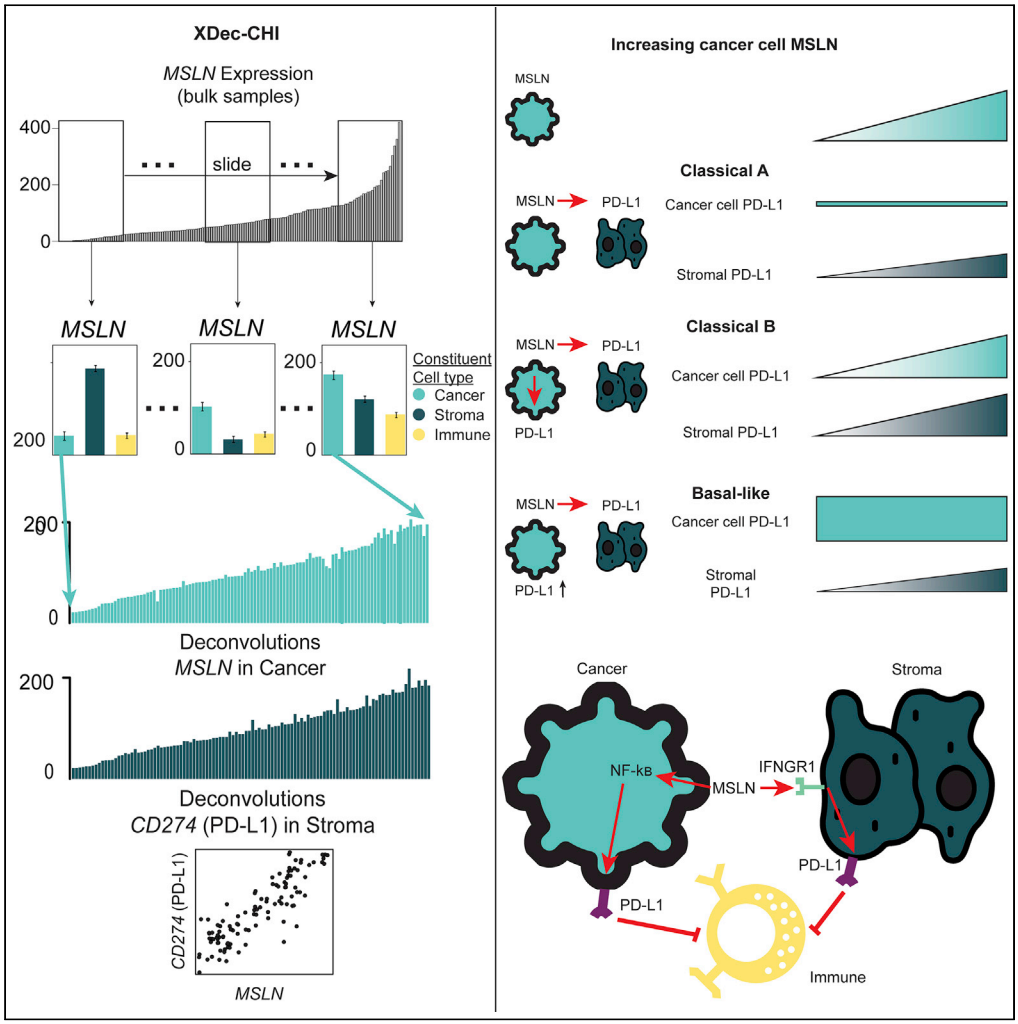
iScience

CellPress
OPEN ACCESS

Article

XDec-CHI reveals immunosuppressive interactions in pancreatic ductal adenocarcinoma



Emily L. LaPlante, Dongliang Liu, Varduhi Petrosyan, Qizhi Yao, Aleksandar Milosavljevic

qizhiyao@bcm.edu (Q.Y.)
amilosav@bcm.edu (A.M.)

Highlights

XDec-CHI deconvolutes cell-type-specific gene interactions from bulk RNAseq data

Subtypes of PDAC upregulate PD-L1 using distinct intra- and inter-cellular pathways

Distinct PD-L1 pathways suggest subtype-specific therapy targets and responses

XDec-CHI software and pre-computed datasets for several solid cancers is provided

LaPlante et al., iScience 25, 105249
October 21, 2022
https://doi.org/10.1016/j.isci.2022.105249

iScience

CellPress
OPEN ACCESS

Article

XDec-CHI reveals immunosuppressive interactions in pancreatic ductal adenocarcinoma

Emily L. LaPlante,[1] Dongliang Liu,[2] Varduhi Petrosyan,[1] Qizhi Yao,[2,3,4,*] and Aleksandar Milosavljevic[1,4,5,6,*]

SUMMARY

Most cancers harbor a diverse collection of cell types including a typically heterogeneous cancer cell fraction. To reconstruct cell-intrinsic and heterotypic interactions driving tumor progression, we combine the XDec deconvolution method with cell-type-specific gene expression correlation analysis into the XDec-CHI method. XDec-CHI identifies intra- and inter-cellular pathways using correlation and places them in the context of specific tumor subtypes, as defined by the state of constituent cancer cells. We make the method web-accessible for analysis of publicly accessible pancreatic ductal adenocarcinoma, breast, head and neck, glioblastoma, and glioma tumors. We apply the method to TCGA and ICGC datasets to identify immune-suppressive interactions within PDAC tumors that are relevant for immunotherapies targeting PD-L1. Subtype-specific interactions derived from correlative analyses validated in co-culture experiments suggest PDAC subtypes have distinct therapeutic weaknesses, with Basal-like and *MSLN*-high Classical B tumors most likely to respond to therapies targeting PD-L1.

INTRODUCTION

Cancer therapy resistance is driven by both cell-intrinsic and heterotypic interactions within the tumor microenvironment (Jacobetz et al., 2013; Olive et al., 2009; Özdemir et al., 2014; Provenzano et al., 2012; Rhim et al., 2014). The highest-resolution method to study the tumor microenvironment *in vivo* is single-cell profiling. However, single-cell profiling is technically demanding, costly, and not readily applicable to large cohorts required to decipher interactions within the microenvironment. It also cannot be applied to standard FFPE samples collected in clinical practice. To address this issue, computational deconvolution methods that estimate tumor composition from bulk methylation or RNA-seq profiles have been developed (Carter et al., 2012; Chan-Seng-Yue et al., 2020; Lurie et al., 2020; Newman et al., 2019; Onuchic et al., 2016; Peng et al., 2019b). The most recent generation of such tools, including the highly popular CIBSERSORTx, utilize single-cell RNA-seq (scRNA-seq) information collected on a small number of samples to create cell-type-specific reference profiles that are subsequently utilized for computational deconvolution of bulk RNA-seq profiles (Newman et al., 2019; Peng et al., 2021; Wang et al., 2019).

Tumor complexity is further exacerbated by the possibility that relevant interactions may occur only within specific tumor subtypes. Moreover, a single tumor may be heterogeneous, with multiple subtypes of cancer cells present (Chan-Seng-Yue et al., 2020; Topham et al., 2021). This heterogeneity of cancer cell states is yet to be tackled by reference-based methods. On the other hand, reference-free methods (Decamps et al., 2020) do not take any advantage of information available from single-cell profiling. Here we propose a "reference optional" strategy that utilizes information from single-cell profiling, if available, to identify informative genes while retaining the power to infer cancer subtype profiles by deconvolution. The strategy is particularly suitable for situations where bulk profiling data is available for a much larger number of tumors than single-cell data, implying that the diversity of tumor types is better represented in bulk profiles. The reference-optional strategy, initially implemented by the EDec algorithm, has indeed shown the potential to recover the diversity of cancer cell states that correspond to clinically relevant subtypes of breast cancer (Onuchic et al., 2016), glioblastoma (Lucero et al., 2020) and pancreatic ductal adenocarcinoma (Lurie et al., 2020).

XDec (Murillo et al., 2019, 2022) is a modification of EDec that deconvolutes RNA-seq profiles. XDec may be characterized as being "reference-optional" as it utilizes scRNA-seq data, when available, to identify genes

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[2]Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA

[3]Center for Translational Research on Inflammatory Diseases (CTRID), Michael E. DeBakey VA Medical Center, Houston, TX, USA

[4]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

[5]Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX, USA

[6]Lead contact

*Correspondence: qizhiyao@bcm.edu (Q.Y.), amilosav@bcm.edu (A.M.)

https://doi.org/10.1016/j.isci.2022.105249

that are variable across cell types and therefore informative for deconvolution. Unlike reference-based methods, XDec does not fix cell-type-specific expression levels for the informative genes. Because it infers the cell-type expression levels for all genes from bulk profiles, we reasoned that XDec would enable the discovery of the diversity of cancer cell states including those not explicitly identified in the reference data. Specifically, we reasoned that XDec may uncover cancer cell states corresponding to subtypes of Pancreatic Ductal Adenocarcinoma (PDAC) that cannot be clearly discerned from the limited number of tumors for which scRNA-seq data are available by harnessing the large number of publicly accessible bulk tumor profiles that are representative of the full spectrum of cancer cell states across the diversity of tumor subtypes.

Going beyond XDec itself, XDec-CHI correlates gene expression levels within and across distinct cell types within the tumor microenvironment. It functions by performing non-negative least squares deconvolution on each dataset in the manner of a sliding window to determine multiple cell-type-specific profiles for each dataset. As input, it uses proportions generated by XDec-SM (Murillo et al., 2022) or, alternatively, other deconvolution algorithms including CIBERSORTx or BLADE (Barbosa et al., 2021; Newman et al., 2019) and bulk RNA-seq profiles. We reasoned that the correlation may point to gene-gene and cellular interactions within the microenvironment that can then be confirmed in experimental models. To empower the community, we create user-friendly web applications that perform this method across multiple pre-deconvoluted cancer dataset (Pancreatic cancer, breast cancer, head and neck cancer, glioblastoma, and glioma) and an R package for performing analysis on other datasets. To demonstrate the power of this method to generate experimentally verifiable discoveries that are relevant for tumors *in vivo*, we apply it to study the heterogeneity of immune pathway activation across subtypes of PDAC.

RESULTS

Detection of cell-intrinsic and heterotypic gene interactions from deconvoluted tumor profiles using XDec-CHI

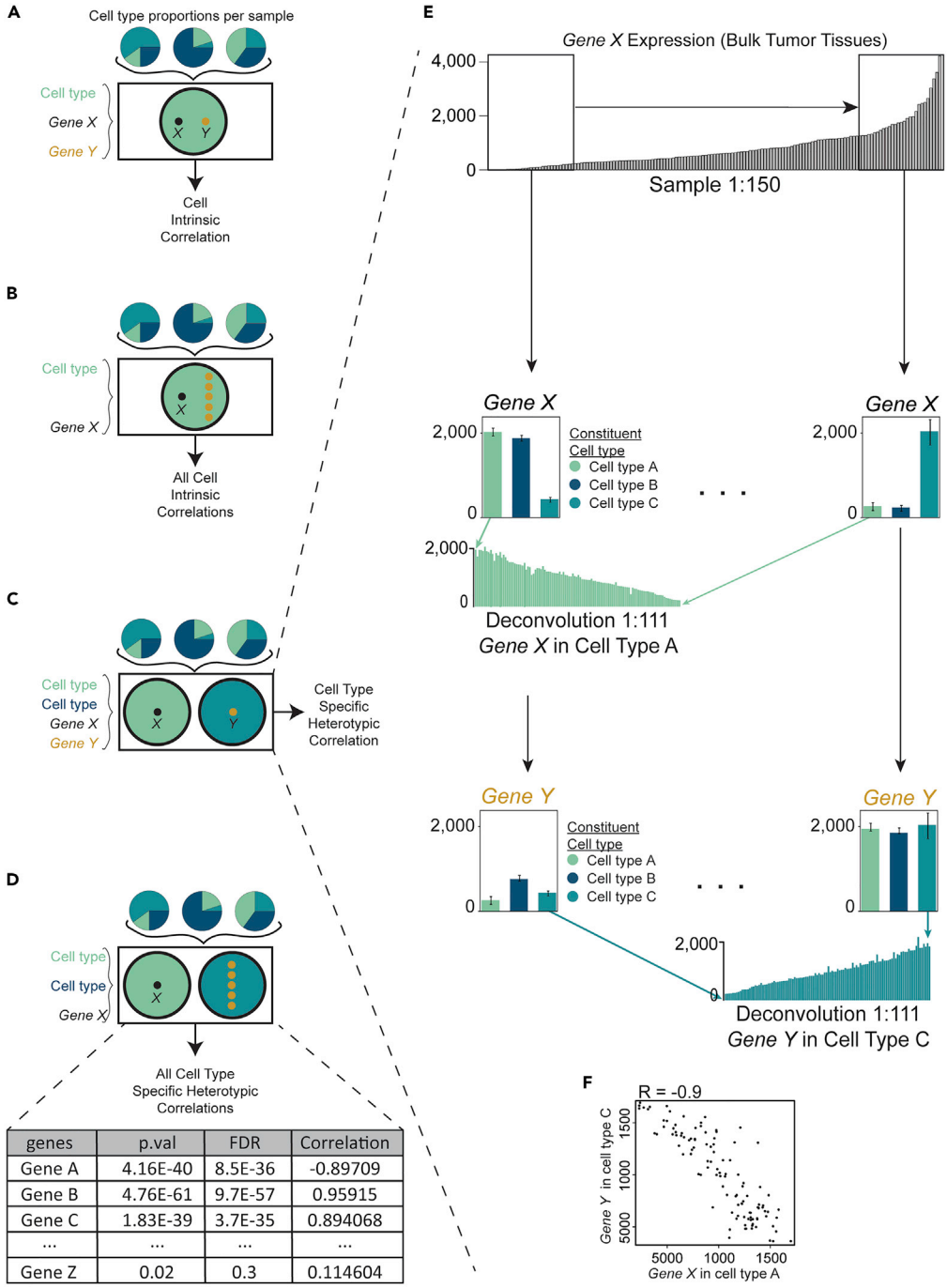
XDec determines cell-type-specific gene expression from bulk RNA-seq data (Murillo et al., 2019, 2022). XDec-CHI extends XDec deconvolution by determining cell intrinsic (Figures 1A and 1B) and heterotypic (Figures 1C and 1D) gene-gene interactions. In both of these settings, we envisioned the need for detecting either a single correlation between two specified genes (Figures 1A and 1C) or detecting all significant correlations between a gene of interest and any other gene (Figures 1B and 1D). XDec-CHI can also start from cell proportions determined by methods other than XDec. The method is therefore not tightly coupled with XDec while benefiting from the potential of XDec to identify subtypes of cancer cells within deconvoluted tumors.

XDec-CHI identifies gene-gene correlation by performing sliding-window deconvolution. Samples are first ordered by bulk expression of a gene of interest, termed "anchor gene." Next, groups of 40 samples are deconvoluted, moving the window by 1 sample so that groups of samples corresponding to consecutive windows overlap by 39 samples (Figure 1E *top*, STAR Methods). Every deconvolution of 40 samples results in cell-type-specific profiles that estimate the expression of all genes across all cell types (Figure 1E *bottom*). The method provides information about the changes in cell-type-specific gene expression levels (Figure 1E *bottom*). Correlations between the gene expression levels can are then visualized (Figure 1F).

Simulation experiments confirm the power of XDec-CHI to detect cell-intrinsic and heterotypic interactions

To evaluate the power of the method as a function of key parameters, we performed simulation experiments. The simulations using varying coefficients of variation between gene pairs showed larger power of the method to detect cell intrinsic than heterotypic correlations presumably because of the larger effect of variability of cell type composition on the ability to accurately assess heterotypic interactions (Figures S1A and S1B).

We next explored the power to detect cell-intrinsic correlations as a function of the strength of correlation between the genes of interest and the proportion of the cell type within tumors. Toward this purpose, we used the Pancreatic Ductal Adenocarcinoma (PDAC) tumor profiles from the ICGC collection. Using *MSLN* and *CD274* genes as a model, we performed a series of simulation experiments. Starting with *MSLN* and *CD274* means and standard deviations observed in the ICGC collection, we simulated additional varying amounts of noise to determine the probability of detecting their interaction within the top 1% or 2% of



genes	p.val	FDR	Correlation
Gene A	4.16E-40	8.5E-36	-0.89709
Gene B	4.76E-61	9.7E-57	0.95915
Gene C	1.83E-39	3.7E-35	0.894068
...	...	...	...
Gene Z	0.02	0.3	0.114604

Figure 1. Sliding window deconvolution allows cell-type-specific correlation

(A) Correlation for *Gene X* and *Gene Y* in cell type A.

(B) Correlations of genes in cell type A that correlate with *Gene X*.

(C) Correlation for *Gene X* in cell type A and *Gene Y* in cell type B.

(D) Correlation of genes in cell type B that correlate with *Gene X* in cell type A.

(E) Schema of XDec sliding window deconvolution. (*Top*) Samples ordered by bulk Anchor Gene expression. (*Bottom*) Each window provides estimated gene expression for all genes across all cell types and looking at a gene across each cell type displays cell-type-specific expression changes.

(F) Exemplar correlation of cell intrinsic correlation based on 1E (*bottom*).

correlations. As expected, stronger correlations could be detected within smaller cell type proportions than weaker correlations (Figure S1C).

XDec is available as an R package and a web application with pre-computed deconvolutions

To empower the community, we implemented XDec-CHI as an R package (https://github.com/BRL-BCM/XDec_CHI) that takes as an input bulk RNA-seq profiles and cell proportions for each sample (identified by XDec or other deconvolution methods) and identifies cell intrinsic or heterotypic gene-gene interactions.

In addition to the package, we have released a web application to enable reproducibility of our findings and empower the community to perform cell-type-specific analysis without running the R package. The web application hosts provide correlation analyses based on pre-computed deconvolution of five cancer types (PDAC, BRCA, HNSC, GBM, and Glioma) and are accessible at https://brl-bcm.shinyapps.io/XDec-CHI_Homepage (Carrero et al., 2019; Lucero et al., 2020; Lurie et al., 2020; Murillo et al., 2022).

The R package and the web application contain two groups of functionalities – first, allowing cell-type-specific correlation of two genes where users provide two genes and the two cell types of interest (Figures 1A, 1C, 1E, and 1F). Second, the user can identify all genes in any specific cell type that correlate with a gene of interest in the same or different cell type (Figures 1B and 1D). Results are displayed listing all correlated genes and the direction of their correlation. This "all gene" comparison can be used to rank genes or perform further analysis on significant sets that are positively or negatively correlated with the gene of interest.

Deconvolution by XDec identifies subtypes and cellular composition of PDAC tumors

Whereas XDec-CHI may utilize as an input cell type proportions inferred by methods other than XDec, we asked if XDec may identify PDAC subtypes of cancer cells within deconvoluted tumors, thus enabling the detection of subtype-specific correlations. Recent PDAC profiling studies have delineated multiple PDAC subtypes (Chan-Seng-Yue et al., 2020; Moffitt et al., 2015; Rashid et al., 2020). As the PDAC single-cell profiling study used as reference information did not categorize cancer cells into different subtypes, we reasoned that reference-optional deconvolution methods such XDec may identify subtype-specific profiles from the large number of publicly accessible bulk profiles even if single-cell profiles are not subtyped. To this end, we performed XDec deconvolution, which uses scRNA-seq references only to identify informative genes and outputs the number of constituent cell types, cell-type-specific expression profiles, and proportions of the cell types within each bulk tumor sample.

Deconvolution of the TCGA PAAD cohort (n = 149) identified nine unique cell profiles. Because the deconvolution did not identify the profiles, their identify was established by correlation to reference single-cell profiles (Figure 2A). The highly specific correlation provided initial evidence for successful deconvolution. To further establish the accuracy of the method, we compared the predicted proportions with those obtained by other methods. The comparison showed a high correlation between our predicted cancer proportions and other methods (ABSOLUTE R = 0.82; EDec R = 0.80) (Figures 2B, S2A, and S2C). In comparison, CIBERSORTx, a reference-based deconvolution method, identified only one cancer profile (ABSOLUTE, R = 0.70; EDec, R = 0.74) (Figures 2C, S2B, and S2D). Taken together, these results validate XDec at the coarse level of resolution afforded by state-of-the-art methods while also demonstrating the power of XDec to increase the resolution beyond current methods by deconvoluting multiple distinct profiles of the cancer cell fraction, which we discuss next.

Of the nine profiles identified by correlation (Figure 2A), two were Cancer epithelial, one of which had higher proportions in Classical patients and the other higher proportions in Basal-like patients. We therefore identified the two as corresponding to distinct PurIST subtypes (Rashid et al., 2020) (Figure 3A). Another classification mechanism, Bailey subtyping, breaks samples into four subtypes (Bailey et al., 2016). We reproduced results showing that Progenitor samples overlap with Classical and Squamous samples overlap with Basal-like (Figure 3B). This concordance with previous findings gave us confidence that our cancer profiles indeed correspond to Classical and Basal-like subtypes.

We next asked if we could go beyond the established Classical/Basal dichotomy and deconvolute the recently proposed subtypes of higher specificity (Chan-Seng-Yue et al., 2020). To do this, we applied the same deconvolution method to the ICGC PACA-CA Cohort (n = 108) used by the Ontario Institute for Cancer Research (OICR) to create a subtyping system that splits Classical and Basal-like tumors into Classical

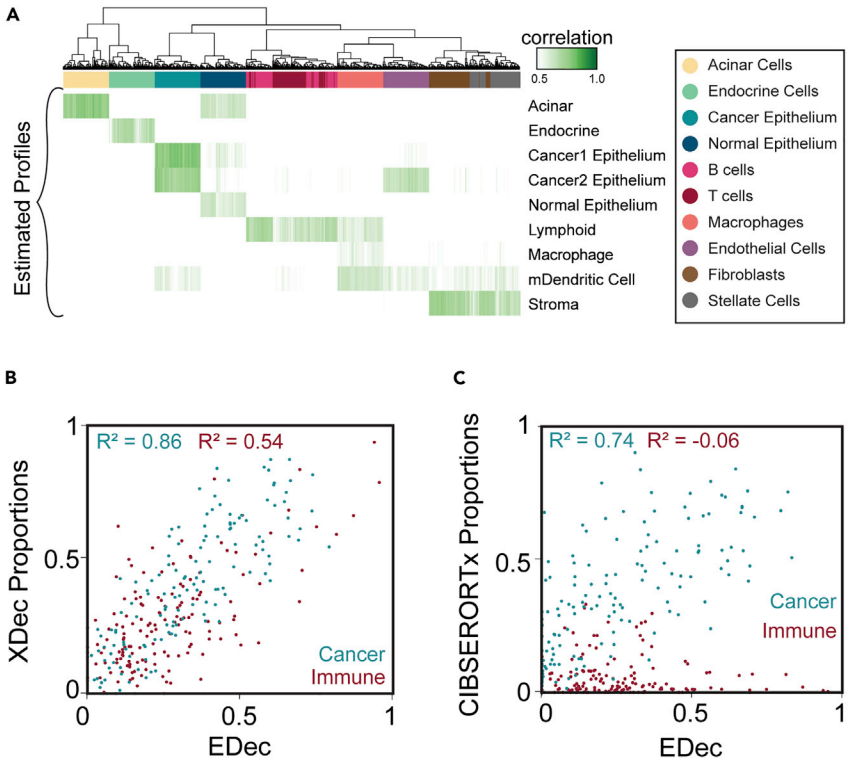


Figure 2. XDec deconvolution correlates well with other methods

(A) Correlation of nine TCGA cell types with pseudo-bulk references over informative genes.

(B) Correlation of XDec estimated tumor and immune proportion with EDec.

(C) Correlation of CIBERSORTx estimated tumor and immune proportion with EDec.

A/B, Basal-like A/B, and Hybrid (Chan-Seng-Yue et al., 2020). We again identified nine cell types and were able to identify cancer profiles enriched in Classical A and Classical B tumors (Figure 3C). However, we did not see a split in Basal-like profiles. The profile enriched in Basal-like A patients is also present in Classical patients at high proportions suggesting it is instead the Hybrid profile (Figure 3D). Both the Chan-Seng-Yue subtyping and PurIST subtyping identify very few Basal-like samples present in this set (18 and 15, respectively). It is therefore likely that we did not see the split in Basal-like patients because of the small number of samples present. Other contributing factor mays be high proportions of the hybrid and immune cell types and the low proportion of cancer cells among individual Basal-like samples. Our final cancerous profile that could not be subtyped is found in low proportion across all samples suggesting a possible cancer stem cell population. To validate our results, we also considered all publicly available datasets that had more than 42 samples in addition to the ones discussed above (PACA-AU and Dijk (Dijk et al., 2020)). Both datasets deconvoluted Classical, Basal-like, and hybrid cancer profiles.

Taken together, these results show that the results of deconvolution of TCGA and ICGC samples show concordance at the coarse level of resolution, and that the discerning power of deconvolution diminishes at finer levels of resolution, as the number of samples per highly specific subtype decreases. The differential performance across datasets may further be exacerbated by batch effects such as different requirements for tumor purity as well as different methods – bulk vs LCM – employed by the studies. The differences extend to the non-cancer fraction, with TCGA deconvoluting a single stromal profile, but ICGC deconvoluting into three stromal profiles. Moreover, likely because of the low proportions of immune cells found in the ICGC cohort (average 8.3%), which is especially pronounced in Classical samples, it is not surprising that we see a single immune profile unlike the three profiles observed in TCGA that averages 26% immune.

After estimating cell type proportions, deconvolution estimates cell -type-specific gene expression. This provided an additional opportunity to validate the results by confirming the expression of cell-type-specific marker genes and to determine expression of other genes (Figures S3 and S4).

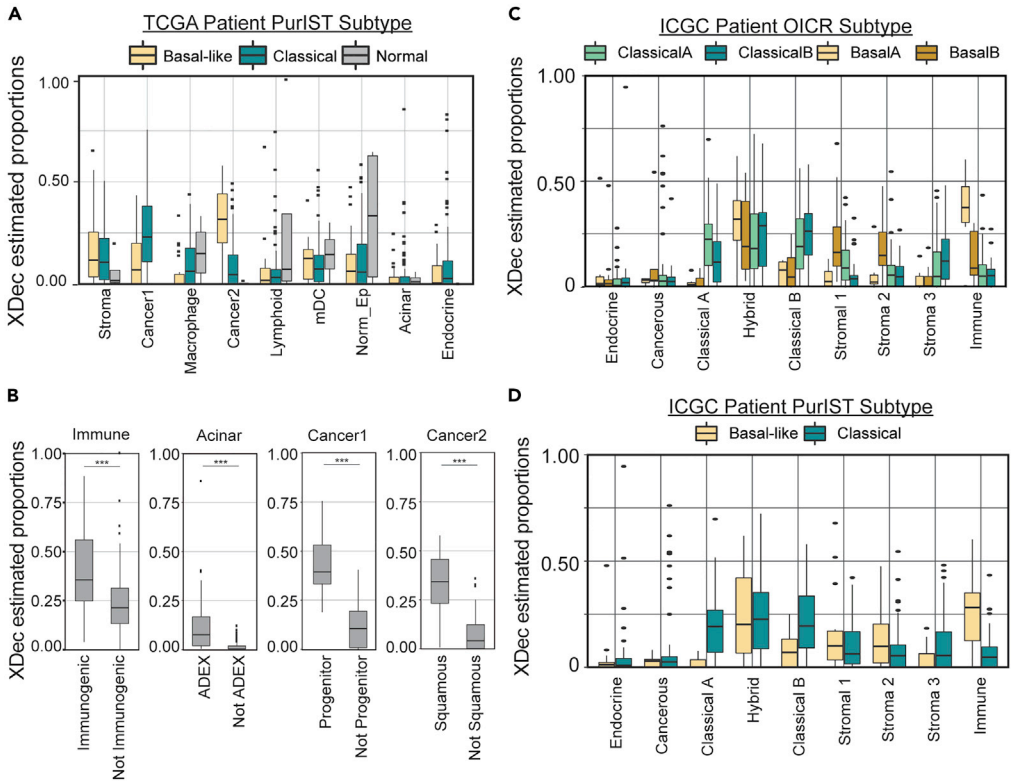


Figure 3. Estimated proportions can identify subtype of cancer cells

Per sample proportion of each TCGA cell type in patients separated by (A) PurIST subtype and (B) Bailey subtypes. Estimated per sample proportion of each ICGC cell type separated by (C) OICR subtype, (D) PurIST subtype. Error bars extend to 1.5 IQR and all others are depicted as outlier points.

XDec-CHI analysis reveals distinct cell-intrinsic immunosuppressive interactions within PDAC subtypes

We next focused on testing the potential of XDec-CHI to identify cell-intrinsic immunosuppressive interactions in PDAC tumors. Given a recent study that showed MSLN and NF-κB as drivers of PD-L1 (Liu et al.), we investigated their cell-type-specific expression of these genes. *In vitro* experiments (Liu et al.) suggested the *MSLN* high subtype (Classical) to also be *CD274* (PD-L1) high, but this did not turn not to be the case *in vivo* (Figures 4A and 4B). We used XDec-CHI to deconvolute the studies using a sliding window approach, which created a single cell-type-specific expression profile for each set of 40 samples. Given multiple measures of expression for each cell type, we could then study gene relationships and potential interactions using correlation analysis. XDec-CHI can be applied to any dataset where cell type proportions and bulk expression are known but we moved forward with XDec as it was able to provide subtype-specific profiles (Figure S5A).

We focused on the relation of cell-type-specific expression of *MSLN* and *RELA (p65, NF-κB)* to immunosuppression via amplification of *CD274* (PD-L1). In the Basal-like tumors, *MSLN* correlated with *RELA* ($R = 0.78$) (Figure 4C) but not with *CD274* (PD-L1) (Figure 4D) that was confirmed in two separate cohorts (Figures S6A, S6B, S6G, and S6I). Whereas this was at odds with *in vitro* studies (Liu et al.), we noted that *CD274* was already upregulated in Basal-like tumors that were shown to be IFNγ high (Espinet et al., 2020; Laise et al., 2020) and express markers of *CD274* high tumors (Asgarova et al., 2018) (Figures 4B and 4G). These results suggest that most Basal-like tumors would still be responsive to therapies targeting PD-L1 and may possibly be synergistic with therapies targeting the IFNγ but not MSLN and RELA pathways.

In contrast to Basal tumors, Classical tumors did not show a correlation between *MSLN* and *RELA*. Like Basal tumors, Classical tumors also lacked a correlation between *MSLN* and *CD274* (Figures 4E and 4F). We asked if the lack of correlations may be related to the recent results suggesting that Classical subtype

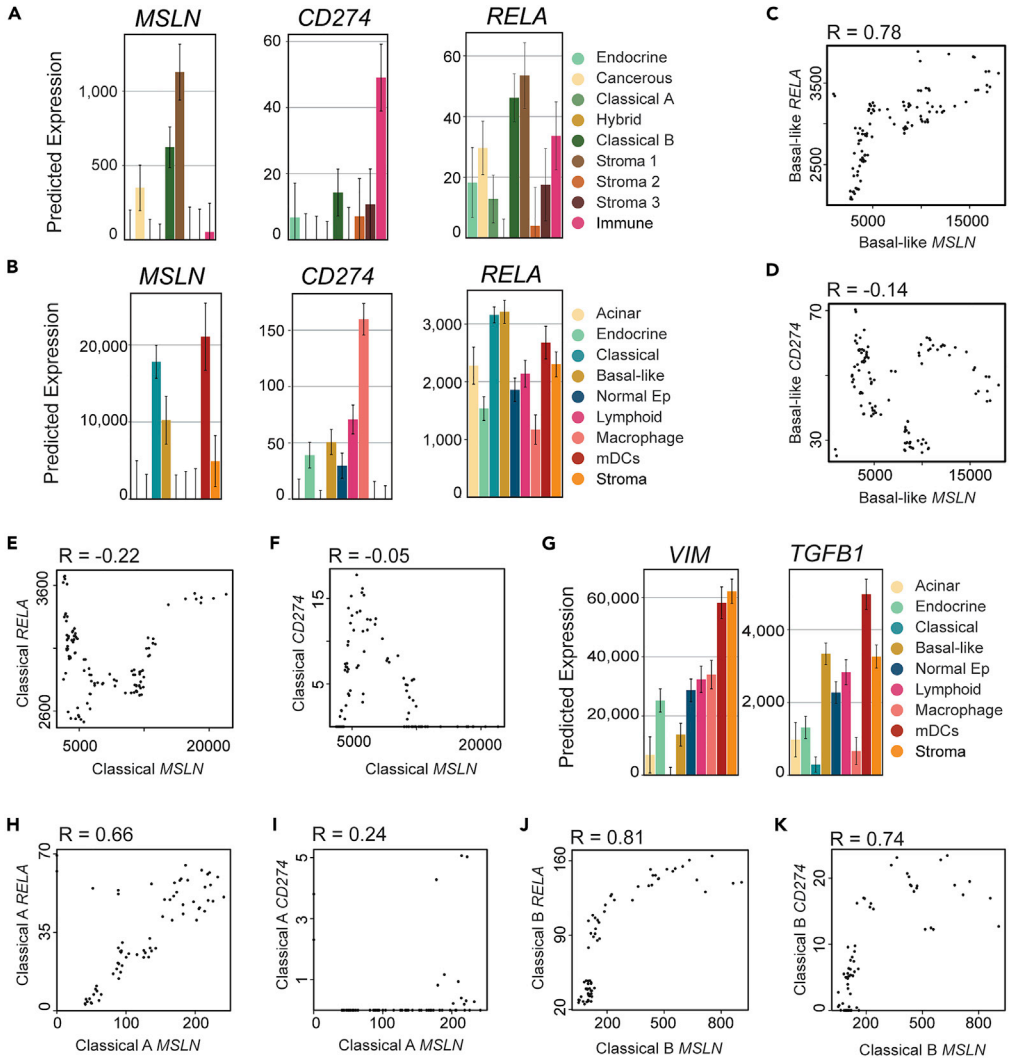


Figure 4. *MSLN* and *CD274* have different relationships in Classical A and Classical B

(A) XDec estimated expression of *MSLN*, *CD274*, and *RELA* in ICGC. Error bars report SE.

(B) XDec estimated expression of *MSLN*, *CD274*, and *RELA* in TCGA. Error bars report SE.

(C) Correlation of Basal-like *MSLN* and *RELA*, (D) *CD274*.

(E) Correlation of Classical *MSLN* and *RELA*, (F) *CD274*.

(G) Estimated expression of genes associated with *CD274* expression.

(H) Correlation of Classical A *MSLN* and *RELA*, (I) *CD274*.

(J) Correlation of Classical B *MSLN* and *RELA*, (K) *CD274*.

is heterogeneous, consisting of Classical A and Classical B subtypes. To address this question we turned to the ICGC tumor profile collection where the Classical A and Classical B subtypes were initially discovered (Chan-Seng-Yue et al., 2020). XDec-CHI analysis of the ICGC collection indeed revealed strong correlation between Classical A *MSLN* and *RELA* but estimated *CD274* expression remained near zero suggesting immune suppression in Classical A tumors by a non-PD-L1 mechanism (Figures 4H and 4I). In contrast, we saw strong positive correlation between Classical B *MSLN*, *RELA*, and *CD274* (*RELA* $R = 0.81$, *CD274* $R = 0.74$) (Figures 4J and 4K). These results highlight potentially distinct immuno-suppressive pathways in Classical A and Classical B subtypes.

We next asked if concordant results can also be obtained by analyzing the TCGA PDAC tumor collection. Consistent with the TCGA collection consisting mostly of the Classical A subtype, the Classical TCGA deconvolution profile appears to be an amalgamation of a majority of Classical A and a minority of Classical B

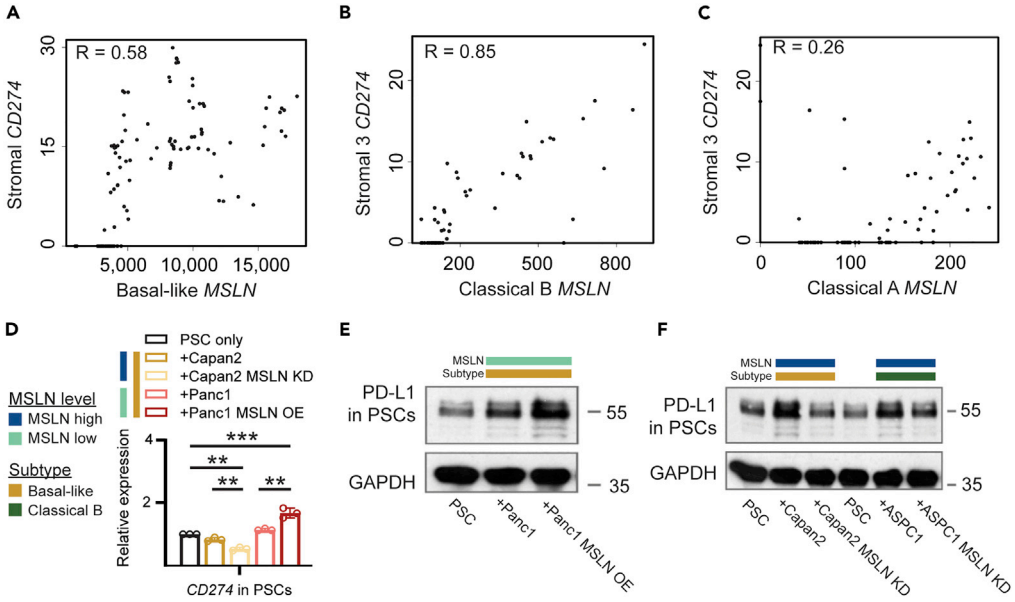


Figure 5. Cancer cell *MSLN* induces stromal *CD274*

Correlation between (A) Basal-like *MSLN* and stromal *CD274*, (B) Classical B *MSLN* and stroma 3 *CD274*, (C) Classical A *MSLN* and stromal *CD274*.

(D) qRT-PCR for stromal *CD274* co-cultured in cancer cells with and without *MSLN* ($n = 3$ replicates). Error bars are standard deviation.

(E) Western blot for stromal PD-L1 in stromal cells alone, stromal cells cultured with cancer cells without and overexpressing *MSLN*.

(F) Western blot for stromal PD-L1 in stromal cells alone, stromal cells cultured with cancer cells with *MSLN* overexpressed knocked down. ANOVA, $^{**}p < 0.01$, $^{***}p < 0.0001$.

with a lack of *MSLN, RELA,* and *CD274* correlation. Whereas we could not validate the *MSLN-CD274* Classical B relationship using the TCGA collection owing to a lack of Classical B profile, we still identified a Classical B specific signal: genes that correlate strongly with CD274 in the Classical TCGA (R > 0.7, 1,258 genes) profile show enrichment for genes upregulated in Classical B tumors compared with Classical A tumors (26/275 gene overlap, p <0.05, OR = 1.6) suggesting the same relationship between Classical B and *CD274* observed in the ICGC collection also exists in TCGA.

XDec-CHI explores the contribution of heterotypic interactions toward an immuno-suppressive microenvironment

Because of the high proportion of stomal cells in PDAC tumors and the known stromal involvement in supporting tumor growth, we performed gene expression correlations between cancer and stromal cells. We found that stromal *CD274* correlated with cancer cell *MSLN* in all subtypes (Classical An $R = 0.26$; Basal-like $R = 0.58$; Classical B $R = 0.85$) (Figures 5A–5C and S6K). As this potentially immuno-suppressive interaction has not been previously observed, we turned to cancer and stromal cell lines in a cancer-stromal co-culture system to validate this interaction. Using Basal-like and Classical B cell lines, we manipulated the amount of *MSLN* present in cancer cell lines and observed changes in pancreatic stellate cells (PSCs) at the RNA and protein level in the co-culture system. *MSLN* was knocked down using shRNA in *MSLN* high cell lines and over-expressed in *MSLN* low cell lines. By increasing *MSLN* in cancer cells we significantly increased stromal *CD274*/PD-L1 expression. The reverse direction could also be validated – by removing *MSLN* from cancer cells we significantly decreased stromal *CD274*/PD-L1 expression (Figures 5D–5F). These results suggest that cancer cell MSLN is necessary and sufficient for stromal PD-L1 expression on the RNA and protein level in both Classical B and Basal-like cell lines. These results further suggest that MSLN may be a good candidate for combination immuno-therapy as it may control stromal PD-L1 expression in Classical B and Basal-like tumors.

We next attempted to identify the gene that may mediate amplification of stromal *CD274* expression upon up-regulation of MSLN in cancer cells. Toward this end, we identified all genes that strongly correlated with

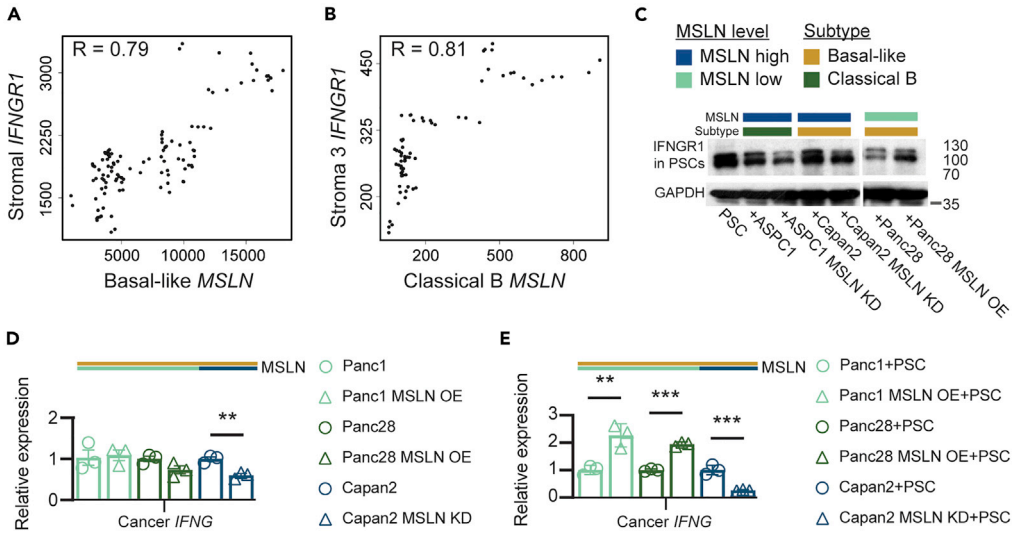


Figure 6. Cancer cell *MSLN* induces stromal *IFNGR1* and *IFNG*

Correlation between (A) Basal-like cancer cell *MSLN* and stromal *IFNGR1* in TCGA and (B) Classical B *MSLN* and stroma 3 *IFNGR1* in ICGC.

(C) Western blot for *IFNGR1* in stromal cells when co-cultured with endogenously MSLN high-expressing cancer cells and *MSLN* knocked down (AsPC1 and Capan2 cells), or endogenously MSLN low-expressing Panc28 cancer cells and overexpresses MSLN.

(D) qRT-PCR for cancer cell *IFNG* expression in PDAC cancer cell lines with *MSLN* overexpressed or knocked down. Error bars are standard deviation.

(E) qRT-PCR for cancer cell *IFNG* expression in PDAC cancer cell lines with *MSLN* overexpressed or knocked down when co-cultured with PSC stromal cells. *T*-test, Error bars are standard deviation. $^{**}p < 0.01$, $^{***}p < 0.0001$.

stromal *CD274* and cancer cell *MSLN* and *RELA* expression (R > 0.6). We identified *IFNGR1* as the only gene present in both sets (TCGA $R = 0.79$; ICGC $R = 0.81$) (Figures 6A and 6B). IFNGR1, a receptor for IFNγ that is known to upregulate PD-L1, has been shown to correlate with PD-L1 expression after activation (Garcia-Diaz et al., 2017). To validate this interaction we again employed the cell co-culture system. We found that stromal IFNGR1 protein levels indeed correlate with cancer cell MSLN levels (Figure 6C).

The increase of IFNGR1 expression would not effectively mediate heterotypic interaction that results in stromal CD274 expression without the presence of the IFNγ ligand. We therefore hypothesized that IFNγ may be produced by cancer cells when co-cultured with PSC cells. We indeed found expression of IFNγ in cancer cell lines (Figure 6D). Moreover, IFNγ was significantly upregulated upon *MSLN* over-expression and down-regulated upon MSLN knockdown suggesting a role for MSLN in regulating IFNγ expression (Figure 6E). These results are consistent with the role of IFNGR1 in mediating the upregulation of stromal PD-L1 upon the activation of MSLN expression. Moreover, the upregulation of IFNγ itself by MSLN suggests a synergistic effect on the upregulation of stromal PD-L1 (Figure 6E).

DISCUSSION

The high failure rate of clinical trials for therapies of cancer and other diseases calls for a critical examination of current strategies for drug discovery. Approximately 10% of drugs succeed in attaining approval and a large portion of these drugs fail owing to a lack of efficacy in humans (Kola and Landis, 2004; Mak et al., 2014; Takebe et al., 2018). Whereas the stage at which drugs drop out varies widely, oncology therapies have one of the lowest success rates at 5% with 70% of therapies failing in stage II clinical trials (Kola and Landis, 2004). One possible source of failure of a candidate drug may be traced to differences between pre-clinical cell-line or mouse models and tumors *in vivo*. Specifically, cell lines propagated *in vitro* do not always model the microenvironment that a cancer cell interacts with *in vivo*. Moreover, mouse models may not be relevant for all subtypes of human tumors (Murillo et al., 2022). The XDec-CHI method enables a bedside-to-bench strategy where relevant interactions are identified *in vivo* before being modeled experimentally. This new strategy will help focus the experimental modeling efforts on the interactions that are relevant for tumors *in vivo*.

The large heterogeneity of cancer cells and drug resistance mechanisms calls for strategies targeted to specific subtypes of tumors. Toward this end, we demonstrate the power of XDec-CHI to identify PDAC subtype-specific intracellular and heterotypic interactions that may lead to immunotherapy resistance. Our results suggest that Basal-like and *MSLN* high Classical B patients will be most likely to respond to therapies targeting PD-L1 and PD-1. Because we find low to no cancer cell expression of PD-L1 in Classical A tumors despite the correlation between *MSLN* and *RELA*, the question is what represses *CD274* in Classical A tumors. Such mechanisms may even point to new approaches to silence PD-L1 in other cancer cells.

We also find that MSLN plays a role in stromal PD-L1 expression via IFNGR1 and IFNγ. Given the large proportion of stromal cells generally present in PDAC, stromal expression of PD-L1 may also affect T cell activation and response to therapeutics. However, we did not determine the mechanism by which cancer cell MSLN affects stromal IFNGR1 expression. Future work should focus not only on identifying the mechanism for cancer cell MSLN to affect stromal IFNGR1 but also identify markers for this interaction.

Taken together, our analysis highlights the potential importance of tailoring therapy to subtypes of PDAC. Only the most recent clinical trials have included stratification by subtype. For example, the Pancreatic Cancer Action Network announced in 2020 they will be applying PurIST to their Precision Promise clinical trial and the PANCREAS trial will treat patients based on subtype. However, PurIST only separates Classical and Basal-like patients. Given our discovery of differences in expression of *CD274* in Classical A and Classical B patients, it will be important going forward to develop additional methods to subtype samples into sub-classes in a single sample manner. Having this information would also be important for retrospective clinical trial analysis and the correct identification of cell lines and patient-derived xenograft models. Subtyping PDAC models will allow knowledge derived from model organisms to be better applied to primary tumors.

Our analyses critically depend on the power of XDec to deconvolute PDAC subtype-specific cancer profiles. This is clearly the key enabling factor in discovering subtype-specific gene–gene interactions and resistance pathways. Not surprisingly, the power to deconvolute subtypes also depends on the dataset itself, as apparent from the summary table (Figure S6K) that indicates interactions discovered across datasets examined here. One illustrative example is the deconvolution of Classical A and B in the ICGC cohort that enabled the discovery of Classical-B-specific interactions such as *MSLN-CD274*. That interaction could not be identified in the TCGA dataset as XDec deconvoluted only one Classical profile. One plausible explanation is that the interaction could not be detected because of the "dilution" of the subtype-specific correlation pattern in the heterogeneous sample population.

Our results also highlight several limitations of the power of the method to identify all links along a causative pathway. Information theory tells us that the correlation of the node at the end of a pathway with the node at the beginning of the same pathway cannot exceed the smallest pairwise correlation along the causative path connecting the two. Therefore, we should expect that any interactions that are between cancer cell MSLN and stromal IFNGR1 would be stronger than or equal to correlations to MSLN. However, as our simulations show, correlation strength is only one of several factors that define the power of our method to detect them. Other important factors affecting the detection power include the proportion of the relevant cell type, number of relevant samples in the dataset, and noise levels in the measurement of gene expression. Despite the power being limited by these factors, our successful validation experiments indicate a low false discovery rate of the method and thus high utility of our correlation analyses for generating hypotheses that validate experimentally.

To ensure the reproducibility of the PDAC findings and enable the wide application of the method by the community without requiring software or data downloads, we created a web application backed by five pre-deconvoluted cancer datasets (PDAC, BRCA, HNSC, GBM, and Glioma). To enable the application of the method to other datasets, we released an R package. Taken together, our tools, methods, and their validation in the context of PDAC will help decipher the contribution of intracellular and heterotypic interactions within complex tissues to human health and disease.

Limitations of the study

As with all deconvolution algorithms, the major limitation is sample number – the authors recommend using windows of 40 samples, which means to return at least three measures of cell-type-specific expression for correlation 42 samples should be included per dataset. In addition, single-cell references are used to select genes for the optional Stage 0 of XDec deconvolution – single-cell sequencing can be expensive

and finding datasets for less studied diseases may be difficult. However, the sliding window deconvolution functions after the second stage of deconvolution, so as long as the criteria are met (cell type proportions have been estimated and enough samples are present) the method will return cell-type-specific values.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Co-culture of pancreatic cancer cells and PSCs
- METHOD DETAILS
  - XDec methodology
  - Sliding window deconvolution and cell types specific correlations
  - Simulations
  - Processing RNA-seq datasets
  - Correlations to outside methods
  - Generation of summary table
  - Classical B gene enrichment in genes correlated to CD274
  - Classical A and classical B subtype identification
  - qRT-PCR
  - Immunoblotting assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105249.

ACKNOWLEDGMENTS

The authors also thank Julie M. Wilson at the Ontario Institute for Cancer Research who helped access the data, Andrew Jackson for resource management, and Oscar Murillo for beta testing the web applications. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government. This project was partially supported by NIH R01 CA183984, VA Merit Award CX001822, and BCM Dan L Duncan Cancer Center Pilot Project Award P30 CA125123 to Dr Yao. This publication was also supported in part by the NIH Common Fund, including NCATS UG3TR002881, NIDA U54DA036134, and NIDA U54DA049098 awarded to Dr Milosavljevic. This study was conducted with the support of the Ontario Institute for Cancer Research through funding provided by the Government of Ontario.

AUTHOR CONTRIBUTIONS

E.L.L. - Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing, Visualization. D.L. - Methodology, Validation, Formal Analysis, Writing. V.P. - Methodology, Software, Writing. Q.Y. - Conceptualization, Resources, Writing, Supervision, Funding Acquisition. A.M. - Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Writing, Supervision, Project Administration, Funding Acquisition.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: January 18, 2022
Revised: July 27, 2022
Accepted: September 26, 2022
Published: October 21, 2022

REFERENCES

Aran, D., Sirota, M., and Butte, A.J. (2015). Systematic pan-cancer analysis of tumour purity. Nat. Commun. *6*, 8971. https://doi.org/10.1038/ncomms9971.

Asgarova, A., Asgarov, K., Godet, Y., Peixoto, P., Nadaradjane, A., Boyer-Guittaut, M., Galaine, J., Guenat, D., Mougey, V., Perrard, J., et al. (2018). PD-L1 expression is regulated by both DNA methylation and NF-kB during EMT signaling in non-small cell lung carcinoma. OncoImmunology *7*, e1423170. https://doi.org/10.1080/2162402x.2017.1423170.

Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.-M., Gingras, M.-C., Miller, D.K., Christ, A.N., Bruxner, T.J.C., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52. https://doi.org/10.1038/nature16965.

Andrade Barbosa, B., van Asten, S.D., Oh, J.W., Farina-Sarasqueta, A., Verheij, J., Dijk, F., van Laarhoven, H.W.M., Ylstra, B., Garcia Vallejo, J.J., van de Wiel, M.A., and Kim, Y. (2021). Bayesian log-normal deconvolution for enhanced in silico microdissection of bulk gene expression data. Nat. Commun. *12*, 6106. https://doi.org/10.1038/s41467-021-26328-2.

Carrero, I., Liu, H.-C., Sikora, A.G., and Milosavljevic, A. (2019). Histoepigenetic analysis of HPV- and tobacco-associated head and neck cancer identifies both subtype-specific and common therapeutic targets despite divergent microenvironments. Oncogene *38*, 3551–3568. https://doi.org/10.1038/s41388-018-0659-4.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421. https://doi.org/10.1038/nbt.2203.

Chan-Seng-Yue, M., Kim, J.C., Wilson, G.W., Ng, K., Figueroa, E.F., O'Kane, G.M., Connor, A.A., Denroche, R.E., Grant, R.C., McLeod, J., et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat. Genet. *52*, 231–240. https://doi.org/10.1038/s41588-019-0566-9.

Decamps, C., Privé, F., Lutsik, P., Jost, D., Waguet, A., HADACA consortium, Lurie, E., Bacher, R., Milosavljevic, A., Scherer, M., et al. (2020). Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software. BMC Bioinf. *21*, 16. https://doi.org/10.1186/s12859-019-3307-2.

Dijk, F., Veenstra, V.L., Soer, E.C., Dings, M.P.G., Zhao, L., Halfwerk, J.B., Hooijer, G.K., Damhofer, H., Marzano, M., Steins, A., et al. (2020). Unsupervised class discovery in pancreatic ductal adenocarcinoma reveals cell-intrinsic mesenchymal features and high concordance between existing classification systems. Sci. Rep. *10*, 337. https://doi.org/10.1038/s41598-019-56826-9.

Espinet, E., Gu, Z., Imbusch, C.D., Giese, N.A., Buscher, M., Safavi, M., Weisenburger, S., Klein, C., Vogel, V., Falcone, M., et al. (2020). Aggressive PDACs show hypomethylation of repetitive elements and the execution of an intrinsic IFN program linked to a ductal cell-of-origin. Cancer Discov. CD-20-1202. https://doi.org/10.1158/2159-8290.cd-20-1202.

Garcia-Diaz, A., Shin, D.S., Moreno, B.H., Saco, J., Escuin-Ordinas, H., Rodriguez, G.A., Zaretsky, J.M., Sun, L., Hugo, W., Wang, X., et al. (2017). Interferon receptor signaling pathways regulating PD-L1 and PD-L2 expression. Cell Rep. *19*, 1189–1201. https://doi.org/10.1016/j.celrep.2017.04.031.

Goldman, M., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2019). The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. Preprint at bioRxiv. https://doi.org/10.1101/326470.

Jacobetz, M.A., Chan, D.S., Neesse, A., Bapiro, T.E., Cook, N., Frese, K.K., Feig, C., Nakagawa, T., Caldwell, M.E., Zecchini, H.I., et al. (2013). Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer. Gut *62*, 112–120. https://doi.org/10.1136/gutjnl-2012-302529.

Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov. *3*, 711–715. https://doi.org/10.1038/nrd1470.

Laise, P., Turunen, M., Maurer, H.C., Curiel, A.G., Elyada, E., Schmierer, B., Tomassoni, L., Worley, J., Alvarez, M.J., Kesner, J., et al. (2020). Pancreatic ductal adenocarcinoma comprises coexisting regulatory states with both common and distinct dependencies. Preprint at bioRxiv. https://doi.org/10.1101/2020.10.27.357269.

Li, M., Bharadwaj, U., Zhang, R., Zhang, S., Mu, H., Fisher, W.E., Brunicardi, F.C., Chen, C., and Yao, Q. (2008). Mesothelin is a malignant factor and therapeutic vaccine target for pancreatic cancer. Mol. Cancer Ther. *7*, 286–296. https://doi.org/10.1158/1535-7163.mct-07-0483.

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J.S., Li, B., and Liu, X.S. (2017). TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res. *77*, e108–e110. https://doi.org/10.1158/0008-5472.can-17-0307.

Liu, D., Poteet, E., Liang, Z., LaPlante, E.L., Brubaker, L., Dhingra, S., Milosavljevic, A., Chen, C., and Yao, Q. Mesothelin exerts an immunosuppressive effect through regulating PD-L1 and remodeling tumor microenvironment in pancreatic ductal adenocarcinoma.

Lucero, R., Zappulli, V., Sammarco, A., Murillo, O.D., Cheah, P.S., Srinivasan, S., Tai, E., Ting, D.T., Wei, Z., Roth, M.E., et al. (2020). Glioma-derived miRNA-containing extracellular vesicles induce angiogenesis by reprogramming brain endothelial cells. Cell Rep. *30*, 2065–2074.e4. https://doi.org/10.1016/j.celrep.2020.01.073.

Lurie, E., Liu, D., LaPlante, E.L., Thistlethwaite, L.R., Yao, Q., and Milosavljevic, A. (2020). Histoepigenetic analysis of the mesothelin network within pancreatic ductal adenocarcinoma cells reveals regulation of retinoic acid receptor gamma and AKT by mesothelin. Oncogenesis *9*, 62. https://doi.org/10.1038/s41389-020-00245-3.

Mak, I.W., Evaniew, N., and Ghert, M. (2014). Review Article Lost in translation: animal models and clinical trials in cancer treatment. Am. J. Transl. Res. *6*, 114.

Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G.H., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat. Genet. *47*, 1168–1178. https://doi.org/10.1038/ng.3398.

Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., et al. (2019). exRNA atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. Cell *177*, 463–477.e15. https://doi.org/10.1016/j.cell.2019.02.018.

Murillo, O.D., Petrosyan, V., LaPlante, E.L., Dobrolecki, L.E., Lewis, M.T., and Milosavljevic, A. (2022). XDec Simplex Map of Breast Cancer Cell States Enables Precise Modeling and Targeting of Breast Cancer. https://doi.org/10.1101/2022.07.06.498858.

Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2.

Olive, K.P., Jacobetz, M.A., Davidson, C.J., Gopinathan, A., McIntyre, D., Honess, D., Madhu, B., Goldgraben, M.A., Caldwell, M.E., Allard, D., et al. (2009). Inhibition of hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. Science *324*, 1457–1461. https://doi.org/10.1126/science.1171362.

Onuchic, V., Hartmaier, R.J., Boone, D.N., Samuels, M.L., Patel, R.Y., White, W.M., Garovic, V.D., Oesterreich, S., Roth, M.E., Lee, A.V., and Milosavljevic, A. (2016). Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. Cell Rep. *17*, 2075–2086. https://doi.org/10.1016/j.celrep.2016.10.057.

Özdemir, B.C., Pentcheva-Hoang, T., Carstens, J.L., Zheng, X., Wu, C.-C., Simpson, T., Laklai, H., Sugimoto, H., Kahlert, C., Novitskiy, S.V., et al. (2014). Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with diminished survival. Cancer Cell *25*, 719–734. https://doi.org/10.1016/j.ccr.2014.04.005.

Peng, J., Sun, B.-F., Chen, C.-Y., Zhou, J.-Y., Chen, Y.-S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.-S., et al. (2019a). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. *29*, 725–738. https://doi.org/10.1038/s41422-019-0195-y.

Peng, J., Han, L., and Shang, X. (2021). A novel method for predicting cell abundance based on

single-cell RNA-seq data. BMC Bioinf. *22*, 281. https://doi.org/10.1186/s12859-021-04187-4.

Peng, X.L., Moffitt, R.A., Torphy, R.J., Volmar, K.E., and Yeh, J.J. (2019b). De novo compartment deconvolution and weight estimation of tumor samples using DECODER. Nat. Commun. *10*, 4729. https://doi.org/10.1038/s41467-019-12517-7.

Provenzano, P.P., Cuevas, C., Chang, A.E., Goel, V.K., Von Hoff, D.D., and Hingorani, S.R. (2012). Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. Cancer Cell *21*, 418–429. https://doi.org/10.1016/j.ccr.2012.01.007.

Raphael, B.J., Hruban, R.H., Aguirre, A.J., Moffitt, R.A., Yeh, J.J., Stewart, C., Robertson, A.G., Cherniack, A.D., Gupta, M., Getz, G., et al. (2017). Integrated genomic characterization of pancreatic ductal adenocarcinoma. Cancer Cell *32*, 185–203.e13. https://doi.org/10.1016/j.ccell.2017.07.007.

Rashid, N.U., Peng, X.L., Jin, C., Moffitt, R.A., Volmar, K.E., Belt, B.A., Panni, R.Z., Nywening, T.M., Herrera, S.G., Moore, K.J., et al. (2020). Purity independent subtyping of tumors (PurIST), A clinically robust, single-sample classifier for tumor subtyping in pancreatic cancer. Clin. Cancer Res. *26*, 82–92. https://doi.org/10.1158/1078-0432.ccr-19-1467.

Rhim, A.D., Oberstein, P.E., Thomas, D.H., Mirek, E.T., Palermo, C.F., Sastra, S.A., Dekleva, E.N., Saunders, T., Becerra, C.P., Tattersall, I.W., et al. (2014). Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer Cell *25*, 735–747. https://doi.org/10.1016/j.ccr.2014.04.021.

Takebe, T., Imai, R., and Ono, S. (2018). The current status of drug discovery and development as originated in United States academia: the influence of industrial and academic collaboration on drug discovery and development. Clin. Transl. Sci. *11*, 597–606. https://doi.org/10.1111/cts.12577.

Topham, J.T., Karasinska, J.M., Lee, M.K.C., Csizmok, V., Williamson, L.M., Jang, G.H., Denroche, R.E., Tsang, E.S., Kalloger, S.E., Wong, H.L., et al. (2021). Subtype-discordant pancreatic ductal adenocarcinoma tumors show intermediate clinical and molecular characteristics. Clin. Cancer Res. *27*, 150–157. https://doi.org/10.1158/1078-0432.ccr-20-2831.

Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. *10*, 380. https://doi.org/10.1038/s41467-018-08023-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PD-L1	Cell Signaling Technology	13684S
IFNGR1	Abcam	Ab134070
GAPDH	Sigma	G8795
Experimental models: Cell lines		
Human pancreatic stellate cells	Dr. Rosa F. Hwang	PSC
Human Panc28 (female)	Dr. Craig Logsdon	RRID:CVCL_3917
Human ASPC1 (female)	ATCC	RRID:CVCL_0152
Human Capan2 (male)	ATCC	RRID:CVCL_0026
Human HPAFII (male)	ATCC	RRID:CVCL_0313
Human MIA PaCa2 (male)	ATCC	RRID:CVCL_0428
Human Panc 1 (male)	ATCC	RRID:CVCL_0428
Oligonucleotides		
Primers: CD274 forward TGGCATTTGCTGAACGCATTT reverse TGCAGCCAGGTCTAATTGTTTT	This paper	N/A
Primers: IFNGR1 forward AGCAGGAAGTCGATTATGA TCCC reverse CTGGCACTGAATCTCGTCACA	This paper	N/A
Primers: IFNG forward TCGGTAACTGACTTGAATGTC CA reverse TCGCTTCCCTGTTTTAGCTGC	This paper	N/A
Primers: GAPDH forward TCGACAGTCAGCCGCATCT reverse CCGTTGACTCCGACCTTCA	This paper	N/A
Software and algorithms		
R 4.0.0	R Core Team, 2020	https://www.R-project.org/
XDec-CHI Code	This paper	https://github.com/BRL-BCM/XDec_CHI
XDec-CHI Web instance	This paper	https://brl-bcm.shinyapps.io/XDec-CHI_Homepage
Other		
TCGA: PAAD	Raphael et al. (2017)	https://xenabrowser.net/datapages/
ICGC: PACA-CA	Chan-Seng-Yue et al. (2020)	https://ega-archive.org/studies/EGAS00001002543
ICGC: PACA-AU	ICGC DCC	https://dcc.icgc.org/
Dijk	EMBL-EBI	https://www.ebi.ac.uk/ena/browser/view/PRJEB27026

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Aleksandar Milosavljevic (amilosav@bcm.edu).

Material availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

All original code has been deposited at https://github.com/BRL-BCM/XDec_CHI and is publicly available as of the date of publication. A web resource which implements the method is available https://brl-bcm.shinyapps.io/XDec-CHI_Homepage. DOIs are listed in the key resources table

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

Human pancreatic stellate cells (PSC) were generously given by Dr. Rosa F. Hwang, Panc28cells were obtained from Dr. Craig Logsdon from MD Anderson Cancer Center (Houston, TX). PDAC cell lines ASPC1, Capan2, HPAFII, MIA PaCa2, and Panc1 (were purchased from ATCC and maintained by the Yao lab. All cells were authenticated by short tandem repeat profiling and were propagated for less than 6 months after resuscitation. The MSLN overexpression stable cell lines of Panc1 and Panc28, as well as MSLN shRNA knockdown cell lines of ASPC1 and Capan2 were established and maintained as previously described ((Li et al., 2008; Lurie et al., 2020). All cell lines were cultured under standard culture conditions (5% CO2, at 37°C) in culture media (ASPC1 in RPMI-1640, Capan2 in McCOY's 5A, PSC, HPAFII, MIA PaCa2, Panc1, and Panc28 cells in DMEM media) supplemented with 10% FBS and 1% penicillin/streptomycin.

Co-culture of pancreatic cancer cells and PSCs

Human PSC cells (1–10×10^5 cells/well) were seeded in 24 well or 6 well culture plates (BD Bioscience) in DMEM medium supplemented with 10% FBS, 1% penicillin/streptomycin. 1:1 proportion of PDAC cells (1–10×10^5 cells/culture insert) were seeded into the culture inserts of 3 μm pore size (BD Bioscience) in appropriate media supplemented with 10% FBS, 1% penicillin/streptomycin. The culture inserts seeded with PDACs were placed into the 24/6-well plates containing PSC cells, and incubation was conducted for 2 days under standard culture conditions (5% CO2, at 37°C).

METHOD DETAILS

XDec methodology

XDec is a reference-optional method which functions in three stages: Stage 0 (optional) identifies informative genes that are differentially expressed between cell types from scRNA-seq data. Any set of cell type specific genes may be provided if users do not want to perform unsupervised identification via reference data. Stage 1 uses iterative matrix factorization to identify the constituent cell types and their proportions in each tumor sample, and Stage 2 uses non-negative least squares optimization to estimate the transcriptomic profile for each cell type. XDec was performed as previously described (Murillo et al., 2019, 2022).

$$y = 1 - e^{-ax},\ a = \frac{1}{max\{x\}} \qquad \text{(Equation 1)}$$

where {x} denotes the set of expression levels x for a specific gene across all samples.

PDAC single cell RNA-seq profiles were downloaded from GSA: CRA001160 (Peng et al., 2019a). Metadata from the study gave cell type identity for each cell. Profiles with low coverage (<300 genes/cell) and cells with coverage in the bottom quartile were removed. For each cell type, cells were ranked by total coverage and then pseudo-bulk profiles were created by summing every 5 profiles. Pseudo-bulk depth was then normalized to match the sample with the highest coverage and transformed into a 0-1 range using Equation 1. 100 pseudo-bulks were created for all cell types resulting in 11 cell type references. Endothelial 1 pseudo-bulks were then removed due to the total Endothelial cell count being over-represented.

Informative genes were chosen by performing the one.vs.rest command in *EDec* R package. 20 up- and downregulated genes were selected for each cell type at $P < 0.00001$. Acinar and normal epithelial cells proved difficult to separate so an addition 20 up- and down-regulated genes were selected at $P < 0.00001$ by using the each.pair command. Genes chosen for multiple cell types were removed leaving 283 genes (Table S3). 269/283 genes were present in ICGC. Cell stability was estimated using the estimate_stablility function as previously described (Onuchic et al., 2016). 9 cell types were selected for both TCGA and PACA-CA datasets. In the case of the PACA-AU dataset, the most stable cell number of 11 produced multiple profiles with no correlation to reference samples so the second most stable cell number of 9

was selected. The Dijk cohort (Dijk et al., 2020) had the most stable cell number of 11 but produced cancer cell profiles that could be mapped to classification using 12 cell types so 12 was selected.

Sliding window deconvolution and cell types specific correlations

Stage 1 was performed on the entire cohort to estimate cell type proportions. Samples were then ordered by bulk expression of a gene of interest (Figure 2B). Stage 2 was then performed on the groups of 40 samples with the lowest expression of the anchor gene which returned estimated gene expression for all genes across all cell types. The window was then moved by one sample (sample 2-41) and Stage 2 was performed, again returning estimated gene expression. This deconvolute and move strategy was repeated until every sample had been deconvoluted. Each cell type specific profile was then concatenated creating cell type specific gene by deconvolution matrixes. Pearson correlation compared estimated expression for genes across all windows using the appropriate cell types.

Simulations

Simulating varying correlation efficient and coefficients of variation

500 simulated mixtures of 3 cell types were generated by (1) generating proportions of each cell type present in the mixture and (2) multiplying a cell type specific expression by the proportion of that cell present in each of the 500 samples. The proportions associated with each of the 3 cell type were generated for each mixture from one of three Dirichlet distributions as previously described (Onuchic et al., 2016). The first Dirichlet distribution had parameters α = [2,2,6], generating proportions with averages 20% for the first cell type, 20% for the second cell type, and 60% for the third cell type. The second and third Dirichlet distributions were α = [4,3,3] and α = [3,3,4] respectively.

500 profiles for Gene X and Gene Y were generated for each cell type. For cell intrinsic interactions, the expression for Gene X and Gene Y were placed into the same cell type and for heterotypic interactions Gene X was placed in cell type 1 and Gene Y was placed in cell type 2. To model different coefficients of variation, Gene X and Gene Y were set equal and the same random gaussian noise was added to each for a correlation of 1. Gaussian noise with a mean of 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, and 2000 to model different coefficient of variations. To model different correlation coefficients, additional gaussian noise was then added to Gene Y. The amount of noise added varied based on the coefficient of variation in order to approximate correlation coefficients 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 (no noise was added for those with perfect correlation). The cell type specific profiles were then multiplied by the proportions generated from the Dirichlet distribution and summed together to create 500 simulated bulk samples.

These 500 simulated bulk samples were then ordered by Gene X and using the proportions and simulated bulk samples sliding window deconvolution was performed. The spearman correlation between cell type specific Gene X and Gene Y were then calculated. This process was completed 1,000 times for both the cell intrinsic and heterotypic simulations. True positive interactions were considered those correlations which landed within 0.1 in either direction of the true correlation and false negatives were those than landed outside. This allowed for the calculation of sensitivity.

Simulated pair ranks against experimental results

XDec was applied to the ICGC cohort and samples were ordered by MSLN for sliding window deconvolution. All gene correlations to MSLN in the Classical B cell type were determined and the correlation strength at n = 1, top 1%, and top 2% were determined.

500 simulated mixtures of 3 cell types were generated by (1) generating proportions of each cell type present in the mixture and (2) multiplying a cell type specific expression by the proportion of that cell present in each of the 500 samples. Four different sets of proportions were generated – Cell 1 proportion centered around 10%, 25%, 50%, and 90% and noise was added using a Dirichlet distribution as previously described (Onuchic et al., 2016). *Gene X* was then placed in Cell 1 with a mean of 210 and standard deviation of 300 (matching the bulk mean and standard deviation of *MSLN* in the ICGC dataset) and *Gene Y* was also placed in Cell 1 with a mean of 8 and standard deviation of 13 (matching the bulk mean and standard deviation of *CD274* in the ICGC dataset). Gaussian noise was added to *Gene Y* from 1 to 55 standard deviations to get a wide range of true cell intrinsic correlations. The cell type specific profiles were then multiplied by the

proportions generated from the Dirichlet distribution and summed together to create 500 simulated bulk samples for each of the average proportions.

Simulated bulks were ordered by *Gene X* and then sliding window deconvolution was performed. The estimated correlation for *Gene X* and *Gene Y* were recorded. For each rank, we reported the smallest true cell intrinsic correlation which resulted in an estimated correlation that fell into that rank based on the experimental data.

Processing RNA-seq datasets

TCGA PAAD level 3 RNA-seq dataset was downloaded from UCSC Xena (Goldman et al., 2019) for samples not flagged during review (Raphael et al., 2017). ICGC PACA-CA fastq files were downloaded from EGAD00001004548 and were aligned to hg38 using STAR v2.5 and expression was calculated using RSEM v1.3. RSEM counts were used for further analysis. Primary PDAC samples that were present in the ICGC database with survival and OICR subtyping information (Chan-Seng-Yue et al., 2020) were used (Table S2). ICGC PACA-AU PDAC sample expression data (exp_seq.PACA-AU.tsv.gz) was downloaded from the ICGC web portal (dcc.icgc.org/releases). The normalized_read_count column was used.

Correlations to outside methods

CIBERSORTx was performed on TCGA samples using 100 single cell profiles for all cell types. Estimated proportions from CIBERSORTx and XDec were correlated to other methods (Carter et al., 2012; Li et al., 2017; Lurie et al., 2020) using spearman correlation. ABSOLUTE PAAD tumor purity was precomputed and downloaded from Aran, Sirota, and Butte 2015(Aran et al., 2015). TIMER estimated the immune score for 6 immune types and all scores were summed for comparison.

Generation of summary table

Sliding window deconvolution was performed on all datasets and correlations between gene pairs were performed. Any correlation above 0.55 was considered present and designated with a check mark.

Classical B gene enrichment in genes correlated to CD274

Using estimated means and standard deviations from Stage 2 we identified differentially expressed genes with a t-test (Onuchic et al., 2016). For the ICGC Classical B cell type we identified genes with FDR<0.05 and fold change < 1. 275/308 genes identified were present in TCGA. We correlated all genes in the Classical cell type in TCGA to Classical *CD274* and identified 1278 genes which correlated at $R > 0.7$. Fisher's exact test was used to test for enrichment.

Classical A and classical B subtype identification

Expression from individual samples (patient or cell line) were combined with ICGC data and transformed using Equation 1. XDec identified the Classical A and B cell profiles and subtype was assigned based on which cell type was in predominant proportion. Method available via https://brl-bcm.shinyapps.io/XDec_PDAC_subtyping (Using the Single Sample option).

qRT-PCR

Total RNA for PSCs co-cultured with PDACs was extracted by TRIzol reagent (Invitrogen, Thermo Fisher Science, USA) according to the product's manual. Briefly, 1 ml TRIzol reagent was added into samples for 5-min incubation at room temperature, followed by addition of 200 μl chloroform and then the samples were separated by centrifugation. The supernatant was incubated with isopropanol and washed with 75% ethanol. Finally, the total RNA pellet was dissolved with DNase/RNase-free water and used as template for cDNA synthesis. The cDNA synthesis was performed with reverse transcription-PCR kit derived from Bio-Rad, and real-time PCR was performed by standard procedures as previously described (Li et al., 2008; Lurie et al., 2020). CD274 and IFNGR1 measurements were normalized to GAPDH in the PSC cells and relative to controls in the overexpression and knockdown comparisons. Primer sequences (5–3′):Human CD274 forward TGGCATTTGCTGAACGCATTT reverse TGCAGCCAGGTCTAATTGTTTT, Human IFNGR1 forward AGCAGGAAGTCGATTATGATCCC reverse CTGGCACTGAATCTCGTCACA, Human IFNG forward TCGGTAACTGACTTGAATGTCCA reverse TCGCTTCCCTGTTTTAGCTGC, Human GAPDH forward TCGACAGTCAGCCGCATCT reverse CCGTTGACTCCGACCTTCA.

Immunoblotting assay

Cell lysates were prepared using RIPA lysis buffer (89900) with protease inhibitor cocktail (78410) and phosphatase inhibitor cocktail (78420) and quantitated for protein abundance using the Pierce BCA Protein Assay Kit (23225) from Thermo Fisher Scientific. PVDF membranes were blocked in 5% nonfat dry milk/TBST solution and incubated at 1:1000 dilution with: PD-L1 mAb, IFNGR1 mAb, and GAPDH mAb. Immunodetection was performed using SuperSignal West Pico (34577) and Femto (34095) Chemiluminescent Substrate according to the manufacturer's instructions (Thermo Fisher Scientifc, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

P values were adjusted for multiple comparisons using the Benjamini & Hochberg method. The qRT-PCR results were expressed as mean ± SD. Experiments were performed at least three times and similar results were obtained. Differences between the groups were evaluated by ANOVA, followed by the Tukey–Kramer test for post hoc analysis. A $P < 0.05$ was considered statistically significant. R 4.0.0 was used for deconvolution and figure generation.