



OPEN

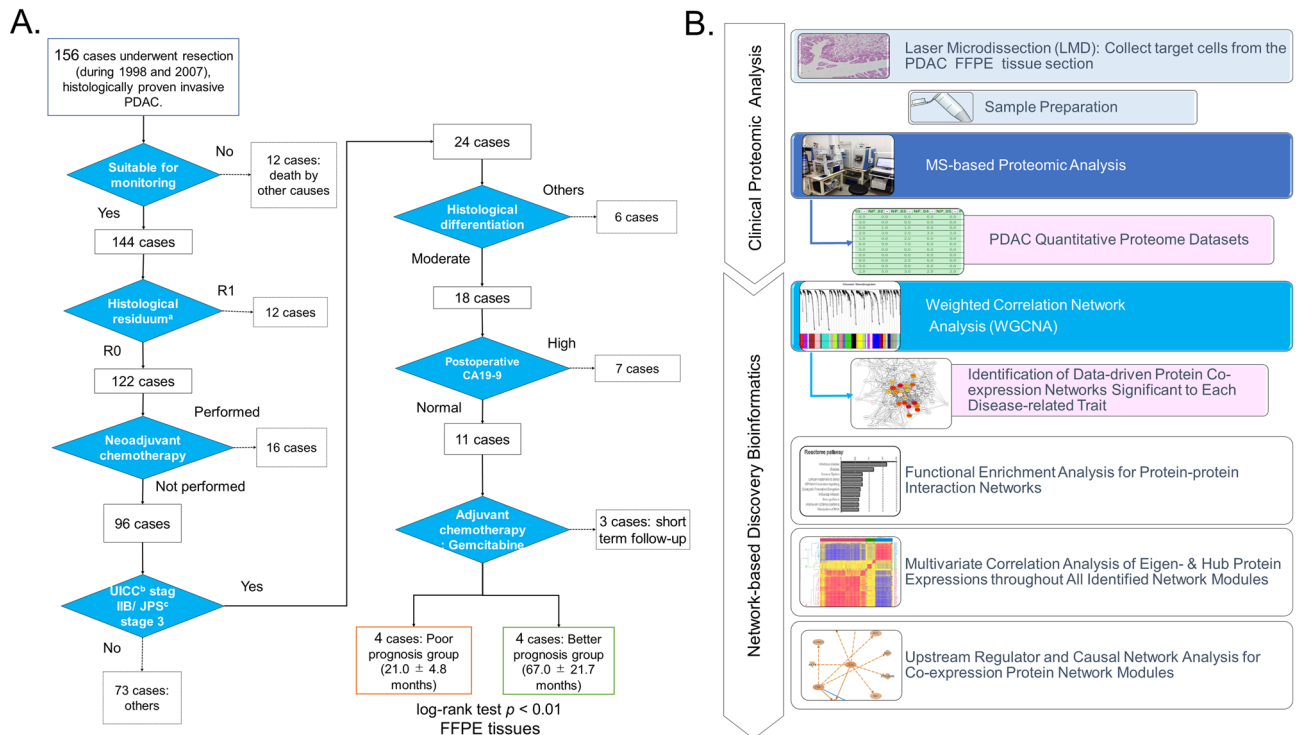
## Disease-related protein co-expression networks are associated with the prognosis of resectable node-positive pancreatic ductal adenocarcinoma

Toshihide Nishimura<sup>1,6,8</sup>✉, Tatsuyuki Takadate<sup>2,8</sup>, Shimpei Maeda<sup>3,8</sup>, Takashi Suzuki<sup>4</sup>, Takashi Minowa<sup>5</sup>, Tetsuya Fukuda<sup>6</sup>, Yasuhiko Bando<sup>6</sup> & Michiaki Unno<sup>2,7</sup>

Pancreatic ductal adenocarcinoma (PDAC) is a multifactorial disease, the molecular profile of which remains unclear. This study aimed at unveiling the disease-related protein networks associated with different outcomes of resectable, node-positive PDAC cases. We assessed laser-microdissected cancerous cells from PDAC tissues of a poor outcome group (POG;  $n = 4$ ) and a better outcome group (BOG;  $n = 4$ ). Noncancerous pancreatic duct tissues ( $n = 5$ ) were used as the reference. We identified four representative network modules by applying a weighted network correlation analysis to the obtained quantitative PDAC proteome datasets. Two network modules that were significant for POG were associated with the heat shock response to hypoxia-related stress; in the latter, a large involvement of the non-canonical Hedgehog pathway (regulated by GLI1), the internal ribosome entry site-mediated cap-independent translation, the inositol requiring enzyme 1- $\alpha$  (IRE1 $\alpha$ )/X-box binding protein 1 pathway of the unfolding protein response (UPR), and the aerobic glycolysis was observed. By contrast, the BOG characteristic module was involved in the inactivation of the UPR pathway via the synoviolin 1-dependent proteasomal degradation of IRE1 $\alpha$ , the activation of SOX2, and the loss of PALB2 (partner and localizer of BRCA2) function, all potentially suppressing malignant tumor development. Our findings might facilitate future therapeutic strategies for PDAC.

Pancreatic cancer is the seventh leading cause of cancer-related deaths worldwide and the most lethal of the malignant ones, with a 5-year survival rate of 9%<sup>1,2</sup>. Pancreatic adenocarcinoma is the most common type of pancreatic cancer (85% of the cases) arising from the exocrine glands of the pancreas, with pancreatic ductal adenocarcinoma (PDAC) being its most common form<sup>3</sup>. Very few PDAC patients (less than 20%) receive curative radical surgical resection, and even in these patients, the recurrence rate is as high as 85%<sup>4</sup>. Patients with the Union for International Cancer Control (UICC) stage IIA (extends to the surrounding organs; node-negative) and IIB (extends to the surrounding organs; node-positive) PDAC are most frequently regarded as candidates for the undertaking of a curative resection<sup>5-7</sup>. UICC stage IIB PDAC still contains wide heterogeneity. Since stage IIB pancreatic cancer is the group with the worst treatment outcome (lymph node metastasis positive) among resectable pancreatic cancers, it is pivotal to unveil the molecular profiles between two PDAC groups with the same histological IIB grade and with the same clinical stage that is treated similarly but resulted differently in better or poor outcomes.

<sup>1</sup>Department of Translational Medicine Informatics, St. Marianna University School of Medicine, Kawasaki, Kanagawa 216-8511, Japan. <sup>2</sup>Department of Surgery, National Hospital Organization Sendai Medical Center, Sendai, Miyagi 983-8520, Japan. <sup>3</sup>Department of Surgery, Saitama Medical Center, Jichi Medical University, Shimotsuke, Tochigi 329-0498, Japan. <sup>4</sup>Department of Pathology and Histotechnology, Tohoku University Graduate School of Medicine, Sendai, Miyagi 980-8574, Japan. <sup>5</sup>Nanotechnology Innovation Station, National Institute for Materials Science, Tsukuba, Ibaraki, Japan. <sup>6</sup>Biosys Technologies, Inc., Tokyo, Tokyo 153-8904, Japan. <sup>7</sup>Department of Surgery, Tohoku University Graduate School of Medicine, Sendai, Miyagi 980-8574, Japan. <sup>8</sup>These authors contributed equally: Toshihide Nishimura, Tatsuyuki Takadate and Shimpei Maeda. ✉email: t-nishimura@marianna-u.ac.jp



**Figure 1.** (A) Strategy followed for the selection of two PDAC groups that have the same histological grade within a clinical-stage and are treated similarly but result in different outcomes. (B) The workflow of the network-based discovery bioinformatic analysis following the clinical proteomic analysis. Notes: <sup>a</sup>histological residuum R0 shows microscopically negative margins; <sup>b</sup>Union Internationalis Contra Cancrum/Union for International Cancer Control (UICC) stage IIB: cancer has spread to nearby lymph nodes and may also have spread to adjacent tissues and organs; <sup>c</sup>Japan Pancreas Society (JPS) stage 3: cancer has not spread into the portal vein, the extra-pancreatic nerve plexus or other organs in UICC stage IIB.

High-accuracy mass spectrometry (MS)-based proteomics has advanced shotgun sequencing and the quantitative analysis of proteins expressed in clinical specimens. The obtained quantitative proteome data can be used to identify key disease-related proteins and therapeutic targets<sup>8</sup>. We have, herein, adopted label-free spectral counting-based semiquantitative MS-based proteomics that was applied on target cells obtained from formalin-fixed paraffin-embedded (FFPE) PDAC tumor specimens by using laser microdissection. The present study aimed at identifying the protein co-expression networks that are significantly associated with a poor and better clinical outcome in patients with resectable, node-positive (UICC stage IIB/JPS stage III) PDAC. Weighted gene co-expression network analysis (WGCNA)<sup>9,10</sup> was applied to quantitative proteome datasets of PDAC. Our case selection strategy and the workflow of the employed network-based discovery bioinformatics analysis (taking place after the MS-based proteomic analysis) are presented in Fig. 1.

## Results

**Proteome datasets of node-positive PDAC.** MS-based proteomic analysis was conducted on FFPE tissue specimens of resectable, node-positive PDAC patients with poor ( $n = 4$ ) and better ( $n = 4$ ) outcomes, based on the patients' survival duration, while, essentially, their clinicopathological backgrounds were otherwise the same. Noncancerous pancreatic duct (NPD) tissue specimens ( $n = 5$ ) were also analyzed as a reference and were obtained from the bile duct and ampulla of Vater carcinoma patients who underwent pancreaticoduodenectomy and whose cancer cells did not extend into the pancreas.

A total of 156 histologically diagnosed PDAC cases underwent pancreatectomy at the Tohoku University Hospital between January 1998 and December 2007. PDAC and TNM were classified according to the JPS (6th edition) and UICC (7th edition) at the time when PDAC cases were selected<sup>11</sup>. Our case selection strategy (Fig. 1A) was the following: (i) we selected 144 cases (by excluding those with perioperative mortality) suitable for monitoring; (ii) we chose 103 cases with microscopically complete resection (R0) and no evidence of para-aortic lymph node metastasis; (iii) among the 87 identified cases that received no neoadjuvant chemotherapy, we focused on those with standardized known prognostic factors, such as pathological stage, histological differentiation, postoperative carbohydrate antigen 19-9 levels, and adjuvant chemotherapy; and (iv) we finally chose eight cases that were subsequently divided into two groups based on the significant difference in their postoperative average survival time as calculated using the Kaplan–Meier method (the log-rank test;  $p = 0.0067$ ): the poor outcome group (POG;  $n = 4$ ;  $21.0 \pm 4.8$  months) and the better outcome group (BOG;  $n = 4$ ;  $67.0 \pm 21.7$  months) (Table 1).

Patient group	Sample ID	Age	Gender	Tumor location	Clinical TNM classification*	UICC stage*	JPS stage*	Differentiation	Residuum	Postoperative CA19-9	Adjuvant chemotherapy	Postoperative survival months
Noncancerous pancreatic duct (NPD) (n = 5)	NPD-01	70	M	Bile duct								
	NPD-02	73	F	Bile duct								
	NPD-03	67	M	Bile duct								
	NPD-04	69	M	Vater								
	NPD-05	75	F	Vater								
				M(60%) / F(40%)								
	Average ± SD	70.8 ± 3.2										
Poor outcome group (POG) (n = 4)	POG-01	53	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	15.9
	POG-03	67	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	17.9
	POG-04	68	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	24.5
	POG-06	79	F	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	25.6
				M(75%) / F(25%)								
	Average ± SD	64.0 ± 10.7										21.0 ± 4.8
Better outcome group (BOG) (n = 4)	BOG-02	74	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	42.8
	BOG-03	57	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	56.6
	BOG-07	48	F	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	76.2
	BOG-08	74	M	Head	T3N1M0	IIB	3	mod	R0	< 37 U/ml	Gemcitabine	92.3
				M(75%) / F(25%)								
	Average ± SD	65.4 ± 12.9										67.0 ± 21.7
Group comparison	t-test p value =	0.691									Log-rank test p value =	0.0067

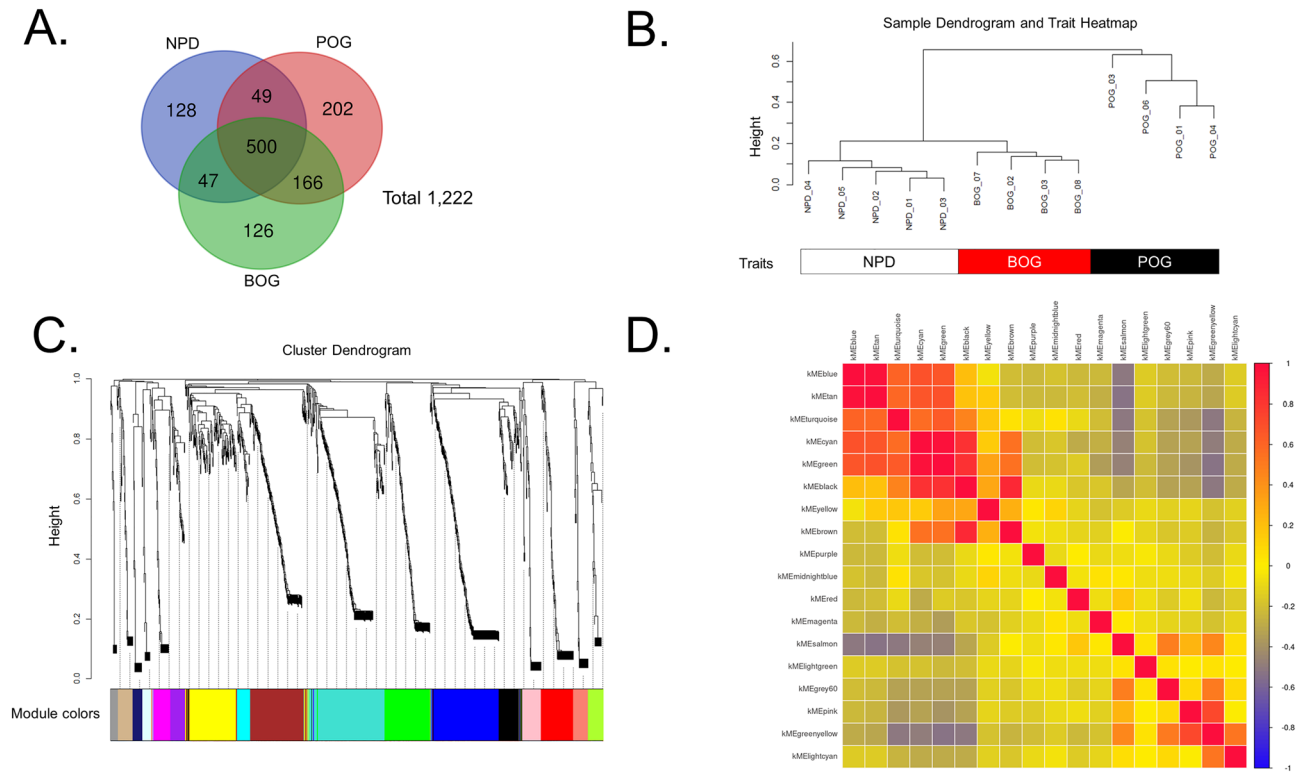
**Table 1.** Clinicopathological information regarding the recruited patients. \*The JPS (6th edition) and UICC (7th edition) were used at the time when PDAC cases were selected<sup>11</sup>.

A total of 1,222 proteins were identified, of which 500 (40.9%) were commonly expressed in the cancerous cells of the POG, the BOG, and the NPD Group. Moreover, 202 (16.5%), 126 (10.3%), and 128 (10.5%) proteins were unique to the POG, the BOG, and the NPD Group, respectively (Fig. 2A).

**Data-driven co-expression protein network by WGCNA.** The hierarchical clustering of the samples according to protein abundance exhibited a clear correlation with the three traits of the POG, the BOG, and the NPD Group (Fig. 2B), in which both the NPD group and the BOG were found to be closely clustered. In particular, 18 protein modules were identified by clustering all identified proteins and constructing weighted protein co-expression networks (Fig. 2C). The WGCNA analysis was performed with a soft threshold power of “8” that was selected to approximate a scale-free topology, a minimum module size of “10,” and a module detection sensitivity (*deepSplit*) of “4.” Correlations between the resultant modules and the traits were obtained to identify the protein modules that were significant to the respective traits. Pairwise correlations between the modules in the connectivity measure (KME) of the module eigen-protein are presented in Fig. 2D.

We identified three significant modules with high and/or moderate correlations ( $r > 0.5$ ) and statistical significances (multiple testing correction using the Benjamini–Hochberg method:  $q < 0.05$ ) with clinical traits (Fig. S1). The WM5 module (green-yellow;  $r = 0.88$ ,  $q = 0.001$ ) was most significantly correlated with the NPD Group. Both the WM7 (black;  $r = 0.76$ ,  $q = 0.025$ ) and WM11 (green;  $r = 0.94$ ,  $q = 2.56 \times 10^{-5}$ ) modules were significantly correlated with POG. Trait correlation analysis often tends to overlook important modules. The four identified WGCNA modules – WM15 (purple), WM16 (red), WM17 (magenta), and WM18 (midnight-blue)—correlated moderately ( $0.4 < r \leq 0.5$ ) with the BOG, but none of these correlations were found to be significant ( $q > 0.05$ ). The statistical over-representative analysis could help in evaluating potential key WGCNA modules with identified proteins uniquely expressed for each trait. The overlaps of the WGCNA-derived modules with 202, 126, and 128 proteins that were found to be uniquely expressed in the POG, the BOG, and the NPD Group, respectively (Fig. 2A), were subsequently assessed by using the over-representation test. Among the four BOG characteristic modules (WM15, WM16, WM17, and WM18), WM16 (red) was the most significant ( $r = 0.49$ ; overlapping  $q = 3.49 \times 10^{-39}$ ) (Fig. S1).

**Functional enrichment analysis of the protein–protein interaction (PPI) networks.** The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database<sup>12</sup> was used to generate the human PPI



**Figure 2.** Gene modules identified by weighted gene co-expression network analysis (WGCNA). **(A)** Venn map of the identified proteins. NPD, the noncancerous pancreatic duct; POG, the poor outcome group; BOG, the better outcome group. **(B)** Sample dendrogram and trait heatmap. **(C)** Protein dendrogram obtained by clustering dissimilarity according to the topological overlap with the corresponding module; the colored rows correspond with the 18 modules identified by dissimilarity according to the topological overlap. **(D)** Pairwise correlations between the modules in the connectivity measure ( $K_{ME}$ ) of the module eigen-protein (correlation coefficient: Pearson; heatmap order: eigenvectors; agglomeration method: complete; the number of clusters: 3).

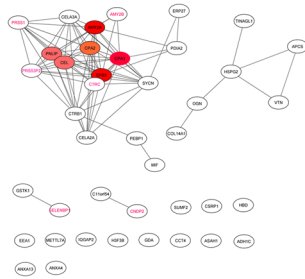
networks for the four WGCNA network modules: WM5 was significant to the NPD Group trait, WM7 and WM11 were significant to the POG trait, and WM16 was significant to the BOG trait. Those PPI networks were reconstructed by using the Cytoscape (version 3.8.2) software (Institute for Systems Biology, Seattle, WA, USA; <https://cytoscape.org/>)<sup>13</sup> (Fig. 3). Top hub proteins were determined by using the *cytoHubba* plugin with maximal clique centrality (MCC)<sup>14</sup>. In data-driven protein co-expression networks, eigen-proteins are indicated in red borders, hub proteins in red to orange fill colors, and some key proteins are denoted in red letters. Top pathway enrichment results for the WGCNA modules are presented in Fig. S2.

The pathways enriched for the WGCNA module that was significant to the NPD Group (Fig. 3A), included the following: (i) digestion as a biological process (GO), (ii) pancreatic secretion as well as protein digestion and absorption as KEGG pathways, and (iii) digestion as a Reactome pathway (Fig. S2). The eigen-protein carboxypeptidase A1 (CPA1) is a member of the carboxypeptidase A family of zinc metalloproteases, the mutations of which have been linked to chronic and hereditary pancreatitis<sup>15</sup>. The hub proteins were CPA1, carboxypeptidase B (CPB1), carboxypeptidase A2 (CPA2), pancreatic amylase alpha 2A (AMY2A), carboxyl ester lipase, and pancreatic lipase (PNLIP), while other key proteins included cationic trypsinogen (PRSS1; trypsin 1), chymotrypsin-C (CTRC), amylase alpha 2B (AMY2B), PRSS3 pseudogene 2 (PRSS3P2; trypsin-2), selenium binding protein 1 (SELENBP1), and carnosine dipeptidase 2 (CNDP2). A set of pancreatic tissue-specific proteins (CPA1, CPA2, CPB1, and CTRC) are carboxypeptidase family members secreted by the pancreas and are downregulated in pancreatic cancer<sup>16</sup>. In Japan, the cumulative incidence of pancreatic cancer for patients with hereditary pancreatitis bearing the PRSS1 and the serine protease inhibitor Kazal 1 (SPINK1) variants was estimated to be 40% until the age of 70 years<sup>17</sup>. Selenium exhibits potent anticarcinogenic properties, and its decreased SELENBP1 expression has often been associated with several cancer types. CNDP2 acts as a tumor suppressor, the upregulation of which leads to an activation of the p38 and JNK/MAPK pathways (thereby leading to cell apoptosis), while its downregulated expression results in an activated ERK/MAPK pathway that can promote cell proliferation<sup>18</sup>.

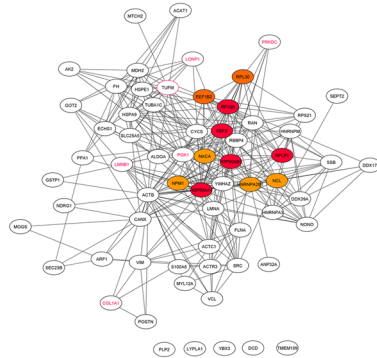
The enriched pathways of the WM7 module significant to POG (Fig. 3B) included the following: (i) the response to chemical stress and the symbiotic process as biological processes (GO); (ii) the carbon metabolism, the tight junction, and the apoptotic pathways as KEGG pathways; and (iii) the disease, the immune system, and the cellular response to stress pathways as Reactome pathways (Fig. S2). The eigen-protein Tu translation elongation factor (EF-Tu) participates in almost all of the mitochondria-mediated protein translation. The upregulation of EF-Tu has been reported in various cancer types, including pancreatic cancers<sup>19</sup>. Hub proteins included the eukaryotic translation elongation factor 2 (EEF2), the HSP90AA1 (Hsp90 $\alpha$ ), the HSP90AB1 (HSP90 $\beta$ ),



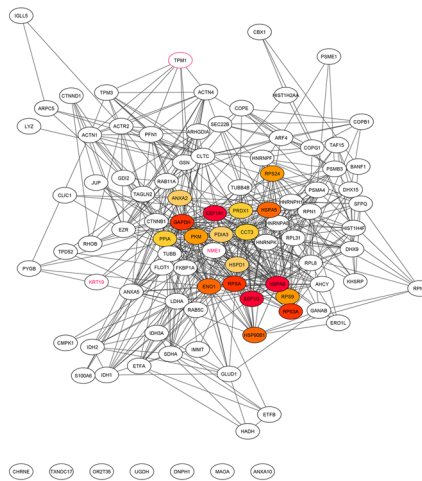
## A. WM5 (green-yellow)



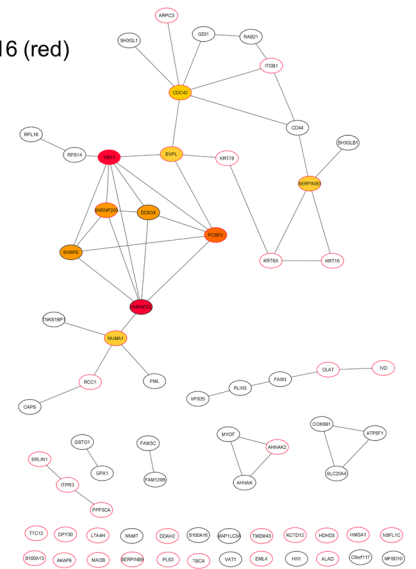
## B. WM7 (black)



## C. WM11 (green)



## D. WM16 (red)



**Figure 3.** Data-driven protein co-expression networks: WM5 (green-yellow) (A), WM7 (black) (B), WM11 (green) (C), and WM16 (red) (D) modules. Circle nodes with a red border and with a fill color ranging from red to orange represent the eigen proteins and/or hub proteins, respectively, for each module. Circle nodes with a red letter also indicate key proteins in the network modules. The top 10 pathways enriched for the protein core networks obtained for biological process (GO), KEGG pathways, and Reactome pathways are presented in Fig. S2 in an order of significance defined by the  $q$ -value.

the 40S ribosomal protein S20 (RPS20), and the 60S acidic ribosomal protein P2 (RPLP2). EEF2 is an essential factor for protein synthesis that has been found overexpressed in numerous types of cancers and that plays an oncogenic role in tumor progression<sup>20</sup>. Both Hsp90 $\alpha$  and HSP90 $\beta$  are molecular chaperones that support the folding of newly synthesized proteins and maintain protein stability under various types of cellular stress. An upregulated HSP90 $\alpha$  expression has been reported in various types of tumors, including pancreatic cancer<sup>21</sup>. Other key proteins include phosphoglycerate kinase 1 (PGK1), lamin-B1 (LMNB1), the mitochondrial Lon protease homolog (LONP1), collagen alpha-1(I) chain (COL1A1), and the DNA-dependent protein kinase catalytic subunit (PRKDC). The glycolytic enzyme PGK1 is involved in the HIF1 $\alpha$  transcription factor network, in which PGK1 together with pyruvate kinase M2 (PKM2) controls the ATP production during aerobic glycolysis in cancer cells (also known as the “Warburg effect”). The PGK1 overexpression is predictive of poor survival in breast, head and neck, cervical, liver, and pancreatic cancers ( $p < 0.001$ )<sup>22</sup>. The mitochondrial ATP-dependent protease LONP1 mediates the selective degradation of misfolded, unassembled, or oxidatively damaged polypeptides, as well as certain short-lived regulatory proteins of the mitochondrial matrix. LONP1 reduces mitochondrial stress to promote cell survival, proliferation, and metastasis in cancer<sup>23</sup>. The fibroblast-deriving COL1A1 is involved in extracellular matrix (ECM) remodeling, tumor cell adhesion, and cell migration<sup>24</sup>. The increased expression of COL1A1 can promote cancer progression and metastasis and has been associated with poor prognosis in numerous cancer types<sup>25</sup>. COL1A1 is one of the downstream targets of the glioma-associated oncogenes (GLI1 and GLI2), is involved in collagen deposition, and is associated with PDAC aggression<sup>24</sup>. Finally, PRKDC plays a major role in nonhomologous end-joining DNA repair, which is an important factor for tumor progression and metastasis; in fact, PRKDC is considered an emerging therapeutic target in cancer<sup>26</sup>.

The enriched pathways of the WM11 (green) module significant to the POG (Fig. 3C) included the following: (i) the establishment of localization in cell, transport, and secretion as a biological process (GO); (ii) the carbon metabolism, the biosynthesis of amino acids, and the protein processing in the endoplasmic reticulum as KEGG pathways; and (iii) the disease, the metabolism of proteins, the neutrophil degranulation, the innate immune system, the eukaryotic translation elongation, and the cellular responses to stress as Reactome pathways (Fig. S2). The eigen-protein tropomyosin-1 (TPM1) is an actin-binding cytoskeletal protein, and its altered expression levels have been closely associated with the rearrangement of microfilament bundles that are responsible for the change of cellular morphology and motility. Hub proteins included the elongation factor 1-alpha 1 (EEF1A1), the 40S ribosomal protein SA (RPSA), the elongation factor 1-gamma (EEF1G), the glyceraldehyde-3-phosphate dehydrogenase (GAPDH), the 40S ribosomal protein S3a (RPS3A), the heat shock cognate 71 kDa protein (HSPA8), the endoplasmic reticulum chaperone BiP (HSPA5; GRP78), endoplasmic reticulum chaperone (HSP90B1), and PKM2. EEF1A1 belongs to the translation factor-related class translation factor GTPase superfamily, and strongly

promotes the heat shock response (HSR), thereby protecting cancer cells from proteotoxic stress (such as oxidative stress and hypoxia). An overexpression of EEF1G has been reported in gastric carcinoma, colon adenocarcinoma, and pancreatic cancer<sup>27</sup>. PKM2 is a key regulator of the Warburg effect in cancer cells. Networks of HSP70 or HSP90 (including HSPA8, HSPA5, and HSP90B1) play important roles in the regulation of energy metabolism as well as in cancer cells' oncogenesis and malignant progression<sup>28</sup>.

The enriched pathways of the WM16 (red) module (Fig. 3D) that are characteristic of the BOG included a symbiotic process, cellular localization, response to the selenium ion, and transport as biological processes (GO) (Fig. S2). The representative eigen- and/or hub proteins included Y-box-binding protein 1 (YBX1), the heterogeneous nuclear ribonucleoprotein D-like (hnRNPD), the poly(rC)-binding protein 2 (PCBP2; hnRNPE2), the U5 snRNP-specific 200 kDa protein (U5-200KD), the DEAD-box helicase family member DBX (DDX3X), and the small nuclear ribonucleoprotein E (SNRPE); all of which are RNA-binding proteins (RBPs). YBX1 mediates the pre-mRNA alternative splicing regulation and is involved in translational regulation by modulating the interaction between the mRNA and the eukaryotic initiation factors. Moreover, YBX1 is significantly overexpressed in PDAC and has been correlated with poor prognosis and reduced survival. HNRNPDL belongs to the heterogeneous nuclear ribonucleoproteins (representing a large family of RBPs), and its aberrant expression has been reported in several cancer types<sup>29</sup>. PDAC is most prominently characteristic of desmoplasia, an abundant fibrotic stroma in which type I collagen proteins are the most abundant and the main component of the ECM. PCBP2 binds to the C-rich region in 3'-UTR of the collagen  $\alpha 1(I)$  mRNA, thereby stabilizing the mRNA and subsequently increasing the type I collagen expression<sup>30</sup>. It has been reported that the combination of gemcitabine with the silencing of PCBP2 can markedly suppress tumor progression in a desmoplastic PDAC orthotopic mouse model<sup>30</sup>. Both ribonucleoproteins U5-200KD and SNRPE are known to engage in the dynamic network of RNA–RNA interactions in the spliceosome machinery. DDX3X is an ATP-dependent RNA helicase that serves multiple functions of cancer (ranging from tumorigenesis to metastasis)<sup>31</sup> and is involved in many cancer-related pathways (including those of P53,  $\beta$ -catenin, and KRAS)<sup>32</sup>.

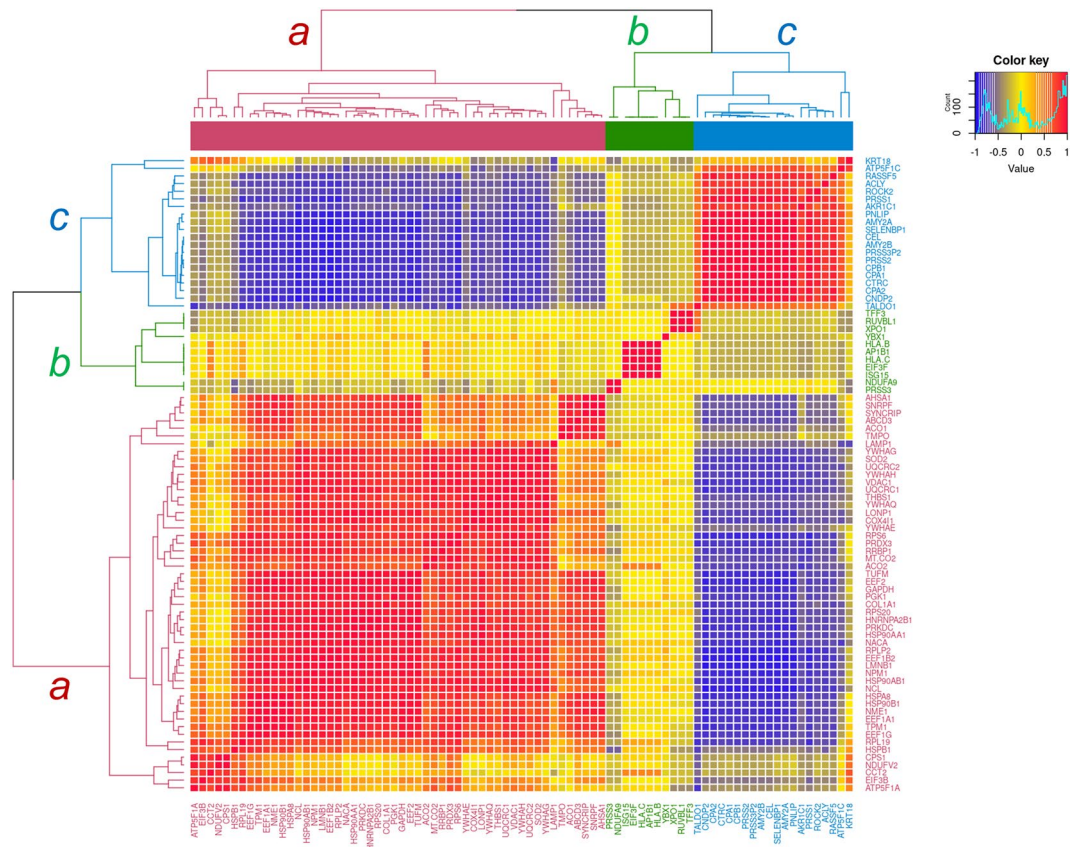
**Multivariate correlation analysis (MVA) of key protein expressions.** The representative 87 key proteins expressed in all 18 modules were subjected to an MVA and were clustered into three groups that corresponded successfully to the POG (*a*), the BOG (*b*), and the NPD Group (*c*), respectively (Fig. 4); interestingly, clusters *b* and *c* were found to be close.

**Upstream regulator and causal network analysis by ingenuity pathway analysis (IPA).** The upstream regulator and causal network analysis together with the downstream annotation were performed for the WGCNA modules<sup>33</sup>, where data were analyzed through the use of IPA (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>). Table 2 summarizes the top upstream regulators, master regulators, canonical pathways, and diseases or functions predicted for the four WGCNA modules.

*Upstream and master regulators predicted for the WM5 module.* Highly activated upstream and/or master regulators predicted for the WM5 module included *NR5A2*, and *PTF1A*, while *GLI1* was highly inhibited (Table 2). *NR5A2* is a nuclear receptor that participates in diverse processes, including bile acid synthesis, the resolution of endoplasmic reticulum stress (ERS), pancreatic development, and acinar differentiation. The pancreas transcription factor 1 alpha (*PTF1A*) is required for the formation of pancreatic acinar and ductal cells. *GlI1* is an oncogenic transcription factor and a critical effector in the Hedgehog pathway and plays a significant role in PDAC progression. The highly inhibited *SPINK1* pancreatic cancer pathway ( $z = -2.813$ ) was most significantly annotated in the canonical pathway.

*Upstream and master regulators predicted for the WM7 and WM11 modules.* Highly activated regulators for the WM7 protein networks included *MYC*, *BRD4*, *YAP1*, *NFE2L2*, *VEGFA*, *STK11*, *HIF1A*, *SP1*, *STAT3*, and *GLI1*, while *LARP1*, *CLPP*, and *RICTOR* were found to be inhibited (Table 2). The *MYC* proto-oncogene protein (*MYC*) binds to the *VEGFA* promoter that promotes the *VEGFA* production, subsequently leading to angiogenesis. The bromodomain-containing protein 4 (*BRD4*) binds acetylated histones and plays an important role in epigenetic regulation<sup>34</sup>. *YAP1* is the critical transcriptional regulator downstream of the Hippo signaling pathway, while the nuclear factor erythroid 2-related factor 2 (*Nrf2* or *NFE2L2*) is the redox master regulator. The serine/threonine-protein kinase *STK11* has been recently suggested to confer protection to cancer cells against metabolic stress and to promote cancer cell survival and invasion, whereas *STK11* was previously considered as a tumor suppressor<sup>35</sup>. The hypoxia-inducible factor 1 subunit alpha (*HIF1A*) is a master transcriptional regulator of the adaptive response to hypoxia. The specificity protein 1 (*SP1*) is a zinc-finger transcription factor, the overexpression of which has been correlated with poor clinical outcomes in various cancer types, including PDAC. *SP1* promotes invasiveness and epithelial-mesenchymal transition (EMT) by cross-talking with *STAT3*, which in turn regulates pathways of tumorigenesis (including those of tumor cell-cycle progression, apoptosis, angiogenesis, metastasis, and immune system evasion)<sup>36</sup>. Finally, the highly upregulated *GLI1* ( $z = 1.98$ ) was indicative of its master regulatory role as a hallmark of PDAC<sup>37</sup>.

By contrast, the La-related protein 1 (*LARP1*) was found to be highly inhibited. *LARP1* is the master regulator of the cap-dependent Top mRNA translation; thereby, our findings strongly suggest the inhibition of protein synthesis via cap-dependent mRNA translation or the activation of the cap-independent, IRES-mediated translation of mRNA subsets that encode oncogenic proteins (including *HIF1 $\alpha$* , *MYC*, and *VEGFA*)<sup>38</sup>. Caseinolytic protease P (*CLPP*) plays a key role in the mitochondrial unfolded protein response and is linked to the regulation of cellular bioenergetics<sup>39</sup>.



**Figure 4.** Multivariate correlation analysis (MVA) for the spectral counting-based expression of 87 eigen- and/or hub proteins and other key proteins expressed among all the modules identified for three traits. Clusters are denoted by *a*, *b*, and *c*. Cluster *a* includes the eigen- and hub proteins, and key proteins in the WGCNA module networks correlated to the trait of poor outcome group (POG), Cluster *b* those to the trait of better outcome group (BOG), and Cluster *c* those to the trait of the noncancerous pancreatic duct (NPD).

Highly activated regulators predicted for the WM11 protein networks included *MYC*, *PCGEM1*, *MYCN*, *KDM8*, *XBPI*, *TGFB1*, *HIF1A*, *NFE2L2*, *ERN1*, *CD38*, *CSF1*, and *SMAD3*, and highly inhibited regulators included *LAR1*, *CLPP*, and *RICTOR*. Thus, both WM7 and WM11 modules share several key regulators (Table 2). The prostate cancer gene expression marker 1 (*PCGEM1*) does not code for a protein, but it is a long noncoding RNA *PCGEM1* (lncRNA *PCGEM1*). Hypoxic cancer cells contain large amounts of exosomal lncRNA *PCGEM1*, which plays a crucial role in proliferation, migration, invasion, drug resistance, and angiogenesis in cancer<sup>40</sup>. The deregulation of *MYCN* (N-MYC), as well as of other *MYC* family oncogenes, is frequently associated with a poor prognosis in many types of cancer. *KDM8* is a histone lysine demethylase/dioxygenase that demethylates H3K36me2 by inducing an epigenetic dysregulation that is implicated in carcinogenesis. The oncogenic histone demethylase *KDM8* has also been reported to form a partnership with *PKM2*, thereby promoting *PKM2* nuclear translocation<sup>41</sup>. The X-box binding protein 1 (*XBPI*) plays a key role in the unfolded protein response (UPR) under ERS<sup>42</sup>. The ER stress response element is present in the promoter region of *HSPA8* (Grp78 or BiP; the master regulator of ER stress) and has been captured by the data-driven WM11 protein networks (Fig. 3C). *ERN1* encodes the endoplasmic reticulum-to-nucleus signaling 1, known as inositol requiring enzyme 1 (IRE1 $\alpha$ ), and the IRE1 $\alpha$ /*XBPI* pathway is one of the major UPR pathways and the most highly conserved ERS pathway<sup>43</sup>. The expression of *CD38* is involved in tumor cell escape from the PD-1/PD-L1 blockade<sup>44</sup> and can be upregulated in response to a PD-L1 antibody therapy<sup>45</sup>.

PDAC is characterized by dense desmoplasia in which the fibrotic stroma contains a high number of activated pancreatic stellate cells (PSCs). The aggressive nature of PDAC is now attributed to cells capable of interplaying the surrounding ECM of the tumor microenvironment to promote disease progression and resistance to therapy. Recently, Steins et al. have shown that PDAC cells induce the secretion of colony-stimulating factor 1 (CSF1), which deactivates stromal PSCs, thereby promoting the development of aggressive subtypes of pancreatic tumors<sup>46</sup>. The EMT-like features are often characteristic of high-grade PDACs. PDAC patients usually harbor *SMAD4* mutations and deletions; in such cases, *SMAD3* is found to be accumulated in the nucleus, and its upregulation has been correlated with the EMT-like features seen in PDAC, regardless of the *SMAD4* status<sup>47</sup>.

Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions			
	Top upstream regulators	z-score	p value of overlap	Top master regulators	z-score	Network bias-corrected p value	Top 5 annotations	z-score	p value	Top annotations	z-score	p value	
WM5 (green-yellow)	<i>PTF1A</i>	1.98	8.67E-08	<i>NR5A2</i>	2.24	0.0001	SPINK1 Pancreatic Cancer Pathway	-2.83	5.01E-24	Synthesis of fatty acid	1.92	0.0096	
	<i>NR5A2</i>	2.17	4.10E-06	<i>PTF1A</i>	2.00	0.0002	Retinol Biosynthesis		0.0002	Fatty acid metabolism	2.16	0.0223	
	<i>GLI1</i>	-2.00	0.0163	<i>CBLC</i>	-2.12	0.0024	Triacylglycerol Degradation		0.0003	Morbidity or mortality	-1.92	0.0240	
				<i>NDRG1</i>	-2.84	0.0036	Pulmonary Healing Signaling Pathway		0.0006				
				<i>NDRG1</i>	-2.31	0.0057	SPINK1 General Cancer Pathway		0.0017				
				<i>plasminogen activator</i>	2.18	0.0066							
				<i>SERPINB2</i>	-2.18	0.0070							
				<i>CDC25A</i>	-2.12	0.0228							
				<i>CHEK1</i>	2.11	0.0272							
				<i>ALDH3A1</i>	-2.07	0.0293							
				<i>Glycoprotein 1B</i>	2.07	0.0333							
				<i>GP9</i>	2.07	0.0334							
				<i>MMP8</i>	2.00	0.0377							
				<i>GLI1</i>	-2.00	0.0395							
				<i>F3-F7</i>	2.00	0.0422							
			<i>RAP2B</i>	2.36	0.0444								

Continued



Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions		
	Top upstream regulators	z-score	p value of overlap	Top master regulators	z-score	Network bias-corrected p value	Top 5 annotations	z-score	p value	Top annotations	z-score	p value
WM7 (black)	MYC	3.31	1.94E-17	LARP1	-2.45	0.0001	Integrin Signaling	1.89	1.91E-06	Cell movement	3.42	7.04E-07
	Insulin	2.63	6.28E-13	CLPP	-2.45	0.0001	Paxillin Signaling	2.24	1.20E-05	Migration of cells	3.33	4.38E-06
	CD3	2.71	3.15E-09	MYC	3.13	0.0001	RHOA Signaling	2.24	2.34E-05	Cell viability	3.10	7.45E-06
	YAP1	2.24	3.40E-09	BRD4	3.27	0.0001	Actin Cytoskeleton Signaling	2.24	5.62E-05	Invasion of cells	2.96	2.71E-07
	CLPP	-2.45	3.35E-08	CTDSP1	-3.27	0.0001	Epithelial Adherens Junction Signaling	2.24	7.08E-05	Synthesis of protein	2.96	3.65E-08
	LARP1	-2.45	1.58E-07	CULAB	-2.86	0.0001	Signaling by Rho Family GTPases	2.45	9.33E-05	Organismal death	-4.38	3.91E-06
	NFE2L2	3.16	9.33E-07	HMGCR	3.92	0.0001	Fcy Receptor-mediated Phagocytosis in Macrophages and Monocytes	2.00	0.0001			
	FGF2	2.08	1.40E-05	RND3	-3.27	0.0001	Death Receptor Signaling	2.00	0.0001			
	VEGFA	2.65	2.81E-05	RPS14	-3.27	0.0001	Regulation of Actin-based Motility by Rho	2.00	0.0003			
	INSR	2.62	0.00020	PFDN5	-2.75	0.0001	RHOGDI Signaling	-2.24	0.0003			
	STK11	2.45	0.00023	FBXL14	-3.40	0.0001	MSP-ROn Signaling In Cancer Cells Pathway	2.00	0.0006			
	ANGPT2	2.17	0.00041	AMBRA1	-2.12	0.0001						
	IL4	2.05	0.00053	FBXO32	-3.14	0.0001						
	MLXIPL	2.00	0.00090	CXCL14	4.00	0.0002						
	SP1	1.99	0.00143	PTPN2	-3.18	0.0002						
	IL5	2.24	0.00152	PCGEM1	2.86	0.0002						
	CEBPB	2.41	0.00191	Mir200	-2.54	0.0002						
	HMGA1	2.00	0.00204	VWF	3.40	0.0003						
	TGFB1	2.01	0.00300	ZEB	2.20	0.0003						
	miR-124-3p (and other miRNAs w/ seed AAG GCAC)	-2.00	0.00465	Insulin	2.32	0.0004						
PRL	2.00	0.00925	USP8	2.67	0.0007							
RICTOR	-2.00	0.00970	LCK/Fyn	2.31	0.0019							
HIF1A	1.98	0.01660	SH2D2A	2.69	0.0019							
GLI1	1.98	0.01950	CD3	2.71	0.0030							
KLF3	-2.00	0.02040	PIK3C3	-2.48	0.0043							
STAT3	2.17	0.04410	growth factor receptor	2.99	0.0045							

Continued

Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions		
	Top upstream regulators	z-score	p value of overlap	Top master regulators	z-score	Network bias-corrected p value	Top 5 annotations	z-score	p value	Top annotations	z-score	p value
WM11 (green)	MYC	3.11	5.56E-16	LARP1	-2.65	0.0001	Remodeling of Epithelial Adherens Junctions	1.00	1.58E-13	Cell proliferation of tumor cell lines	3.32	2.63E-11
	CLPP	-3.00	9.34E-12	KDM8	2.65	0.0001	BAG2 Signaling Pathway	1.00	1.17E-06	Necrosis	-2.93	1.35E-13
	PCGEM1	2.80	9.41E-12	UBA1	-2.45	0.0001	Integrin Signaling	1.89	2.51E-06	Cell death of tumor cell lines	-2.93	6.75E-10
	MYCN	2.21	1.88E-10	MYCN	2.14	0.0001	Regulation of Actin-based Motility by Rho	2.24	7.76E-06	Cell survival	2.78	5.77E-09
	UBA1	-2.41	7.00E-10	CLPP	-3.00	0.0001	Actin Cytoskeleton Signaling	1.89	6.31E-05	Cell viability	2.85	2.91E-08
	KDM8	2.62	2.90E-08	PCGEM1	2.83	0.0001	Fcy Receptor-mediated Phagocytosis in Macrophages and Monocytes	2.00	0.0006	(Cellular Response to Therapeutics) Sensitivity of carcinoma cell lines	-2.41	2.44E-06
	LARP1	-2.65	7.75E-08	MYC	2.89	0.0001	Leukocyte Extravasation Signaling	2.00	0.0011	Cell viability of tumor cell lines	2.67	5.90E-06
	XBP1	2.95	9.67E-08	CREBZF	-3.00	0.0001	RHOA Signaling	2.00	0.0016	Cell death of osteosarcoma cells	-2.45	1.93E-05
	IL15	2.22	1.04E-07	ZMIZ1	3.66	0.0001	Synaptogenesis Signaling Pathway	2.24	0.0018	Cell-cell contact	2.93	1.60E-04
	TGFB1	2.66	1.47E-06	Max-Myc	2.65	0.0001	RHOGDI Signaling	-1.34	0.0018	Organismal death	-4.65	2.58E-04
	IL5	3.00	4.65E-06	CUL4B	-3.02	0.0001	NAD Signaling Pathway	2.00	0.0033	Apoptosis of colorectal cancer cell lines	-2.51	0.0008
	EGFR	2.00	1.09E-05	RAD21	3.02	0.0001						
	CD3	2.19	1.99E-05	SP1-c-Myc	3.29	0.0001						
	PRL	2.75	3.23E-05	USP10	-2.16	0.0001						
	RICTOR	-2.83	3.56E-05	PCGEM1	3.29	0.0001						
	MLXIPL	2.24	4.45E-05	H2AX	3.10	0.0001						
	SLC13A1	-2.24	4.94E-05	Importin alpha/beta	-2.61	0.0001						
	TSC2	-2.22	7.36E-05	MAP3K12	3.31	0.0002						
	HIF1A	2.22	7.75E-05	Ep300/Pcaf	2.29	0.0002						
	CTNNB1	-2.16	9.66E-05	RASAL1	-2.95	0.0002						
	miR-1-3p (and other miRNAs w/seed GGA AUGU)	-2.40	0.0002	XBP1	3.00	0.0003						
	NFE2L2	2.92	0.0002	p70 S6k	2.20	0.0003						
	ERN1	2.15	0.0004	HDAC10	3.16	0.0004						
	EGF	2.18	0.0005	TRIM28	-3.36	0.0006						
	CD38	2.22	0.0007	SIRT7	-3.48	0.0008						
	CSF1	2.43	0.0010	TCR	2.10	0.0008						
	ESRRA	2.20	0.0012	Jmy-p300	2.45	0.0008						
	SMAD3	2.20	0.0015	miR-483-3p (miRNAs w/seed CAC UCCU)	-3.61	0.0013						
AKT1	2.24	0.0019	RNF2	-2.29	0.0015							
GLI1	0.33	0.0005										

Continued

Module ID (color)	Upstream regulators			Causal networks			Canonical pathways			Diseases or functions		
	Top upstream regulators	z-score	p value of overlap	Top master regulators	z-score	Network bias-corrected p value	Top 5 annotations	z-score	p value	Top annotations	z-score	p value
WM16 (red)	MYC	1.89	1.86E-05	PALB2	-2.04	0.0010	fMLP Signaling in Neutrophils	2.00	0.0005	Apoptosis of carcinoma cell lines	-1.89	0.0001
	CST5	-2.65	2.05E-05	NABP1	2.20	0.0012	RAC Signaling	2.00	0.0007	Apoptosis of tumor cell lines	-3.00	0.0003
	SYVN1	2.24	7.35E-05	ZEB1	2.45	0.0012	RHO GDI Signaling	-2.00	0.0034	Migration of tumor cell lines	2.63	0.0005
	IL4	2.59	0.0002	ARG2	-2.29	0.0027	Senescence Pathway	2.00	0.0105	Cell death of tumor cell lines	-2.62	0.0005
	HRAS	2.19	0.0002	CD47	2.12	0.0050				Cell movement of tumor cell lines	2.60	0.0006
	ATG7	1.99	0.0002	TPOR dimer	2.86	0.0066				Growth of tumor	2.44	0.0007
	HGF	2.16	0.0008	MAP3K12	2.50	0.0088				Phagocytosis	2.53	0.0013
	Vegf	2.45	0.0010	JINK1/2	2.04	0.0099				Cell viability	2.31	0.0044
	JUN	2.00	0.0041	CST5	-2.65	0.0003				Insulin sensitivity	-1.98	0.0050
	SOX2	2.39	0.0042	SYVN1	2.24	0.0013						
	TP63	2.17	0.0159									
	ERBB2	1.95	0.0239	(GLI1	1.18	0.0124)						
IGF1	2.16	0.0352										

**Table 2.** Representative master regulators predicted to be activated or inhibited ( $|z\text{-value}| > 2.0$ ) and upregulated ( $1.5 < z\text{-value} < 2.0$ ) are briefly summarized for the four identified WGCNA modules; top annotations of canonical pathways and diseases or functions are also provided.

*Upstream and master regulators predicted for the WM16 module.* Highly activated regulators for the WM16 protein networks included MYC, SYVN1, HRAS, *hepatocyte growth factor (HGF)*, and SOX2, while highly suppressed regulators included CST5 and PALB2. Interestingly, GLI1 was greatly restricted ( $z = 1.18$ ) (Table 2). Synoviolin 1 (SYVN1) is a transmembrane E3 ubiquitin-protein ligase that accepts ubiquitin specifically from the ER-associated ligase and transfers it to substrates, thereby promoting their degradation. Cancer cells harboring a p53 mutation (~70% of PDAC cases) are resistant to anticancer chemotherapy and are characterized by aggressive phenotypes. p53 regulates the ER function in response to stress. Namba et al. have demonstrated that the p53 function loss upregulates IRE1 $\alpha$  that subsequently targets XBP1<sup>48</sup>. The latter enhances the activation of the IRE1 $\alpha$ -XBP1 pathway, thereby providing a response to ERS and unfolding protein stress<sup>48</sup>. The ER membrane protein homeostasis is maintained by ER-associated degradation<sup>49</sup>, while SYVN1 promotes the ubiquitination and degradation of IRE1 $\alpha$ . Moreover, SYVN1 is a favorable prognostic marker in head and neck cancer ( $p < 0.001$ ) (<https://www.proteinatlas.org>). Inhibiting the activation of the IRE1 $\alpha$ /XBP1 pathway to maintain the ER function through an SYVN1-dependent proteasomal degradation of IRE1 $\alpha$  could be a promising modality for treating cancer cases lacking the p53 function<sup>48</sup>.

Oncogenic KRAS mutations are predominant in PDAC (Fig. S3)<sup>50</sup> and are thought to be the driver mutations in PDAC. However, two other RAS family members, NRAS and HRAS, can be activated in PDAC by the oncogenic KRAS. Weyandt et al. have shown that the loss of wild-type HRAS increases tumor load and reduces the survival in an oncogenic KRAS-driven PDAC mouse model<sup>51</sup>. They have also examined those results by tracing the earliest stages of pancreatic cancer and have suggested that the wild-type HRAS is tumor suppressive during these early stages<sup>51</sup>. The paracrine HGF is strongly secreted from PSCs. Yan et al. have shown (by using the pancreatic cancer cell-line Panic-1) that the paracrine HGF (via its receptor c-MET activation) can induce a YAP nuclear translocation and an HIF1 $\alpha$  stabilization, thereby promoting the expression of cancer stem cell pluripotency markers (including the sex-determining region Y (SRY)-Box2; SOX2) and tumorsphere formation<sup>52</sup>. Xu et al. have reported that the paracrine HGF can also activate the c-MET/PI3K/AKT pathway to induce the EMT and inhibit the apoptosis in pancreatic cancer cells, thereby enhancing gemcitabine chemoresistance<sup>53</sup>.

SOX2 is a key regulator of cancer stemness in PDAC<sup>54</sup>. Interestingly, Wuebben et al. have demonstrated that inducible overexpression of SOX2 in engineered PDAC cell lines inhibits growth in vitro and reduces tumorigenicity, while an inducible knockdown of SOX2 has been shown to reduce the PDAC growth both in vitro and in vivo<sup>55</sup>. SOX2 functions seem to act as a molecular rheostat for the control of the growth, tumorigenicity, and drug response of PDAC cells, while the latter seems to highly depend on the expression of optimal levels of SOX2<sup>55</sup>. Thus, the activation of SOX2 ( $z = 2.39$ ) predicted for the BOG in this study might imply that the overexpression of SOX2 restricts the growth of PDAC cells, while the endogenous intermediate levels of SOX2 lead to maximum tumor growth.

Interestingly, the inhibition of both the CST5 and the PALB2 was predicted for the BOG. p53 directly induces cystatin-D (CST5), which functions (in most cases) as a tumor suppressor by promoting the mesenchymal-epithelial transition<sup>56</sup>. Contrary to expectations, the survival analysis of the CST5 mRNA expression data of pancreatic cancer patients ( $n = 176$ ) from The Cancer Genome Atlas (TCGA) database indicated that the high CST5 expression can be correlated with an unfavorable overall survival rate ( $p < 0.05$ ) (<https://www.proteinatlas.org>), which might suggest that CST5 acts as a tumor-promoting factor in pancreatic cancer.

The partner and localizer of BRCA2 (PALB2) plays a critical role in homologous recombination repair and is also considered a susceptibility gene for pancreatic cancer<sup>57</sup>. Ge et al. have demonstrated that the PALB2-knock-down can significantly decrease PDAC cell migration (but not cell proliferation) and that the overall survival is negatively correlated with the PALB2 expression<sup>58</sup>. We performed a web-based survival analysis (KMplot) for the Pan-cancer mRNA RNA-seq data of PDAC ( $n = 177$ ) of the TCGA database and confirmed a significant negative association between the high PALB2 expression and the overall survival (log-rank test  $p = 2.7 \times 10^{-5}$ ; hazard ratio: 2.38) (Fig. S4)<sup>59</sup>.

A PALB2 mutation is expected to disrupt the BRCA1 and BRCA2 interactions that are critical to DNA double-strand break repair. Global genomic sequencing has identified a biallelic inactivation of PALB2 in a patient who had advanced gemcitabine-resistant pancreatic cancer; the patient later received mitomycin C (a DNA damaging agent) based on a personalized therapy and responded well for 36+ months (when the expected median survival was 3 months)<sup>60</sup>. Poly(ADP-ribose) polymerases (PARPs) are DNA damage sensors and key regulators of single-stranded DNA break repair. Clinical and preclinical studies of talazoparib, a PARP inhibitor (PARPi), delivered to PALB2-deficient solid tumors have suggested that the PARPi can exert a synthetic lethal effect in PALB2-deficient tumors, thereby recommending that the PALB2 status should be assessed for securing the best clinical outcome for a patient<sup>61</sup>.

## Discussion

The WGCNA analysis following the MS-based proteomic analysis has identified protein co-expression networks that are significantly and characteristically associated with the POG and the BOG. The MVA for key proteins has successfully exhibited three clusters that corresponded to the three traits studied (Fig. 4).

Oncogenic KRAS induces the redox master regulator NFE2L2, followed by activation of HIF1A, MYC, and MYCN. Both the WM7 and WM11 protein networks were mainly associated with activation of the HSR pathways, the switching to an IRES-mediated mRNA translation, the Warburg effect, and the UPR under hypoxia-related stress and ERS. *GLI1*, *NFE2L2*, and *HIF1A* were identified as key regulators of the WM7 module. *GLI1* and the highly activated STAT3 and BRD4 imply the involvement of a non-canonical Hedgehog pathway. Those of the WM11 module, besides NFE2L2 and HIF1A, include lncRNA *PCGEM1*, *KDM8*, *XBPI*, *ERN1* (IRE1 $\alpha$ ), *CD38*, and *CSF1*. The activation of both XBP1 and IRE1 $\alpha$ , together with the upregulated Grp78, strongly suggests an upregulated unfolding protein response via the IRE1 $\alpha$ /XBP1 pathway. The aggressive natures of pancreatic tumors are associated with the deactivation of stromal PSCs induced by CSF1 as well as with the mesenchymal nature of PDAC cells induced by SMAD3. An integrative network of representative upstream and master regulators, along with their target molecules in datasets, as predicted for both the WM11 and WM7 modules is presented in Fig. 5.

The WM16 (red) module was representative of the BOG. Lack of p53 function (due to its frequent mutations) induces the expression of IRE1 $\alpha$ , and subsequently, the signaling pathway of the IRE1 $\alpha$ /XBP1 axis is activated to respond to the ER stress (that contributes to the malignant phenotypes, including those of chemoresistance and metastasis). The association of IRE1 $\alpha$  with SYVN1 leads to the degradation of IRE1 $\alpha$ , which promotes ER organelle homeostasis. Highly activated levels of SOX2 might also induce the reduction of tumor growth through its molecular rheostat function. Moreover, the highly inactivated PALB2 predicted for BOG suggests that the loss of PALB2 function might be also associated with the suppression of malignant tumor progression and/or chemoresistance; a suggestion that remains to be clarified. An integrative network of representative upstream and master regulators, along with their target molecules in datasets, as predicted for the WM16 module is presented in Fig. 6.

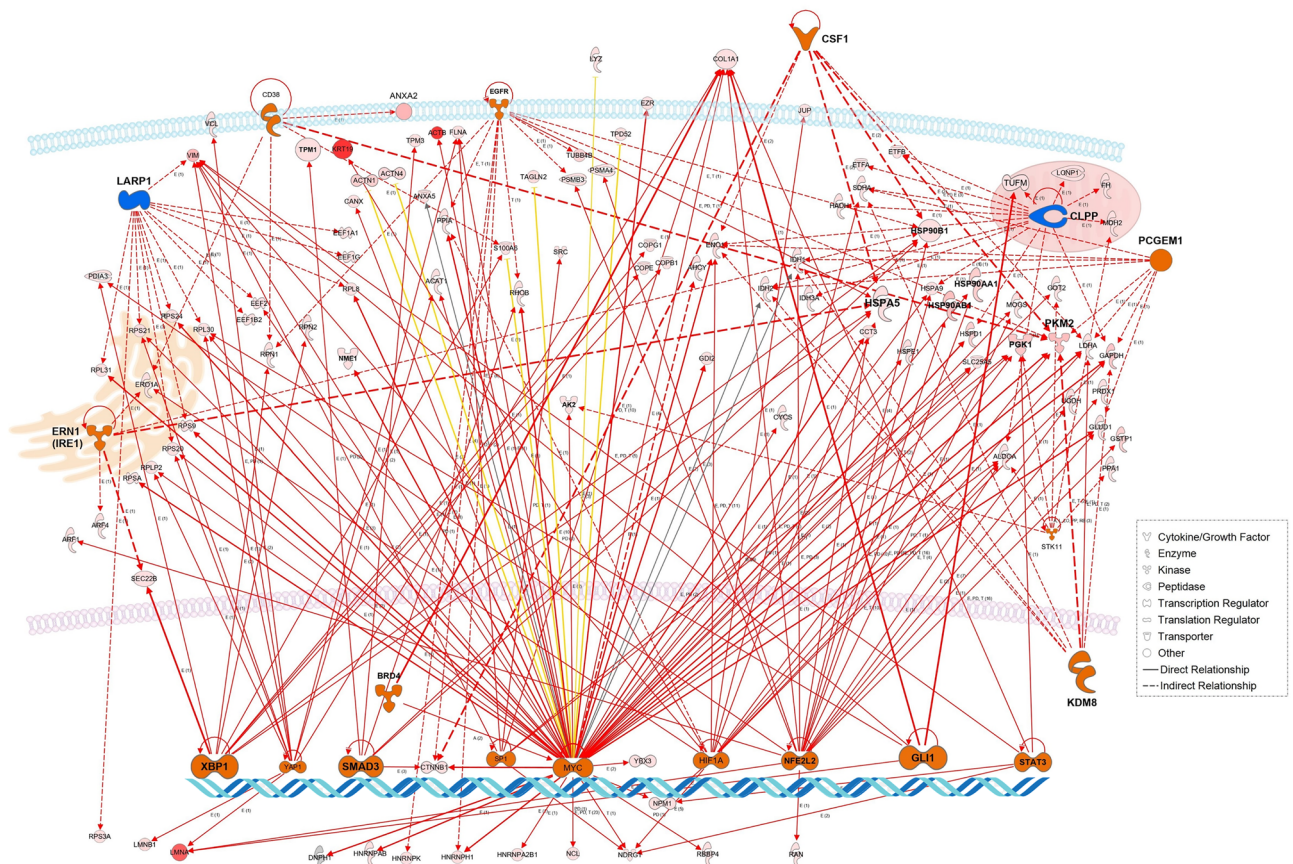
The limitation of this study is the number of patients examined due to the only eight cases available in our hospital, which were with the same histological IIB grade, with the same clinical stage, and treated similarly but resulted differently in better or poor outcomes. We plan to verify the results of this study by using a larger sample size of the external cohort being accumulated in the future.

In conclusion, we have successfully applied WGCNA to clinical proteomics datasets. Our results revealed data-driven co-expression networks and their upstream and master regulators associated with the poor and better outcome PDAC groups. Considering that a limitation of this study was the limited number of patients examined, a future larger-sample cohort study that would include a genomic alteration analysis and investigate data-driven proteogenomic networks might provide even more clinically meaningful insight into the proteogenomic landscape of PDAC.

## Materials and methods

**PDAC FFPE tissue specimens and sample preparation.** This study was approved by the Tohoku University Ethics Committee (2006-119). The FFPE tissues were obtained from individual patients along with their informed consent, and the study adhered to the Helsinki Declaration. Resected pancreatic tissues were fixed in 4% paraformaldehyde and routinely processed for paraffin sectioning. For tissue microdissection, 10- $\mu$ m-thick sections from the FFPE tumor blocks were cut and placed on DIRECTOR™ slides (OncoPlex Diagnostics Inc., Rockville, MD, USA). The sections were then deparaffinized and stained with hematoxylin by using standard histological methods before dissection. Microdissection was performed by using a Leica LMD6000 (Leica Microsystems GmbH, Wetzlar, Germany). A total area of 8 mm<sup>2</sup> (with approximately 30,000 tumor cells) was directly transferred from the FFPE sections, via laser dissection, into the cap of a 200- $\mu$ L low-binding PCR tube.



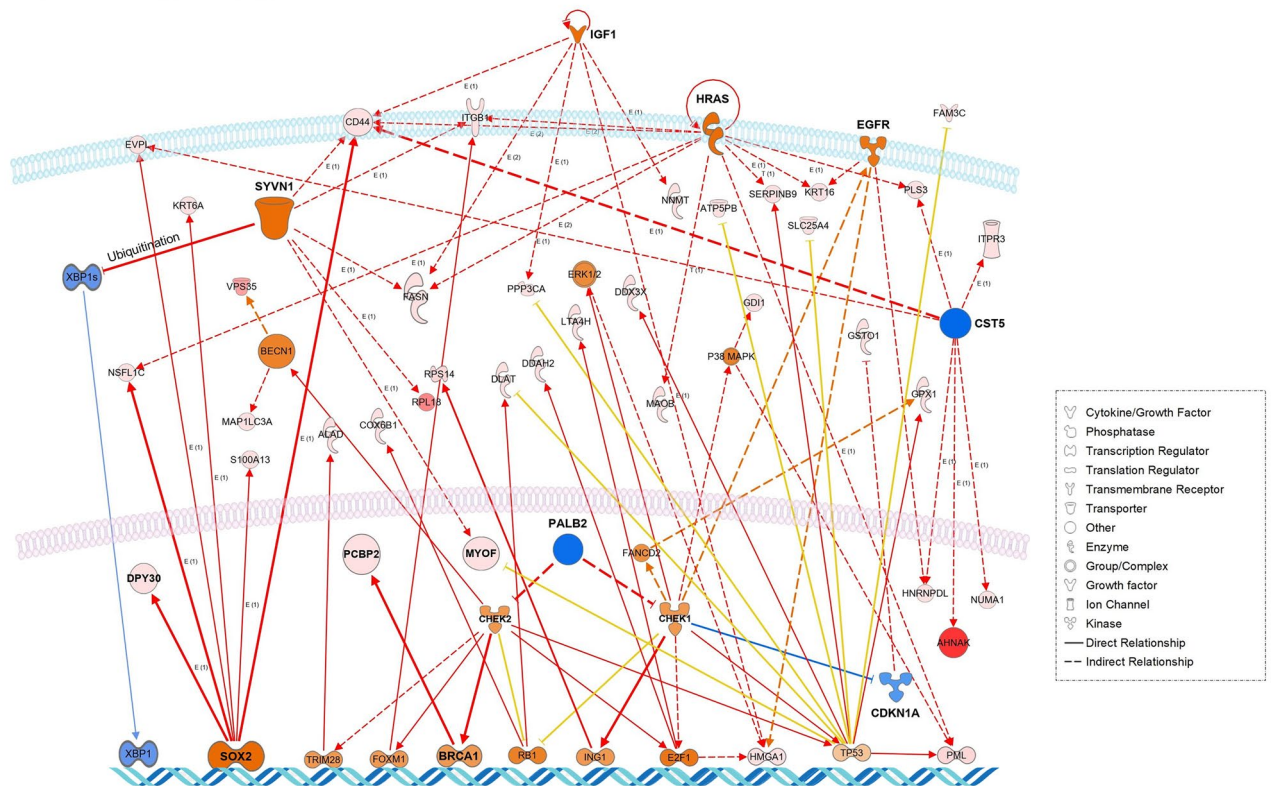


**Figure 5.** Integrative network of representative upstream and master regulators, along with their target molecules in datasets, as predicted for the WM11 and WM7 modules (significant to the poor outcome group; POG).

Proteins were extracted and digested with trypsin by using the Liquid Tissue™ MS Protein Prep kits (OncoPlex Diagnostics Inc.) according to the manufacturer's instructions<sup>62</sup>. Briefly, dried microdissection pellets were suspended in 20  $\mu$ L of Liquid Tissue buffer and heated at 95 °C for 90 min, and then cooled on ice for 3 min before the addition of 0.1  $\mu$ g of trypsin to each tube. The tubes were then incubated at 37 °C overnight. Dithiothreitol was added to a final concentration of 10 mM, and the samples were heated for 5 min at 95 °C. The digested samples were dried, then resuspended in 50  $\mu$ L of a 2% acetonitrile aqueous solution containing 0.1% trifluoroacetic acid and stored at -20 °C until analysis. The pathological specimens were independently reviewed by two pathologists (S. M. and M. U.).

**Global proteomics by LC-MS/MS.** A label-free spectral counting-based quantitative proteomic analysis was conducted through LC-MS/MS<sup>63</sup>. The digested samples (5  $\mu$ L for a single run) were analyzed in triplicate by LC-MS/MS by using a reverse-phase LC interfaced with an LTQ-Orbitrap hybrid MS (Thermo Fisher Scientific, CA) and a nano-electrospray ionization device (AMR Inc., Tokyo, Japan). The LC system consisted of Paradigm MS4B (Michrom BioResources, CA), a trap cartridge (0.3 mm  $\times$  5.0 mm, CERI, Tokyo, Japan), a peptide Cap-Trap cartridge (2.0  $\times$  0.5 mm<sup>2</sup> inside diameter), and an analytical column (L-column Micro, 150  $\times$  0.2 mm<sup>2</sup> L-C18, 3  $\mu$ m, 12 nm; Chemical Evaluation Research Institute, Tokyo, Japan) fitted with an emitter tip (FortisTip, OmniSeparo-TJ, Hyogo, Japan). An aliquot of samples was loaded into the trap and was then washed with solvent A (2% acetonitrile aqueous solution containing 0.1% formic acid) to concentrate the peptides in the trap and desalt them. Subsequently, the trap was connected to the separation column, and the peptides were eluted from the whole column with a 0.1% formic acid aqueous solution and acetonitrile, by a linear 5–40% acetonitrile concentration gradient over 70 min, at a flow rate of 1  $\mu$ L/min.

**LC-MS/MS analysis.** All MS/MS spectral data were searched against the *Homo sapiens* entries in the Swiss-Prot database (Release 57.13, 20,349 entries) by using the Mascot software (version 2.1.1; Matrix Science, London, UK). This search considered tryptic peptide candidates, and the formylation of lysine and the oxidation of methionine were considered variable modifications. The peptide mass tolerance was 20 ppm, the fragment mass tolerance was 0.8 Da, and the trypsin specificity was applied with a maximum of two missed cleavages. A *p* value lower than 0.05 was considered to be statistically significant in the protein identification<sup>64</sup>. The expressions of



**Figure 6.** Integrative network of representative upstream and master regulators, along with their target molecules in datasets, as predicted for the WM16 module (characteristic of the better outcome group; BOG).

the identified proteins were assessed by spectral count-based protein quantification. The spectral count is the number of MS/MS spectra assigned to each protein.

**WGCNA, a weighted correlation network analysis.** The similarity in protein expression patterns for all protein pairs was calculated according to their pairwise Pearson's correlation coefficient; i.e., the similarity between proteins  $i$  and  $j$  was defined as  $(1 - r_{ij})/2$ , where  $r_{ij}$  is the Pearson's correlation coefficient of the protein expression patterns between these two proteins. We performed a network topology analysis for various soft-thresholding powers ranging from 1 to 100 to choose an optimal value of balance between independence and mean connectivity. A topological overlap matrix (TOM) that considers topological similarities between a pair of proteins in the network was then generated from the resultant scale-free co-expression network. We generated a tree that clustered proteins in its branches by hierarchical clustering through the use of dissimilarity according to TOM ( $1 - \text{TOM}$ ), and protein modules were determined by using a dynamic tree-cutting to trim the branches<sup>9</sup>.

The modules that were summarized by the first principal component are referred to as eigen proteins in the text, as they express the highest connectivity in the module. Module membership, defined as the correlation between the protein expression profile and the module eigen-protein, was measured with values ranging from 0 to 1, with "0" representing a gene that is not part of the module and "1" representing high connectivity with the module. Subsequently, the module-trait association was determined by using the correlation between the module eigen-protein and the three clinical traits: the POG, the BOG, and the NPD group. A protein module was summarized by the top hub protein (referred to as "eigen-protein") with the highest connectivity in the module. The WGCNA analysis was performed using the WGCNA R-package<sup>9</sup> that is implemented in RStudio.

**Protein-PPI network construction.** We used the STRING database (version 11.5) to construct a PPI network for a protein module<sup>12</sup>. STRING networks were calculated under the criteria for linkage with experiments, databases, text mining, and co-expression, by using the default settings (medium confidence score: 0.400; network depth: 0 interactions). Functional enrichment results were obtained for canonical pathways, with a  $p < 0.05$ . Proteins in a protein module were mapped in the PPI network from the STRING database to produce the results of the enrichment analysis regarding the biological process (GO), the KEGG pathways (hsa), and the Reactome pathways (HAS). Protein networks were subsequently exported to Cytoscape (version 3.8.2)<sup>13</sup> from the STRING database. We then identified the hub proteins in each module according to their intramodular connectivity, and their correlation with module eigen proteins. The proteins inside the co-expression modules exhibit high connectivity, and the proteins within the same module may play similar roles. The top 10 high-degree proteins were identified by using the *cytoHubba* plugin<sup>14</sup>. The top-ranked proteins in each module were considered to be



hub proteins; hereby referred to as “highly connected proteins.” Functional enrichment results were obtained for canonical pathways by considering a  $p$  value of  $< 0.05$  to be statistically significant.

The multivariate correlation analysis (MVA) of semiquantitative key protein expressions was performed by using the JMP software (SAS Institute, Cary, NC, USA), and which result was visualized using Intervene Shiny App (<https://intervene.shinyapps.io/intervene/>)<sup>65</sup>.

**Upstream regulator and causal network analysis by IPA.** Upstream regulators, causal networks, and canonical pathways were predicted by employing the IPA software (<http://resources.qiagenbioinformatics.com/getting-started-guides/citing-ingenuity-products.pdf>)<sup>33</sup>. Quantile-normalized protein expression data of the selected modules were used as input datasets. Both the upstream regulators and causal networks ( $p < 0.05$ ) predicted from the WGCNA network modules were significantly associated with the three clinical traits (POG, BOG, and NPD group), where the activation and the inhibition of a predicted network were defined by  $z$ -values that were  $> 2.0$  and  $< -2.0$ , respectively. The upregulation was defined by  $z$ -values being  $> 1.5$  and  $< 2.0$ , while the downregulation was defined by  $z$ -values being  $> -2.0$  and  $< -1.5$ .

## Data availability

The unfiltered MS datasets generated and analyzed in this study have been deposited in the ProteomeXchange (<http://proteomecentral.proteomexchange.org>) and jPOST (<https://repository.jpostdb.org>), with the dataset identifiers PXD032681 and JPST001535, respectively.

Received: 9 April 2022; Accepted: 25 August 2022

Published online: 29 August 2022

## References

- Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Thomas, A. G. & Awasthi, N. Targeted therapy for pancreatic cancer: Lessons learned and future opportunities. *Dig. Med. Res.* **4**, 32 (2021).
- Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of pancreatic cancer: Global trends, etiology and risk factors. *World J. Oncol.* **10**, 10–27 (2019).
- Lee, J. *et al.* Management of recurrent pancreatic cancer after surgical resection: A protocol for systematic review, evidence mapping, and meta-analysis. *BMJ Open* **8**, e017249 (2018).
- Matsuno, S. *et al.* Pancreatic Cancer Registry in Japan: 20 years of experience. *Pancreas* **28**, 219–230 (2004).
- Kim, R. *et al.* A single institution review of adjuvant therapy outcomes for resectable pancreatic adenocarcinoma: Outcome and prognostic indicators. *J. Gastrointest. Surg.* **14**, 1159–1169 (2010).
- Lim, J. E., Chien, M. W. & Earle, C. C. Prognostic factors following curative resection for pancreatic adenocarcinoma: A population-based, linked database analysis of 396 patients. *Ann. Surg.* **237**, 74–85 (2003).
- Nishimura, T. *et al.* Current status of clinical proteogenomics in lung cancer. *Expert Rev. Proteom.* **16**, 761–772 (2019).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
- Yamada, K. *et al.* Protein co-expression networks identified from HOTAIR lesions of ER+HER2-Ki-67high luminal breast carcinomas. *Sci. Rep.* **11**, 1705 (2021).
- Hirata, K. *et al.* Current status of surgery for pancreatic cancer. *Dig. Surg.* **24**, 137–147 (2007).
- Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
- Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Chin, C. H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(Suppl 4), S11 (2014).
- Tamura, K. *et al.* Mutations in the pancreatic secretory enzymes CPA1 and CPB1 are associated with pancreatic cancer. *Proc. Natl. Acad. Sci. USA* **115**, 4767–4772 (2018).
- Song, Y. *et al.* Label-free quantitative proteomics unravels carboxypeptidases as the novel biomarker in pancreatic ductal adenocarcinoma. *Transl. Oncol.* **11**, 691–699 (2018).
- Masamune, A. *et al.* Nationwide survey of hereditary pancreatitis in Japan. *J. Gastroenterol.* **53**, 152–160 (2018).
- Zhang, Z. *et al.* Underexpressed CNDP2 participates in gastric cancer growth inhibition through activating the MAPK signaling pathway. *Mol. Med.* **20**, 17–28 (2014).
- Xi, H. Q. *et al.* Expression and clinicopathologic significance of TUFM and p53 for the normal–adenoma–carcinoma sequence in colorectal epithelia. *World J. Surg. Oncol.* **15**, 90 (2017).
- Song, Y. *et al.* Elevated eukaryotic elongation factor 2 expression is involved in proliferation and invasion of lung squamous cell carcinoma. *Oncotarget* **7**, 58470–58482 (2016).
- Zuehlke, A. D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8–16 (2015).
- He, Y. *et al.* PGK1-mediated cancer progression and drug resistance. *Am. J. Cancer Res.* **9**, 2280–2302 (2019).
- Lee, Y. G. *et al.* LONP1 and ClpP cooperatively regulate mitochondrial proteostasis for cancer cell survival. *Oncogenesis* **10**, 18 (2021).
- Li, M. *et al.* Microenvironment remodeled by tumor and stromal cells elevates fibroblast-derived COL1A1 and facilitates ovarian cancer metastasis. *Exp. Cell Res.* **394**, 112153 (2020).
- Chakravarthy, D. *et al.* Palmatine suppresses glutamine-mediated interaction between pancreatic cancer and stellate cells through simultaneous inhibition of survivin and COL1A1. *Cancer Lett.* **419**, 103–115 (2018).
- Mohiuddin, I. S. & Kang, M. H. DNA-PK as an emerging therapeutic target in cancer. *Front. Oncol.* **9**, 635 (2019).
- Hassan, M. K., Kumar, D., Naik, M. & Dixit, M. The expression profile and prognostic significance of eukaryotic translation elongation factors in different cancers. *PLoS ONE* **13**, e0191377 (2018).
- Yang, S., Haiyan, X. H. & Cao, L. Recent advances in heat shock proteins in cancer diagnosis, prognosis, metabolism and treatment. *Biomed. Pharmacother.* **142**, 112074 (2021).
- Zhang, P. *et al.* Oncogenic heterogeneous nuclear ribonucleoprotein D-like promotes the growth of human colon cancer SW620 cells via its regulation of cell-cycle. *Acta Biochim. Biophys. Sin.* **50**, 880–887 (2018).
- Li, Y. *et al.* Silencing PCBP2 normalizes desmoplastic stroma and improves the antitumor activity of chemotherapy in pancreatic cancer. *Theranostics* **11**, 2182–2200 (2021).

31. Chen, H. H., Yu, H. I., Cho, W. C. & Tarn, W. Y. DDX3 modulates cell adhesion and motility and cancer cell metastasis via Rac1-mediated signaling pathway. *Oncogene* **34**, 2790–2800 (2015).
32. Wu, D. W. *et al.* DDX3 enhances oncogenic KRAS-induced tumor invasion in colorectal cancer via the  $\beta$ -catenin/ZEB1 axis. *Oncotarget* **7**, 22687–22699 (2016).
33. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. IPA: Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).
34. Jiao, F. *et al.* Caveolin-2 is regulated by BRD4 and contributes to cell growth in pancreatic cancer. *Cancer Cell Int.* **20**, 55 (2020).
35. Li, T. T. & Zhu, H. B. LKB1 and cancer: The dual role of metabolic regulation. *Biomed. Pharmacother.* **132**, 110872 (2020).
36. Huang, C. & Xie, K. Crosstalk of Sp1 and Stat3 signaling in pancreatic cancer pathogenesis. *Cytokine Growth Factor Rev.* **23**, 25–35 (2012).
37. Kasai, K. GLI1, a master regulator of the hallmark of pancreatic cancer. *Pathol. Int.* **66**, 653–660 (2016).
38. Silvera, D., Formenti, S. C. & Schneider, R. J. Translational control in cancer. *Nat. Rev. Cancer* **10**, 254–266 (2010).
39. Cormio, A., Sanguedolce, F., Pesce, V. & Musicco, C. Mitochondrial caseinolytic protease P: A possible novel prognostic marker and therapeutic target in cancer. *Int. J. Mol. Sci.* **22**, 6228 (2021).
40. Yun, B. D. *et al.* Oncogenic role of exosomal circular and long noncoding RNAs in gastrointestinal cancers. *Int. J. Mol. Sci.* **23**, 930 (2022).
41. Chen, T. J. *et al.* Mutations in the PKM2 exon-10 region are associated with reduced allosterity and increased nuclear translocation. *Commun. Biol.* **2**, 105 (2019).
42. Chen, S. *et al.* The emerging role of XBP1 in cancer. *Biomed. Pharmacother.* **127**, 110069 (2020).
43. Robinson, C. M. *et al.* An emerging role for the unfolded protein response in pancreatic cancer. *Cancers (Basel)* **13**, 261 (2021).
44. Chen, L. *et al.* CD38-mediated immunosuppression as a mechanism of tumor cell escape from PD-1/PD-L1 blockade. *Cancer Discov.* **8**, 1156–1175 (2018).
45. Dwivedi, S., Rendón-Huerta, E. P., Ortiz-Navarrete, V. & Montaña, L. F. CD38 and regulation of the immune response cells in cancer. *J. Oncol.* **2021**, 6630295 (2021).
46. Steins, A. *et al.* High-grade mesenchymal pancreatic ductal adenocarcinoma drives stromal deactivation through CSF-1. *EMBO Rep.* **21**, e48780 (2020).
47. Yamazaki, K. *et al.* Upregulated SMAD3 promotes epithelial-mesenchymal transition and predicts poor prognosis in pancreatic ductal adenocarcinoma. *Lab. Invest.* **94**, 683–691 (2014).
48. Namba, T. *et al.* Loss of p53 enhances the function of the endoplasmic reticulum through activation of the IRE1 $\alpha$ /XBP1 pathway. *Oncotarget* **6**, 19990–20001 (2015).
49. Ruggiano, A., Foresti, O. & Carvalho, P. Quality control: ER-associated degradation: Protein quality control and beyond. *J. Cell Biol.* **204**, 869–879 (2014).
50. Cao, L. *et al.* Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* **184**, 5031–5052 (2021).
51. Weyandt, J. D. *et al.* Wild-type Hras suppresses the earliest stages of tumorigenesis in a genetically engineered mouse model of pancreatic cancer. *PLoS ONE* **10**, e0140253 (2015).
52. Yan, B. *et al.* Paracrine HGF/c-MET enhances the stem cell-like potential and glycolysis of pancreatic cancer cells via activation of YAP/HIF-1 $\alpha$ . *Exp. Cell Res.* **371**, 63–71 (2018).
53. Xu, J., Liu, S., Yang, X., Cao, S. & Zhou, Y. Paracrine HGF promotes EMT and mediates the effects of PSC on chemoresistance by activating c-Met/PI3K/Akt signaling in pancreatic cancer in vitro. *Life Sci.* **263**, 118523 (2020).
54. Mamun, M. A., Mannoor, K., Cao, J., Qadri, F. & Song, X. SOX2 in cancer stemness: Tumor malignancy and therapeutic potentials. *J. Mol. Cell Biol.* **12**, 85–98 (2020).
55. Wuebben, E. L. *et al.* SOX2 functions as a molecular rheostat to control the growth, tumorigenicity and drug responses of pancreatic ductal adenocarcinoma cells. *Oncotarget* **7**, 34890–34906 (2016).
56. Hünten, S. & Hermeking, H. p53 directly activates cystatin D/CST5 to mediate mesenchymal-epithelial transition: A possible link to tumor suppression by vitamin D3. *Oncotarget* **6**, 5842–15856 (2015).
57. Jones, S. *et al.* Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* **324**, 217 (2009).
58. Ge, O. *et al.* PALB2 upregulation is associated with a poor prognosis in pancreatic ductal adenocarcinoma. *Oncol. Lett.* **21**, 224 (2021).
59. Lánckzy, A. & Györfy, B. Web-based survival analysis tool tailored for medical research (KMplot): Development and implementation. *J. Med. Internet Res.* **23**, e27633 (2021).
60. Villarroel, M. C. *et al.* Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. *Mol. Cancer Ther.* **10**, 3–8 (2011).
61. Wu, S. *et al.* Molecular mechanisms of PALB2 function and its role in breast cancer management. *Front. Oncol.* **10**, 301 (2020).
62. Prieto, D. A. *et al.* Liquid tissue: Proteomic profiling of formalin-fixed tissues. *Biotechniques* **38**, S32–S35 (2005).
63. Kawase, H. *et al.* Differential LC-MS-based proteomics of surgical human cholangiocarcinoma tissues. *J. Proteome Res.* **8**, 4092–4103 (2009).
64. Takadate, T. *et al.* Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. *Int. J. Cancer* **132**, 1368–1382 (2013).
65. Khan, A. & Mathelier, A. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinform.* **18**, 287 (2017).

## Acknowledgements

Funding from the Grant-in-Aid for Scientific Research (B), Grant Number JP16669931, is gratefully acknowledged.

## Author contributions

T.N., S.M., and M.U. conceptualized this study; T.N., S.M., and M.U. initiated and managed the collaboration; T.T., S.M., and T.S. surveyed the cases that underwent surgical pancreatic ductal carcinoma resection and confirmed the clinicopathological information; T.T., S.M., and M.U. selected the FFPE tissue blocks of both pancreatic ductal adenocarcinomas and noncancerous pancreatic ducts; T.T. and T.S. prepared the slides; S.M. and U.M. reviewed the pathological specimens; T.T., S.M., T.F., Y.B., and T.M. performed laser microdissection, sample preparation, and MS-based proteomic analysis; T.N. designed and performed bioinformatics analysis for protein expression data; T.N. wrote the main manuscript text and prepared Tables 1 and 2, Figs. 1–6, and Supplementary Information File 1; T.N., S.M., and M.U. wrote the first draft of the manuscript; all authors reviewed the manuscript and agreed with the submission.



### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19182-9>.

**Correspondence** and requests for materials should be addressed to T.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022