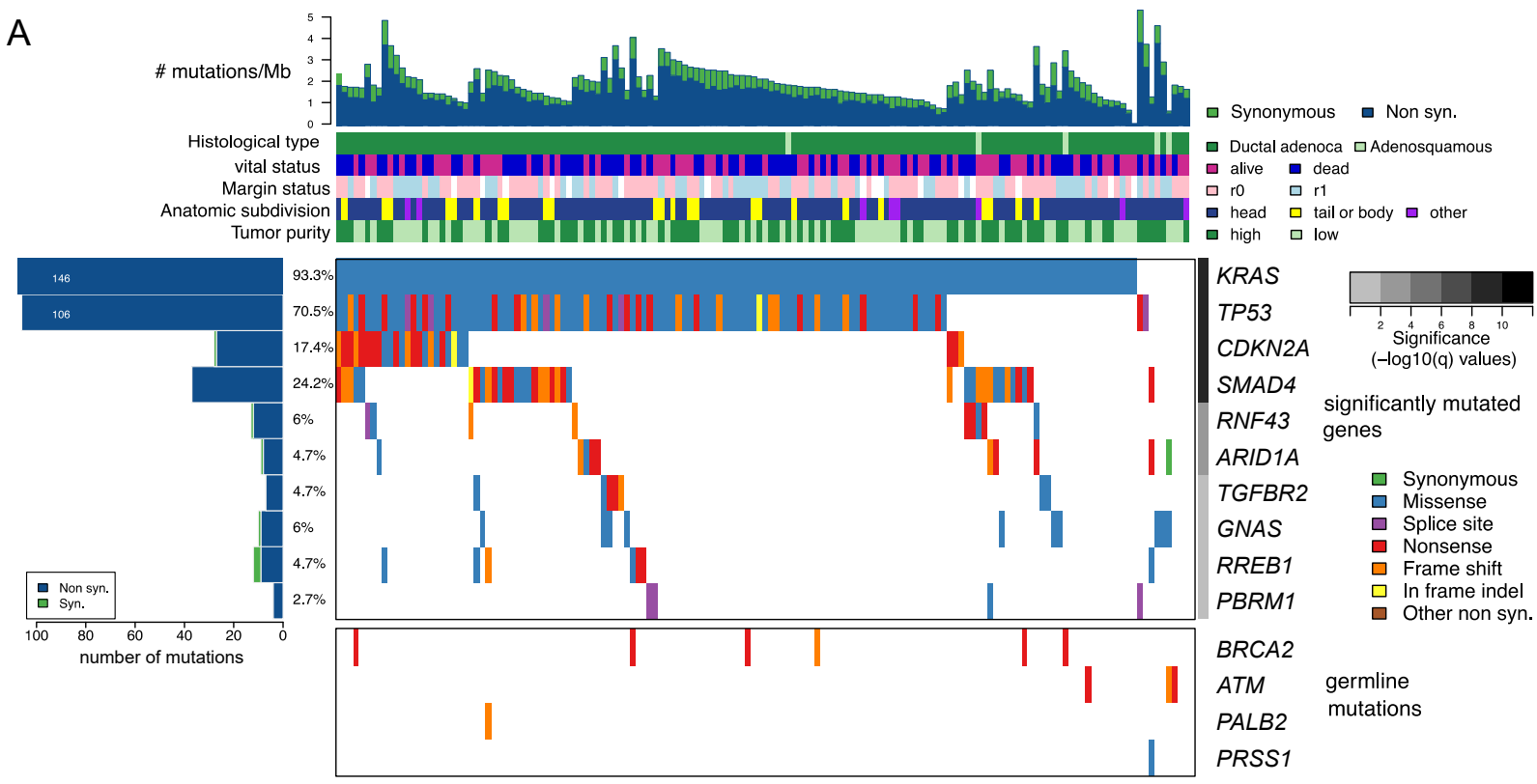

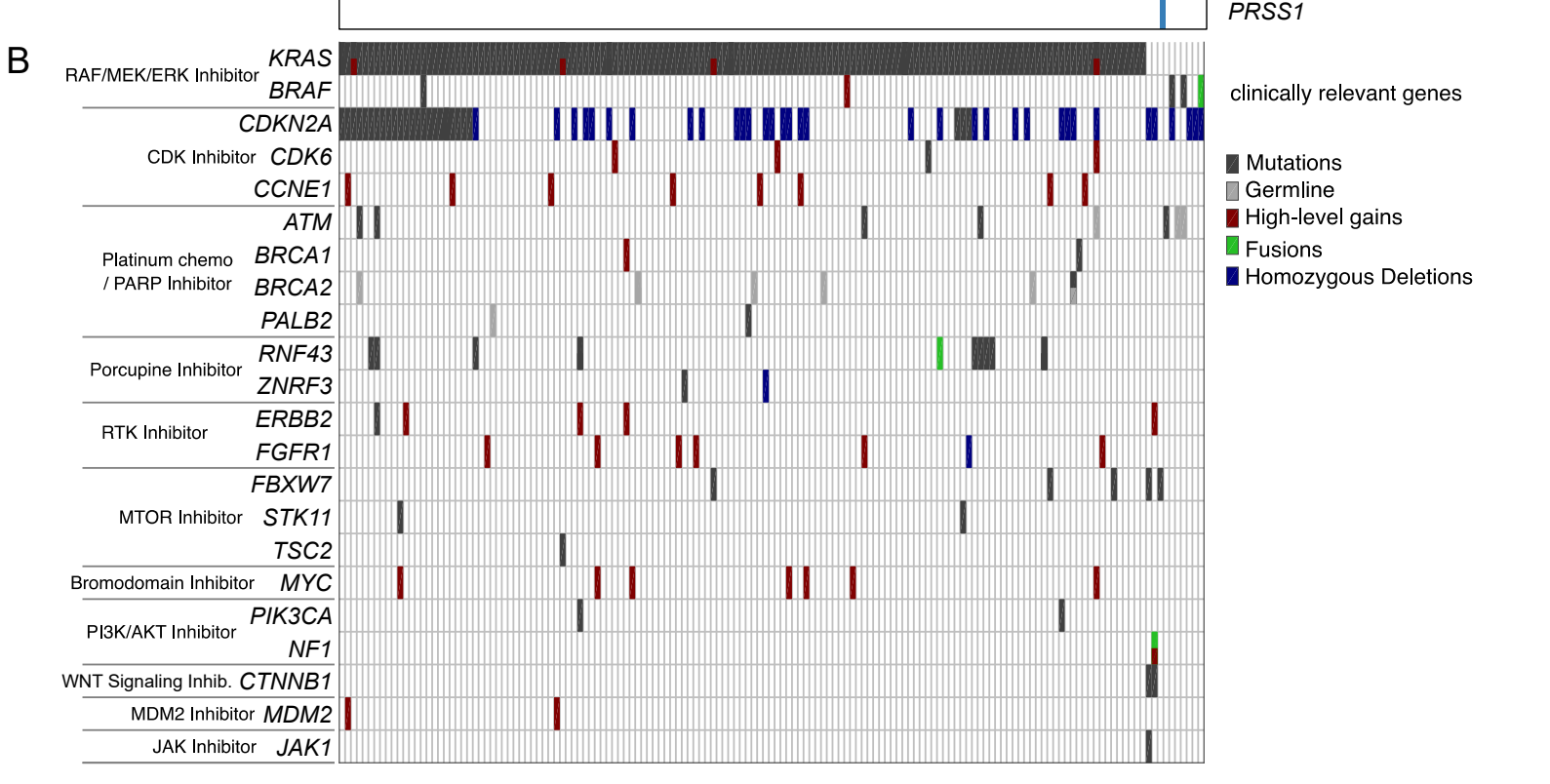

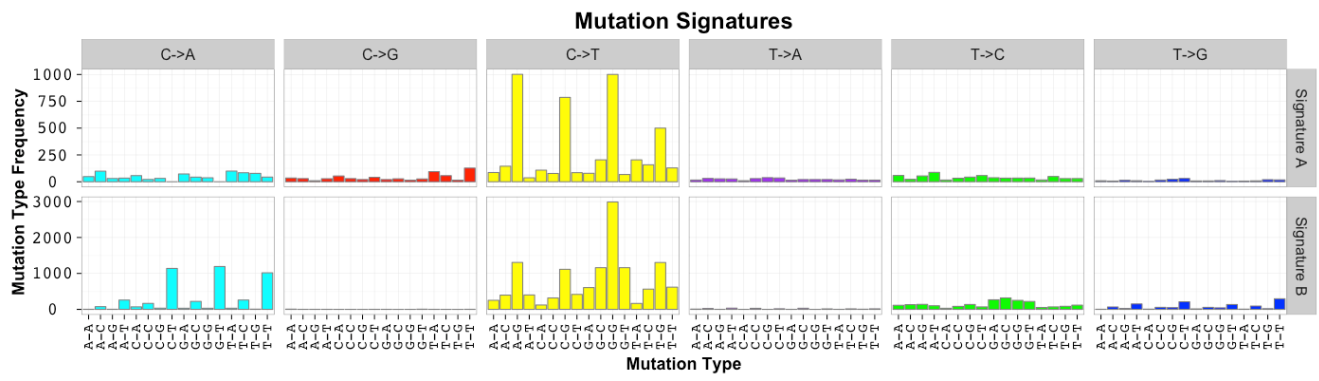

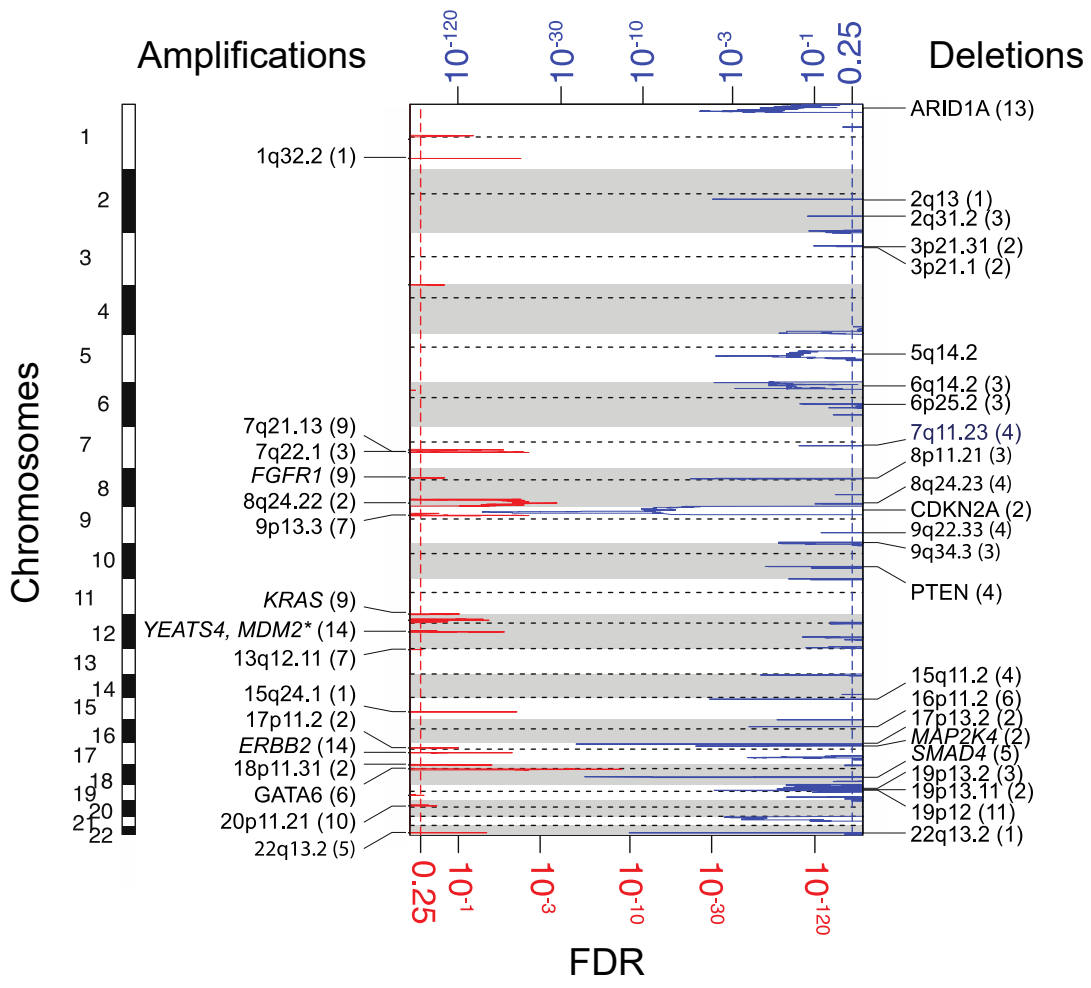

Figure S1, related to Figure 1

(A) CoMut plot displaying multiple tracks with all 149 non-hypermutated samples displayed as columns and overall number of mutations/Mb and clinicopathologic data at the top. Mutation type is indicated in the legend. Clinical tracks are plotted as in Figure 1. Significantly mutated genes with q value ≤ 0.1 that were identified by MutSig2CV analysis of exome sequencing data are listed vertically in order of q value. The percentage of PDAC samples with a mutation detected by automated calling is noted at the left. Samples are arranged to emphasize mutual exclusivity among mutations. Syn, synonymous. Germline data for analysis of known PDAC inherited susceptibility genes as described in main text. (B) Clinically relevant alterations with potential therapeutic relevance are shown, with genes grouped according to therapy class shown on the left. (C) Lego plot showing identification of two signatures: Signature A, resembling COSMIC signature 1; Signature B, resembling COSMIC signature 14. (D) GISTIC peaks for high-level amplifications and deletions corresponding to Supplemental Table 4.

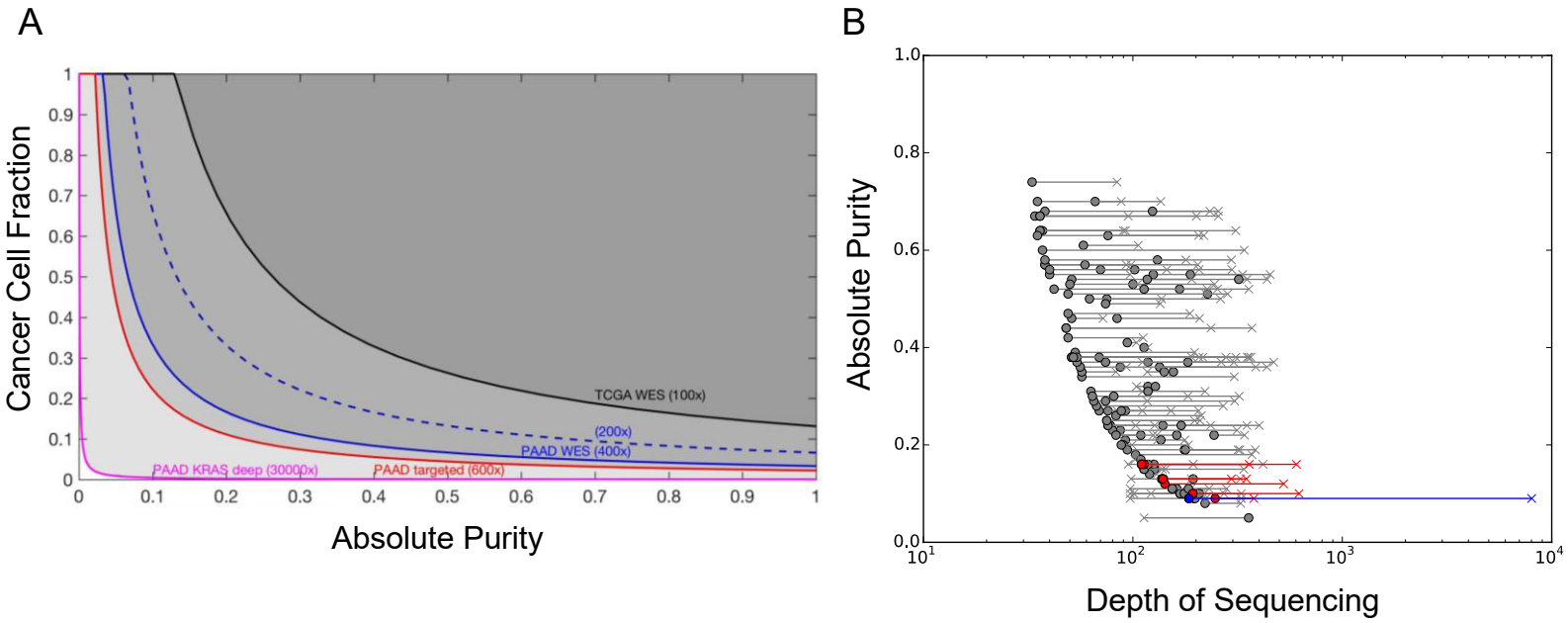

Figure S2, related to Figure 2

(A) Model demonstrating that greater sequencing depth improves the sensitivity to detect mutations at lower CCF and purity values. ABSOLUTE tumor purity (X-axis) and mutation cancer cell fraction (CCF, Y-axis) are plotted. Colored lines depict constant detection power = 0.95 for standard TCGA WES sequencing (~100x, black) and for the three data types utilized in this study: PDAC WES data (~400x, blue solid), PDAC targeted sequencing (~600x, red), and PDAC *KRAS* mutation hotspot deep sequencing data (~30000x, pink). The actual WES coverage at observed *KRAS* mutations averaged closer to 200x (blue dashed), rather than the mean coverage across all coding regions of ~400x. Each curve defines a region of CCF and purity values to the right of that line in which the demonstrated sequencing depth would be predicted to detect mutations with the indicated power. Importantly, the combined depth of coverage across multiple modalities used in this project enabled high-confidence detection of mutations, including subclonal mutations that would have been missed at lower sequencing depths. Iteratively downsampling WES data to evaluate the number of *KRAS* mutations observed at 4 or more reads that we would have missed by sequencing to conventional TCGA depths of ~100X (data not shown) suggested that 11% (15/139) of *KRAS* mutations would have been missed at conventional lower sequencing depths.

(B) Analysis of the depth of sequencing required to detect clonal *KRAS* mutation in our samples. Absolute purity (Y-axis) and depth of sequencing (X-axis) are shown for each clonal *KRAS* mutant sample in the cohort. Filled circles demonstrate the estimated local depth of sequencing coverage required at the *KRAS* locus to detect clonal *KRAS* mutations with 95% power given the tumor purity and local copy number in each of our samples (Carter et al., 2012). “x” demonstrate the actual coverage achieved at the *KRAS* locus for that particular sample in which we detected 4 or more reads. Gray, mutations detected in WES; Red, mutations detected in targeted panel that were not detected with WES; Blue, mutations detected in *KRAS* hotspot deep coverage sequencing but not in targeted panel sequencing or WES. As expected, we observed a general trend of increasing predicted depth of sequencing required to detect *KRAS* mutation in lower purity samples (filled circles). For most samples, the actual depth of sequencing performed at the *KRAS* locus with deep WES (mean ~200X at *KRAS* locus, ~400X across all exons) was greater than this predicted depth needed to detect a mutation; thus, most *KRAS* mutations in our cohort were identified by deep WES (gray “x”). However, even with deep WES, approximately 6% (8/139) of *KRAS* mutations were not detected by deep WES but were identified only through deeper targeted *KRAS* sequencing.

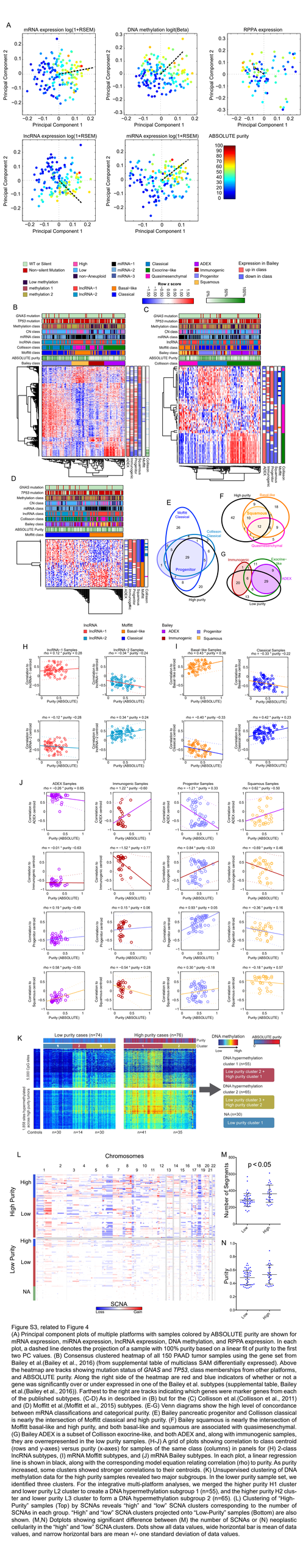

Figure S3, related to Figure 4

(A) Principal component plots of multiple platforms with samples colored by ABSOLUTE purity are shown for mRNA expression, miRNA expression, lncRNA expression, DNA methylation, and RPPA expression. In each plot, a dashed line denotes the projection of a sample with 100% purity based on a linear fit of purity to the first two PC values. (B) Consensus clustered heatmap of all 150 PAAD tumor samples using the gene set from Bailey et al.(Bailey et al., 2016) (from supplemental table of multiclass SAM differentially expressed). Above the heatmap are tracks showing mutation status of *GNAS* and *TP53*, class memberships from other platforms, and ABSOLUTE purity. Along the right side of the heatmap are red and blue indicators of whether or not a gene was significantly over or under expressed in one of the Bailey et al. subtypes (supplemental table, Bailey et al.(Bailey et al., 2016)). Farthest to the right are tracks indicating which genes were marker genes from each of the published subtypes. (C-D) As in described in (B) but for the (C) Collisson et al.(Collisson et al., 2011) and (D) Moffitt et al.(Moffitt et al., 2015) subtypes. (E-G) Venn diagrams show the high level of concordance between mRNA classifications and categorical purity. (E) Bailey pancreatic progenitor and Collisson classical is nearly the intersection of Moffitt classical and high purity. (F) Bailey squamous is nearly the intersection of Moffitt basal-like and high purity, and both basal-like and squamous are associated with quasimesenchymal. (G) Bailey ADEX is a subset of Collisson exocrine-like, and both ADEX and, along with immunogenic samples, they are overrepresented in the low purity samples. (H-J) A grid of plots showing correlation to class centroid (rows and y-axes) versus purity (x-axes) for samples of the same class (columns) in panels for (H) 2-class lncRNA subtypes, (I) mRNA Moffitt subtypes, and (J) mRNA Bailey subtypes. In each plot, a linear regression line is shown in black, along with the corresponding model equation relating correlation (rho) to purity. As purity increased, some clusters showed stronger correlations to their centroids. (K) Unsupervised clustering of DNA methylation data for the high purity samples revealed two major subgroups. In the lower purity sample set, we identified three clusters. For the integrative multi-platform analyses, we merged the higher purity H1 cluster and lower purity L2 cluster to create a DNA hypermethylation subgroup 1 (n=55), and the higher purity H2 cluster and lower purity L3 cluster to form a DNA hypermethylation subgroup 2 (n=65). (L) Clustering of "High-Purity" samples (Top) by SCNAs reveals "high" and "low" SCNA clusters corresponding to the number of SCNAs in each group. "High" and "low" SCNA clusters projected onto "Low-Purity" samples (Bottom) are also shown. (M,N) Dotplots showing significant difference between (M) the number of SCNAs or (N) neoplastic cellularity in the "high" and "low" SCNA clusters. Dots show all data values, wide horizontal bar is mean of data values, and narrow horizontal bars are mean +/- one standard deviation of data values.

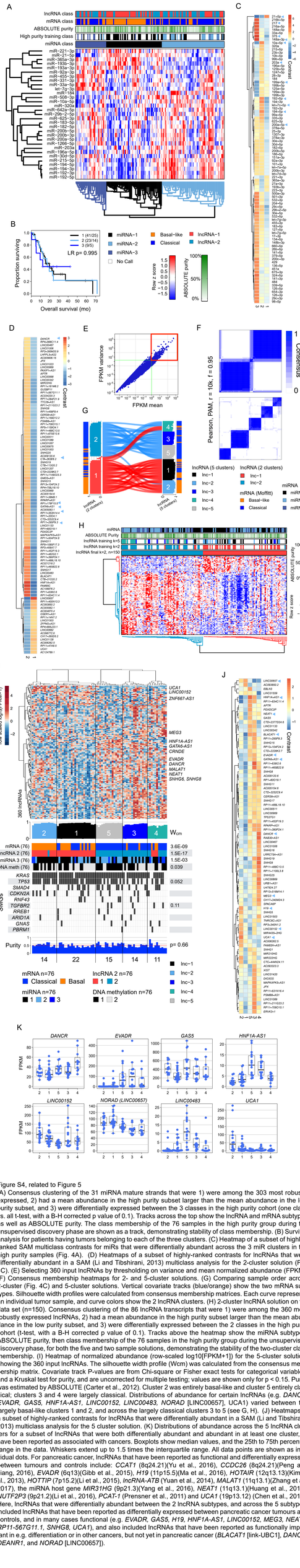

Figure S4, related to Figure 5

(A) Consensus clustering of the 31 miRNA mature strands that were 1) were among the 303 most robustly expressed, 2) had a mean abundance in the high purity subset larger than the mean abundance in the low purity subset, and 3) were differentially expressed between the 3 classes in the high purity cohort (one class vs. all t-test, with a B-H corrected p value of 0.1). Tracks across the top show the lncRNA and mRNA subtypes as well as ABSOLUTE purity. The class membership of the 76 samples in the high purity group during the unsupervised discovery phase are shown as a track, demonstrating stability of class membership. (B) Survival analysis for patients having tumors belonging to each of the three clusters. (C) Heatmap of a subset of highly-ranked SAM multiclass contrasts for miRs that were differentially abundant across the 3 miR clusters in the high purity samples (Fig. 4A). (D) Heatmaps of a subset of highly-ranked contrasts for lncRNAs that were differentially abundant in a SAM (Li and Tibshirani, 2013) multiclass analysis for the 2-cluster solution (Fig. 4C). (E) Selecting 360 input lncRNAs by thresholding on variance and mean normalized abundance (FPKM). (F) Consensus membership heatmaps for 2- and 5-cluster solutions. (G) Comparing sample order across 2-cluster (Fig. 4C) and 5-cluster solutions. Vertical covariate tracks (blue/orange) show the two mRNA subtypes. Silhouette width profiles were calculated from consensus membership matrices. Each curve represents an individual tumor sample, and curve colors show the 2 lncRNA clusters. (H) 2-cluster lncRNA solution on full data set (n=150). Consensus clustering of the 86 lncRNA transcripts that were 1) were among the 360 most robustly expressed lncRNAs, 2) had a mean abundance in the high purity subset larger than the mean abundance in the low purity subset, and 3) were differentially expressed between the 2 classes in the high purity cohort (t-test, with a B-H corrected p value of 0.1). Tracks above the heatmap show the mRNA subtypes, ABSOLUTE purity, then class membership of the 76 samples in the high purity group during the unsupervised discovery phase, for both the five and two sample solutions, demonstrating the stability of the two-cluster class membership. (I) Heatmap of normalized abundance (row-scaled log10[FPKM+1]) for the 5-cluster solution, showing the 360 input lncRNAs. The silhouette width profile (Wcm) was calculated from the consensus membership matrix. Covariate track P-values are from Chi-square or Fisher exact tests for categorical variables, and a Kruskal test for purity, and are uncorrected for multiple testing; values are shown only for $p < 0.15$. Purity was estimated by ABSOLUTE (Carter et al., 2012). Cluster 2 was entirely basal-like and cluster 5 entirely classical; clusters 3 and 4 were largely classical. Distributions of abundance for certain lncRNAs (e.g. *DANCR*, *EVADR*, *GAS5, HNF1A-AS1*, *LINC00152, LINC00483, NORAD* [LINC00657], UCA1) varied between the largely basal-like clusters 1 and 2, and across the largely classical clusters 3 to 5 (see G, H). (J) Heatmaps of a subset of highly-ranked contrasts for lncRNAs that were differentially abundant in a SAM (Li and Tibshirani, 2013) multiclass analysis for the 5 cluster solution. (K) Distributions of abundance across the 5 lncRNA clusters for a subset of lncRNAs that were both differentially abundant and abundant in at least one cluster, or have been reported as associated with cancers. Boxplots show median values, and the 25th to 75th percentile range in the data. Whiskers extend up to 1.5 times the interquartile range. All data points are shown as individual dots. For pancreatic cancer, lncRNAs that have been reported as functional and differentially expressed between tumours and controls include: *CCAT1* (8q24.21)(Yu et al., 2016), *CCDC26* (8q24.21)(Peng and Jiang, 2016), *EVADR* (6q13)(Gibb et al., 2015), *H19* (11p15.5)(Ma et al., 2016), *HOTAIR* (12q13.13)(Kim et al., 2013), *HOTTIP* (7p15.2)(Li et al., 2015), *lncRNA-ATB* (Yuan et al., 2014), *MALAT1* (11q13.1)(Zhang et al., 2017), the miRNA host gene *MIR31HG* (9p21.3)(Yang et al., 2016), *NEAT1* (11q13.1)(Huang et al., 2017), *NUTF2P3* (9p21.2)(Li et al., 2016), *PCAT-1* (Prensner et al., 2011) and *UCA1* (19p13.12) (Chen et al., 2016). Here, lncRNAs that were differentially abundant between the 2 lncRNA subtypes, and across the 5 subtypes, included lncRNAs that have been reported as differentially expressed between pancreatic cancer tumours and controls, and in many cases functional (e.g. *EVADR, GAS5, H19, HNF1A-AS1, LINC00152, MEG3, NEAT1, RP11-567G11.1, SNHG8, UCA1*), and also included lncRNAs that have been reported as functionally important in e.g. differentiation or in other cancers, but not yet in pancreatic cancer (*BLACAT1* [link-UBC1], *DANCR, DEANR1*, and *NORAD* [LINC00657]).

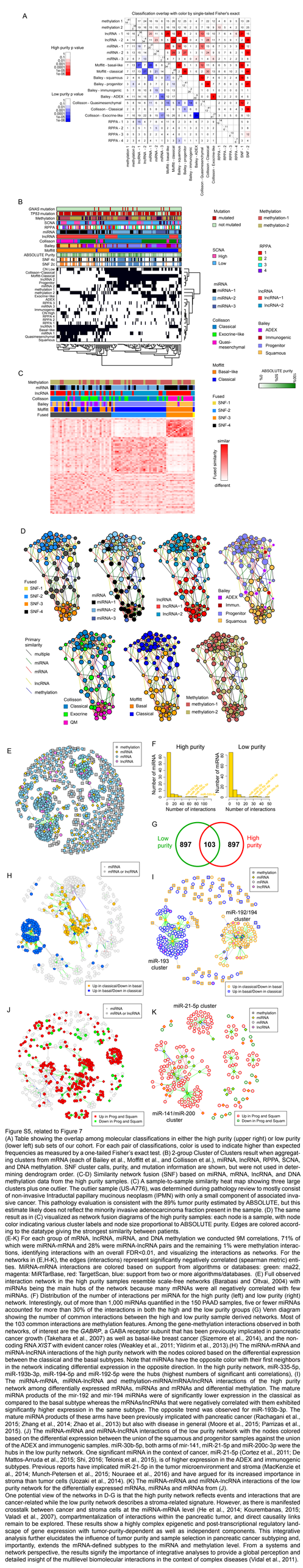

Figure S5, related to Figure 7

(A) Table showing the overlap among molecular classifications in either the high purity (upper right) or low purity (lower left) sub sets of our cohort. For each pair of classifications, color is used to indicate higher than expected frequencies as measured by a one-tailed Fisher's exact test. (B) 2-group Cluster of Clusters result when aggregating clusters from mRNA (each of Bailey et al., Moffitt et al., and Collisson et al.), miRNA, lncRNA, RPPA, SCNA, and DNA methylation. SNF cluster calls, purity, and mutation information are shown, but were not used in determining dendrogram order. (C-D) Similarity network fusion (SNF) based on miRNA, mRNA, lncRNA, and DNA methylation data from the high purity samples. (C) A sample-to-sample similarity heat map showing three large clusters plus one outlier. The outlier sample (US-A776), was determined during pathology review to mostly consist of non-invasive Intraductal papillary mucinous neoplasm (IPMN) with only a small component of associated invasive cancer. This pathology evaluation is consistent with the 89% tumor purity estimated by ABSOLUTE, but this estimate likely does not reflect the minority invasive adenocarcinoma fraction present in the sample. (D) The same result as in (C) visualized as network fusion diagrams of the high purity samples: each node is a sample, with node color indicating various cluster labels and node size proportional to ABSOLUTE purity. Edges are colored according to the datatype giving the strongest similarity between patients.

(E-K) For each group of mRNA, lncRNA, miRNA, and DNA methylation we conducted 9M correlations, 71% of which were miRNA-miRNA and 28% were miRNA-lncRNA pairs and the remaining 1% were methylation interactions, identifying interactions with an overall FDR<0.01, and visualizing the interactions as networks. For the networks in (E,H-K), the edges (interactions) represent significantly negatively correlated (spearman metric) entities. MiRNA-mRNA interactions are colored based on support from algorithms or databases: green: rna22, magenta: MiRTarBase, red: TargetScan, blue: support from two or more algorithms/databases. (E) Full observed interaction network in the high purity samples resemble scale-free networks (Barabasi and Oltvai, 2004) with miRNAs being the main hubs of the network because many mRNAs were all negatively correlated with few miRNAs. (F) Distribution of the number of interactions per miRNA for the high purity (left) and low purity (right) network. Interestingly, out of more than 1,000 miRNAs quantified in the 150 PAAD samples, five or fewer miRNAs accounted for more than 30% of the interactions in both the high and the low purity groups (G) Venn diagram showing the number of common interactions between the high and low purity sample derived networks. Most of the 103 common interactions are methylation features. Among the gene-methylation interactions observed in both networks, of interest are the *GABRP*, a GABA receptor subunit that has been previously implicated in pancreatic cancer growth (Takehara et al., 2007) as well as basal-like breast cancer (Sizemore et al., 2014), and the non-coding RNA *XIST* with evident cancer roles (Weakley et al., 2011; Yildirim et al., 2013).(H) The miRNA-mRNA and miRNA-lncRNA interactions of the high purity network with the nodes colored based on the differential expression between the classical and the basal subtypes. Note that miRNAs have the opposite color with their first neighbors in the network indicating differential expression in the opposite direction. In the high purity network, miR-335-5p, miR-193b-3p, miR-194-5p and miR-192-5p were the hubs (highest numbers of significant anti correlations). (I) The miRNA-mRNA, miRNA-lncRNA and methylation-miRNA/mRNA/lncRNA interactions of the high purity network among differentially expressed mRNAs, miRNAs and mRNAs and differential methylation. The mature miRNA products of the mir-192 and mir-194 miRNAs were of significantly lower expression in the classical as compared to the basal subtype whereas the mRNAs/lncRNAs that were negatively correlated with them exhibited significantly higher expression in the same subtype. The opposite trend was observed for miR-193b-3p. The mature miRNA products of these arms have been previously implicated with pancreatic cancer (Rachagani et al., 2015; Zhang et al., 2014; Zhao et al., 2013) but also with disease in general (Moore et al., 2015; Parrizas et al., 2015). (J) The miRNA-mRNA and miRNA-lncRNA interactions of the low purity network with the nodes colored based on the differential expression between the union of the squamous and progenitor samples against the union of the ADEX and immunogenic samples. miR-30b-5p, both arms of mir-141, miR-21-5p and miR-200c-3p were the hubs in the low purity network. One significant miRNA in the context of cancer, miR-21-5p (Cortez et al., 2011; De Mattos-Arruda et al., 2015; Shi, 2016; Telonis et al., 2015), is of higher expression in the ADEX and immunogenic subtypes. Previous reports have implicated miR-21-5p in the tumor microenvironment and stroma (MacKenzie et al., 2014; Munch-Petersen et al., 2015; Nouraee et al., 2016) and have argued for its increased importance in stroma than tumor cells (Uozaki et al., 2014). (K) The miRNA-mRNA and miRNA-lncRNA interactions of the low purity network for the differentially expressed mRNAs, miRNAs and mRNAs from (J).

One potential view of the networks in D-G is that the high purity network reflects events and interactions that are cancer-related while the low purity network describes a stroma-related signature. However, as there is manifested crosstalk between cancer and stroma cells at the miRNA-mRNA level (He et al., 2014; Kourembanas, 2015; Valadi et al., 2007), compartmentalization of interactions within the pancreatic tumor, and direct causality links remain to be explored. These results show a highly complex epigenetic and post-transcriptional regulatory landscape of gene expression with tumor-purity-dependent as well as independent components. This integrative analysis further elucidates the influence of tumor purity and sample selection in pancreatic cancer subtyping and, importantly, extends the mRNA-defined subtypes to the miRNA and methylation level. From a systems and network perspective, the results signify the importance of integrative analyses to provide a global perception and detailed insight of the multilevel biomolecular interactions in the context of complex diseases (Vidal et al., 2011).