2012).

15 Hester, C. A. *et al.* Comparative outcomes of adenosquamous carcinoma of the pancreas: An analysis of the National Cancer Database. *Journal of surgical oncology* **118**, 21-30, doi:10.1002/jso.25112 (2018).

16 Brody, J. R. *et al.* Adenosquamous carcinoma of the pancreas harbors KRAS2, DPC4 and TP53 molecular alterations similar to pancreatic ductal adenocarcinoma. *Mod Pathol* **22**, 651-659, doi:10.1038/modpathol.2009.15 (2009).

17 Kardon, D. E., Thompson, L. D., Przygodzki, R. M. & Heffess, C. S. Adenosquamous carcinoma of the pancreas: a clinicopathologic series of 25 cases. *Mod Pathol* **14**, 443-451, doi:10.1038/modpathol.3880332 (2001).

18 Andricovich, J. *et al.* Loss of KDM6A Activates Super-Enhancers to Induce Gender-Specific Squamous-like Pancreatic Cancer and Confers Sensitivity to BET Inhibitors. *Cancer Cell* **33**, 512-526.e518, doi:10.1016/j.ccell.2018.02.003 (2018).

19 Overholtzer, M. *et al.* A nonapoptotic cell death process, entosis, that occurs by cell-in-cell invasion. *Cell* **131**, 966-979, doi:10.1016/j.cell.2007.10.040 (2007).

20 Yamaguchi, K. & Enjoji, M. Adenosquamous carcinoma of the pancreas: a clinicopathologic study. *Journal of surgical oncology* **47**, 109-116 (1991).

21 Mackay, H. L. *et al.* Genomic instability in mutant p53 cancer cells upon entotic engulfment. *Nature communications* **9**, 3070, doi:10.1038/s41467-018-05368-1 (2018).

22 Schleger, C., Verbeke, C., Hildenbrand, R., Zentgraf, H. & Bleyl, U. c-MYC activation in primary and metastatic ductal adenocarcinoma of the pancreas: incidence, mechanisms, and clinical significance. *Mod Pathol* **15**, 462-469, doi:10.1038/modpathol.3880547 (2002).

23 Wirth, M., Mahboobi, S., Kramer, O. H. & Schneider, G. Concepts to Target MYC in Pancreatic Cancer. *Mol Cancer Ther* **15**, 1792-1798, doi:10.1158/1535-7163.mct-16-0050 (2016).

24 Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nature genetics* **50**, 1189-1195, doi:10.1038/s41588-018-0165-1 (2018).

25 Liu, C. *et al.* The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. *Nature medicine* **20**, 596-598, doi:10.1038/nm.3548 (2014).

26 McDonald, O. G. *et al.* Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nature genetics* **49**, 367-376, doi:10.1038/ng.3753 (2017).

27 Lomberk, G. *et al.* Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. *Nature communications* **9**, 1978, doi:10.1038/s41467-018-04383-6 (2018).

28 Sausen, M. *et al.* Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. *Nature communications* **6**, 7686, doi:10.1038/ncomms8686 (2015).

29 Dawkins, J. B. *et al.* Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer research* **76**, 4861-4871, doi:10.1158/0008-5472.Can-16-0481 (2016).

30 Simone, C. G. *et al.* Characteristics and outcomes of adenosquamous carcinoma of the pancreas. *Gastrointestinal cancer research : GCR* **6**, 75-79 (2013).

31 Sun, Q. *et al.* Competition between human cells by entosis. *Cell research* **24**, 1299-1310, doi:10.1038/cr.2014.138 (2014).

32 de la Cova, C., Abril, M., Bellosta, P., Gallant, P. & Johnston, L. A. Drosophila myc regulates organ size by inducing cell competition. *Cell* **117**, 107-116 (2004).

33 Claveria, C., Giovinazzo, G., Sierra, R. & Torres, M. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500**, 39-44, doi:10.1038/nature12389 (2013).

34 Hamann, J. C. *et al.* Entosis Is Induced by Glucose Starvation. *Cell reports* **20**, 201-210, doi:10.1016/j.celrep.2017.06.037 (2017).

35 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

36 Engstrom, P. G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature methods* **10**, 1185-1191, doi:10.1038/nmeth.2722 (2013).

37 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

38 Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep*, doi:10.15252/embr.201846255 (2018).

39 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

40 Makohon-Moore, A. P. *et al.* Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nature genetics* **49**, 358-366, doi:10.1038/ng.3764 (2017).

41 Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033-1037, doi:10.1126/science.aat7171 (2018).

42 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

43 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498, doi:10.1038/ng.806 (2011).

44 Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. & Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinformatics (Oxford, England)* **30**, 2813-2815, doi:10.1093/bioinformatics/btu376 (2014).

45 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).

46 Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic acids research* **44**, e131, doi:10.1093/nar/gkw520 (2016).

47 Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 14330-14335, doi:10.1073/pnas.1616440113 (2016).

48 Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948-962, doi:10.1016/j.cell.2013.10.011 (2013).

49 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218, doi:10.1038/nature12213 (2013).

50 Reiter, J. G. *et al.* Reconstructing metastatic seeding patterns of human cancers. *Nature communications* **8**, 14114, doi:10.1038/ncomms14114 (2017).

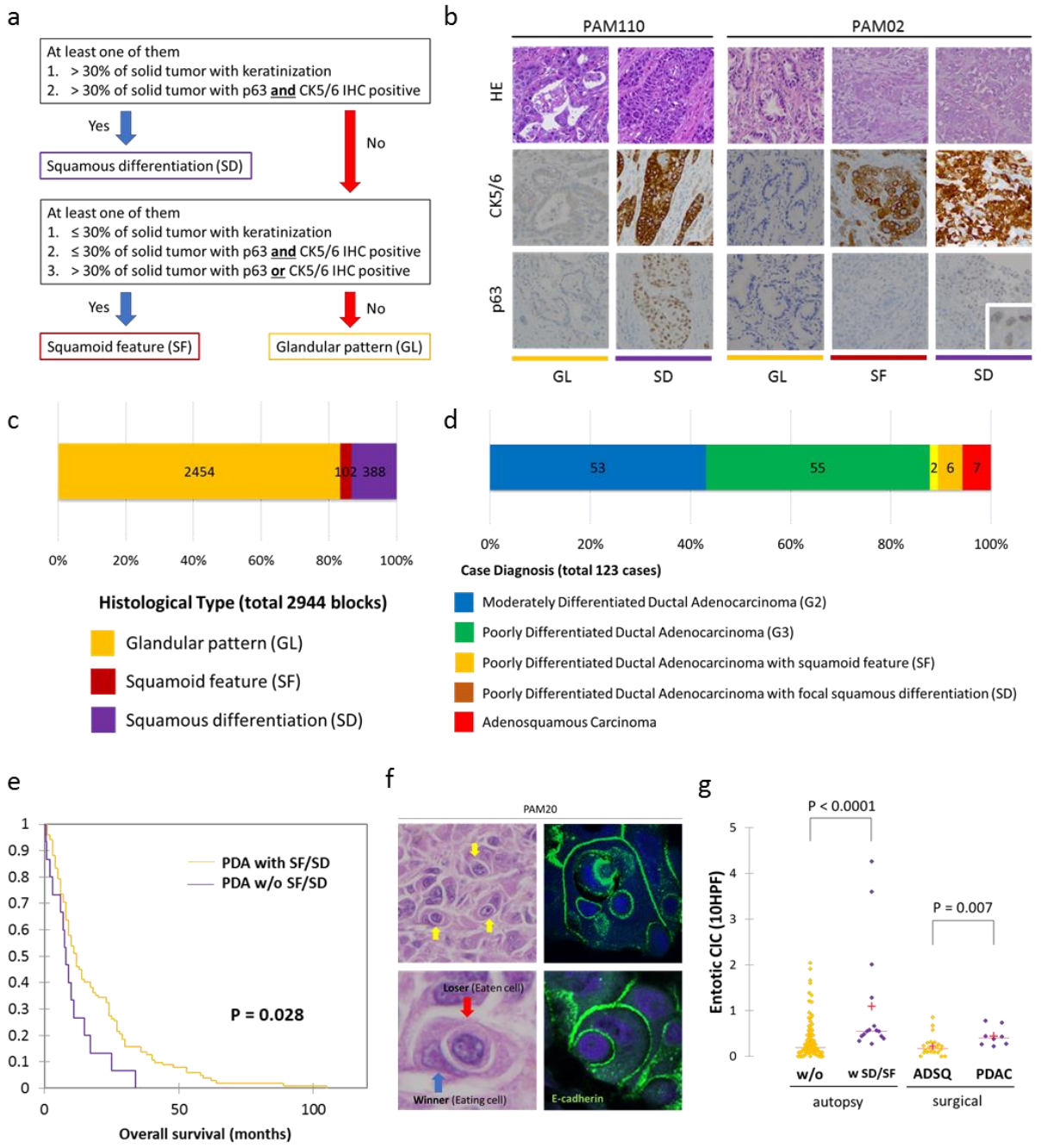


Figure 1

627

628

Figure 1. Morphologic Heterogeneity for Squamous Features in Pancreatic Ductal Adenocarcinoma. a. Schematic for classification of sections. Squamous differentiation or squamoid feature was determined for each block in all cases based on the combination of histomorphologic features and p63 and CK5/6 immunohistochemistry (IHC). **b.** Representative histomorphologic and immunocytochemical images of glandular pattern (GL), squamoid feature (SF) and squamous differentiation (SD) in patients PAM02 and PAM110. SD areas showed solid growth pattern with both CK5/6 and p63 positivity, while SF areas showed CK5/6 positive labeling but are negative for p63. **c. Summary of block diagnoses. d. Postmortem case diagnoses**. Seven cases corresponded to adenosquamous carcinoma (ASC), four cases showed focal (<30%) squamous differentiation (SD), and two cases showed focal squamoid features (SF). **e. Kaplan-Meier analysis of PDAC with or without squamous/squamoid features.** PDAC with SF/SD showed poorer prognosis than PDACs without SF/SD. **f. Representative histomorphologic and immunofluorescent images of entotic CIC in patient PAM20.** A clearly defined 'moonshape' host nucleus, intervening vacuolar space and internalized cell is identified. Immunofluorescent images clear e-cadherin membranous labeling of the winners (eating cells) and losers (eaten cells). **g. Average number of entotic cell in cell structures (entotic CIC) in PDAC with or without SF/SD.** Both autopsy cases (with or without SF/SD) and surgical cases (ADSQ or PDAC) have a higher level of entotic CICs.

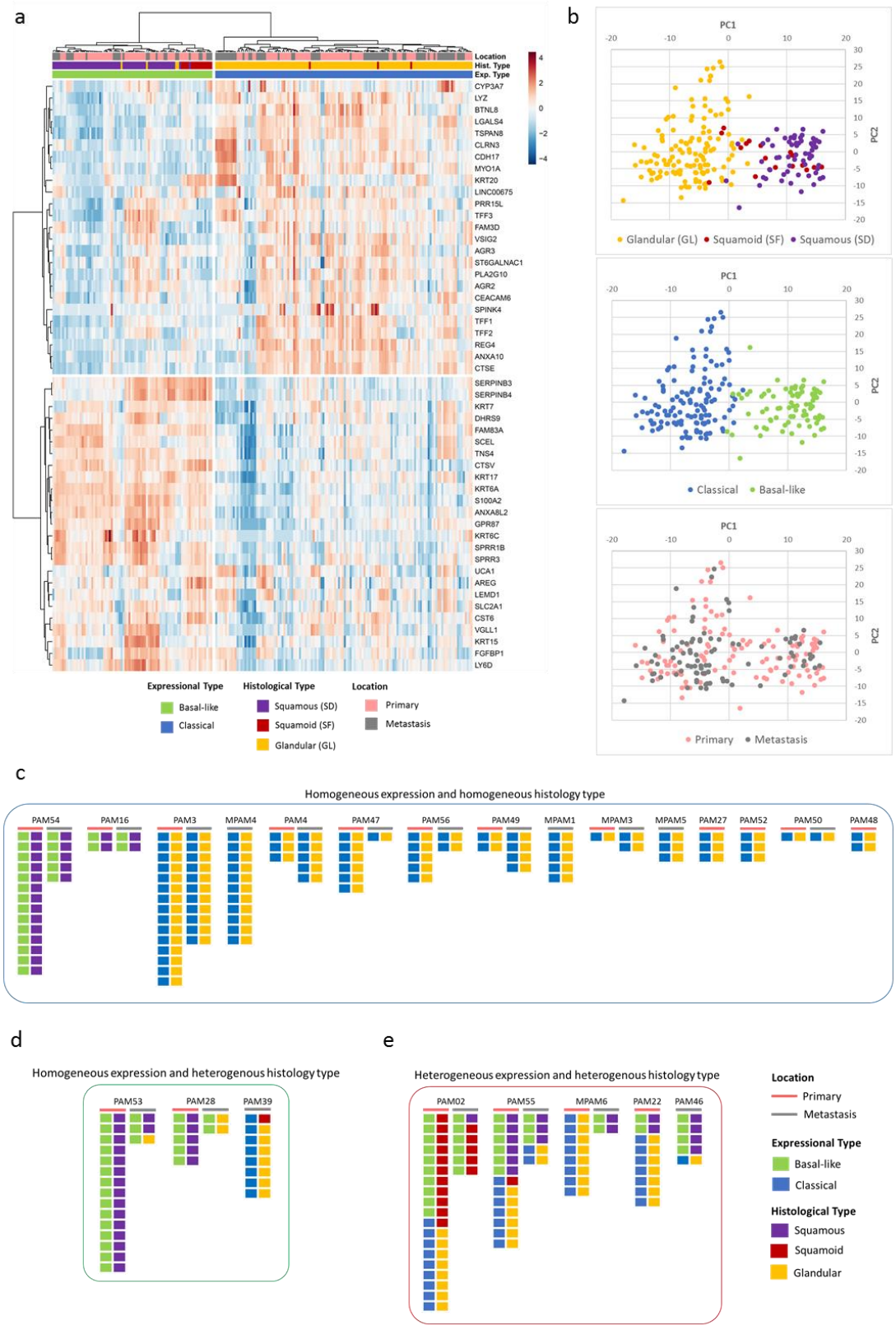

a
Location
Hist. Type
Exp. Type
CYP3A7
LYZ
BTNL8
LGALS4
TSPAN8
CLRN3
CDH17
MYO1A
KRT20
LINC00675
PRR15L
TFF3
FAM3D
VSIG2
AGR3
ST6GALNAC1
PLA2G10
AGR2
CEACAM6
SPINK4
TFF1
TFF2
REG4
ANXA10
CTSE
SERPINB3
SERPINB4
KRT7
DHRS9
FAM83A
SCEL
TNS4
CTSV
KRT17
KRT6A
S100A2
ANXA8L2
GPR87
KRT6C
SPRR1B
SPRR3
UCA1
AREG
LEMD1
SLC2A1
CST6
VGLL1
KRT15
FGFBP1
LY6D
Expressional Type
Basal-like
Classical
Histological Type
Squamous (SD)
Squamoid (SF)
Glandular (GL)
Location
Primary
Metastasis
b
PC1
PC2
Glandular (GL)
Squamoid (SF)
Squamous (SD)
Classical
Basal-like
Primary
Metastasis
c
Homogeneous expression and homogeneous histology type
PAM54
PAM16
PAM3
MPAM4
PAM4
PAM47
PAM56
PAM49
MPAM1
MPAM3
MPAM5
PAM27
PAM52
PAM50
PAM48
d
Homogeneous expression and heterogenous histology type
PAM53
PAM28
PAM39
e
Heterogeneous expression and heterogenous histology type
PAM02
PAM55
MPAM6
PAM22
PAM46
Location
Primary
Metastasis
Expressional Type
Basal-like
Classical
Histological Type
Squamous
Squamoid
Glandular

Figure 2

648

Figure 2. Transcriptional Heterogeneity for Squamous Features in Pancreatic Ductal Adenocarcinoma. RNA sequencing (RNA-seq) of was performed on snap-frozen tissues of 214 unique samples from 27 patients including five with ASC and five with focal SF/SD. RNA-seq data were used to classify each of the 214 samples into “Basal-like” and “Classical” tumors (Moffitt et al. Nature Genetics, 2015). Both the heatmap (**a**) and PCA plots (**b**) indicate a strong correlation of SF/SD morphology with a “Basal-like” transcriptional signature, and GL morphology with a “Classical” transcriptional signature. **c-e.** Integrated analysis of transcriptional subtype with unique block diagnosis indicates intratumoral heterogeneity for both transcriptional signatures and histomorphologic features.

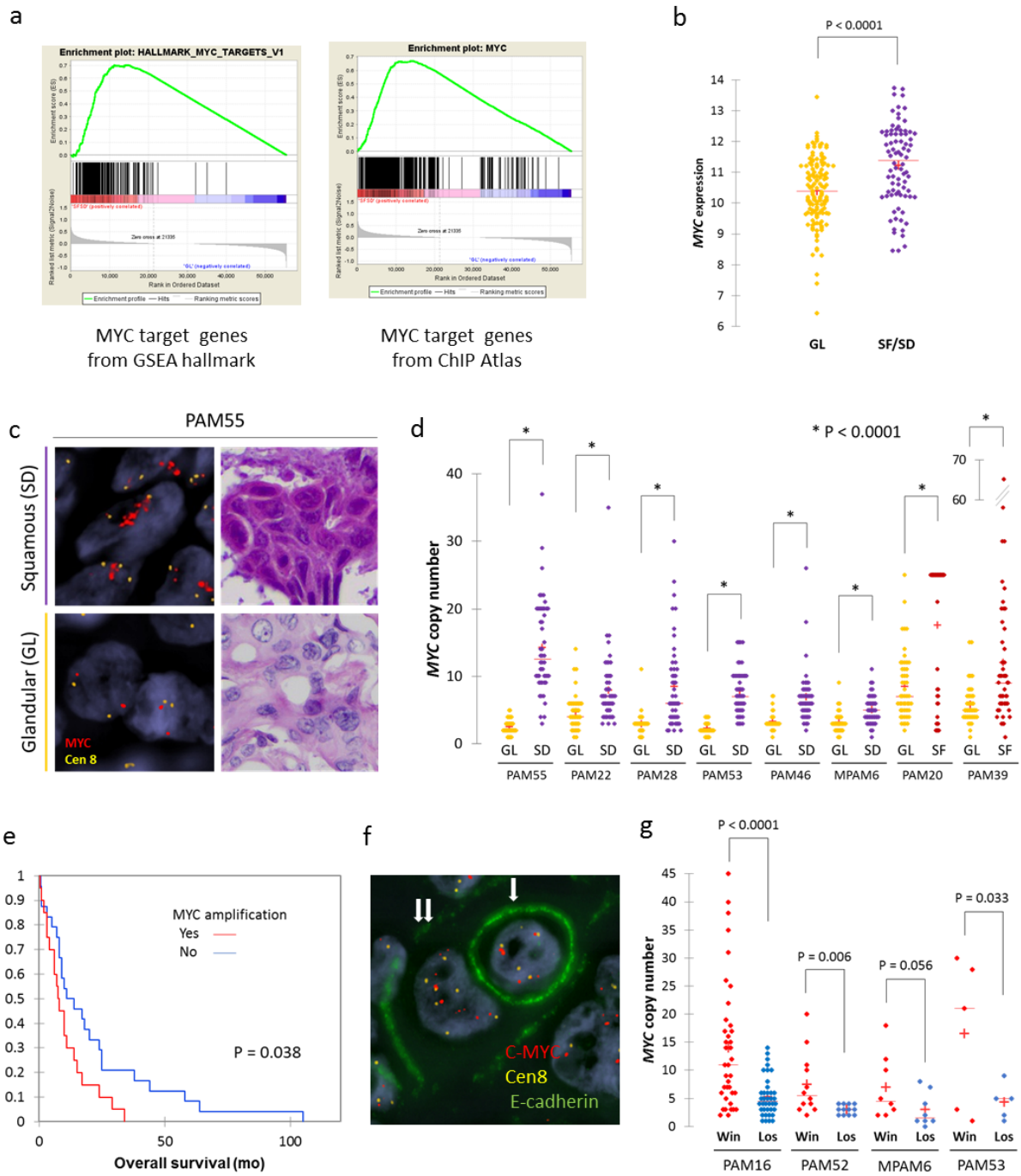

a
Enrichment plot: HALLMARK_MYC_TARGETS_V1
Enrichment plot: MYC
MYC target genes
from GSEA hallmark
MYC target genes
from ChIP Atlas
b
P < 0.0001
MYC expression
GL
SF/SD
c
PAM55
Squamous (SD)
Glandular (GL)
MYC
Cen 8
d
* P < 0.0001
MYC copy number
GL SD GL SD GL SD GL SD GL SD GL SD GL SF GL SF
PAM55 PAM22 PAM28 PAM53 PAM46 MPAM6 PAM20 PAM39
e
MYC amplification
Yes
No
P = 0.038
Overall survival (mo)
f
C-MYC
Cen8
E-cadherin
g
P < 0.0001
P = 0.006
P = 0.056
P = 0.033
MYC copy number
Win Los Win Los Win Los Win Los
PAM16 PAM52 MPAM6 PAM53

Figure 3

658

Figure 3. Squamous Features in Pancreatic Ductal Adenocarcinoma Correspond to Enhancement of MYC. a. Gene-set enrichment analysis (GSEA) using hallmark gene sets and transcription factor target gene sets collected from ChIP-Atlas identify MYC target genes as the top ranked gene set in SF/SD (see also Supplemental Tables 6 and 7). **b.** Normalized MYC RNA transcript abundance is significantly higher in SF/SD samples than in GL samples. **c.** Representative images of MYC-FISH analysis in SF/SD and GL regions. **d.** Analysis of *MYC* copy number in eight cases indicates that *MYC* is significantly amplified in SF/SD regions compared to GL regions in the same carcinoma. **e.** Kaplan-Meier analysis indicating patients whose carcinomas have *MYC* high (>= 6) copy number have a worse outcome than carcinomas with low *MYC* copy number. **f.** Representative images of entosis (single arrow: loser (eaten cell), double arrows: winner (eating cell)). **g.** Winner cells have higher *MYC* copy number than loser cells.

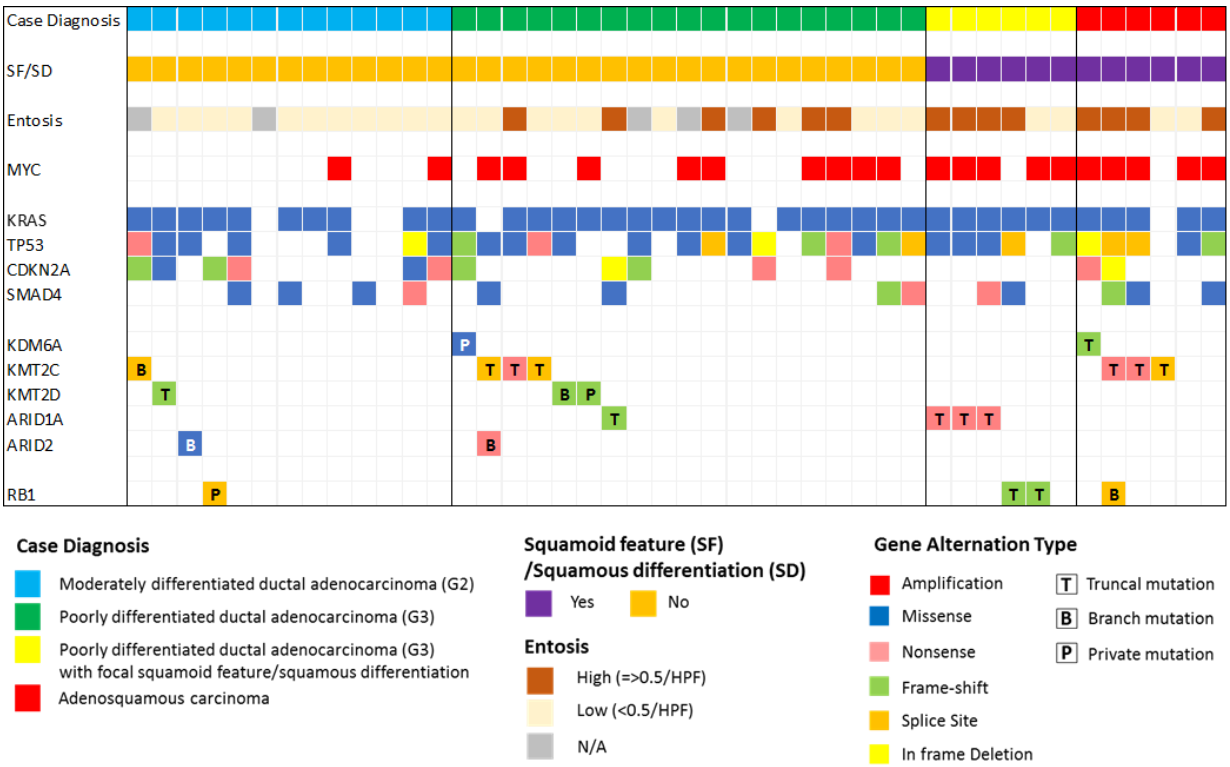


Figure 4

Figure 4. Genomic Landscape of End Stage Pancreatic Ductal Adenocarcinomas with and without Squamous Features. Oncoprint illustrating the driver gene somatic alterations of 44 cases with respect to their histologic and immunolabeling profiles. Truncal mutations in chromatin modifier genes are significantly enriched in PDACs with focal SF/SD and ASCs, whereas MYC amplification is seen in both high grade PDACs (G3) and in PDACs with SF/SD.

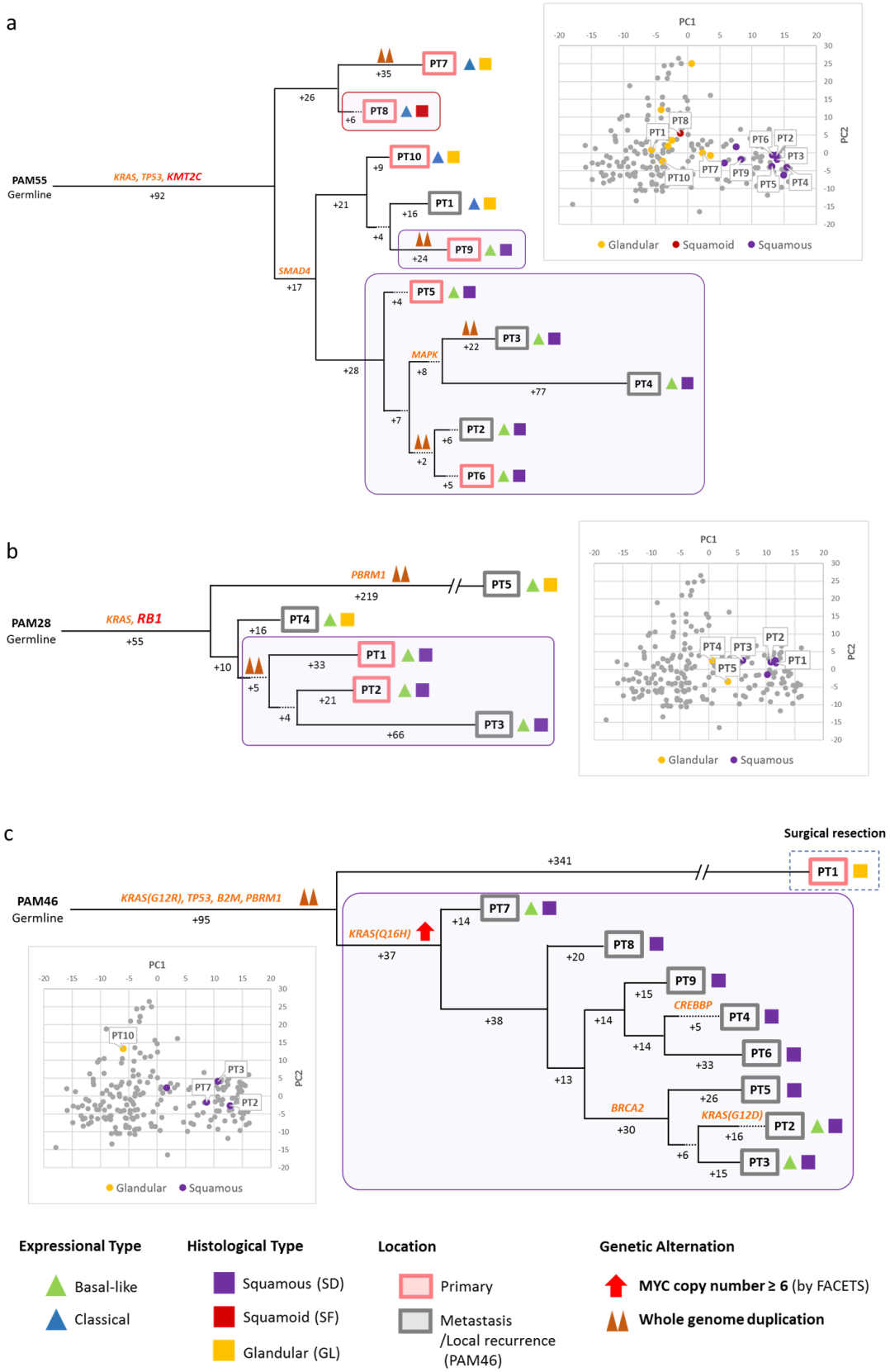

a
PAM55
Germline
KRAS, TP53, KMT2C
+92
+26
+35
PT7
+6
PT8
+9
PT10
+21
+16
PT1
+4
+24
PT9
SMAD4
+17
+28
+4
PT5
+22
PT3
MAPK
+8
+77
PT4
+7
+6
PT2
+2
+5
PT6
PC1
PC2
Glandular
Squamoid
Squamous
b
PAM28
Germline
KRAS, RB1
+55
PBRM1
+219
PT5
+16
PT4
+10
+33
PT1
+5
+21
PT2
+4
+66
PT3
PC1
PC2
Glandular
Squamous
c
Surgical resection
PAM46
Germline
KRAS(G12R), TP53, B2M, PBRM1
+95
+341
PT1
KRAS(Q16H)
+37
+14
PT7
+38
+20
PT8
+15
PT9
CREBBP
+5
PT4
+14
+14
PT6
+33
+13
+26
PT5
BRCA2
+30
KRAS(G12D)
+16
PT2
+6
+15
PT3
PC1
PC2
PT10
PT3
PT7
PT2
Glandular
Squamous
Expressional Type
Basal-like
Classical
Histological Type
Squamous (SD)
Squamoid (SF)
Glandular (GL)
Location
Primary
Metastasis
/Local recurrence
(PAM46)
Genetic Alternation
MYC copy number ≥ 6 (by FACETS)
Whole genome duplication

Figure 5

678

Figure 5. Integration of Transcriptomic and Morphologic Features with Phylogenetic Patterns in Pancreatic Ductal Adenocarcinoma. Red and purple outlines indicate samples that have SF/SD based on RNAseq (triangles) and/or histology (squares). The predicted timing of somatic alterations in driver genes, whole genome duplication and MYC amplification are also shown. Mutations in chromatin modifier genes are in red font, all others in orange. Principal Components Analysis plots are shown for each patient as well. **a.** Phylogenetic tree of patient PAM55. Truncal driver genes are notable for a *KMT2C* somatic alteration. SF/SD in this carcinoma have arisen as three independent subclones: in primary tumor sample PT8, in primary tumor sample PT9, and in the subclone giving rise to primary tumor samples PT5 and PT6 and metastases PT2-PT4. **b**. Phylogenetic tree of patient PAM28. Truncal driver genes include a deleterious mutation in *RB1*. SF/SD in this carcinoma have occurred due to one subclone that gave rise to the primary tumor samples PT1 and PT2 and the metastasis PT3. **c**. Phylogenetic tree of patient PAM46. No mutations in chromatin modifier genes were identified. *MYC* amplification (>6 copies) was detected in all samples of the local recurrence but not the original resected primary tumor PT1.

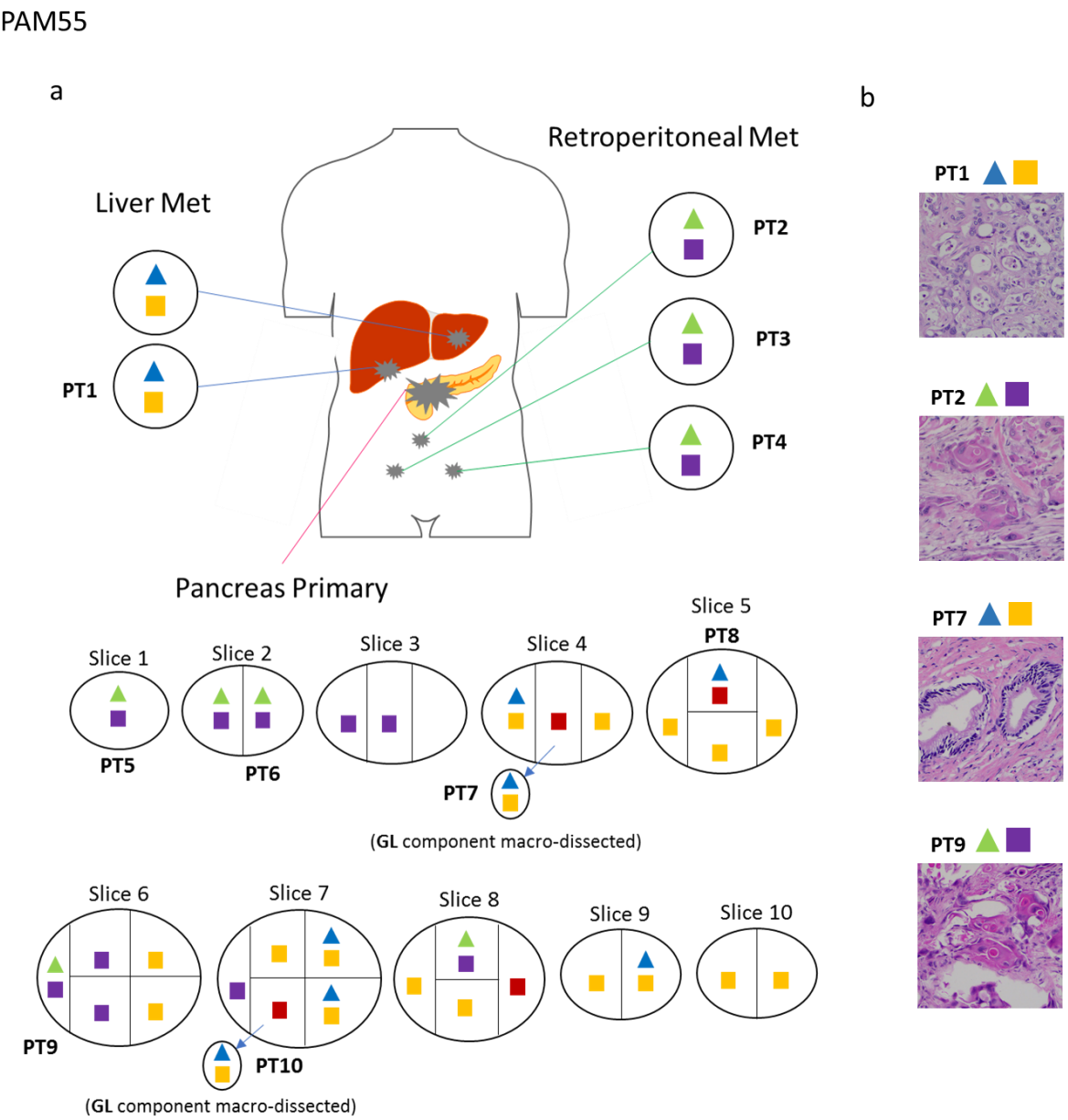
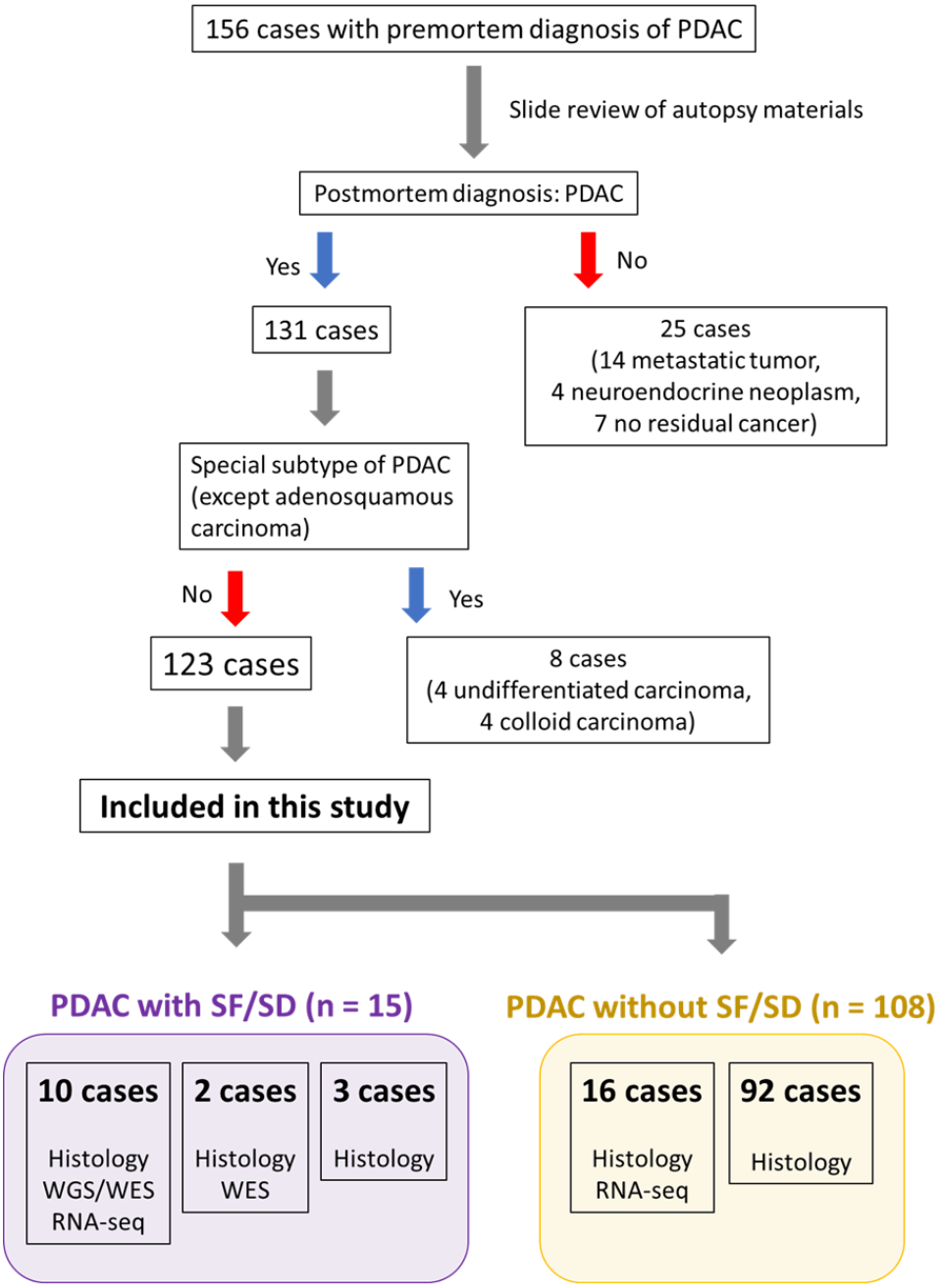


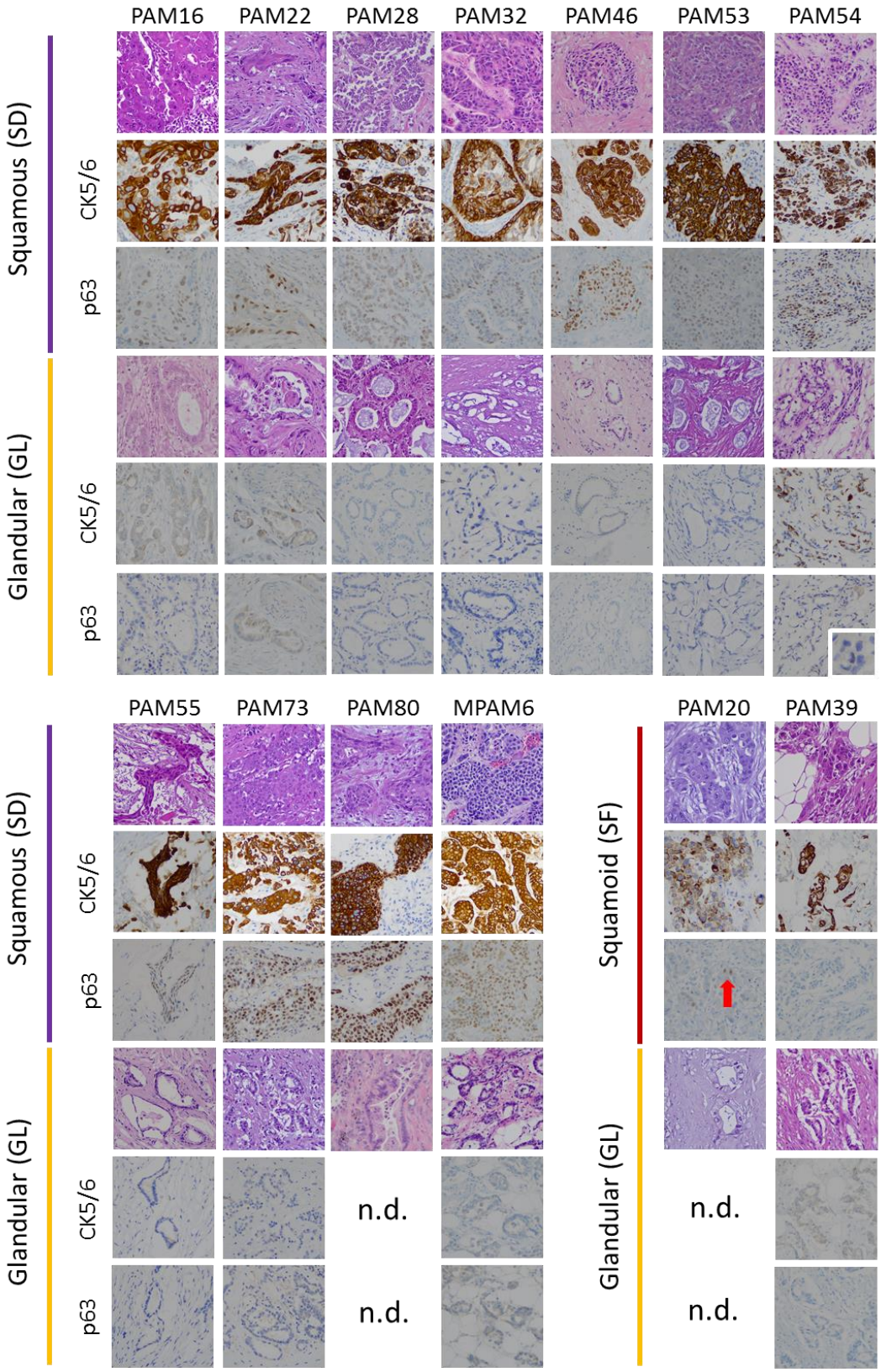
Figure 6

Figure 6. Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity. (a) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (b) Representative histologic images of representative tumors in the same patient.



Ex. Data Figure 1

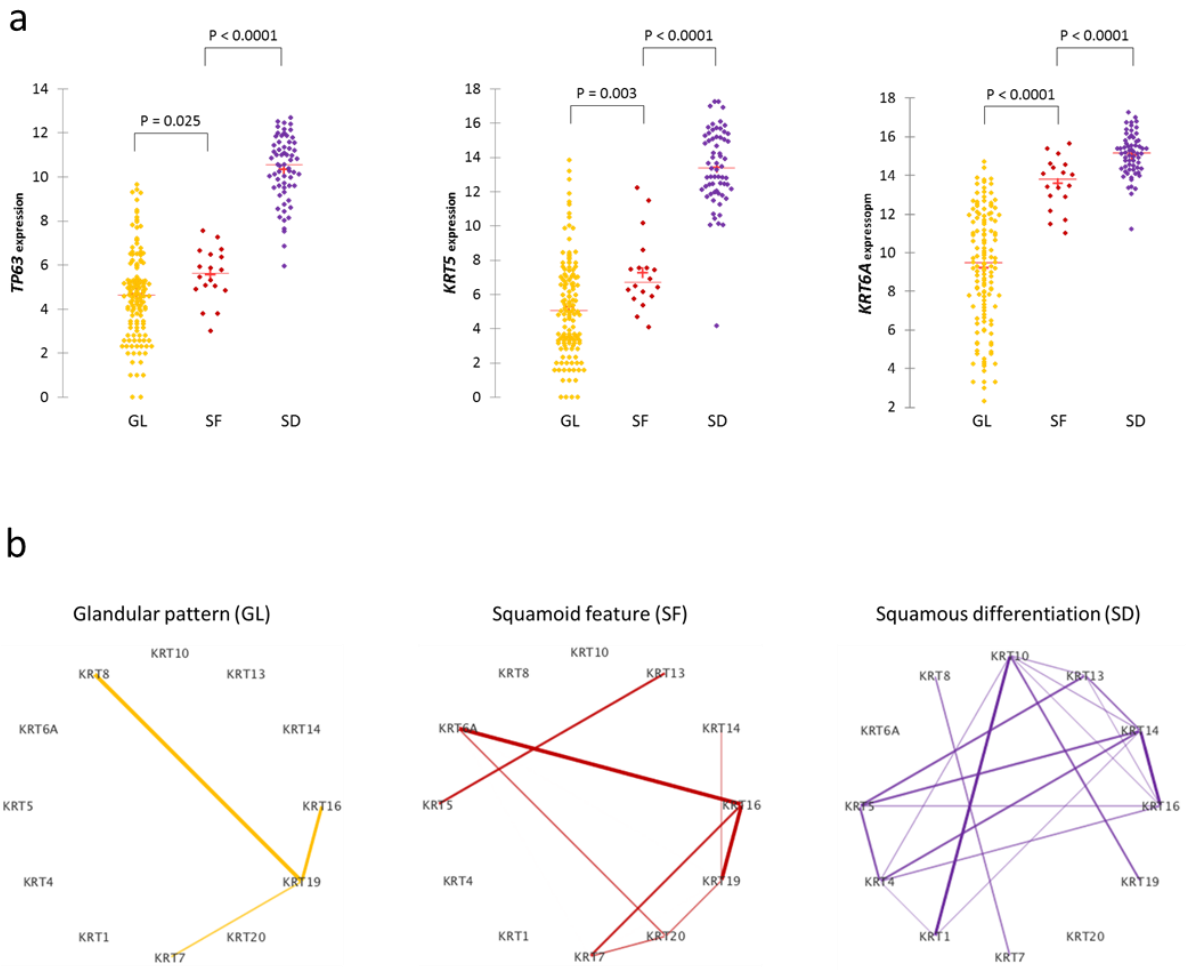
Extended Data Figure 1: Schematic of case selection for current study.



Ex. Data Figure 2

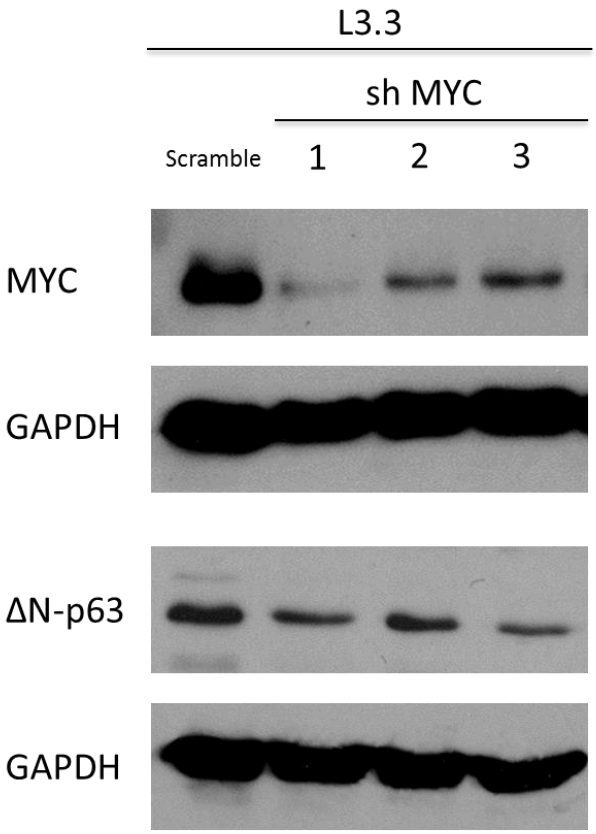
701

Extended Data Figure 2: Immunolabelling for glandular and squamous components in 13 Representative PDACs. All regions with squamous differentiation (SD) showed positivity for CK5/6 and p63, whereas no labeling was observed in regions with glandular morphology (GL). In two PDACs the neoplastic cells stained positive for CK5/6 but were negative for p63 and thus classified as having squamoid features. n.d., no immunolabeling performed.



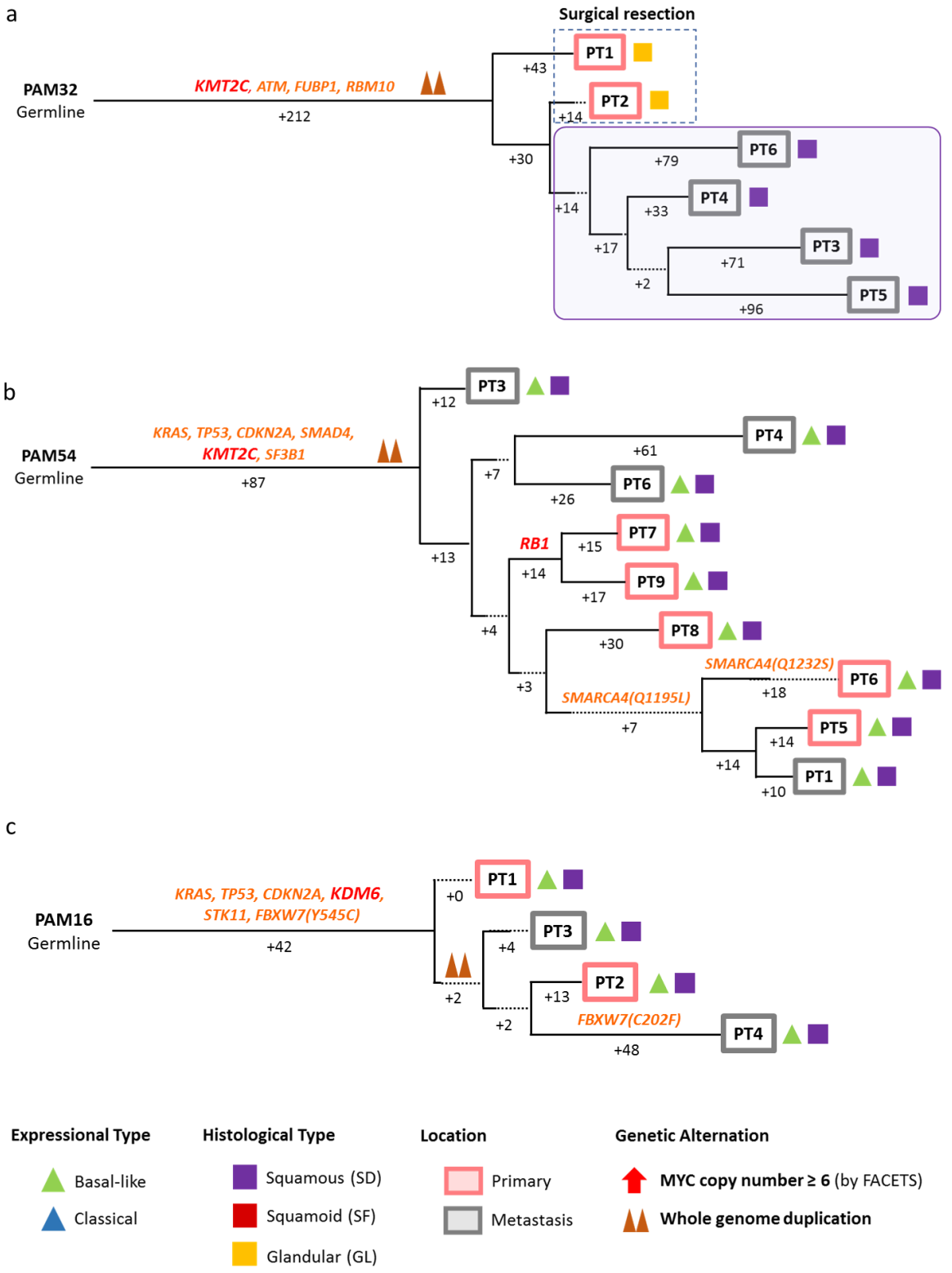
Ex. Data Figure 3

Extended Data Figure 3: mRNA Expression of squamous markers in samples with glandular growth pattern (GL), squamoid features (SF) and squamous differentiation (SD). a. mRNA expression of *TP63, KRT5* and *KRT6A*. SD have higher expression *of TP63, KRT5* and *KRT6A* than GL. SF have intermediate expression pattern between SD and GL. **b. KRT network based on mRNA expression.** In GL, KRT19 (normally expressed in ductal epithelia) is a hub in pancreas cancer. In SF, KRT6A and KTR5 (normally expressed in squamous epithelium) have some interaction. In SD, stratified squamous epithelium keratins (KRT4, KRT5, KRT14, KRT15) and heavy weight keratins (KRT1 and KRT10) were expressed in the network.



Ex. Data Figure 4

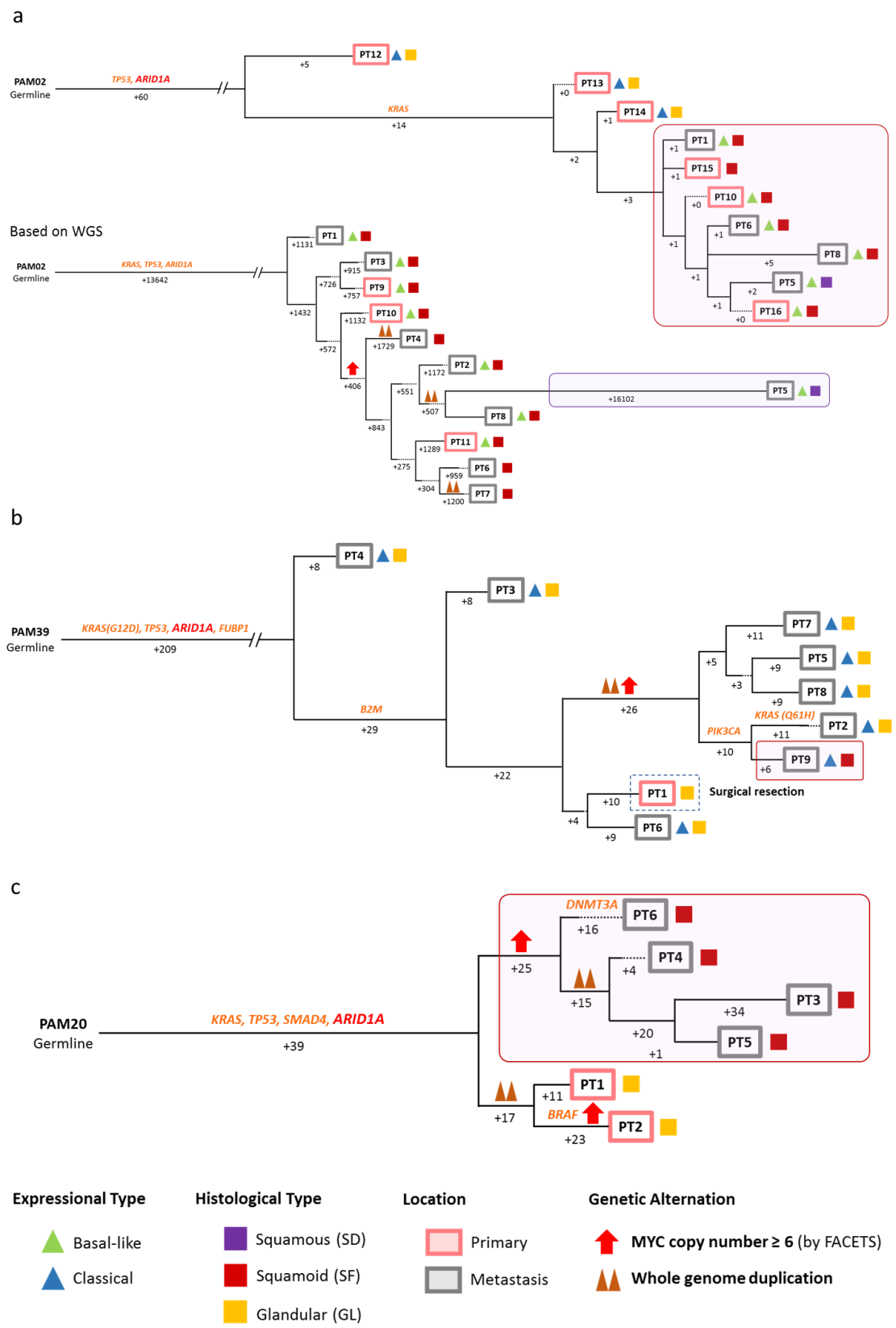
Extended Data Figure 4: Knockdown of MYC (c-MYC) protein using three independent siRNAs shows downregulation of ΔN-p63 protein (L3.3 cell line).



Ex. Data Figure 5

721

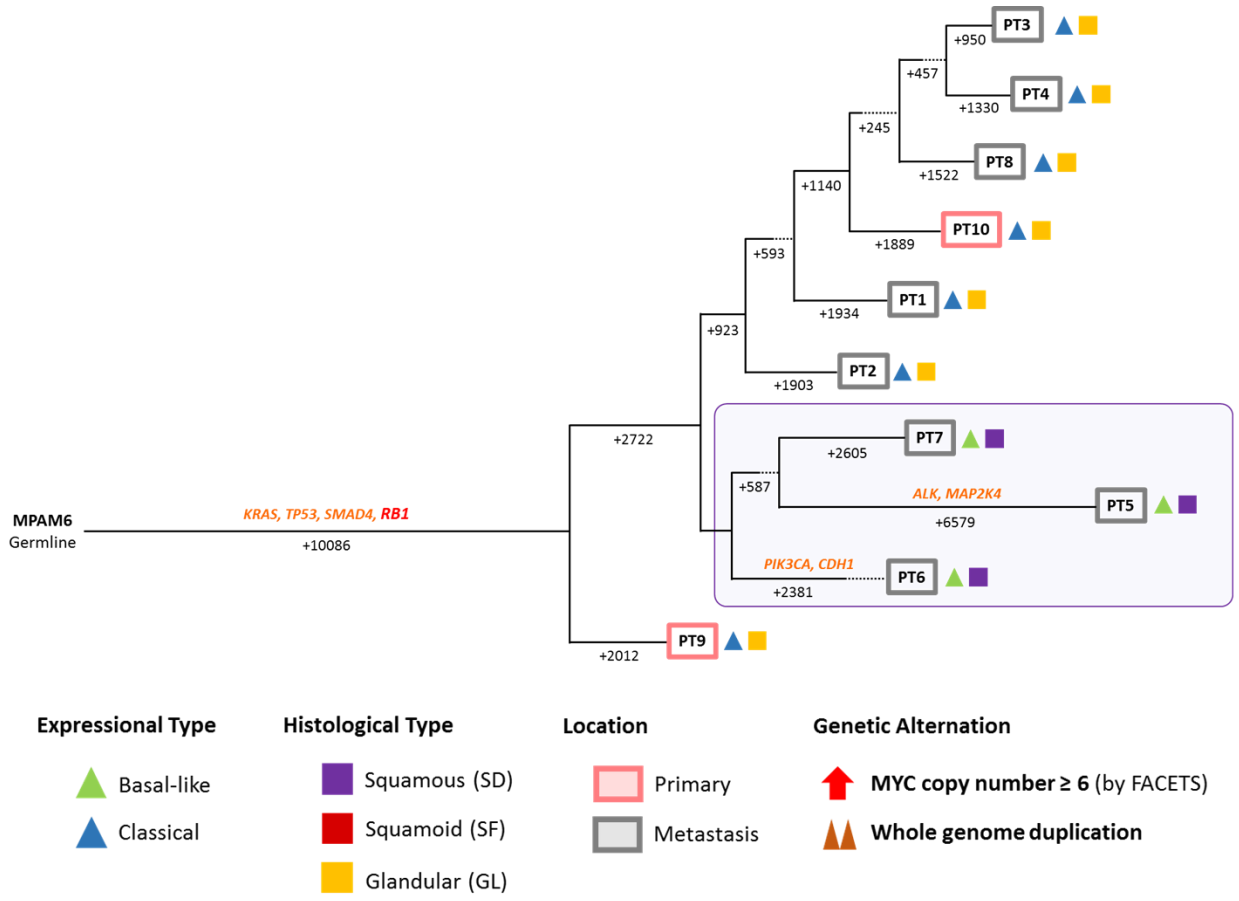
Extended Data Figure 5: Integration of Transcriptomic and Morphologic Features with Phylogenetic Patterns in Pancreatic Ductal Adenocarcinomas with *KMT2C* and *KDM6A* mutant cases. Red and purple outlines indicate samples that have SF/SD based on RNAseq (triangles) and/or histology (squares). The predicted timing of somatic alterations in driver genes on the evolutionary tree, the timing of whole genome duplication, and the number of mutations per branch are also shown. Mutations in chromatin modifier genes are in red font, all others in orange. **a.** Phylogenetic tree of patient PAM32. Truncal driver genes are notable for a *KMT2C* somatic alteration. **b.** Phylogenetic tree of patient PAM54. Truncal driver genes are notable for a *KMT2C* somatic alteration. **c.** Phylogenetic tree of patient PAM16. Truncal driver genes are notable for a *KDM6A* somatic alteration.



Ex. Data Figure 6

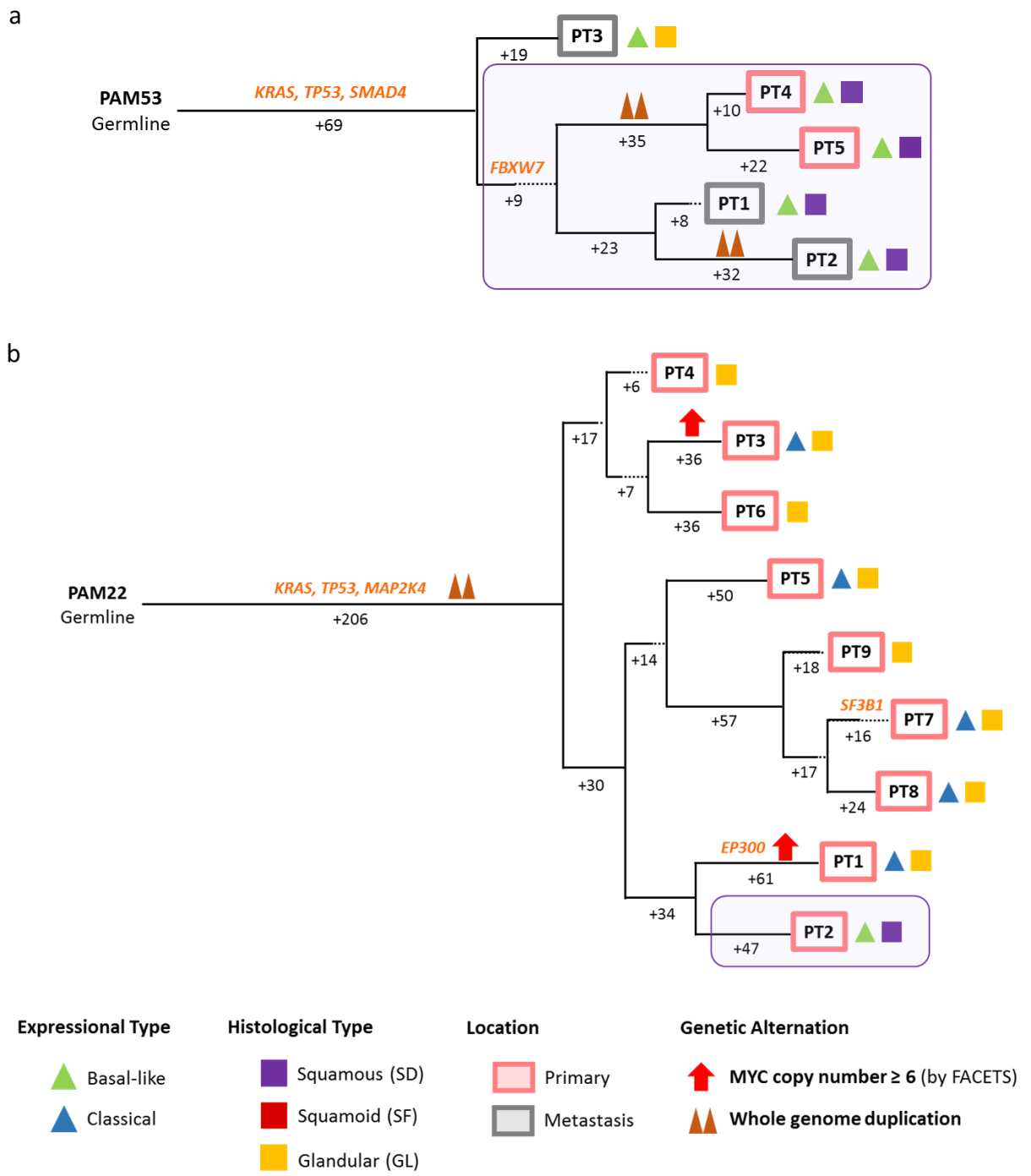
732

Extended Data Figure 6: Integration of Transcriptomic and Morphologic Features with Phylogenetic Patterns in Pancreatic Ductal Adenocarcinoma with *ARID1A* mutant cases. Red and purple outlines indicate samples that have SF/SD based on RNAseq (triangles) and/or histology (squares). The predicted timing of somatic alterations in driver genes on the evolutionary tree, the timing of whole genome duplication, and the number of mutations per branch are also shown. Mutations in chromatin modifier genes are in red font, all others in orange. Phylogenetic tree of patient PAM02 (a), PAM39 (b) and PAM20 (c). Truncal driver genes are notable for a *ARID1A* somatic alterations.



Ex. Data Figure 7

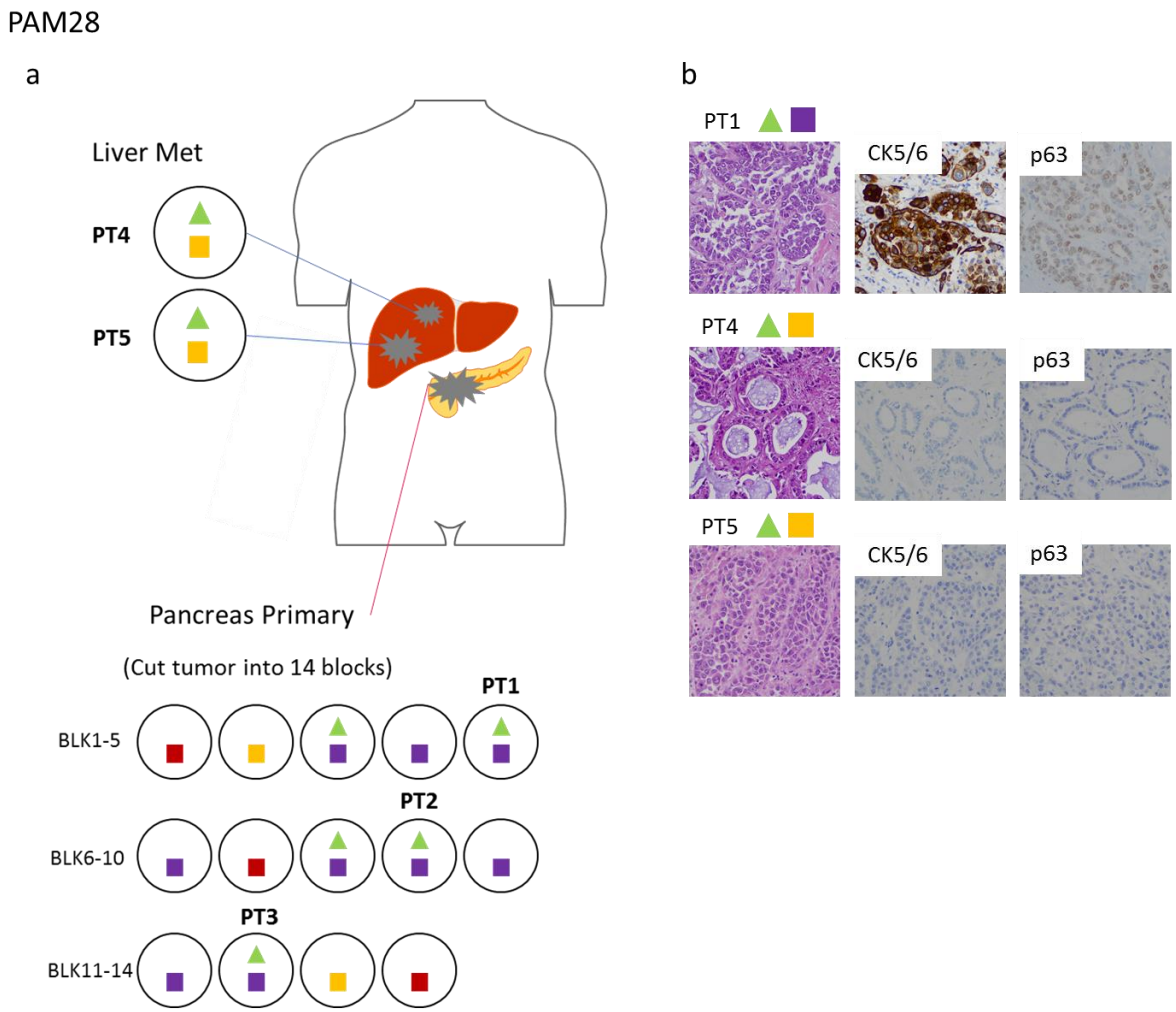
Extended Data Figure 7: Integration of Transcriptomic and Morphologic Features with Phylogenetic Patterns in Pancreatic Ductal Adenocarcinoma with *RB1* mutation. Red and purple outlines indicate samples that have SF/SD based on RNAseq (triangles) and/or histology (squares). The predicted timing of somatic alterations in driver genes on the evolutionary tree, the timing of whole genome duplication, and the number of mutations per branch, are also shown. Mutations in chromatin modifier genes are in red font, all others in orange. Truncal driver genes are notable for a *RB1* somatic alteration.



Ex. Data Figure 8

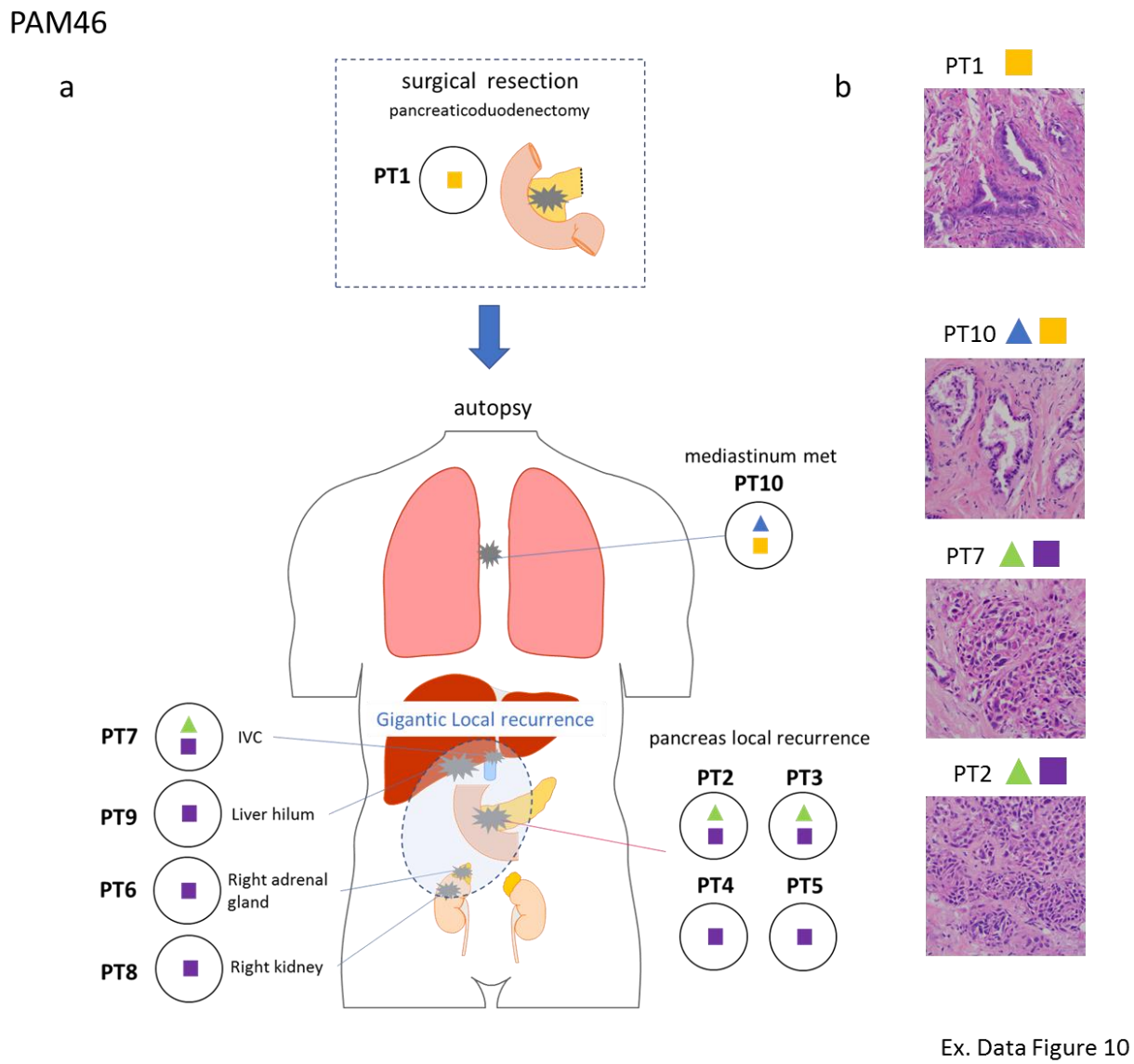
750

Extended Data Figure 8: Integration of Transcriptomic and Morphologic Features with Phylogenetic Patterns in Pancreatic Ductal Adenocarcinoma without Detectable Mutations in Chromatin Modifier Genes. Red and purple outlines indicate samples that have SF/SD based on RNAseq (triangles) and/or histology (squares). The predicted timing of somatic alterations in driver genes on the evolutionary tree, whole genome duplication, MYC amplification, and the number of mutations per branch are also shown. Mutations in chromatin modifier genes are in red font, all others in orange. Phylogenetic tree of patient PAM53 (a) and PAM22 (b).

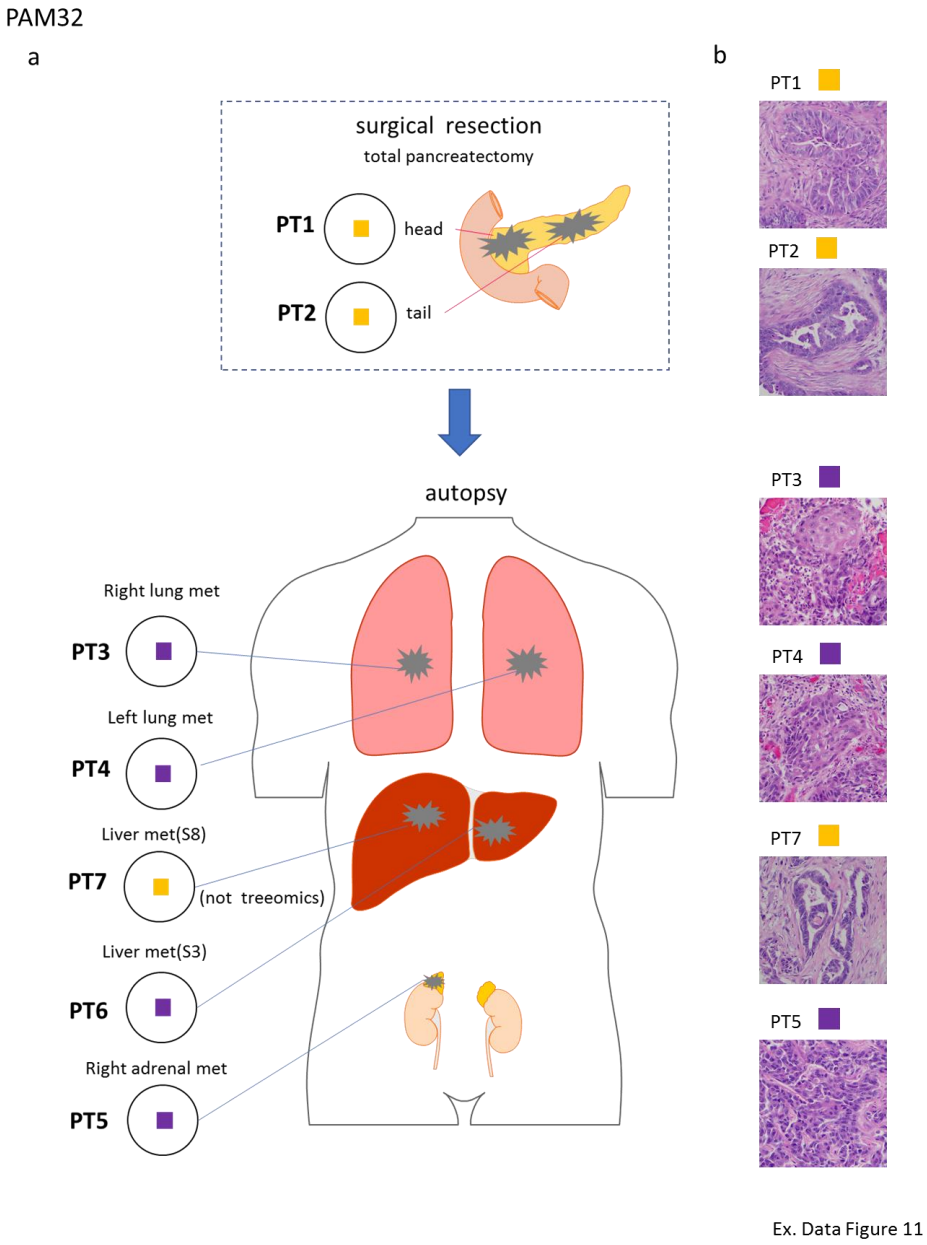


Ex. Data Figure 9

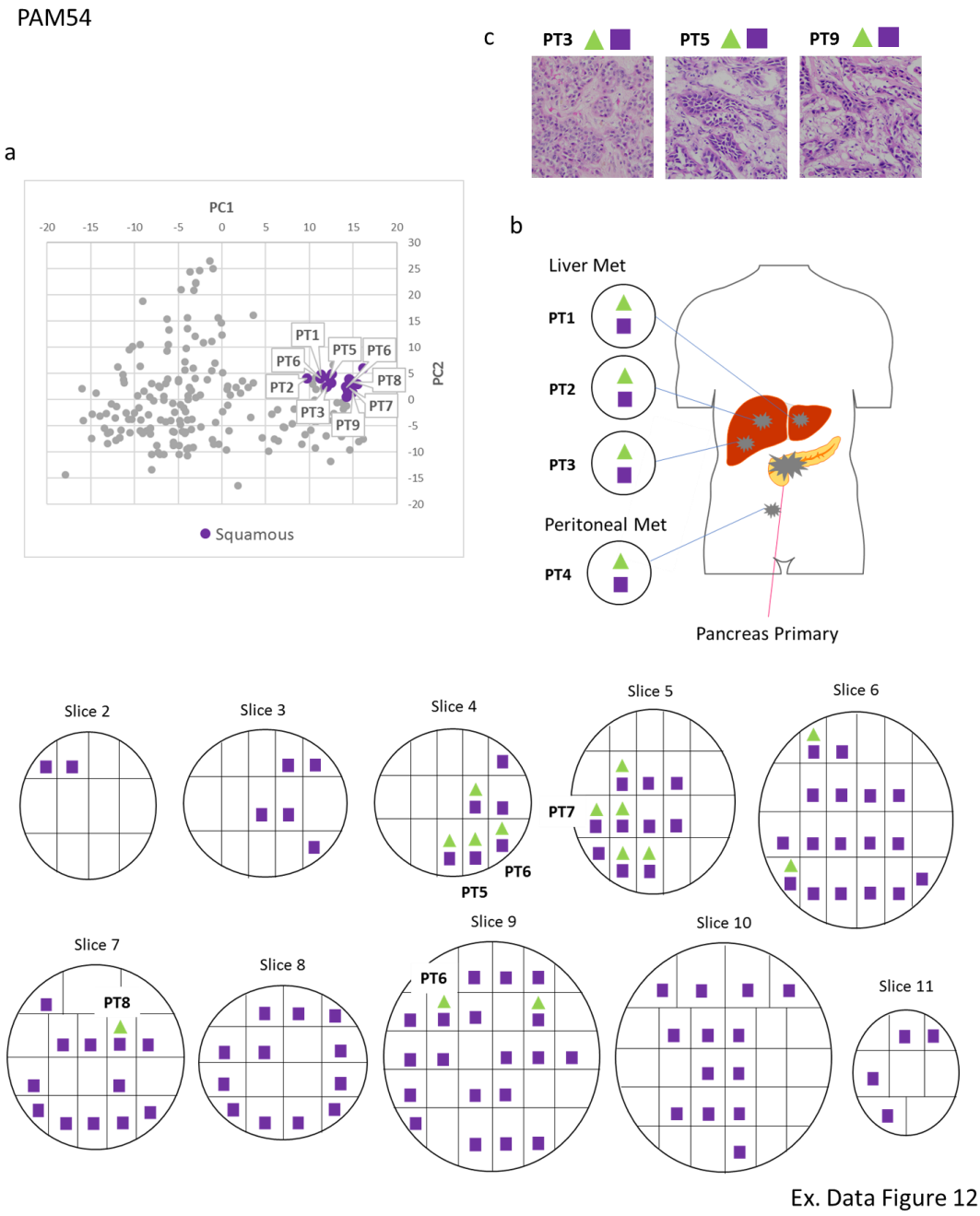
Extended Data Figure 9: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM28. (a) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (b) Representative histologic images and immunohistochemical labeling of representative tumors in the same patient.



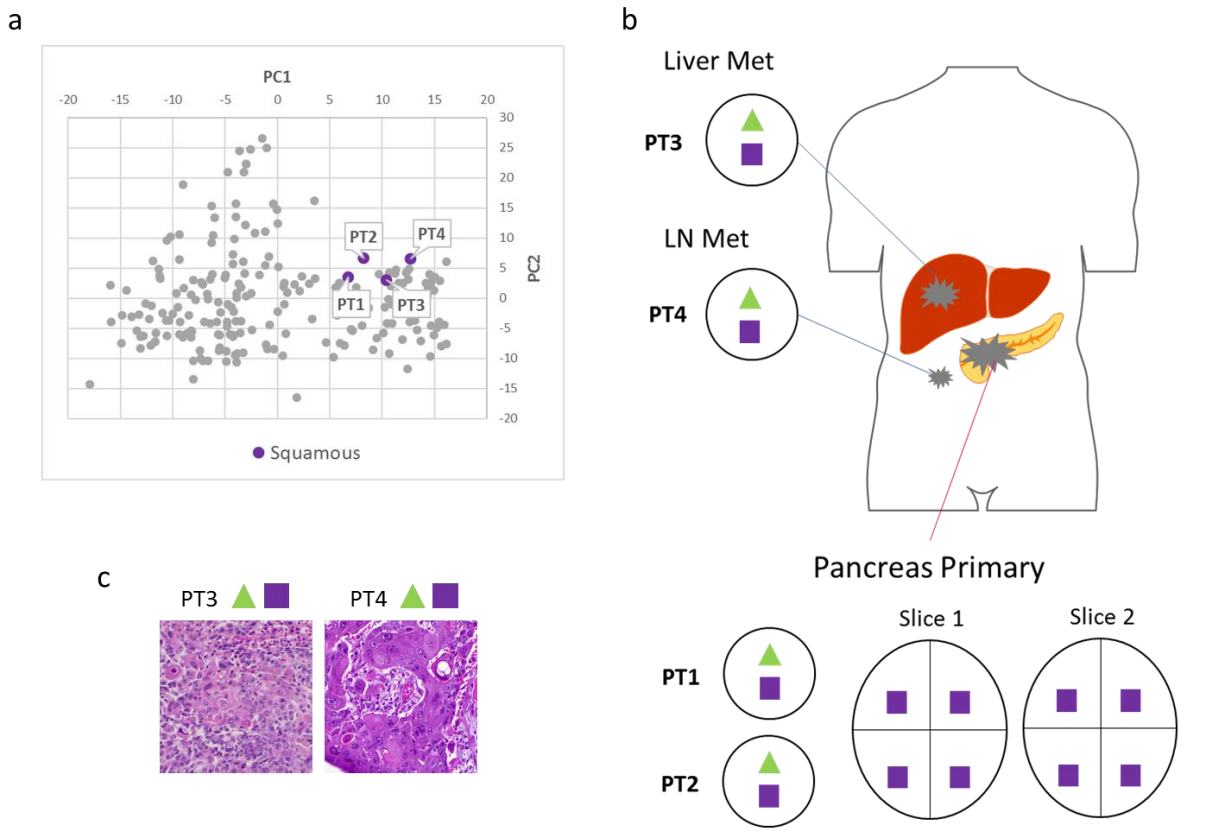
Extended Data Figure 10: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM46. (a) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (b) Representative histologic images of representative tumors in the same patient.



Extended Data Figure 11: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM32. (a) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (b) Representative histologic images and immunohistochemical labeling of representative tumors in the same patient.

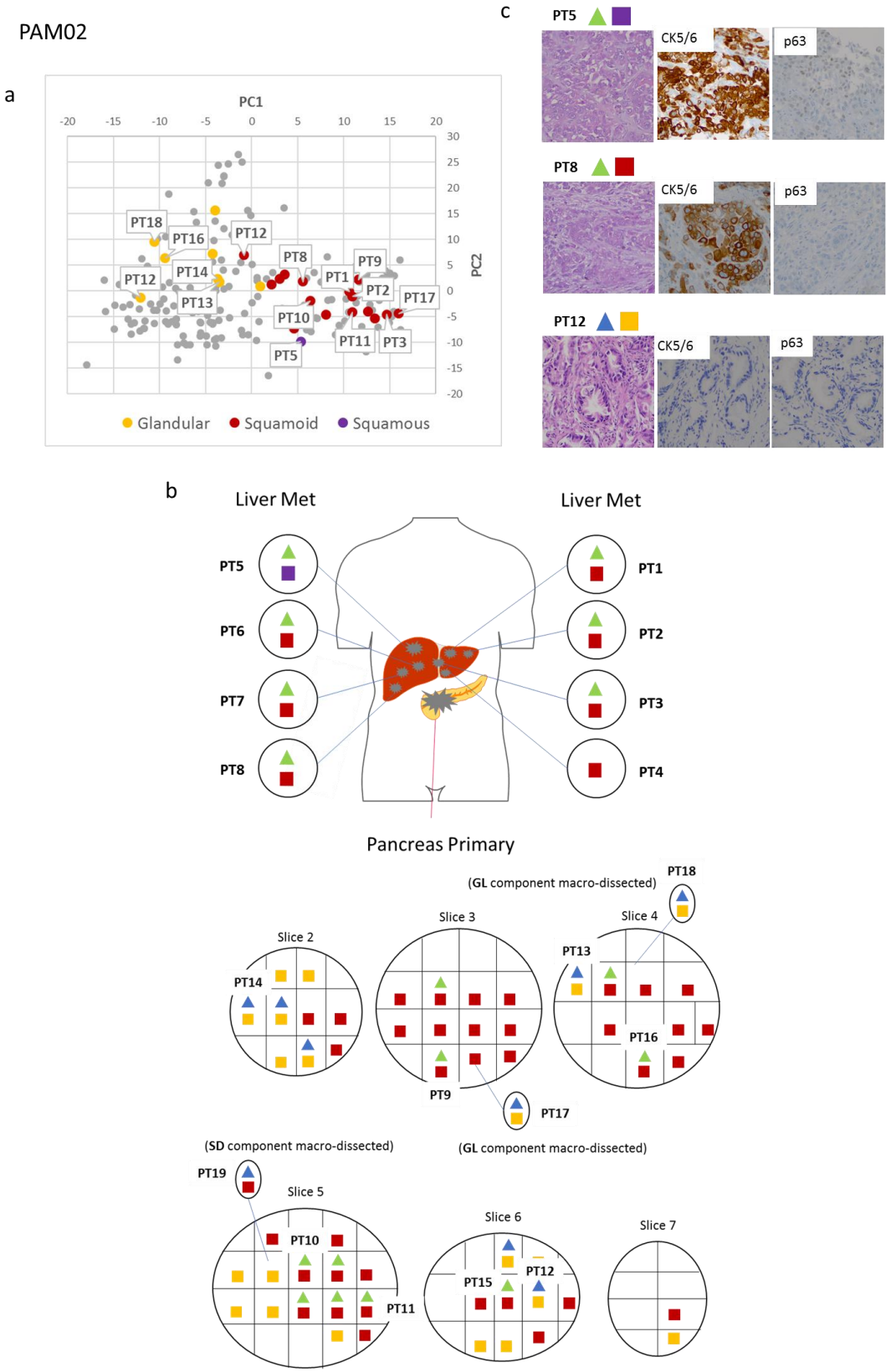


Extended Data Figure 12: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM54. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of representative tumors in the same patient.



Ex Data Figure 13

Extended Data Figure 13: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM16. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of representative tumors in the same patient.

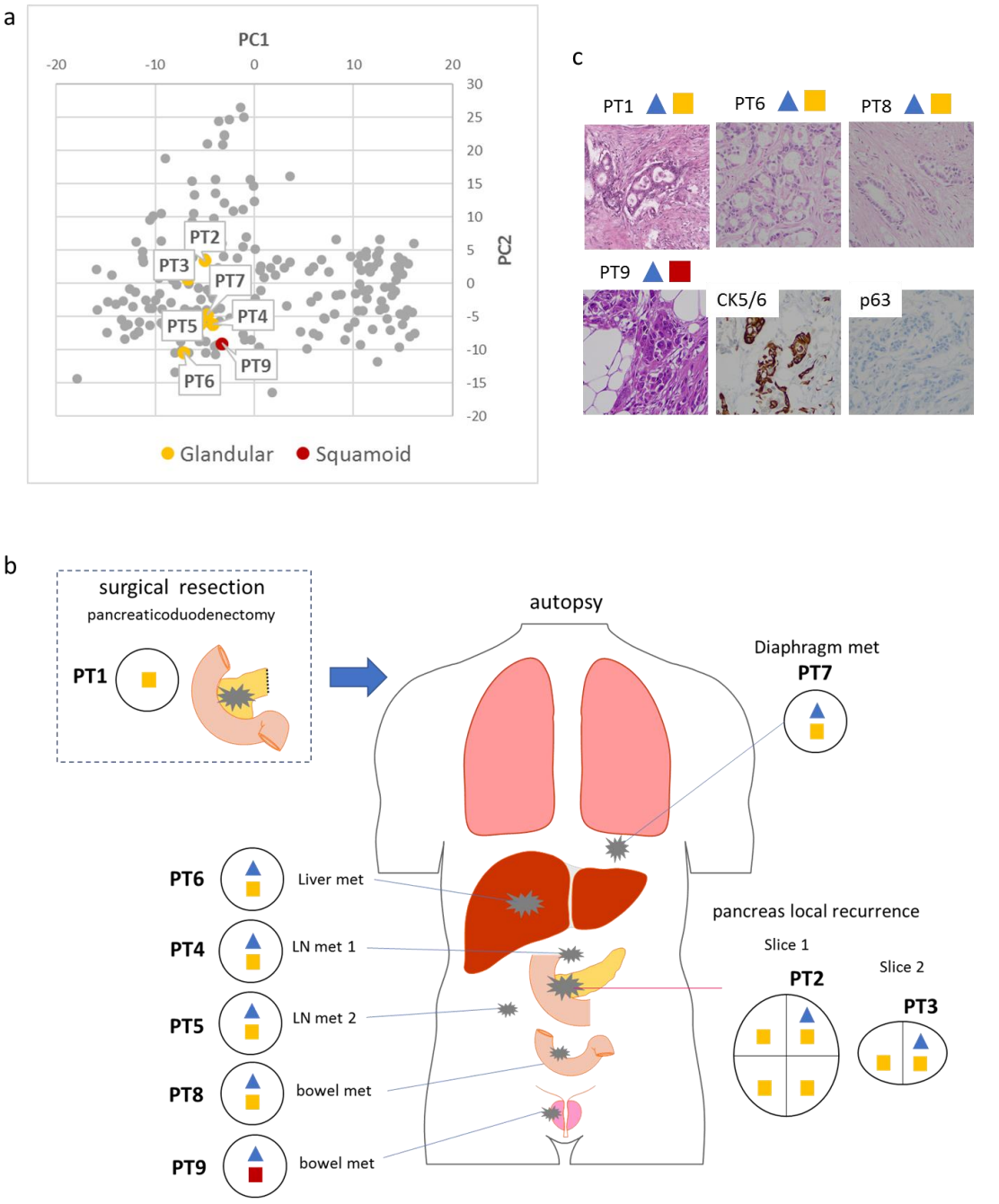


Ex Data Figure 14

792

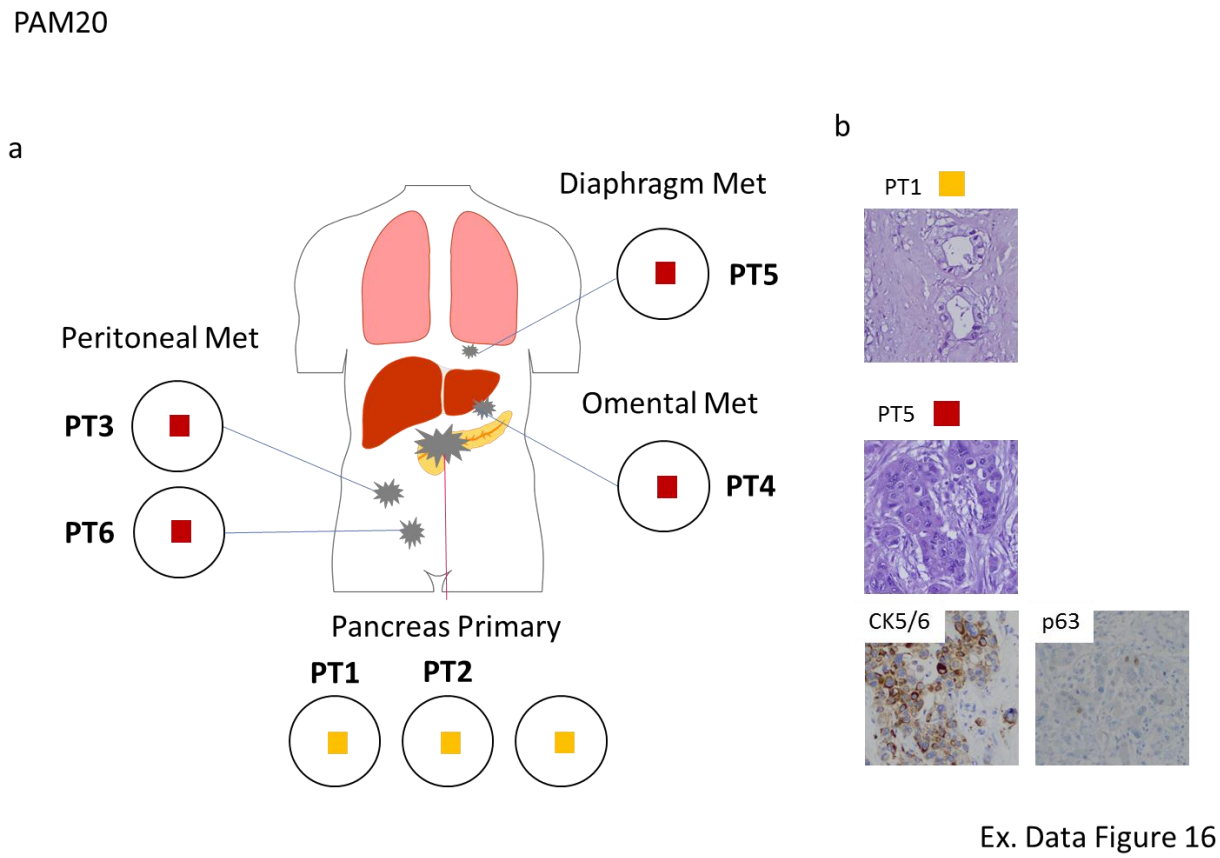
Extended Data Figure 14: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM02. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of representative tumors in the same patient.

PAM39



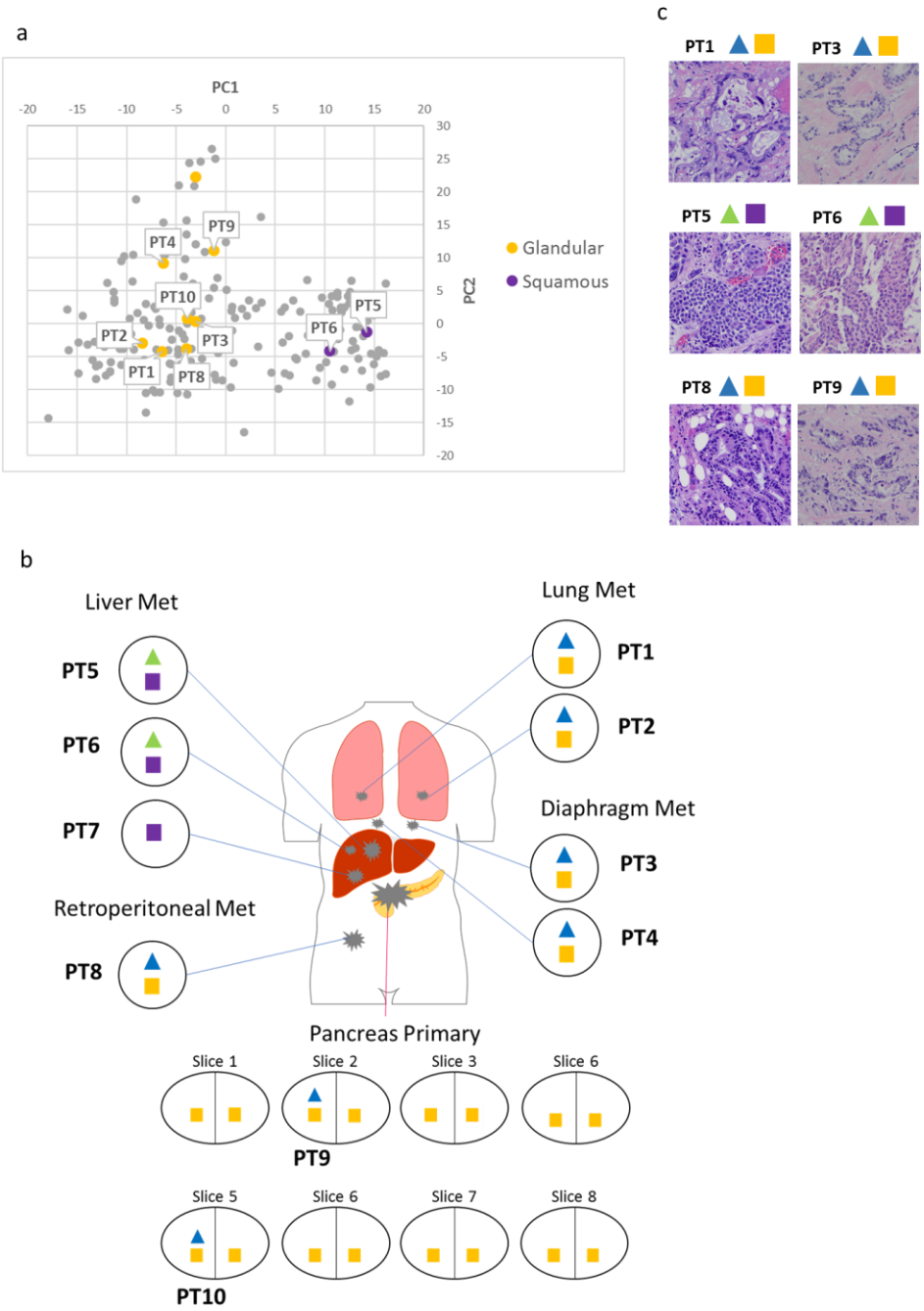
Ex. Data Figure 15

Extended Data Figure 15: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM39. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of representative tumors in the same patient.



Extended Data Figure 16: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM16. (a) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (b) Representative histologic images and immunohistochemical labeling of representative tumors in the same patient.

MPAM6

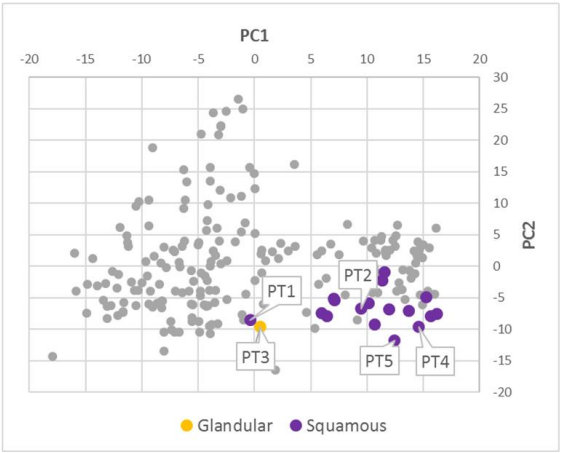


Ex. Data Figure 17

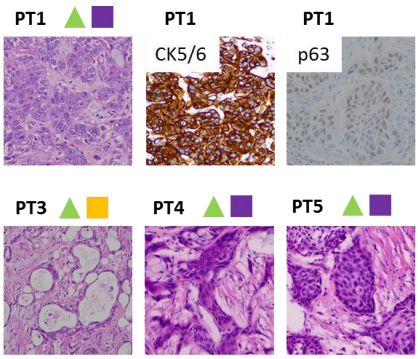
Extended Data Figure 17: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in MPAM6. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of representative tumors in the same patient.

PAM53

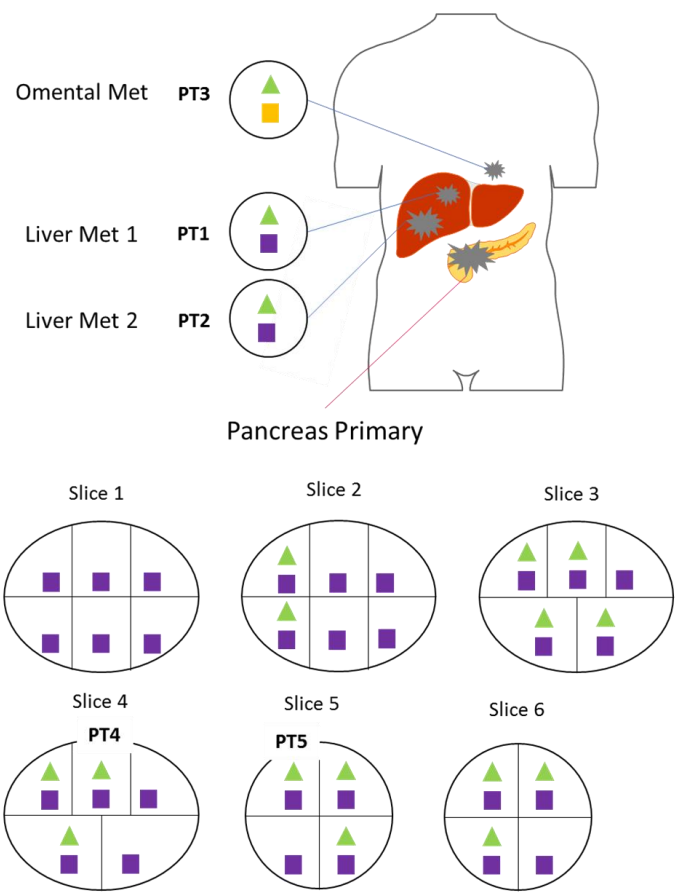
a



c



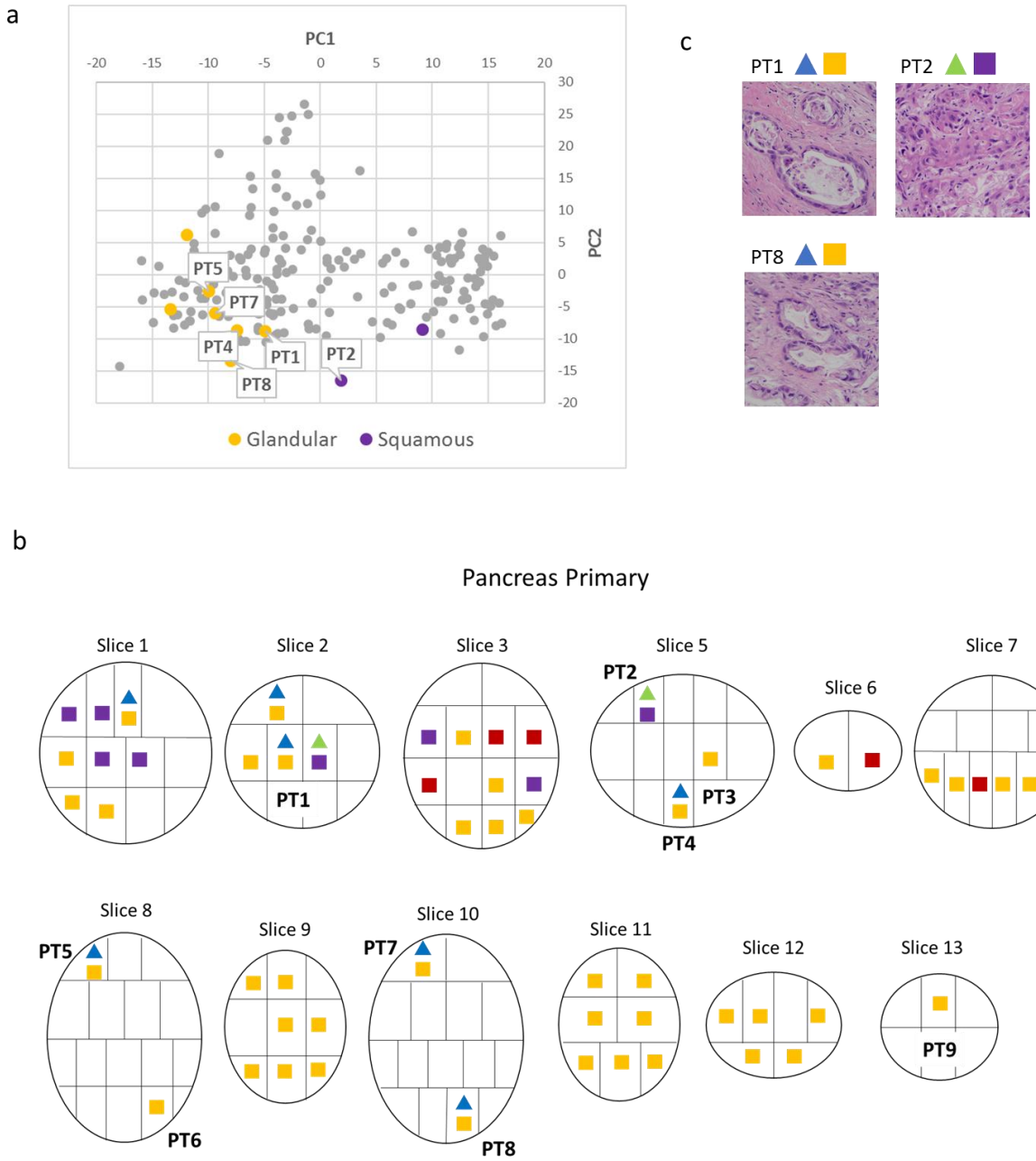
b



Ex. Data Figure 18

Extended Data Figure 18: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM53. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor or distant sites and their corresponding transcriptional and histological subtypes. (c) Representative histologic images and immunolabeling of representative tumors in the same patient.

PAM22



Ex. Data Figure 19

Extended Data Figure 19: Relationship of Anatomical Location to Morphologic and Transcriptional Heterogeneity in PAM22. (a) Principal component analysis using Moffit 50 gene set. (b) Shown is the spatial location of each sample within the primary tumor and their corresponding transcriptional and histological subtypes. (c) Representative histologic images of the tumor in the same patient.