



OPEN

Weighted gene co-expression network analysis reveals key stromal prognostic markers in pancreatic cancer

G. Mantini¹✉, A. Agostini^{1,2}, M. Tufo¹, S. Rossi¹, M. Kulesko¹, C. Carbone², L. Salvatore^{2,3}, G. Tortora^{2,3}, G. Scambia^{4,5} & L. Giacò¹

In recent years, it has been shown that stroma compartment can favor tumor proliferation and aggressiveness. Although extensive research with network analyses such as Weighted Gene Co-expression Network Analysis (WGCNA) has been conducted on pancreatic cancer and its stromal components, WGCNA has not previously been applied to isolate and identify genes associated with the abundance of stroma and survival outcome from bulk RNA data. We investigated the gene expression profile and clinical information of 140 pancreatic ductal adenocarcinoma patients from TCGA. Network analysis was performed using WGCNA and four modules were found to be associated to patients' clinical traits. Specifically, one module of 2459 genes, was associated to stromal sample content. Subsequently, those genes were further analyzed for survival association through log-rank test and Cox regression. HPGDS and ITGA9-AS1 emerged as significant indicators of favorable prognosis while KCMF1 and YARS1 were implicated in poorer prognostic outcomes. Importantly, HPGDS was found to be stromal-specific in the TMA cohort of Human Protein Atlas. Single sample GSEA showed that the stromal module is enriched for stromal signature of Moffitt and Puleo. These findings suggest that we uncovered a stromal specific signature through WGCNA and found putative prognostic markers.

Keywords WGCNA, PDAC, Stroma, Biomarkers, Survival

Pancreatic cancer ranks as the fourth leading cause of cancer-related death in men and third in female worldwide¹.

Pancreatic ductal adenocarcinoma (PDAC), accounts for 90% of pancreatic cancer cases. Despite the increase of five-year survival up to 13% in the last years², PDAC usually results in poor survival with limited options for treatment, as most affected individuals present advanced disease at the time of diagnosis. Identifying early biomarkers is therefore critical for detecting recurrence, guiding prognosis, and tailoring treatment.

Currently, carbohydrate antigen 19-9 (CA19-9) is the only FDA approved biomarker associated with PDAC and PDAC poor survival³⁻⁶.

However, CA19-9 may not be so effective or accurate in prognostic evaluation. Therefore, there is an urgent need for advanced analytical methods to identify novel prognostic markers that are more specific to this type of cancer.

The tumor stroma plays a critical and complex role in tumor development, progression, and resistance to therapy. PDAC is known for its dense stromal reaction, which makes up a significant portion of the tumor mass (often more than 50%) and contributes to its aggressive nature and poor prognosis.

Cancer-associated fibroblasts (CAFs), a major component of the stroma, are key drivers of extracellular matrix (ECM) remodeling, facilitating tumor invasion, metastasis, and resistance to chemotherapy. The stromal environment also fosters an immunosuppressive microenvironment by recruiting immune cells such as regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), and macrophages, which suppress anti-tumor immune responses. Complex bidirectional signaling between stromal cells and tumor cells further accelerates tumor progression, including providing metabolic support to sustain the survival of cancer cells⁷.

¹Bioinformatics Research Core Facility, Gemelli Science and Technology Park (GSTeP), Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy. ²Medical Oncology, Comprehensive Cancer Center, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy. ³Medical Oncology, Catholic University of the Sacred Heart, Rome, Italy. ⁴Department of Woman, Child and Public Health, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy. ⁵Institute of Obstetrics and Gynecology, Catholic University of the Sacred Heart, Rome, Italy. ✉email: giulia.mantini@policlinicogemelli.it

Bulk RNA sequencing primarily reflects stromal content rather than epithelial cell type due to the high stromal content (nearly 70%) in PDAC tissue. Recent statistical approaches, like WGCNA, have been developed to extrapolate important features from high-throughput data, e.g. pinpointing genes/proteins/variants associated to clinical outcome⁸. While prior studies have applied WGCNA to the PAAD-TCGA cohort⁵⁻¹², they often overlooked the stromal content.

Here we analyze the PAAD-TCGA cohort using WGCNA on bulk RNA-seq data and clinical information from 140 patients, with an emphasis on the stromal contribution to PDAC prognosis.

Results

PDAC gene modules are associated to clinical traits

Normalized gene expression data were acquired from 140 naive-treatment patients that underwent PDAC surgical resection as described in Cao et al.¹³.

Clinicopathological characteristics of the cohort are shown in **Supplementary Table S1**. The median age of patient cohort is 65 years old with balanced gender distribution.

Data preprocessing resulted in a gene expression matrix of 21,793 genes and 140 PAAD patients.

Samples were highly correlated to each other indicating a minimum gene expression variance between samples (see **Supplementary Fig. S1**).

To identify genes that are co-expressed and associated with stromal content, a WGCNA was conducted. Initially, a co-expression network was constructed where nodes represent genes and edges represent the correlation values between gene pairs. The network was constructed based on a selected soft-threshold; edges below this threshold were discarded. For instance, to include only positive correlations, edges with negative correlations were removed.

To determine an appropriate threshold, a range of 20 soft-thresholding powers (from 1 to 20) was tested, evaluating each resulting network for its fit to a scale-free topology model. The first threshold with a correlation of 0.9 to the scale-free topology network was selected. Based on this criterion, a soft threshold of 7 was chosen for this study (Fig. 1A).

Next, using hierarchical clustering, genes sharing high correlations were grouped together into modules. Each module represents a cluster of highly interconnected genes.

A total of 32 gene modules were identified after WGCNA. Next, each module was tested for association with clinical traits, such as patient survival rates and stromal content, by correlating the eigenvector of each module to the clinical data (Fig. 1B). For example, to find if any of those modules were associated to sample stromal content, the eigenvector of each module was correlated to the percentage of stromal content in each sample.

Four modules were found to be significantly associated to clinical traits using thresholds correlation > 0.4 and p-value < 0.05.

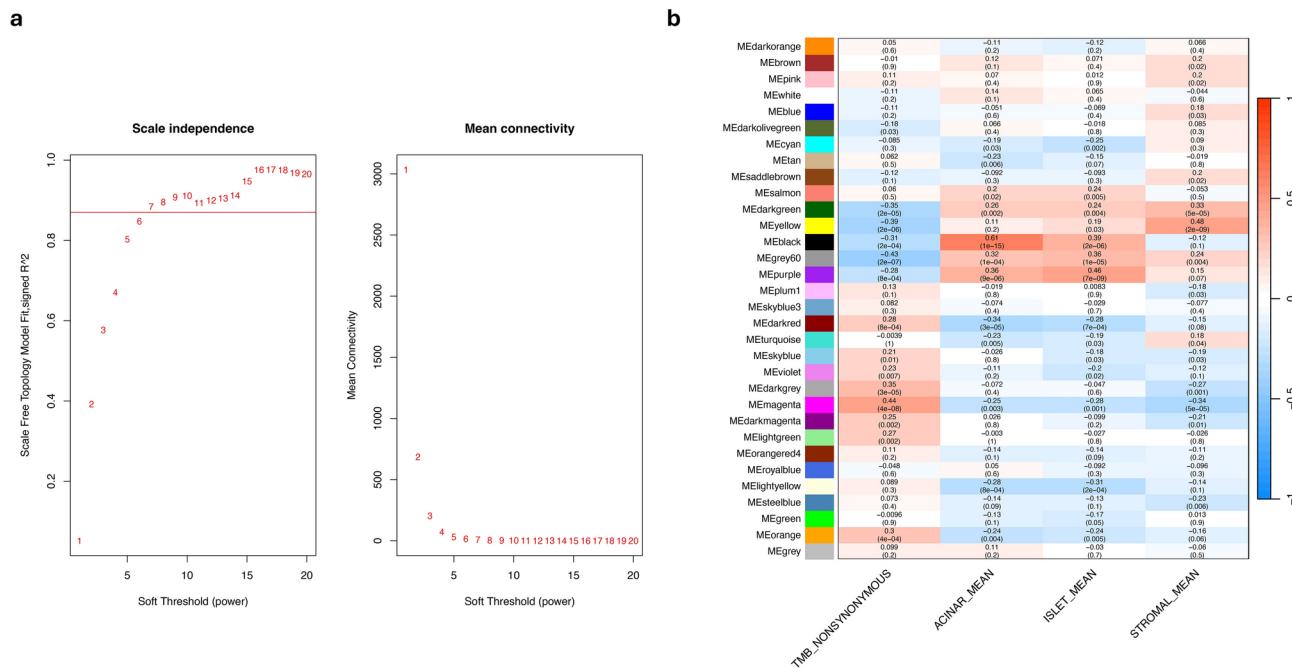


Fig. 1. Choice of soft thresholds parameter for network construction. **a)** Correlation of resulting network with scale free topology network for each selected threshold (on the left); mean connectivity of resulting network for each selected threshold (on the right). **b)** Association between modules and clinical traits was assessed by correlating each module with clinical traits (TMB, Acinar mean, Islet mean, and Stromal mean), providing both correlation coefficients and p-values. Positive correlations are indicated in red, while negative correlations are shown in blue.

MEblack, MEpurple, MEmagenta and MEyellow were found to be positively associated to stromal mean, acinar mean, islet mean, and Tumor Mutational Burden (TMB), respectively (Table 1).

Biological enrichment on modules associated to clinical traits

Next, the gene ontology of those modules was investigated using GO, KEGG and Reactome databases.

The module associated to acinar content (MEblack) revealed pathways related to pancreatic secretion, metabolic and extracellular processes that are typical of the acinar function.

Interestingly, the module associated with islet mean (MEpurple) resulted to be enriched for insulin processing, regulation of beta-cell development, growth hormone synthesis, secretion and type II diabetes mellitus which is in line with islets of Langerhans function.

Notably, the magenta module (MEmagenta) associated with TMB, is enriched for mismatch repair pathways, p53 signaling pathways and cell cycle regulation which are key pathways frequently related to TMB.

The gene ontology mining on the module associated to stromal content (MEyellow) showed several pathways related to PDAC stroma such as Hippo signaling, TGF- β signaling and WNT signaling.

The Hippo pathway influences the activation and the function of CAFs, which are critical to produce extracellular matrix (ECM) components and the modulation of the tumor microenvironment.

TGF- β is a major cytokine in the tumor microenvironment that promotes fibrosis by stimulating the production of extracellular matrix (ECM) components by cancer-associated fibroblasts (CAFs).

WNT signaling pathway is involved in the regulation of cell proliferation, differentiation, and migration. It also plays a role in the communication between cancer cells and stromal cells, contributing to the formation of a supportive microenvironment for tumor growth.

Gene ontology mining on biological processes, KEGG and Reactome databases for all significant associated modules are shown in **Supplementary Fig. S2**. This study aims to identify new prognostic markers for PDAC survival focusing specifically on the stromal compartment. Consequently, our focus will now shift exclusively to the genes within the yellow module (which was found to be associated to sample stromal content) to find putative prognostic markers for resectable PDAC patients.

Survival analysis on module associated to stromal content

The yellow module included 2459 genes. A log-rank test was performed on each gene, identifying four genes associated to PDAC survival: KCMF1, YARS1, HPGDS and ITGA9-AS1 (see **Supplementary Table S2**).

High expression of KCMF1 and YARS1 were found to be associated to poor PDAC prognosis while low expression of HPGDS and ITGA9-AS1 were found to be associated to good PDAC prognosis (Fig. 2).

The association of the four prognostic genes was also validated in four independent cohorts^{14–17} with each gene showing a significant association with patient survival in at least two of the four validation datasets (see **Supplementary Table S3**).

Next, univariate, and multivariate Cox regression analysis were performed to define the effective risk of those markers and adjusted by stage. KCMF1 (HR = 1.88; p-value = 0.033), YARS1 (HR = 2.03; p-value = 0.02) and HPGDS (HR = 0.51; p-value = 0.016) resulted as good independent predictors for poor and good survival in this PDAC cohort (Fig. 3).

Additionally, HPGDS was confirmed to be a stromal gene by immunohistochemistry (IHC) as shown in Fig. 4. High expression of stromal HPGDS is shown in panel A and B of Fig. 4, while low expression of HPGDS is shown in Fig. 4C-D. The tissue microarrays (TMA) are taken by Human Protein Atlas (HPA) database.

Prostaglandin-D synthase (HPGDS) is a sigma class glutathione-S-transferase family member. The enzyme catalyzes the conversion of PGH2 to PGD2 and plays a role in the production of prostanoids in the immune system and mast cells. The presence of this enzyme can be used to identify the differentiation stage of human megakaryocytes.

Stromal module genes enriched for stromal signature

Furthermore, to determine if the yellow module was truly enriched for stromal genes, we conducted a single-sample Gene Set Enrichment Analysis (ssGSEA) using established PDAC subtypes from Moffitt and Puleo. As illustrated in Fig. 5, the yellow module is indeed enriched for stromal signature genes.

Discussion

Nowadays, it is known that the Tumor Microenvironment (TME) affects the progression and poor prognosis of patients with pancreatic cancer. Although emerging techniques, such as single-cell RNA sequencing (scRNA-seq), can profile and distinguish individual tumor cell types—thus helping to disentangle the tumor

Module	N genes	Clinical trait	Correlation	p-value
MEyellow	2459	STROMAL_MEAN	0.48	2.002e ⁻⁰⁹
MEblack	1087	ACINAR_MEAN	0.61	1.296e ⁻¹⁵
MEpurple	988	ISLET_MEAN	0.46	7.187e ⁻⁰⁹
MEmagenta	1000	TMB	0.44	3.661e ⁻⁰⁸

Table 1. Modules significantly associated to clinical traits. Table summary of the four modules (yellow, black, purple, and magenta) associated to clinical traits, including corresponding number of genes, correlation values and p-values.

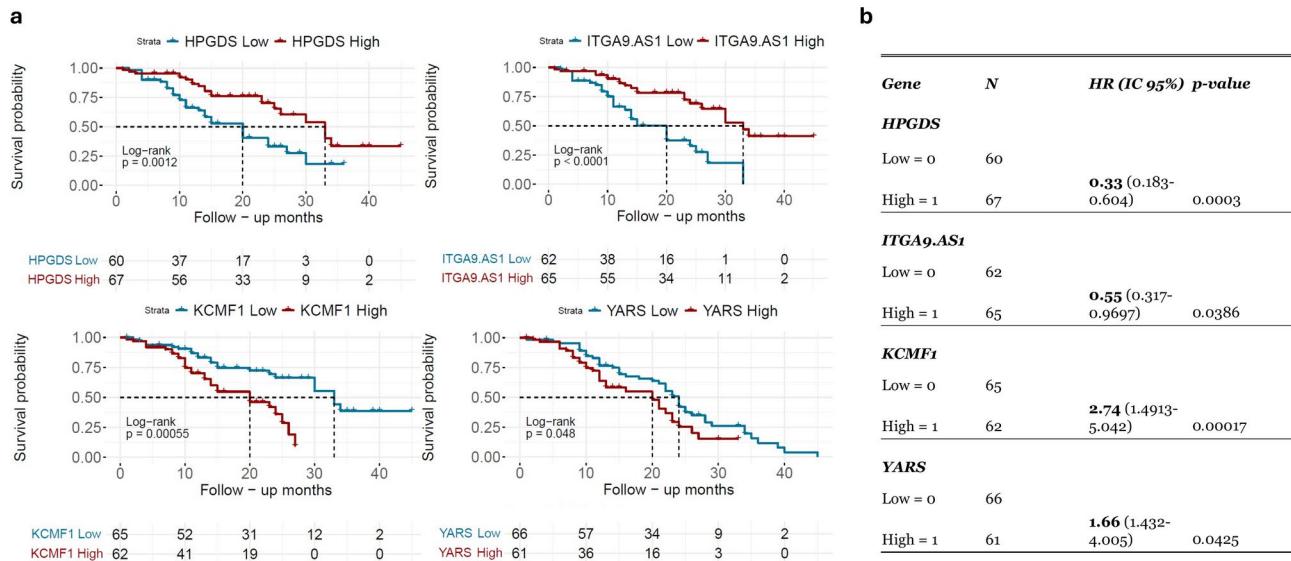


Fig. 2. Survival analyses on module associated to stromal content. **a)** KM curves of significant stromal prognostic markers. **b)** Univariate cox-regression test on significant prognostic markers.

microenvironment (TME)—bulk RNA sequencing remains the most widely used method in clinical settings due to its faster preparation times and lower costs. The publication of the PAAD-TCGA RNAseq dataset has allowed for an in-depth analysis of this type of tumor through various analytical methods. WGCNA has been applied several times to this dataset but has never been used to extract a stromal signature and identify stromal prognostic markers for PDAC.

In this study, the PAAD-TCGA gene expression dataset was used to identify novel prognostic stromal markers for PDAC patients. All patients underwent surgical resection. The WGCNA approach was used to dissect the bulk gene expression profile and identify gene expression signature associated to sample stromal content. WGCNA is a user-friendly and comprehensive tool that has already been extensively applied in cancer research^{18,19}.

Gene co-expression modules obtained in this study were associated to several clinical traits such as TMB, acinar means, islet mean, and stromal content.

We retained to analyze only the module associated to stromal content as for our research focus. The module associated to stromal content (also called yellow module) was functionally enriched for E2F transcription factor which is involved in cell cycle regulation and DNA synthesis. Moreover, KEGG enrichment on yellow module revealed enrichment of TGF-beta signaling and WNT signaling pathways highlighting a possible interplay between these biological processes.

Next, all genes belonging to stromal module were tested for survival association.

We identified four putative stromal-associated prognostic markers for PDAC. High expression of YARS1 and KCMF1 was associated with poor prognosis, while high expression of ITGA9-AS1 and HPGDS was linked to a favorable prognosis. Given the heterogeneity of this cohort, all Cox regression analyses were adjusted for disease stage. The prognostic significance of the four candidate genes: KCMF1, YARS1, ITGA9-AS1 and HPGDS was also validated in four independent datasets of PDAC patients.

KCMF1, which encodes a highly conserved zinc-finger protein, is found to be upregulated in human pancreatic cancer, particularly in Panc1 cells that exhibit a more metastatic phenotype²⁰. Upregulation of nuclear KCMF1 is observed in preneoplastic lesions and several epithelial malignancies, including pancreatic cancer in both mice and humans. In vitro, KCMF1 promotes the proliferation, migration, and invasion of HEK-293 and Panc1 cells²⁰.

YARS1 encodes an aminoacyl-tRNA synthetase involved in protein metabolism, and recent studies have highlighted its role in the metabolic behavior of PDAC²¹⁻²⁵. In a study of Wang and colleagues²⁶, YARS1 was found to be significantly associated with survival outcomes in bladder cancer, with higher expression correlating with poorer prognosis.

Of note, in the study presented here, high expression of HPGDS was found to be associated to patients with good prognosis. High HPGDS expression might help inhibit processes that would otherwise promote tumor progression, such as ECM remodeling and TAF activation²⁷. Moreover, elevated HPGDS expression in PDAC may influence the tumor microenvironment by modulating Th17 cell activity. Specifically, increased PGD2 production could suppress the pro-inflammatory Th17 subsets associated with autoimmunity, thereby reducing inflammation and potentially inhibiting tumor progression. This immunomodulatory effect might contribute to the improved patient outcomes observed with higher HPGDS levels²⁸.

Drugs or compounds that increase PGD2 levels or stabilize its activity might therefore serve as therapeutic options to create a less favorable environment for cancer progression.

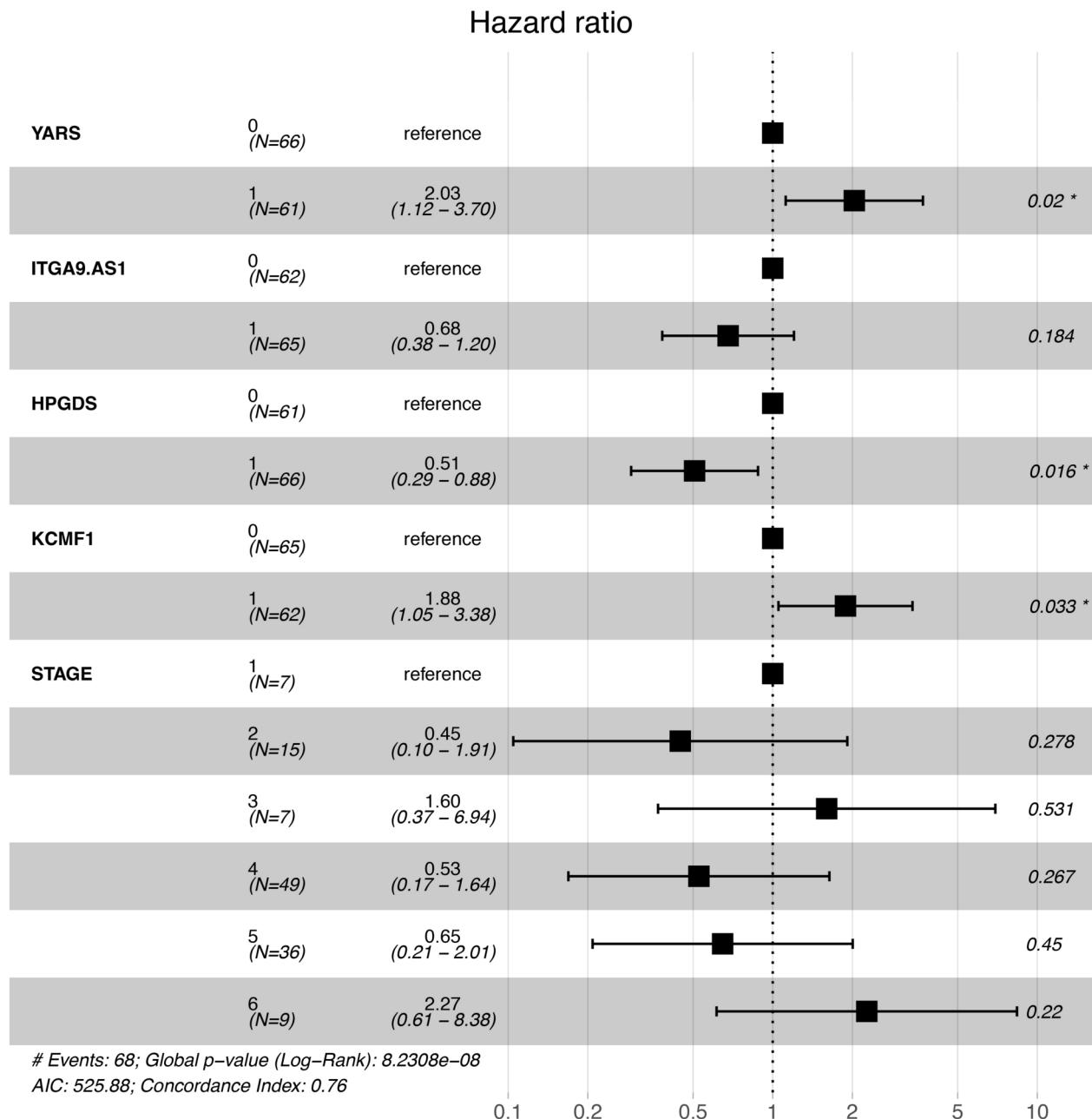


Fig. 3. Cox regression adjusted for stage. Multivariate cox-regression of four putative prognostic markers (YARS1, KCMF1, ITGA9-AS1 and HPGDS) adjusted for stage.

ITGA9-AS1, a long non-coding RNA (lncRNA) associated with the ITGA9 gene, may contribute to better prognosis in pancreatic ductal adenocarcinoma (PDAC) and other cancers by influencing gene expression and cellular pathways critical to tumor suppression and microenvironment regulation²⁹. The biological mechanism of ITGA9-AS1 was elucidated in a study of elderly NSCLC by Liu et al.³⁰ where it was verified that the expression of ITGA9-AS1 elevates ITGA9 expression by competitively binding to miR-4765 and recruiting HNRNPU to stabilize the 3' UTR of ITGA9 mRNA in NSCLC cells. In a clinical setting, CRISPR/Cas systems can be adapted to target lncRNAs by activating their expression and favor patient prognosis. Additionally, delivery systems, such as nanoparticles or viral vectors, can be used to deliver lncRNA-targeting agents or mimics directly to tumor cells, improving specificity and reducing off-target effects.

To validate our methodology and confirm the presence of our identified stromal-associated genes, we applied ssGSEA to compare these genes against previously defined stromal signatures. We found that genes in the yellow module were enriched in stromal signatures from Moffitt³¹ and Puleo³², thus supporting our research hypothesis.

Interestingly, none of the genes identified as putative prognostic markers were present in these published signatures, suggesting that we have potentially uncovered new stromal markers relevant to PDAC survival.

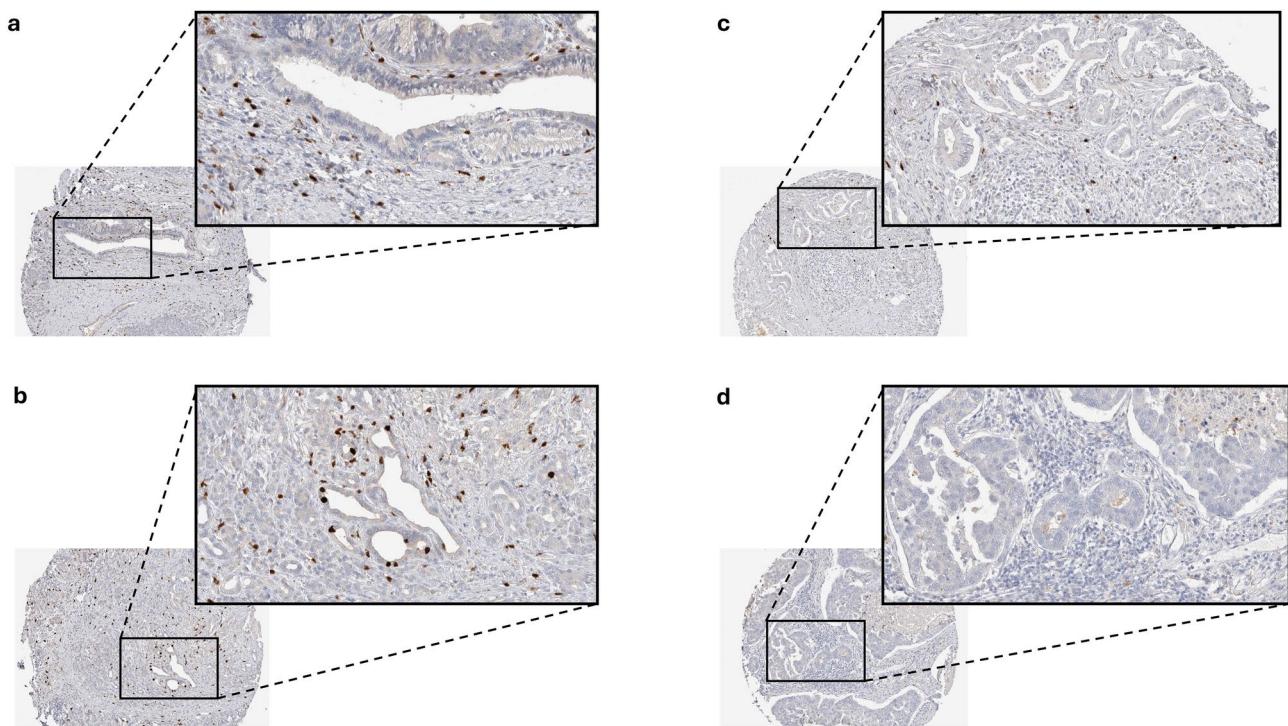


Fig. 4. Validation of stromal expression of HPGDS. **(a-b)** HPGDS expression in stromal compartment of PDAC tissue. **(c-d)** Low expression of HPGDS in stromal compartment of PDAC tissue. Images are taken by Human Protein Atlas.

Although emerging techniques, such as single-cell RNA sequencing (scRNA-seq), can profile and distinguish individual tumor cell types—thus helping to disentangle the tumor microenvironment (TME)—bulk RNA sequencing remains the most widely used method in clinical settings due to its faster preparation times and lower costs.

While our study has identified new prognostic biomarkers in the PDAC stroma, it is important to acknowledge certain limitations. As an initial exploratory analysis using bulk measurement techniques, it lacks functional validation of these biomarkers. Moreover, emerging methods like single-cell RNA sequencing hold promise for addressing new cell-type specific biomarkers.

Overall, our results demonstrate that co-expression networks can effectively extrapolate tumor-stroma-specific biology and underlying biological mechanisms, prompting four novel biomarkers: KCMF1, YARS1, HPGDS and ITGA9-AS1.

Methods

The results shown here are in whole based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. Gene expression data of PAAD-TGCA cohort was obtained from www.linkedomics.org³³. Bulk RNA-seq was performed in paired-end mode on Illumina HiSeq platform. Data were obtained in Rsem normalization.

Patient and sample information was obtained from www.cBioPortal.org³⁴.

Specifically, the stromal, acinar, and islet means were calculated using a transcriptomics-based deconvolution method, which was further validated with a DNA methylation-based tumor deconvolution approach, as detailed in the original publication¹³. TMB values were obtained from the corresponding mutation-profiled samples provided by The Cancer Genome Atlas Research Network³⁵.

Data preprocessing and filtering were performed in R (version 4.3.0).

The analysis was conducted from a starting matrix of 28,057 genes and 140 samples. Duplicated genes were removed obtaining 27,966 genes.

Normalized gene expression data were filtered retaining genes having counts > 1 in at least 70% of all samples obtaining a final dataset of 21,793 genes and 140 samples.

Categorical data of clinical information were transformed into numerical factors to be able to perform clinical traits association studies.

WGCNA³⁶ was performed using a soft-power threshold of 7. Dynamic clustering and cut-tree dynamic functions were used to determine the first group of modules. Next, module eigengenes were used to merge modules with similar expression profiles.

Enrichment for gene ontology terms on detected modules was performed using gprofiler2 R library using GO, KEGG³⁷ and REACTOME³⁸ databases. Significance for enriched terms was set at FDR < 0.2.

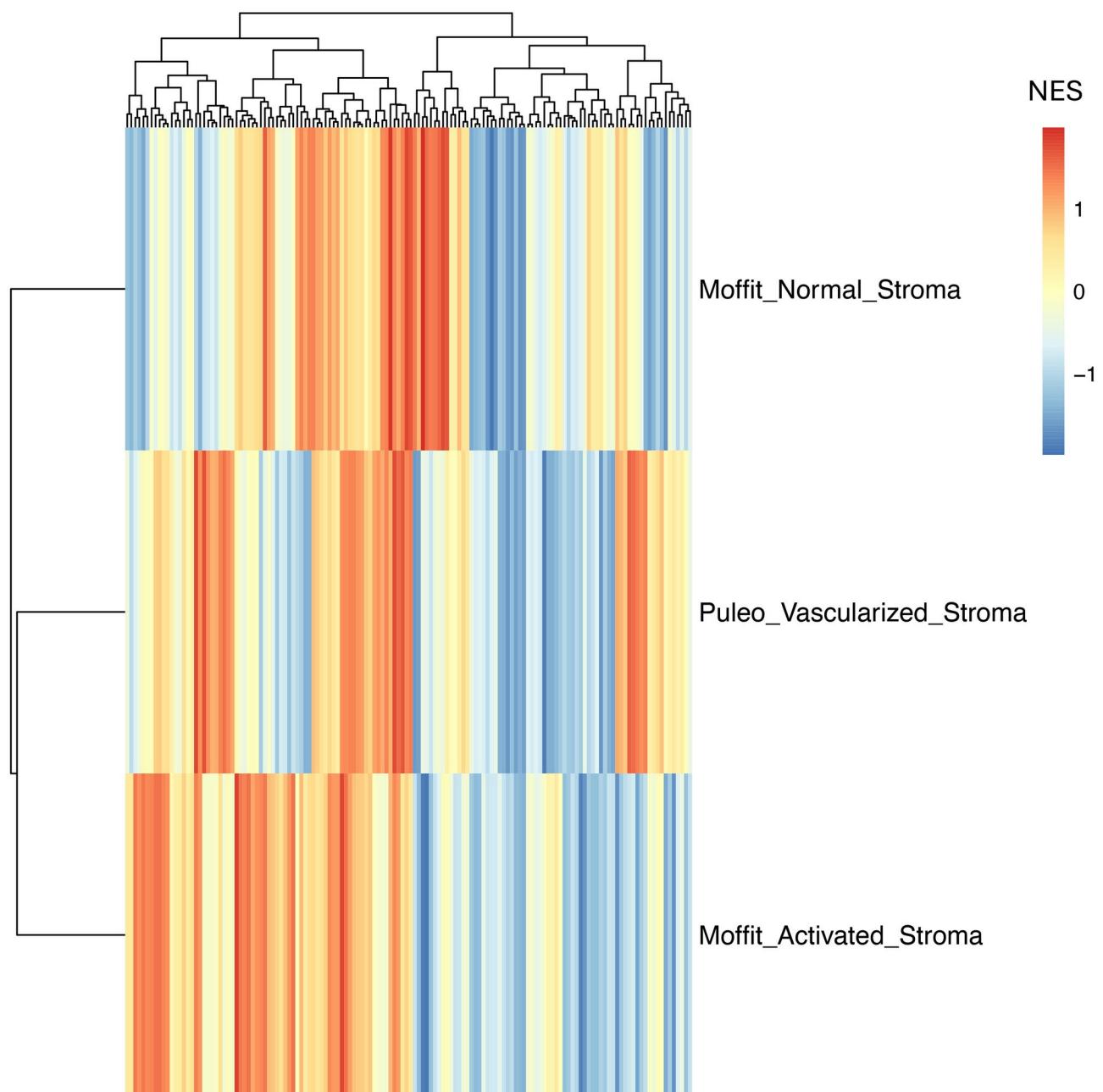


Fig. 5. Validation of stromal signatures in module associated to stromal content. ssGSEA of stromal genes identified by WGCNA and enriched for known stromal signatures of Moffitt and Puleo.

Survival analyses were performed on 132 patients. Eight patients were excluded having only 30 days of follow up. Analysis was performed using `survfit` function in R (`survminer` R library).

Log-rank test was used to test significant differences between patients with high gene expression and patients with low gene expression using median gene expression as threshold. P-values were corrected using BH.

The difference between curves was measured with Cox regression model using `survival` R library with patient stage correction.

To validate that the stromal module was enriched with PDAC stromal genes, we extracted established gene signatures from Moffitt and Puleo. We then performed a single-sample Gene Set Enrichment Analysis (ssGSEA R library) using these signatures on our stromal module.

Data availability

The data reported in this paper are available through www.cbioportal.org (clinical information) and www.linkedomics.org (RSEM gene expression matrix).

Received: 30 August 2024; Accepted: 6 December 2024

Published online: 30 December 2024

References

1. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA A Cancer J. Clinic* **74**, 12–49 (2024).
2. Cancer Facts & Figures 2024. (1930).
3. Boeck, S., Stieber, P., Holdenrieder, S., Wilkowski, R. & Heinemann, V. Prognostic and therapeutic significance of carbohydrate antigen 19–9 as tumor marker in patients with pancreatic cancer. *Oncology* **70**, 255–264 (2006).
4. Daamen, L. A. et al. Postoperative surveillance of pancreatic cancer patients. *Eur. J. Surg. Oncol.* **45**, 1770–1777 (2019).
5. Dayimu, A. et al. Clinical and biological markers predictive of treatment response associated with metastatic pancreatic adenocarcinoma. *Br. J. Cancer* **128**, 1672–1680 (2023).
6. Kondo, N. et al. Prognostic impact of perioperative serum CA 19–9 levels in patients with resectable pancreatic cancer. *Ann. Surg. Oncol.* **17**, 2321–2329 (2010).
7. Musiu, C. et al. Cellular collusion: cracking the code of immunosuppression and chemo resistance in PDAC. *Front. Immunol.* **15**, 1341079 (2024).
8. Wu, C. et al. A selective review of multi-level omics data integration using variable selection. *High-Throughput* **8**, 4 (2019).
9. Lu, H. et al. The immune escape signature predicts the prognosis and immunotherapy sensitivity for pancreatic ductal adenocarcinoma. *Front. Oncol.* **12**, 978921 (2022).
10. Nisar, M. et al. Integrated analysis of microarray and RNA-Seq data for the identification of hub genes and networks involved in the pancreatic cancer. *Front. Genet.* **12**, 663787 (2021).
11. Tan, Z. et al. Analysis of immune-related signatures related to CD4+ T cell infiltration with gene co-expression network in pancreatic adenocarcinoma. *Front. Oncol.* **11**, 674897 (2021).
12. Zhao, R., Wei, W. & Zhen, L. WGCNA-based identification of potential targets and pathways in response to treatment in locally advanced breast cancer patients. *Open Med.* **18**, 20230651 (2023).
13. Cao, L. et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* **184**, 5031–5052.e26 (2021).
14. Zhang, G. et al. Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin. Cancer Res.* **19**, 4983–4993 (2013).
15. Stratford, J. K. et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS Med.* **7**, e1000307 (2010).
16. Yang, S. et al. A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. *Cancer Res.* **76**, 3838–3850 (2016).
17. Ohara, Y. et al. SERPINB3-MYC axis induces the basal-like/squamous subtype and enhances disease progression in pancreatic cancer. *Cell Rep.* **42**, 113434 (2023).
18. Guo, C., Xie, B. & Liu, Q. Weighted gene co-expression network analysis combined with machine learning validation to identify key hub biomarkers in colorectal cancer. *Funct. Integr. Genomics* **23**, 24 (2023).
19. Luo, J. et al. Coexpression network analysis identified MT3 as a hub gene that promotes the chemoresistance of oral cancer by regulating the expression of YAP1. *BMC Oral Health* **23**, 954 (2023).
20. Beilke, S. et al. The zinc-finger protein KCMF1 is overexpressed during pancreatic cancer development and downregulation of KCMF1 inhibits pancreatic cancer development in mice. *Oncogene* **29**, 4058–4067 (2010).
21. Ali, A. et al. Metabolic pathways as a novel landscape in pancreatic ductal adenocarcinoma. *Cancers* **14**, 3799 (2022).
22. Daemen, A. et al. Metabolite profiling stratifies pancreatic ductal adenocarcinomas into subtypes with distinct sensitivities to metabolic inhibitors. *Proc. Natl. Acad. Sci. USA* <https://doi.org/10.1073/pnas.1501605112> (2015).
23. De Santis, M. C., Bockorny, B., Hirsch, E., Cappello, P. & Martini, M. Exploiting pancreatic cancer metabolism: challenges and opportunities. *Trends Mol. Med.* **30**, 592–604 (2024).
24. Miyashita, Y. et al. Metabolic syndrome is linked to the incidence of pancreatic cancer. *eClinicalMedicine* **67**, 102353 (2024).
25. Pervin, J. et al. Clinically impactful metabolic subtypes of pancreatic ductal adenocarcinoma (PDAC). *Front. Genet.* **14**, 1282824 (2023).
26. Wang, Y. et al. Unveiling the role of YARS1 in bladder cancer: A prognostic biomarker and therapeutic target. *J. Cell. Mol. Med.* **28**, 1–20 (2024).
27. Song, Q. et al. Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. *Commun. Biol.* **6**, 760 (2023).
28. Bouch, R. J. et al. Distinct inflammatory Th17 subsets emerge in autoimmunity and infection. *J. Exp. Med.* **220**, e20221911 (2023).
29. Zhang, Y.-L. et al. Integrin α 9 suppresses hepatocellular carcinoma metastasis by rho GTPase signaling. *J. Immunol. Res.* **2018**, 1–11 (2018).
30. Liu, W. et al. ITGA9-AS1 up-regulates ITGA9 by targeting miR-4765 and recruiting HNRNPU to affect the proliferation and apoptosis of non-small cell lung cancer cells. *Cell. Mol. Biol. (Noisy-le-grand)* **69**, 22–28 (2023).
31. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
32. Puleo, F. et al. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterol.* **155**, 1999–2013.e3 (2018).
33. Vasaikar, S. V., Straub, P., Wang, J. & Zhang, B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* **46**, D956–D963 (2018).
34. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
35. Raphael, B. J. et al. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* **32**, 185–203.e13 (2017).
36. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
37. Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
38. Milacic, M. et al. The reactome pathway knowledgebase 2024. *Nucleic Acids Res.* **52**, D672–D678 (2024).

Acknowledgements

Ministero della Salute – Ricerca Corrente 2024

Author contributions

G.M. conceived the research idea and outlined the methods; G.M., S.R., and M.K. drafted the manuscript; M.T. and G.M. conducted the analysis; A.A. and L.G. provided consultation on methods. S.R., M.K., C.C., L.S., G.T., G.S. and L.G. have revised the manuscript and agreed on its content.

Declarations

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/s41598-024-82563-9>.

Correspondence and requests for materials should be addressed to G.M.

Reprints and permissions information is available at 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024