

Deeper insights into long-term survival heterogeneity of Pancreatic Ductal Adenocarcinoma (PDAC) patients using integrative individual- and group-level transcriptome network analyses

Archana Bhardwaj^a, Claire Josse^{b,c}, Daniel Van Daele^d, Christophe Poulet^{b,h}, Marcela Chavez^e, Ingrid Struman^f,
Kristel Van Steen^a

^aGIGA-R Centre, BIO3 – Medical Genomics, University of Liège, Liège, Belgium

^bLaboratory of Human Genetics, GIGA Research, University Hospital (CHU), Liège, Belgium

^cMedical Oncology Department, CHU Liège, Liège, Belgium

^dDepartment of Gastro-enterology, University Hospital (CHU), Liège, Belgium

^eDepartment of Medicine, Division of Hematology, University Hospital (CHU), Liège, Belgium

^fGIGA-R Centre, Laboratory of Molecular Angiogenesis, University of Liège, Liège, Belgium

^hLaboratory of Rheumatology, GIGA-R, University Hospital (CHULiege), Liège, Belgium

Contact Detail: Corresponding author

Archana Bhardwaj

University of Liège

GIGA - Quartier Hôpital

Avenue de l'Hôpital, 11

4000 Liège, Belgium

Email : a.bhardwaj@uliege.be

Summary

Background: Pancreatic ductal adenocarcinoma (PDAC) is categorized as the seventh leading cause of cancer mortality worldwide. Its predictive markers for long-term survival are not well known. Therefore, it is interesting to delineate individual-specific perturbed genes when comparing long-term (LT) and short-term (ST) PDAC survivors, and to exploit the integrative individual- and group-based transcriptome profiling.

Method: Using a discovery cohort of 19 PDAC patients from CHU-Liège (Belgium), we first performed differential gene expression (DGE) analysis comparing LT to ST survivor. Second, we adopted unsupervised systems biology approaches to obtain gene modules linked to clinical features. Third, we created individual-specific perturbation profiles and identified key regulators across the LT patients. Furthermore, we applied two gene prioritization approaches: random walk-based Degree-Aware disease gene prioritizing (DADA) method to develop PDAC disease modules; Network-based Integration of Multi-omics Data (NetICS) to integrate group-based and individual-specific perturbed genes in relation to PDAC LT survival.

Findings: We identified 173 differentially expressed genes (DEGs) in ST and LT survivors and five modules (including 38 DEGs) showing associations to clinical traits such as tumor size and chemotherapy. DGE analysis identified differences in genes involved in metabolic and cell cycle activity. Validation of DEGs in the molecular lab suggested a role of REG4 and TSPAN8 in PDAC survival. Individual-specific omics changes across LT survivors revealed biological signatures such as focal adhesion and extracellular matrix receptors, implying a potential role in molecular-level heterogeneity of LT PDAC survivors. Via NetICS and DADA we not only identified various known oncogenes such as CUL1, SCF62, EGF, FOSL1, MMP9, and TGFB1, but also highlighted novel genes (TAC1, KCNH7, IRS4, DKK4).

Interpretation: Our proposed analytic workflow shows the advantages of combining clinical and omics data as well as individual- and group-level transcriptome profiling. It suggested novel potential transcriptome marks of LT survival heterogeneity in PDAC.

Funding: Télévie-FRS-FNRS

Keywords: PDAC, long-term survival, RNA-seq expression, individual- versus group-level signatures

Introduction

Pancreatic ductal adenocarcinoma (PDAC) accounts for 90% of pancreatic tumors.¹ It is the 4th leading cause of cancer-related death worldwide, while remaining the most lethal among digestive cancers.² PDAC has a complex and dense tumor microenvironment that poses a significant barrier to treatment administration.³ Various factors shape the outcome for complex diseases leading to perturbations of a complex intracellular network.⁴ Disease-relevant genes typically do not operate on their own but may be connected to each other and known disease associated genes of interest.⁵ Network approaches that allow integration with regulatory factors are required to fully map complex diseases, including PDAC.

For PDAC, the overall survival (OS) of patients may be coupled to the mutational status of *KRAS* (Kirsten rat sarcoma viral oncogene) as well as several morphological features.⁶ Also, multiple miRNAs and transcription factors influence metastasis and OS time of PDAC patients.^{7,8} Due to the high lethality of PDAC, intensive research is needed to unravel roots of causes for PDAC survival in general and long-term (LT) versus short-term (ST) survival in particular. In the literature, several criteria for LT and ST survival exist : ST (resp. LT) as surviving ≤ 8 (resp. ≥ 8 months)⁹; LT survival as ≥ 10 years¹⁰; ST (<14 months) and very long-term (≥ 10 years) of survival.¹¹ Very little information is available about regulatory mechanisms involved in the context of <12 months and ≥ 36 months of survival within European populations. We aim to fill this gap and to explore PDAC survival mechanisms by making use of genomics data and by integrating a variety of gene prioritization methods.

Multiple questions are of interest, including ‘How do LT and ST PDAC survivors differ from each other’ and ‘Which survival group is most heterogeneous in terms of transcriptome signatures’. PDAC is featured with intra-tumoral heterogeneity.¹² In general, heterogeneity poses a significant challenge to personalized treatments for PDAC.¹³ Previous classification studies paved the path to a better classification of patients with PDAC based on molecular pathology information¹⁴, molecular features¹⁵ and defined five PDAC subtypes, showing associations with patient outcomes.¹⁶ The identification of subgroups by looking into a perturbed profile of each individual might be another interesting approach. Typically, such (molecular) subtyping analyses require relatively large sample sizes. Alternative and more elaborate approaches are required, better exploiting and combining individual-level and group level profiling, to address the aforementioned questions.

Pathological findings with tumor cells suggest an abundance of gene regulatory networks (GRNs) in humans for various cancers including, breast^{17,18}, prostate¹⁹, and PDAC cancer.²⁰ Network biology approaches have the potential to identify key regulators that are responsible for molecular heterogeneity giving rise to LT and ST PDAC survivor subgroups. Weighted Gene Co-Expression Network Analysis (WGCNA) is such an approach and enables the identification of gene modules and their associations with clinical measurements²¹. For the identification of PDAC key regulators, more work is needed to exploit gene connectivity with earlier identified disease genes via the use of protein interaction networks (PPIs).

The current study teases out PDAC survival associated genes, with a focus on LT survivors (≥ 36 months survival; in contrast to ST survival defined as ≤ 12 months survival) and individual-to-individual differences in whole

transcriptome profiles. To this end, we introduced and implemented a flexible and interpretable omics integrative analysis framework, involving a series of group-level and individual-level viewpoints.

Methods

The study's analytic workflow is depicted in Figure 1 and described in more detail in appendix pp 2-5.

Results

Patients characteristics

All patients were divided into ST (≤ 12 months) and LT (≥ 36 months) survival groups (resp. STS and LTS), as summarized in Figure 2A. A total of 19 patients, comprising 10 STS and 9 LTS, met our inclusion criteria (appendix pp 2;9-11).

Differential gene expression analysis and functional follow-up

RNA was extracted from FFPE tissues and a quality check was performed for paired-end sequencing. The long non-coding gene *MIR205HG* was the topmost differentially down-regulated in the LTS group (p-value=0.008), while the protein coding gene *GKNI*, which encodes for gastrokine1, was the topmost differential up-regulated in LTS (p-value=1.25E-05). The gene ontology analysis linked the genes with the highest expression levels in LTS to the digestive system and lowest expression levels in LTS to the carboxylic binding activity (Figure 2B). Specific domain structures of genes play a significant role in gene regulation and expression. The conserved domain analysis resulted in 112 genes containing at least one domain (Figure 2C; appendix pp 3). Sixteen genes contained an Ig domain, followed by a V-set domain. Later, a unique set of domains was identified in both up- (IgC, IG_like, Trypsin) and down-regulated genes (F-Box, IRS, PKc_MAPKK, IgV) suggesting these genes' regulatory role in PDAC survival mechanisms. Fifty survival genes were identified from all DEGs. We observed 22 DEG genes containing at least one domain that overlapped with the survival gene set (Figure 2D). RT-qPCR confirmed the differential expression observed in LTS versus STS for the genes represented in appendix pp 33. Among them, DEGs *REG4* and *TSPAN8* were validated in lab via RT-qPCR analysis (appendix pp 5:section 3.1).

Group-level survival heterogeneity: Significant clinically relevant modules and their corresponding 3D architectures

All 19 samples with clinical information and gene expression were included in WGCNA (appendix pp 3). Genes with similar expression were grouped into gene modules via average linkage hierarchical clustering. By use of a dynamic tree-cutting algorithm, a total of 96 distinct co-expression modules were identified. Correlated modules were merged with a cut-off height of 0.25, resulting in 35 modules containing 66 to 2010 genes per module. Module M34 was the smallest module consisting of 66 genes, whereas M8 was the largest module comprising 2010 genes (Figure 1:C;

Figure 3A). The identified 35 modules covered 97 percent of the 18880 input genes. For those 35 modules, we derived the corresponding module eigengenes (MEs).

Association of clinical features with dysregulated genes may help to clarify genes important for disease development. All identified DEGs (173 in total) were distributed in 25 modules. Five modules had a significant correlation with clinical phenotypes (adjusted p-value <0.05): M7, M9, M15, M30, and M34 (appendix pp 34). Module M9 was found to be significantly associated with tumor size ($r^2=0.72$, adjusted p-value=0.01) and T stage ($r^2 = 0.68$, adjusted p-value=0.03). M9 consisted of the highest number of DEGs (27 genes). Two other modules, M7 ($r^2=0.73$, adjusted p-value=0.01) and M30 ($r^2=0.71$, adjusted p-value=0.02), were negatively associated with time between surgery and chemotherapy clinical traits. M30 contained 10 DEGs. Module M34 was significantly associated with tumor size by imagery ($r^2=0.67$, adjusted p-value<0.05). Interestingly, two modules were significantly associated with chemotherapy: a positive association for M15 ($r^2=0.68$, adjusted p-value=0.04) and a negative association for M9 ($r^2=-0.68$, adjusted p-value<0.04). Gene-gene interactome networks were developed for clinically relevant modules M7, M15 and M34 (Figure 3B). The overlap between DEGs and genes in five modules (M7, M9, M15, M30, M34) is shown in a Venn-Diagram (Figure 3C), and identified 27, 10, and 1 gene as part of M9, M30, and M34, respectively.

Group-level survival heterogeneity: Functional analysis of clinically relevant modules

Clinically pertinent gene modules were functionally analyzed in Cytoscape with the ClueGO plug-in (appendix pp 3-4) to visualize functionally grouped network. Module M9 was linked to 33 significant pathways (adjusted p-value <0.05) distributed over ten groups, such as extracellular matrix (ECM) organization (86 genes) and collagen formation (37 genes) (Figure 4A). Genes regulating the cell cycle and modulating ECM at molecular or cellular levels have been linked to cancer drug targeting and cancer cell plasticity.³² M7, also negatively associated with chemotherapy, contained 91 significant pathways, distributed into three groups, such as proteasome (4 genes) and the regulation of RAS by GAPs (5 genes) (Figure 4B). Module M15, positively associated with ‘chemotherapy’, was enriched with 11 significant pathways distributed into five groups (Figure 4C). More detail is given in appendix (pp 5: section 3.2)

Individual-specific survival heterogeneity: Quantification of heterogeneity between individual transcriptome profiles

To assess heterogeneity in LT survival patients, we constructed individual perturbation expression profiles (PEEPs)²² (appendix pp 4). It resulted in 6336 significantly perturbed genes across LT PDAC survivors (Figure 1:D; Figure 5A). The frequency of disrupted genes in each LT survivor L_i ($i= 1, \dots, 9$) was L1:12, L2:1412, L3:43, L4:474, L5:179, L6:319, L7:957, L8:150 and L9:2789 (Figure 5A). Various genes were uniquely perturbed in one LTS patient only. Only one group-wise DEG was shared among 3 LT survival subjects, namely TNNI3. Also, at most six DEGs (IRS4, KLRC3, CLDN18, NPY, CNTN6, TAC1) were common to 2 out of 9 patients. Hence, for the majority of perturbed genes shared among LT survivors, no evidence was found about them being differentially expressed in a group comparison between LT and ST survivors. Among genes other than significant DEGs, only one was common to 7 out of 9 individuals: *NOSTRIN*, associated with nitric oxide pathways. No other genes were shared by 8 or all 9 LTS. Five out of 9 LTS patients shared *DTYMK* as perturbed gene in their individual transcriptome profile or PEEP. Six genes

(*PDXDC1*, *ATF7IP2*, *LIN7C*, *JTB*, *TTL*, *DVL2*), which regulate the ERG signal transduction pathways, were retained in 4 out of 9 LTS patients, were significantly involved in transcriptional mis-regulation in cancer (adjusted p-value=0.025). There were respectively 41 and 180 genes conserved, in 3 and 2 out of 9 LT survivors (Figure 5A; appendix pp 5: section 3.3).

Individual-specific survival heterogeneity: Functional pathway and domain analysis in long-term PDAC survivors

We furthermore examined the extent to which the LT survivors reflected disruptions in KEGG and Reactome pathways, and identified pathways that were significantly enriched in at least one LT individual. In-depth analysis revealed 18 pathways in at least two LT survivors (Figure 5B; appendix pp 12-18). Thus, 175 pathways were uniquely perturbed in an LT PDAC survivor i.e. not shared among LT survivors (appendix pp 5: section 3.3).

Specific domain structures of genes play a significant role in gene regulation and expression. Hence, we also investigated the domain structures of perturbed genes in PEEPs of LTS to understand their potential regulatory mechanism in LT survival. A total of 47 enriched domains (adjusted p-value<0.05) were identified (appendix pp 19-20). The distribution of motifs that were commonly shared by 2 out of 9 LT survivors are shown in Figure 5B. For each LT survivor, we constructed two hierarchical trees based on the genes potentially involved in multiple domains and pathways, one for each for LT survivor. Resulting trees were statistically compared to identify the common branches. For instance, nine genes (indicated with pink, blue and green) were involved in common branches in LT2 (Figure 5C). More detail is given in appendix (pp 5: section 3.3).

Exploitation of gene connectivity: systems views

Gene connectivity via reference networks can further highlight interesting gene clusters linked to LT survivors. In a first approach, we developed a disease module via DADA^{5,23}, which uses the human protein interactome network structure to prioritize disease genes, while at the same time removing possible biases induced by gene degree distributions (appendix pp 4). The disease module hypothesis proposes that disease regulatory genes should form one or a few large connected components in a human interactome. In this study, we restricted our seed genes (i.e., genes that play significant roles in PDAC according to the prior biological knowledge) to PDAC survival (*SMAD4*, *CDKA2*, and *KRAS*) and PDAC responsiveness based on a literature search and as identified from the DisGeNET database²⁴ (appendix pp 21-22). Only the top 1% of DADA ranked genes were retained (Figure 6A I-IV; Figure 1:J), leading to 70 genes. Only one DADA top gene was also previously identified as DEG (*DKK4*), as shown in (Figure 6C). We also looked at the overlap between DADA-based 1% top-ranked genes and perturbed genes as highlighted by the PEEPs of individuals belonging to the long-term survival PDAC patient group (Figure 1:L). There were 23 genes in total. None of these common genes had previously been identified as DEGs. Out of 23, we identified 7 DADA top-ranked genes in common to clinical gene modules as identified before (Figure 6C; Figure 1:K; appendix pp 24-28). Only a single gene was shared by at least (actually exactly) three LT subjects, namely *GLI2*. Three genes (*RAC1*, *FOSL1*, and *EGF*) were shared by two out of 9 LT survivor PEEPs. Furthermore, three genes (*JAG2*, *TGFA*, *HDAC1*) were uniquely perturbed in a LT survivor (Figure 6B).

In a second approach, we integrated individual-specific gene perturbation information (from PEEPs) with group-level DEG findings. For this, we used NetICS, which further allows unraveling inter- and intra-patient gene expression heterogeneity (appendix pp 4-5; Figure 1:E). Also in this approach a ranked list of genes was generated. The ranks are based on the gene scores, acquired through network diffusion algorithms (Figure 6D, appendix pp 23). Similar to the DADA approach, we focused on the top 1% of ranked genes for each LT survival patient, leading to 500 genes. Those 500 genes constituted a subset of PEEP genes. Only 13 genes out of 500 were also DEGs, including 6 genes that were additionally linked to clinical disease modules (Figure 6D). Among these 13 DEGs, *TNNI3* was NetICS top ranked, and was shared in its significance by 3 out of 9 LT survivors. It was also associated with the M7 module of clinical relevance (Figure 6D; Figure 1:G; appendix pp 24-28). Notably, *NOSTRIN*, a unique to NetICS gene (i.e., not highlighted by any other method shown in Figure 6D) was common to 7 out of 9 LT subjects. Furthermore, we found 14 genes common to DADA and NetICS gene prioritization methodologies (Figure 6E; Figure 1:J; appendix 40), involving the pathways such GPCR, Notch signaling pathway and many others. This common gene set did not include *TNNI3* nor *NOSTRIN*. The percentage of LTS PEEP genes not included in the top 1% DADA gene list is 27% (384/1440) and is similar to the percentage of LTS PEEP genes not included in the top 1% NetICS gene list (263/963).

Discussion

Identifying molecular PDAC cancer drivers is critical for implementing precision medicine in clinical practice. Typically, the optimization and fine tuning of gene prioritization methods require large datasets²⁵. Despite the small sample size of this study, we identified genes showing associations with multiple clinical traits²⁶, and derived plausible links between long-term survival of patients and genes, pathways and protein domains by exploiting multiple approaches, including the combination of individual-level with group-level information in integrated analysis workflows. Throughout the entire study, we have relied on several statistical approaches to determine statistical significance with small samples.

PDAC accounts for over 90% of pancreatic cancer and is a lethal malignancy with very high mortality rates. The gene regulatory landscape of PDAC is defined by four mutational “mountains” (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*), which are the main drivers of PDAC²⁷. Cancer diseases are heterogeneous at different scales: group level, individual level, tumor type, cell level. This study reports on PDAC gene expression differences in patients who survived ≥ 36 months (long term LT) or < 12 months (short term ST). Via advanced genomic profiling of PDAC survivors, we aimed to obtain more insights into LTS-relevant mechanisms that contribute to PDAC heterogeneity.

In this work, we identified known PDAC driver genes associated with survival, including *ROBO2*, *ZG16B*, and *PLXNA1*^{28,29} (appendix pp 2). A thorough investigation of gene expression differences between LT and ST PDAC survivors highlighted gene involvement in immune responses (*CEACAM20*, *C6orf13*, *IRS4*, *CXCL17*), cell cycle (*SPDYE3*, *HLA-DQA2*, *CLDN*) and metabolic pathways (*GBA3*, *LIPN*), further highlighting the importance of these pathways in PDAC disease survival^{30,31}. All of these findings evidence that genes linked to immune responses could be useful in effective therapies for PDAC survival³². We also identified a downstream target of *KRAS* (*MUC16*) as

DEG, supporting *KRAS* implications in survival³³. Also, we observed modifications of *GKN1*, *KRT6*, and *ANKRD43* gene expressions in LTS, known to induce apoptosis and metastasis cancer^{34,35}. A previous study showed *REG4* as a serological marker for PDAC³⁶. Very little information exists though about the role of *TSPAN8* in PDAC. However *TSPAN8* promote cancer cell stemness via activation of Hedgehog signaling³⁷. Furthermore, validation of DEGs via experimental work suggested a role of *REG4* and *TSPAN8* in PDAC survival mechanisms. Conserved domain database represents the curated information on conserved domain architectures of various proteins that shows implication in tumor initiation, tumor progression, angiogenesis and metastasis. The presence of multiple immunogenic domains (IGV, V-SET) in identified DEGs further supports recent activities towards cancer therapy³⁸, and in-depth investigation of immunity cycles in relation to long-term survival in PDAC patients.

Systems biology approaches can provide immediate functional insights by revealing interactions between genes³⁹. A motivation for WGCNA is that genes functioning together are regulated or co-expressed together⁴⁰. Ballouz and cauthor⁴¹ suggested a minimal of 20 samples to predict meaningful functional connectivity. This forced us to pool STS and LTS together for WGCNA analysis on 19 patients and to link thus identified gene modules to clinical traits with non-parametric statistics whenever appropriate. Multiple studies have indicated an association of early survival in PDAC to tumor size⁴². Additionally, multiple targets have been identified in the form of DEGs being associated with numerous traits such as tumor size, and the time between surgery and chemotherapy. In our study, identified several clinically relevant WGCNA gene modules (e.g., a gene module associated to time between surgery and chemotherapy with DEGs *LYZ*, *DKK4*, *CA14*, *NASE7*, *TSPAN8*, *GKN1*, *GKN2*, *SNORD116-18*, *DKK4*), which warrants further exploration on increased sample sizes in the future. Notably, *TSPAN8* serves as a prognostic marker in other cancer types as well³⁷. Apart from time between surgery and chemotherapy, time to surgery may play an important role in PDAC as well (waiting for more than 30 days for surgery after diagnosis has been associated to an increase in tumor size)⁴³. DEG *DKK4* (also top 1% DADA gene) is the least studied protein from the Dickkopf (*DKK*) family, which includes *DKK3*⁴⁴ and *DKK1*⁴⁴. The fact that *DKK4* did not appear in NetICS's prioritization gene list, nor in PEEPs of LTS, seems to suggest that *DKK4* may be more promising in relation to controlling the survival of patients with PDAC rather than explaining individual heterogeneity among long-term PDAC survivors.

The identification of prognostic factors is complicated in the presence of individual-to-individual heterogeneity⁴⁵. Detailed individual-specific omics profiling may be required to provide novel insights into LT survival in pancreatic cancer disease⁴⁶. DEGs alone are unlikely to fully characterize individual (LT) survival, as observed for other complex traits²². Previous studies^{14,47,48} emphasized the existence of subgrouping of PDAC patients in general, based on expression profiling of samples. Our study showed that any LTS patient only exhibits a small fraction of group-wise DEGs in their PEEP profiles and shows a deep level of gene expression heterogeneity. Notably, several genes were uniquely perturbed in an LT survivor, which strengthens our belief that LTS patients exhibit more abundant levels of heterogeneity. Careful inspection of PEEPs across LT survivors highlighted biological signatures: focal adhesion⁴⁹, and ECM receptors⁵⁰. Interestingly multiple PDAC responsive pathways⁵¹ were enriched across several LT survivors and led to further subgrouping of LT survivors. Understanding these pathways may provide novel insight into the LT survival mechanism in PDAC. PEEP analysis identified *FCGR3A*, a potential biomarker in PDAC⁵². Two genes,

NOSTRIN and *ADGRG6*, were shared by 66% of LTS, and have been reported before to be associated with PDAC survival.^{40,53}

Drugs bind to their target proteins and perturb the transcriptome of a cancer cell⁵⁴. In our study, analytic functional analysis of individual PEEPs helped to decode homogeneity patterns within LTS. Heterogeneity at the gene level may go hand in hand with homogeneity at the pathway level as different gene perturbations may lead to disruptions in the same molecular pathway. The use of network-centric approaches resulted in various oncogenes such as *CUL1*, a central component of *SCF*⁵⁵, *EGF*, *FOSL1*⁵⁶, *MMP9*⁵⁷, and *TGFBI*³¹, already known as anticancer targets. Different transcription factors (*GLI2* and *GL3*) were identified, linked to the *KRAS* mechanism of pancreatic tumorigenesis⁵⁸. Identified Immunogenic gene (*CDON*) and regulatory gene (*HDAC1*) targets could play significant roles in the future immunotherapeutic strategies in long-term PDAC survivors⁴⁶. *CD8* revealed in our study is in line with recent studies in which *CD8* expression profiling was linked to an immunologic subtype of PDAC with favorable survival⁵⁹. These results indicate the advantages of adopting integrative analysis pipeline that combining knowledge about network-driven disease modules with individual-specific gene perturbation profiling even for small sample sizes,. Unlike DEG-oriented therapeutic target selection for cancers, commonly used to date, we promote the exploitation of analytic frameworks in which multiple network-centric approaches are used for the identification of patient-specific therapeutic targets. This will boost cancer prognosis and treatment in the context of personalized medicine.

Conclusion

For the first time in PDAC patients, we demonstrated and applied an integrative analytic workflow that combines clinical and omics data, as well as individual- and group-level transcriptome profiling. We showed the utility of network-based approaches, disease modules and multi-scale functional analyses (gene, protein domain, pathway), leading to the identification of known oncogenes and previously unreported marks contributing to heterogeneity in long-term PDAC survival.

Contributions

A.B under supervision of K.V.S performed the detailed BIOINFORMATICS data analyses. D.V.D. and M.C. designed the protocol to obtain patient materials and supervised the extraction of clinical information from CHU Liège databases. A.B. and K.V.S wrote the manuscript with input from all co-authors. C.J performed the nucleic acid extractions. C.J, I.S and C.P participated in the scientific discussions related to the conception of this manuscript.

Declaration of Interests

The authors declare no competing interests

Acknowledgments

We thank the Biobanque of Liège University Hospital and the GIGA Genomics Platform for sequencing. A.B, I.S and K.V.S acknowledge funding by Télévie 2015 “PDAC-xome: Exome sequencing in PDAC” (convention n° 7.4629.15), Télévie 2016 “Drivers and markers in pancreatic cancer” (convention n° 7.4502.16), and FRS-FNRS – CDR 2017 “SysMedPC” (convention n° J.0061.17).

Accession Numbers

Data deposited in GEO with accession number GSE150043.

Figure Legends

Figure 1: Flexible and interpretable omics integrative framework for RNA-seq data collected on two groups of patients, exemplified on PDAC ST/LT survival. RNA-seq quality-controlled data are inputted for A) Survival analysis; B) Group-based differential analysis via DESeq2⁶⁰; C) Weighted gene co-expression network analysis WGCNA²¹; D) Individual-based differential analysis (appendix pp 2-5); E) Genes are ranked based on the integration of individual and group-based differentially expressed genes via NetICS⁶¹; F-H) NetICS specific top 1% ranked genes are traced back in multiple previous analyses (A through E); I) DADA⁵ analysis starting from disease genes; J-L) DADA specific top 1% ranked genes are traced back in previous analyses (A through E).

Figure 2: Overall Kaplan–Meier survival analysis of the ST and LT PDAC cohorts: A) Patient characteristic data for a selection of PDAC relevant traits are shown as mixed bar and heat map plot. P1 to P13 refer to patient specific clinical traits analyzed in this study (selective data has been shown in plot; full details given in appendix pp 9-11). P1 indicates Tumor stage (from 1 to 4). P3, P5, P6, P7, P12 and P13 indicate the frequency of number of nodes analyzed, time between surgery and chemotherapy (in days), disease free survival, OS (in months), tumor size by imagery (in mm) and Time between imagery and surgery, respectively. Remaining P2, P4, P8, P9, P10, P11, refers to status of N stage, surgical margin invaded by tumor cells, vascular resection, re-hospitalization after surgery, vascular contact, and artery contact, respectively. Here 0 and 1 indicate no and yes, reps. P7 clinical trait denotes overall survival and was used for the development of the Kaplan-Meier survival curves for short-term (ST) and long-term (LT) PDAC Survivors (STS: S1 to S10; LTS: L1 to L9); B) Identification of significant gene ontology of associated up and down-regulated DEGs and their relevant functions. Up and down-regulated genes are highlighted with red and green dots, respectively. The size of data points increases with increased significance (uncorrected for multiple testing – see appendix pp 3); C) Top-ranked conserved domains in differentially expressed gene sets; D) Venn-diagram showing the number of identified genes that are common to or different in multiple first-line analysis strategies (CDD:

conserved domain database analysis, DGE: differential gene expression analysis, SA: survival analysis (appendix pp 3)).

Figure 3: Clinical relevance of gene co-expression modules: A) Heatmap indicating the number of genes involved in each WGCNA-derived gene module; B) Network topology of three modules (M7, M15, M34), where nodes are genes and connections among nodes represent gene-gene interactions. In each network, the gene names are indicated in the circular layout as derived from Cytoscape.⁶²; C) Venn diagram indicating the common genes between the identified significant DEGs and the five previously identified clinically relevant modules.

Figure 4: Functional follow-up of clinically relevant gene expression modules: A) Ten groups for module M9 comprising 33 significantly linked pathways; B) Three groups identified in the M7 modules; C) Depiction of the five groups identified in M15; For A-C, redundant groups with >50% overlap were merged. Each node in the network represents an enriched term; the size of each node follows the extent of enrichment significance. Connection among different nodes are based on kappa scores (≥ 0.4), as available from ClueGO.

Figure 5: Genomic distributions of differentially expressed genes (DEGs) and PEEPs related to PDAC survivors using Circos plots and functional profiles of perturbation data: A) first outermost circle labeled with numbers represent chromosomes (same colors); the outermost track represents DEGs (up-regulated and down-regulated DEGs as scattered points); the nine innermost circles refer to the z-score for each LT survivor (LTS: ranging from LT1 to LT9) as scattered points. We have indicated perturbed genes only for chromosome 1 to 22 via track 2 to track 10 (outer to inner); B) Enriched KEGG pathways (P1 to P19 (out of 193)) and motifs common to at least 2 out of 9 LT individuals, shown via Circos Table Viewer (appendix pp 12-20). Each link refers to an LT survivor and a significantly enriched pathway (adjusted p-value < 0.05)/enriched motif based on the perturbed gene set found in that individual (data for LT2, LT7 and LT9 are shown). Uniquely enriched pathways across LT survivors are given in appendix pp 12-18; C) Visual comparison of two dendrograms developed from genes linked to enriched pathway and motif profiles. Similar sub-trees are connected with lines of the same color, while tree branches leading to distinct sub-trees are indicated with dashed lines.

Figure 6: Exploitation of gene connectivity for LT PDAC survivor gene prioritization: A) DADA-oriented multi-step disease module identification: PDAC seed gene selection (I), restriction to top 1% of ranked genes (II-III) and intersection of retained gene list with individual perturbation gene expression profiles for LT survivors (IV); B) DADA-derived top-ranked genes found in at least one, two, or three LT survivors, indicated in green, orange and pink, respectively; C) Common genes to DADA and other gene prioritization approaches: DEGs, clinically relevant WGCNA gene modules, and PEEPs; D) Same as C) but with NetICS instead of DADA; E) Venn diagram showing the overlap between genes prioritized via NetICS and DADA. Common genes to top 1% NetICS individual gene lists and top 1% DADA genes are highlighted via arrows in C) and D).

References

- 1 Hidalgo M, Cascinu S, Kleeff J, *et al.* Addressing the challenges of pancreatic cancer: future directions for improving outcomes. *Pancreatology* 2015; 15: 8–18.
- 2 Valle S, Martin-Hijano L, Alcalá S, Alonso-Nocelo M, Sainz Jr B. The ever-evolving concept of the cancer stem cell in pancreatic cancer. *Cancers* 2018; 10: 33.
- 3 Olive KP, Jacobetz MA, Davidson CJ, *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*; 2009; 324 : 1457-1461
- 4 Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* 2011; 12: 56-68.
- 5 Erten S, Bebek G, Ewing RM, Koyutürk M. DADA: Degree-Aware Algorithms for Network-Based Disease Gene Prioritization. *BioData Min* 2011; 4: 19–19.
- 6 Schlitter AM, Segler A, Steiger K, *et al.* Molecular, morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes. *Scientific reports* 2017; 7: 1-12.
- 7 Liu B, Yang H, Taher L, *et al.* Identification of prognostic biomarkers by combined mRNA and miRNA expression microarray analysis in pancreatic cancer. *Translational oncology* 2018; 11: 700–14.
- 8 Xiong G, Huang H, Feng M, *et al.* MiR-10a-5p targets TFAP2C to promote gemcitabine resistance in pancreatic ductal adenocarcinoma. *Journal of Experimental & Clinical Cancer Research* 2018; 37: 76.
- 9 Duconseil P, Gilibert M, Gayet O, *et al.* Transcriptomic analysis predicts survival and sensitivity to anticancer drugs of patients with a pancreatic adenocarcinoma. *The American journal of pathology* 2015; 185: 1022–1032.
- 10 Stark AP, Sacks GD, Rochefort MM, *et al.* Long-term survival in patients with pancreatic ductal adenocarcinoma. *Surgery* 2016; 159: 1520–1527.
- 11 Chen R, Pan S, Ottenhof NA, *et al.* Stromal galectin-1 expression is associated with long-term survival in resectable pancreatic ductal adenocarcinoma. *Cancer biology & therapy* 2012; 13: 899–907.
- 12 Rao M, Gao M, Delgado AP, *et al.* PO-277 Single-cell RNA-seq analysis of human pancreatic ductal adenocarcinoma. BMJ Publishing Group Limited, 2018.
- 13 Zhao L, Zhao H, Yan H. Gene expression profiling of 1200 pancreatic ductal adenocarcinoma reveals novel subtypes. *BMC cancer* 2018; 18: 603.
- 14 Bailey P, Chang DK, Nones K, *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016; 531: 47-52.
- 15 Peran I, Madhavan S, Byers SW, McCoy MD. Curation of the pancreatic ductal adenocarcinoma subset of The Cancer Genome Atlas is essential for accurate conclusions about survival related molecular mechanisms. *Clinical Cancer Research* 2018; 24:3813-3819.
- 16 Puleo F, Nicolle R, Blum Y, *et al.* Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterology* 2018; 155: 1999-2013

- 17 Presson AP, Yoon NK, Bagryanova L, *et al.* Protein expression based multimarker analysis of breast cancer samples. *BMC cancer* 2011; 11: 230.
- 18 Wirapati P, Sotiriou C, Kunkel S, *et al.* Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Research* 2008; 10: R65.
- 19 Wang L, Tang H, Thayanithy V, *et al.* Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer research* 2009; 69: 9490-9497.
- 20 Giulietti M, Occhipinti G, Principato G, Piva F. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. *Cellular Oncology* 2016; 39: 379–88.
- 21 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008; 9: 559.
- 22 Menche J, Guney E, Sharma A, *et al.* Integrating personalized gene expression profiles into predictive disease-associated gene pools. *NPJ systems biology and applications* 2017; 3: 1-10.
- 23 Sharma A, Kitsak M, Cho MH, *et al.* Integration of Molecular Interactome and Targeted Interaction Analysis to Identify a COPD Disease Network Module. *Scientific Reports* 2018; 8: 1-14.
- 24 Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 2019. <https://doi.org/10.1093/nar/gkz1021>. D845-D855.
- 25 Bailey MH, Tokheim C, Porta-Pardo E, *et al.* Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018; 173: 371-385
- 26 Kim JH, Byun SJ, Park SG, Oh YK, Baek SK. Interval between Surgery and Radiation Therapy Is an Important Prognostic Factor in Treatment of Rectal Cancer. *Cancer Res Treat* 2012; 44: 187–194.
- 27 Haider S, Wang J, Nagano A, *et al.* A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. *Genome medicine* 2014; 6: 105.
- 28 Biankin AV, Waddell N, Kassahn KS, *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 2012; 491: 399-405.
- 29 Zhu T, Gao Y-F, Chen Y-X, *et al.* Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. *Oncotarget* 2017; 8: 21281-21289.
- 30 Osaki M, Oshimura M and, Ito H. PI3K-Akt pathway: its functions and alterations in human cancer. *Apoptosis* 2004; 9: 667–676.
- 31 Glazer ES, Welsh E, Pimiento JM, Teer JK, Malafa MP. TGFβ1 overexpression is associated with improved survival and low tumor cell proliferation in patients with early-stage pancreatic ductal adenocarcinoma. *Oncotarget* 2017; 8: 999-1006.
- 32 Paniccchia A, Merkow J, Edil BH, Zhu Y. Immunotherapy for pancreatic ductal adenocarcinoma: an overview of clinical trials. *Chin J Cancer Res* 2015; 27: 376–391.
- 33 Fan Z, Fan K, Yang C, *et al.* Critical role of KRAS mutation in pancreatic ductal adenocarcinoma. *2018* 2018; 7: 1728–1736.
- 34 Altieri F, Di Stadio CS, Federico A, *et al.* Epigenetic alterations of gastrokine 1 gene expression in gastric cancer. *Oncotarget* 2017; 8: 16899-16911.

- 35 Yao H, Yang Z, Liu Z, *et al.* Glypican-3 and KRT19 are markers associating with metastasis and poor prognosis of pancreatic ductal adenocarcinoma. *Cancer biomarkers* 2016; 17: 397–404.
- 36 Takehara A, Eguchi H, Ohigashi H, *et al.* Novel tumor marker REG4 detected in serum of patients with resectable pancreatic cancer and feasibility for antibody therapy targeting REG4. *Cancer science* 2006; 97: 1191–1197.
- 37 Lin X, Bi Z, Hu Q, *et al.* TSPAN8 serves as a prognostic marker involving Akt/MAPK pathway in nasopharyngeal carcinoma. *Annals of translational medicine* 2019; 7.
- 38 Shurin MR. Immunological targets for cancer therapy: new recognition. *ImmunoTargets and therapy* 2018; 7: 83-85.
- 39 Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A* 2006; 103: 17973–17978.
- 40 Wang J, Xia S, Arand B, *et al.* Single-Cell Co-expression Analysis Reveals Distinct Functional Modules, Co-regulation Mechanisms and Clinical Outcomes. *PLoS computational biology* 2016; 12: e1004892.
- 41 Ballouz S, Verleyen W, Gillis J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 2015; 31: 2123–2130.
- 42 Hur C, Tramontano AC, Dowling EC, *et al.* Early Pancreatic Ductal Adenocarcinoma Survival Is Dependent on Size: Positive Implications for Future Targeted Screening. *Pancreas* 2016; 45: 1062-1066.
- 43 Marchegiani G, Andrianello S, Perri G, *et al.* Does the surgical waiting list affect pathological and survival outcome in resectable pancreatic ductal adenocarcinoma? *HPB* 2018; 20: 411–417.
- 44 Zhou L, Husted H, Moore T, *et al.* Suppression of stromal-derived Dickkopf-3 (DKK3) inhibits tumor progression and prolongs survival in pancreatic ductal adenocarcinoma. *Science Translational Medicine* 2018; 10: eaat3487.
- 45 Stone B, Thibodeau BJ, Baschnagel AM, *et al.* Can Gene Expression Profiling Identify Pancreatic Ductal Adenocarcinoma Patients with Short or Long-Term Prognosis? *JOURNAL OF THE PANCREAS* 2018; 19: 118–125.
- 46 Dal Molin M, Wood LD. Very long-term survival in pancreatic cancer. *Aging (Albany NY)* 2015; 7: 360.
- 47 Moffitt RA, Marayati R, Flate EL, *et al.* Virtual microdissection identifies distinct tumor-and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature genetics* 2015; 47: 1168.
- 48 Waddell N, Pajic M, Patch A-M, *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015; 518: 495-501.
- 49 Decaup E, Rochotte J, Pyronnet S, Bousquet C, Jean C. Focal Adhesion Kinase: A promising therapeutic target in pancreatic adenocarcinoma. *Clinics and research in hepatology and gastroenterology* 2017; 41: 246–248.
- 50 Weniger M, Honselmann K, Liss A. The extracellular matrix and pancreatic cancer: a complex relationship. *Cancers* 2018; 10: 316.
- 51 McCleary-Wheeler AL, McWilliams R, Fernandez-Zapico ME. Aberrant signaling pathways in pancreatic cancer: a two compartment view. *Mol Carcinog* 2012; 51: 25–39.

- 52 Le Large TYS, Bijlsma MF, Kazemier G, van Laarhoven HWM, Giovannetti E, Jimenez CR. Key biological processes driving metastatic spread of pancreatic cancer as identified by multi-omics studies. In: *Seminars in cancer biology*. Elsevier, 2017: 153–169.
- 53 Birnbaum DJ, Finetti P, Lopresti A, *et al.* A 25-gene classifier predicts overall survival in resectable pancreatic cancer. *BMC medicine* 2017; 15: 170.
- 54 Isik Z, Baldow C, Cannistraci CV, Schroeder M. Drug target prioritization by perturbed gene expression and network information. *Sci Rep* 2015; 5: 17417–17417.
- 55 Xie C-M, Wei W, Sun Y. Role of SKP1-CUL1-F-box-protein (SCF) E3 ubiquitin ligases in skin cancer. *J Genet Genomics* 2013; 40: 97–106.
- 56 Vallejo A, Perurena N, Guruceaga E, *et al.* An integrative approach unveils FOSL1 as an oncogene vulnerability in KRAS-driven lung and pancreatic cancer. *Nature communications* 2017; 8: 1-14.
- 57 Awasthi N, Mikels-Vigdal AJ, Stefanutti E, *et al.* Therapeutic efficacy of anti-MMP9 antibody in combination with nab-paclitaxel-based chemotherapy in pre-clinical models of pancreatic cancer. *Journal of cellular and molecular medicine*; 2019; 23: 3878-3887.
- 58 Rajurkar M, De Jesus-Monge WE, Driscoll DR, *et al.* The activity of Gli transcription factors is essential for Kras-induced pancreatic tumorigenesis. *Proceedings of the National Academy of Sciences* 2012; 109: E1038–E1047.
- 59 Danilova L, Ho WJ, Zhu Q, *et al.* Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival. *Cancer immunology research* 2019; 7: 886–895.
- 60 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 2014; 15: 550.
- 61 Dimitrakopoulos C, Hindupur SK, Häfliger L, *et al.* Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics* 2018; 34: 2441–2448.
- 62 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003; 13: 2498–2504.

Appendix

Deeper insights into long-term survival heterogeneity of Pancreatic Ductal Adenocarcinoma (PDAC) patients using integrative individual- and group-level transcriptome network analyses

Archana Bhardwaj^a, Claire Josse^{b,c}, Daniel Van Daele^d, Christophe Poulet^{b,h}, Marcela Chavez^e, Ingrid Struman^f, Kristel Van Steen^a

^aGIGA-R Centre, BIO3 – Medical Genomics, University of Liège, Liège, Belgium

^bLaboratory of Human Genetics, GIGA Research, University Hospital (CHU), Liège, Belgium

^cMedical Oncology Department, CHU Liège, Liège, Belgium

^dDepartment of Gastro-enterology, University Hospital (CHU), Liège, Belgium

^eDepartment of Medicine, Division of Hematology, University Hospital (CHU), Liège, Belgium

^fGIGA-R Centre, Laboratory of Molecular Angiogenesis, University of Liège, Liège, Belgium

^hLaboratory of Rheumatology, GIGA-R, University Hospital (CHU Liege), Liège, Belgium

Contact Detail: Corresponding author

University of Liège

GIGA - Quartier Hôpital

Avenue de l'Hôpital, 11

4000 Liège, Belgium

Appendix Method:

Data and Methods

1. Data and data preparation

1.1 Patient selection, ethical statement, and criteria to maximize the definition of STS and LTS

All aspects of the study comply with the Declaration of Helsinki. PDAC patients from Liege University Hospital were recruited on the basis of an opt-out methodology, from 2007 to 2014, giving to N=96 pancreas tissue. Tissues were obtained from the University of liege Biobank, Belgium. The study was approved by the local institutional ethical board (“Comité d’éthique hospital-faculties universités de Liège (707)”) under the file number B707201627153. Among them, 36 had OS < 12 months or > 36 months, as selected survival criteria. We performed RNA extraction from those 36 samples and processed for RNA quality check.

1.2 RNA extraction, library preparation, sequencing

Tumor areas were determined by a certified pathologist and were manually macro-dissected from the FFPE tissues. RNA was extracted using an All Prep DNA/RNA/miRNA Universal kit (Qiagen, Belgium) according to the manufacturer’s protocol. The next manipulations described in the paragraphs were performed by the GIGA-Genomics facility. The RNA quality (N=36) was assessed using a BioAnalyzer (Agilent, Belgium), and the proportion of RNA with a length higher than 200 bases (DV200) was measured. Only 19 out of 36 met a suitable RNA quality, allowed for sequencing. TruSeq® RNA Access Library Prep Kit (Cat. No. RS-301-2001 and RS-301-2002) (Illumina, The Netherlands) was used to prepare libraries, and next-generation sequencing was performed on a NextSeq500 apparatus (Illumina, The Netherlands), in paired-end 2 x 75bp high output mode.

We performed a series of transcriptome computational analyses to better understand patient heterogeneity between LT and ST survivors. After quality control and adaptor trimming with Trimmomatic¹, sequence data were mapped to the Genome Reference Consortium GRCh38 assembly using STAR v2.5.2². Read counts for known genes were generated using the function HTSeq-count v0.6.1p³ and data were normalized in DESeq2 v1.20.0⁴ as shown in Figure 1.

1.3 Clinical features of Patients

Various clinical and pathological parameters of patients (N=19) were included in the analysis. In particular, we collected the following pathological clinical data: age, sex, tumor size, number of lymph nodes evaluated, tumor grade, sugey magins invaded by tumor cells, time between surgery and chemotherapy (in days), time between surgery and relapse (in months), disease-free survival (DFS), vascular resection, time in hospital after surgery (in days), re-hospitalization 6 months after surgery, vascular contact, artery contact, and chemotherapy as shown in Figure 2A.

1.4 Group based DEGs analysis: Differential Gene analysis and functional follow-up

We used DEseq2⁴ for the identification of differentially expressed genes (DEG), with the thresholds log₂ fold change ≥ 2 and ≤ -2 , to indicate up-regulation and down-regulation, respectively (Figure 1:B). Significance was assessed at an unadjusted p-value < 0.05 in LT vs. ST group comparison⁵. We used the ClusterProfiler v3.8.1⁶ package to predict various GO processes enriched in differentially expressed genes (DEGs). To identify the protein domain in DEGs, we used batch CD-Search⁷. Identified DEG were analyzed for detection of survival genes, with a log-rank test in a Kaplan–Meier survival model⁸ (Figure 1:A-B). For each gene, patients were classified into two groups, the high-expression group (H) and the low-expression group (L), using the expression median of the gene as a cutoff using the survminer⁹ (v. 0.4.6) R package.

2. Methods

The entire workflow is described in Figure 1. Specific details regarding group-level and individual-specific analyses are given in Sections 2.1 and 2.2, respectively.

2.1 Group-level survival heterogeneity: WGCNA for gene module prediction and assessment of clinical relevance

The minimum sample size to run weighted gene co-expression network analysis (WGCNA) is at least 15. Therefore, WGCNA v1.63¹⁰ was used on pooled ST and LT PDAC survival patients to generate a transcriptional network from the normalized expression data. The weighted coefficient β was selected based on scale-free topology criteria. The adjacency coefficient α was computed using the power to measure the correlation strength between two genes. The adjacency matrix was created based on α , which was subsequently transformed into a topological overlap matrix (TOM). The distance measure $\text{dissTOM} = 1 - \text{TOM}$, served as input to perform average linkage hierarchical clustering (with DynamicTreeCut¹¹), giving rise to gene co-expression modules. Gene modules were shown as branches of the resulting pruned tree. It was followed by the calculation of module eigengenes (MEs), which are defined as the 1st linear principal component of each co-expression module. The hierarchical clustering of MEs was performed to study associations between modules. Approximate non-parametric association tests were used to investigate the association between MEs and PDAC clinical traits. In effect, we used two methods to identify modules related to clinical progression traits. First, within-module gene significance was identified for every module and all available clinical traits. Average gene significance for a module was defined as “module significance”, following recommendations of¹². Second, rank-based correlation (r) was performed among each ME with the multiple clinic pathological characteristics available in this study (adjusted p-value for 0.05 MEs). We used parametric (Pearson correlation coefficient) and non-parametric (Spearman rank) tests for each continuous and categorical data, respectively. In order to assess the functional relevance of clinically associated modules, we used ClueGO¹³, a Cytoscape plug-in in order to visualize the non-redundant biological terms for genes in a functionally comparative network from multiple clusters. Non-redundancy was assessed via two-sided hypergeometric testing for enrichment/depletion (Bonferroni adjusted p-value < 0.05). Cytoscape 5.0¹⁴ was used for visualizing gene interaction networks (Figure 1:C).

2.2 Individual-specific survival heterogeneity: Quantification of heterogeneity between individual transcriptome profiles, with functional and clinical relevance

We used principles of the PEPPER¹⁵ method to construct personalized gene expression perturbation profiles for each of N=19 PDAC subjects. PEPPER requires a target class of individuals and a reference class (Figure 1:D). In this study, we took LT PDAC survivors as target group and considered ST survivors as reference (i.e., the most abundant group in real-life). The approach captures the extent to which gene *i* is perturbed in subject *j* via a Z-score. This Z-score indicates how many standard deviations the individual's gene expression is away from the mean value of the reference group. As a threshold, we used $|z|=2$. Positive z-scores > 2 would indicate up-regulation, negative z-scores < -2 would indicate down-regulation. Given the small sample sizes to work with in this study, we reshuffled the ST/LT group labels¹⁶ 500 times, and repeated the experiment. Note that under the null hypothesis, none of the individual LT survivor profiles would be markedly different from average ST survivor profiles and thus LT/ST survivor status would be exchangeable on the basis of individual transcriptome profiles. We used shinyCircos¹⁷ R package to develop circos plot for identified PEEPs. Functional follow-up analyses included checking for enrichment of KEGG pathways, and verifying motif enrichment via ToppGene Suite¹⁸ (multiple testing adjusted p-value < 0.05). Also, patient-specific one-way hierarchical clustering and dendrograms were developed on the basis of the frequency of perturbed genes in identified domains and pathways. Both dendrograms were subsequently compared using the R version 1.12.0 of the dendextend¹⁹ R package". For deeper insights, two-way clustering via the superbiclust package in R (RcmdrPlugin.BiclustGUI²⁰) version 1.1) was used, enabling the application of the Bimax²¹ algorithm to jointly cluster LT survivors and either one of three levels of biological information, namely gene, pathway and motif levels. For each analysis, a higher level (super) biclustering was obtained by constructing a hierarchical tree depicting Jaccard similarity between Bimax clusters.

In the aforementioned PEEPs analyses (PEEP: an individual perturbation expression profile against a reference), no notion of gene-connectivity was used. However, gene connectivity via reference networks can further highlight interesting gene clusters linked to LT survivors. Here, we considered physical interaction data as available from ConsensusPathDB²², and obtained 373,101 links between N=19,117 genes. Starting with genes in pathways that already have been implied in PDAC via²³, and supplementing these genes with searches in the DisGeNet database²⁴ (search term = "Pancreatic Diseases"), resulted in 53 seed genes (Figure 1:I; appendix pp 19-20). We then used DADA's module detection algorithm¹³ to augment the initial list of 53 seed genes and to identify PDAC disease modules. The top 1 percent highest ranked genes were considered to form a disease module. Significantly perturbed genes (in LT survivor PEEPs) were mapped on the identified disease module. This allowed putting LT survival individual specific genes in the context of gene connectivity and gene neighborhoods. All DADA top 1 percent genes were checked for their retrieval in previous analyses (Figure 1:J-L). As an alternative approach to exploit gene interaction network structure, we adapted NetICS²⁵, an approach initially intended to prioritize cancer genes on a directed functional interaction network. It uses an individual-specific list of genes via bidirectional network diffusion of two layers of information (Figure 1:E). As first layer we took the individual-specific significant genes as highlighted in the LT PDAC survival PEEPs analyses before (instead of mutant genes per sample in the original NetICS implementation). As second layer we took groups-specific DEGs (Section 2.1). Individual-specific gene ranks (for LT survivors) were aggregated via NetICS methodology into an overall ranked list of genes, with restart

probability of 0.4. The top 1% percent ranked genes were retained. Similar to follow-up of DADA top-ranked genes, we checked for the frequency of NetICS derived top-ranked genes that were also retrieved in former analyses (Figure 1: F-H).

3. Results

3.1 Potential candidate genes

Gene *XKR5*, showed a significant increase in survival in long-term (LT) patients with lower expression compared to short-term survivors (appendix pp 27). We observed a similar pattern for the genes *GATD3B*, *CYP27C1*, and *miR-765* (appendix pp 28-30). These results highlight the potential of the identified genes in further understanding molecular underpinnings of PDAC survival.

3.2 Functional analysis of clinically relevant modules

Five modules were identified as clinical relevant module via WGCNA analysis (appendix 3). In module M34, we found three significant Reactome pathways distributed into three groups: the effects of PIP2 hydrolysis (4 genes), the deactivation of the beta-catenin transactivating complex (3 genes) and the VEGFA-VEGFR2 pathway (4 genes) (data not shown). In M30, we found two significant pathways: apoptotic cleavage of cell adhesion proteins (4 genes) and o-linked glycosylation (11 genes) (data not shown). Individuals (LT1, LT3, LT4, LT5, LT6) did not show significant enrichment in any KEGG/Reactome pathway.

3.3 Biclustering of functional profiles

To analyse the effect of perturbed genes in PEEPs (gene is significantly perturbed or not) in LT survivor's heterogeneity, two-way clustering (biclustering) highlighted 64 gene clusters (appendix pp 33-34). The largest cluster (cluster 15) consisted of 363 genes. Deeper hierarchical clustering of these identified clusters (appendix pp 4) grouped cluster 7, 36, 37, 42, 47,48, 50, 53, 55 into single super cluster (appendix pp 34) with overrepresentation of cancer specific pathways such as mTOR pathways and NOD-like signaling pathways as highlighted via orange box in appendix (pp 36).

Two-way hierarchical clustering, based on the presence/absence of enriched pathways across LT survivors (LT2, LT7, LT8, LT9) revealed three clusters (appendix pp 38). First two clusters (C1 and C2) showed enriched pathways in two LTS only. C1 consisted of 14 pathways was collectively enriched in L7 and L9, and highlighted a strong association with cancer-related pathways. C2 showed enrichment of 13 pathways between L9 and L2 such as Proteoglycans in cancer and EPH-Ephrin signaling. Smallest cluster C3 consisted of 8 pathways across three LTS survivors i.e LT2, LT7, LT9. Deeper hierarchical clustering groups C2 and C3 into single supercluster based on similar pathways profiles. Likewise, two-way hierarchical clustering (biclustering) based on motif enrichment profiles

(present or absent) across all LT survivors resulted in four clusters (appendix pp 39). The first cluster (C1), represented by LT7 and LT9, was enriched with six domains. The second cluster (C2), active in LT2 and LT7, was enriched with 7 domains. The third clusters (C3) involved enrichment of 7 domains shared two among LT survivors (appendix pp 17-18). The fourth cluster (C4) was largely shared by three LT survivors (LT2, LT7, and LT9). This cluster involved 5 domains: IPR013032, PS01186, IPR000742, PS00022, and IPR009030. Deeper hierarchical clustering groups C1 and C4 into single supercluster based on similar protein domains profiles. Deeper hierarchical clustering groups C1 and C4 into single supercluster based on similar protein domains profiles. More in-depth analysis revealed a common gene set between cluster 24 obtained from gene-level clustering and cluster 1 (C1) derived from pathway-level biclustering (appendix pp 36, 39). Similarly, cluster 25 derived from gene level analysis showed overlap with cluster 2 (C2) derived from pathway-level biclustering.

References

- 1 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 2014; **30**: 2114–20.
- 2 Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15–21.
- 3 Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**: 166–9.
- 4 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 2014; **15**: 550.
- 5 Zhang W, Liu S, Zhan H, Yan Z, Zhang G. Transcriptome sequencing identifies key pathways and genes involved in gastric adenocarcinoma. *Molecular medicine reports* 2018; **18**: 3673–82.
- 6 Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* 2012; **16**: 284–7.
- 7 Marchler-Bauer A, Bo Y, Han L, *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research* 2016; **45**: D200–3.
- 8 Therneau T. A Package for Survival Analysis in S. version 2.38. 2015.
- 9 Kassambara A, Kosinski M. survminer: Drawing survival curves using “ggplot2,” 2018. URL <https://CRAN.R-project.org/package=survminer> R package version 04 2018; **3**.
- 10 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008; **9**: 559.
- 11 Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008; **24**: 719–20.
- 12 Liu R, Cheng Y, Yu J, Lv Q-L, Zhou H-H. Identification and validation of gene module associated with lung cancer through coexpression network analysis. *Gene* 2015; **563**: 56–62.
- 13 Erten S, Bebek G, Ewing RM, Koyutürk M. DADA: Degree-Aware Algorithms for Network-Based Disease Gene Prioritization. *BioData Min* 2011; **4**: 19–19.
- 14 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003; **13**: 2498–504.
- 15 Menche J, Guney E, Sharma A, *et al.* Integrating personalized gene expression profiles into predictive disease-associated gene pools. *NPJ systems biology and applications* 2017; **3**: 10.

- 16 Anderson MJ. Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian journal of fisheries and aquatic sciences* 2001; **58**: 626–39.
- 17 Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 2018; **34**: 1229–31.
- 18 Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* 2009; **37**: W305–11.
- 19 Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 2015; **31**: 3718–20.
- 20 De Troyer E, Otava M. RcmdrPlugin.BiclustGUI: 'Rcmdr' Plug-in GUI for Biclustering. 2016.
- 21 Prelić A, Bleuler S, Zimmermann P, *et al.* A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics* 2006; **22**: 1122–9.
- 22 Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic acids research* 2012; **41**: D793–800.
- 23 McCleary-Wheeler AL, McWilliams R, Fernandez-Zapico ME. Aberrant signaling pathways in pancreatic cancer: a two compartment view. *Mol Carcinog* 2012; **51**: 25–39.
- 24 Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 2019. <https://doi.org/10.1093/nar/gkz1021>.
- 25 Dimitrakopoulos C, Hindupur SK, Häfliger L, *et al.* Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics* 2018; **34**: 2441–8.
- 26 Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corrplot'. *Statistician* 2017; **56**: 316–24.

Table S1. Complete list of all sample (patients) and their clinical features as considered in the present study.

Patient ID	Group*	Age at the surgery (years)	Sex ^a	Tumor size (cm) ^b (measured after surgery)	T	N	Number of nodes analyzed
S1	ST	74	F	3	4	1	25
S2	ST	71	M	3.5	3	0	12
S3	ST	72	F	3.2	3	0	7
S4	ST	73	M	4	3	1	7
S5	ST	54	F	3.2	4	1	12
S6	ST	80	M	4	3	1	8
S7	ST	80	M	4	3	0	16
S8	ST	63	F	3	4	1	14
S9	ST	66	M	5	3	1B	16
S10	ST	66	M	5	4	1	4
L1	LT	78	F	2.5	3	1	13
L2	LT	58	M	3.2	3	0	33
L3	LT	71	F	2	3	0	10
L4	LT	83	M	4	3	0	14
L5	LT	67	M	2.2	3	1	25
L6	LT	62	M	5	3	1	7
L7	LT	58	M	2	1	1	8
L8	LT	77	M	4.5	4	1b	4
L9	LT	55	M	0.6	1	0	0

Complete list of all sample (patients) and their clinical features analyzed in the present study (Table S1 continue)

Patient ID	ratio node (invaded/analyzed)	ratio node % (invaded/analysed)	Sugey Magins invaded by tumor cells**	Time between surgery and chemotherapy (in days)	Time between surgery and relapse (months) (DFS) ^c	Time between surgery and death (months) OS ^d	Time between relapse and death (months)	Death	Vascular resection
S1	2/25	0.08	1	82	NA	3.09	NA	1	1
S2	0/12	0.00	0	NA	5.00	6.31	1.31	1	1
S3	0/7	0.00	0	71	NA	7.23	NA	1	0
S4	5/7	0.71	0	54	5.42	8.32	2.89	1	0
S5	7/12	0.58	0	63	2.30	8.42	6.11	1	0
S6	2/8	0.25	1	74	4.50	8.61	4.11	1	0
S7	0/16	0.00	0	NA	6.87	9.66	2.79	1	0
S8	4/14	0.28	1	43	4.87	9.66	4.80	1	1
S9	10/16	0.62	0	53	7.10	9.76	2.66	1	0
S10	0/4	0.00	0	41	9.37	10.22	0.85	1	0
L1	4/13	0.31	0	55	25.25	NA	NA	1	0
L2	0/33	0.00	2	NA	17.16	40.70	23.54	1	1
L3	0/10	0.00	1	NA	NA	NA	NA	1	0
L4	0/14	0.00	0	NA	NA	48.72	NA	1	0
L5	2/25	0.08	0	16	NA	NA	NA	0	0
L6	1/7	0.14	0	71	37.71	60.52	22.81	1	0
L7	3/8	0.37	0	NA	NA	NA	NA	0	0
L8	3/4	0.75	0	66	NA	NA	NA	0	0
L9	0	NA	0	NA	NA	NA	NA	0	0

Complete list of all sample (patients) and their clinical features analyzed in the present study (Table S1 continue)

Patient ID	Time in hospital after surgery (in days)	Re-hospitalisation 6 months after surgery**	Vascular contact (if tumor contact a vein (often portal vein)**	artery contact (if tumor contact an artery) **	Tumor size by imagery (mm) ^e	Time between imagery and surgery (days)	Chemotherapy (CTH)**	CTH > 3 months	Number of CTH cures	Dose CTH received in % of the maximal theoretical value
S1	42	0	1	0	42	7	1	0	1	100
S2	14	0	0	0	76	121	0	NA	NA	NA
S3	16	0	NA	NA	NA	NA	1	0	NA	NA
S4	22	3	1	0	22	6	1	0	1	100
S5	16	0	1	0	29	6	1	0	4	100
S6	24	0	1	1	34	5	1	0	6	80
S7	21	1	1	1	27	3	0	NA	NA	NA
S8	18	0	1	0	NA	31	1	0	7	100
S9	24	0	1	1	19	21	1	0	3	100
S10	17	0	1	0	NA	92	1	0	2	NA
L1	22	0	1	0	15	9	1	0	4	100
L2	14	1	NA	NA	NA	NA	1	1	NA	NA
L3	11	0	1	0	28	22	0	NA	NA	NA
L4	22	0	1	0	NA	27	0	NA	NA	NA
L5	17	0	1	0	28	33	1	0	NA	NA
L6	17	0	0	0	NA	63	1	0	4	75
L7	13	0	1	1	21	28	NA	NA	NA	NA
L8	28	0	1	0	41	28	1	0	18	100
L9	21	0	1	0	21	28	0	NA	NA	NA

* ST=short-term PDAC survivor (ST; death \geq 3 months et $<$ 12 months) ; LT= long-term PDAC survivor (LT; alive or death \geq 36 months) ; ^aF= Female ; M=Male ; ^bcm = centimeters ; ^cDFS=Disease free survival ; ^dOS= Overall survival ; ^emm= millimeters ; NA =Not available; **(1=yes;0=no)

Table S2. Pathways enriched in LT survivors based on the Individual specific PEEPs annotated using ToppGene Suite¹⁷ (section 2.2). For all identified significant pathways, corresponding Bonferroni corrected p-values are shown. Identified pathways are grouped into various categories based on functional annotation. highlighted based on their involvement either in three LT survivors; two LT survivors or one LT survivor.

Pathway Name	LT2	LT7	LT9	Category
Apoptosis			1.20E-02	Cell regulation
Caspase cascade in apoptosis			7.85E-03	
Cell Cycle			8.98E-03	
Cell Cycle. Mitotic			5.48E-03	
Cell junction organization			9.15E-03	
Cell surface interactions at the vascular wall			4.99E-06	
Cell-Cell communication			3.04E-06	
ECM-receptor interaction	3.53E-02	3.01E-02	3.22E-05	
Endocytosis			3.05E-02	
Mitotic Anaphase			1.06E-02	
Mitotic Metaphase and Anaphase			1.19E-02	
Mitotic Prometaphase			1.10E-03	
Programmed Cell Death			6.95E-03	
Regulation of actin dynamics for phagocytic cup formation			2.88E-03	
Semaphorin interactions			9.86E-05	
Axon guidance	2.25E-07	3.01E-02	3.92E-19	Developmental Biology
EPH-ephrin mediated repulsion of cells	3.52E-02			
L1CAM interactions			2.37E-02	
Osteoclast differentiation			1.91E-05	
Other semaphorin interactions			5.48E-03	
Scavenging by Class A Receptors		3.24E-03		
AGE-RAGE signaling pathway in diabetic complications			1.53E-06	Disease
Amoebiasis			1.42E-04	
Amyotrophic lateral sclerosis (ALS)			4.32E-02	
Angiogenesis			1.81E-04	

Arrhythmogenic right ventricular cardiomyopathy (ARVC)			1.66E-02	
Chagas disease (American trypanosomiasis)			1.96E-04	
HTLV-I infection			1.34E-04	
Hypertrophic cardiomyopathy (HCM)			2.01E-04	
Inflammatory bowel disease (IBD)			9.65E-04	
Influenza A			2.68E-02	
Legionellosis			1.43E-03	
Leishmaniasis			3.34E-04	
Malaria			1.29E-03	
Measles			3.05E-02	
Pathogenic Escherichia coli infection			2.10E-04	
PDGFR-beta signaling pathway			2.08E-02	
Pertussis			1.46E-07	
Primary immunodeficiency			2.30E-02	
Prolonged ERK activation events			1.99E-02	
RHO GTPases Activate Formins			2.01E-05	
Salmonella infection			1.50E-03	
SHC1 events in EGFR signaling			2.30E-02	
SRP-dependent cotranslational protein targeting to membrane			7.42E-03	
Toll-Like Receptors Cascades			1.75E-07	
TRIF-mediated TLR3/TLR4 signaling			4.72E-03	
Tuberculosis			4.00E-05	
Assembly of collagen fibrils and other multimeric structures		3.70E-04	5.76E-12	Extracellular matrix organization
Collagen biosynthesis and modifying enzymes		3.01E-04	6.80E-06	
Collagen chain trimerization		3.01E-04	3.74E-07	
Collagen degradation			1.49E-02	
Collagen formation		1.41E-03	4.36E-11	
Degradation of the extracellular matrix			3.53E-04	
ECM proteoglycans			1.03E-04	
Elastic fibre formation	8.69E-03			

Ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans	3.18E-03	3.01E-02	1.47E-03	
Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins	3.23E-03		4.51E-05	
Extracellular matrix organization	2.25E-07	1.40E-09	2.38E-19	
Genes encoding collagen proteins		2.66E-04	9.34E-08	
Genes encoding structural components of basement membranes	5.52E-03		3.51E-04	
Integrin cell surface interactions		1.16E-02	3.63E-06	
Laminin interactions			1.91E-05	
Molecules associated with elastic fibres	3.57E-02			
Non-integrin membrane-ECM interactions			8.72E-05	
Activated TLR4 signalling			1.83E-04	
Adaptive Immune System			2.25E-04	
Adhesion and Diapedesis of Granulocytes			2.41E-03	
Adhesion and Diapedesis of Lymphocytes			2.93E-03	
amb2 Integrin signaling			4.70E-03	
Bacterial invasion of epithelial cells			3.46E-02	
Cells and Molecules involved in local acute inflammatory response			1.68E-02	
Chemokine signaling pathway			2.94E-06	
Classical Complement Pathway			2.41E-03	
Complement and coagulation cascades			8.27E-05	
Complement Pathway			4.69E-02	
CTL mediated immune response against target cells			5.04E-03	
CXCR4-mediated signaling events			1.67E-04	
Cytokine Signaling in Immune system			1.04E-16	
Cytokine-cytokine receptor interaction			1.43E-03	
DAP12 interactions			3.75E-06	
DAP12 signaling			2.82E-05	
Dilated cardiomyopathy			3.04E-06	
Direct p53 effectors			1.06E-03	
Diseases of signal transduction			1.94E-02	
Fc gamma R-mediated phagocytosis			1.12E-03	

Immune

FCERI mediated MAPK activation			2.13E-03
Fcgamma receptor (FCGR dependent phagocytosis			2.32E-03
Hematopoietic cell lineage			1.06E-02
IL-2 Receptor Beta Chain in T cell Activation			3.05E-02
IL12 signaling mediated by STAT4			9.13E-04
IL12-mediated signaling events			3.13E-04
IL4-mediated signaling events			9.64E-03
Inactivation of Gsk3 by AKT causes accumulation of b-catenin in Alveolar Macrophages			2.61E-02
Inflammation mediated by chemokine and cytokine signaling pathway			8.51E-04
Innate Immune System			1.11E-14
Interleukin-1 signaling			8.72E-05
Interleukin-10 signaling			6.00E-03
Interleukin-2 signaling			9.95E-03
Interleukin-3, 5 and GM-CSF signaling			2.13E-03
Interleukin-4 and 13 signaling			3.30E-05
Leukocyte transendothelial migration			2.13E-07
Monocyte and its Surface Molecules			2.93E-03
MyD88 dependent cascade initiated on endosome			5.48E-03
MyD88-independent TLR3/TLR4 cascade			4.72E-03
MyD88:Mal cascade initiated on plasma membrane			2.32E-03
Natural killer cell mediated cytotoxicity			3.44E-03
NCAM signaling for neurite out-growth			1.48E-03
Neutrophil degranulation			1.26E-05
NOD-like receptor signaling pathway			7.77E-03
PLK1 signaling events			1.43E-03
Regulation of actin cytoskeleton			4.33E-10
Signaling by EGFR			1.41E-04
Syndecan-4-mediated signaling events			6.29E-06
T cell activation			3.74E-05
T cell receptor signaling pathway			2.84E-03
T Cell Signal Transduction			9.42E-03

T Cytotoxic Cell Surface Molecules			2.41E-03	
T Helper Cell Surface Molecules			1.64E-04	
TCR signaling in naive CD4+ T cells			4.82E-05	
Th1 and Th2 cell differentiation			1.38E-03	
Th17 cell differentiation			9.43E-06	
TNF signaling pathway			2.13E-03	
Toll Like Receptor 2 (TLR2 Cascade			2.32E-03	
Toll Like Receptor 3 (TLR3 Cascade			4.72E-03	
Toll Like Receptor 4 (TLR4 Cascade			8.90E-06	
Toll Like Receptor 7/8 (TLR7/8 Cascade			5.48E-03	
Toll Like Receptor 9 (TLR9 Cascade			1.06E-02	
Toll Like Receptor TLR1:TLR2 Cascade			2.32E-03	
Toll Like Receptor TLR6:TLR2 Cascade			2.32E-03	
Toxoplasmosis			3.52E-05	
TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation			1.27E-02	
Aurora B signaling			4.13E-02	other
Binding and Uptake of Ligands by Scavenger Receptors			2.63E-02	
Chylomicron-mediated lipid transport		2.93E-02		
GPVI-mediated activation cascade			1.10E-03	
Hemostasis			2.62E-16	Signal transduction
ARMS-mediated activation			3.29E-02	
Cell adhesion molecules (CAMs			2.42E-04	
Downstream signal transduction			2.26E-04	
EGF receptor signaling pathway			1.02E-02	
EPH-Ephrin signaling	3.53E-02		6.89E-04	
Fc epsilon receptor (FCER1 signaling			2.20E-04	
Focal adhesion	3.57E-02	3.24E+03	1.95E-11	
Frs2-mediated activation			1.67E-02	
Gastrin-CREB signaling pathway via PKC and MAPK			4.35E-03	
GRB2 events in EGFR signaling			2.30E-02	
Hippo signaling pathway			1.98E-02	

HIV-1 Nef: negative effector of Fas and TNF			1.25E-02
Insulin receptor signalling cascade			2.23E-02
integrin signaling			1.10E-03
Integrin signaling pathway	3.57E-02	1.16E-02	6.85E-04
Interleukin signaling pathway			4.72E-03
MAPK family signaling cascades			1.11E-03
MAPK1/MAPK3 signaling			8.25E-03
MET activates PTK2 signaling			9.28E-03
MET promotes cell motility			1.85E-02
mTOR signaling pathway			2.41E-02
NGF signaling via TRKA from the plasma membrane			2.16E-03
Pathways in cancer	8.93E-01	1.95E-02	1.40E-13
Phagosome			1.45E-05
Phospholipase D signaling pathway			7.42E-03
PI3K-Akt signaling pathway	9.26E-02	3.76E+03	4.13E-07
Platelet activation			2.21E-04
Platelet activation. signaling and aggregation			8.93E-06
Proteoglycans in cancer	9.26E-02		7.31E-08
RAF/MAP kinase cascade			2.30E-02
Rap1 signaling pathway			7.70E-04
Ras Pathway			2.68E-02
Resolution of Sister Chromatid Cohesion			9.15E-03
RET signaling			4.35E-03
RHO GTPase Effectors			2.43E-06
Separation of Sister Chromatids			1.97E-02
Shigellosis			4.19E-05
Signaling by Interleukins			3.53E-15
Signaling by Leptin			9.73E-03
Signaling by MET			1.30E-04
Signaling by PDGF			2.80E-06
Signaling by Rho GTPases			1.75E-07

Signaling by SCF-KIT			5.53E-05	
Signaling by VEGF			1.56E-03	
Signaling events mediated by focal adhesion kinase			1.10E-03	
Signaling by NGF			1.26E-05	
Signaling to ERKs			2.37E-02	
Signaling to p38 via RIT and RIN			1.51E-02	
Small cell lung cancer			2.42E-04	
SOS-mediated signaling			2.30E-02	
Staphylococcus aureus infection			5.76E-06	
Syndecan-2-mediated signaling events			3.50E-02	
Tight junction			1.34E-04	
Validated transcriptional targets of AP1 family members Fra1 and Fra2			9.02E-03	
VEGFA-VEGFR2 Pathway			7.65E-03	
VEGFR2 mediated cell proliferation			3.31E-02	

Table S3. Protein domains enriched in LT survivors based on the Individual specific PEEPs annotated using ToppGene Suite¹⁷ (section 2.2). For all identified significant protein domains, corresponding Bonferroni corrected p-values are shown. Identified protein domains are ordered based on similar annotation profile as given below. (Only 3 LT survivors showed enriched IPR domains).

Domain ID	LT2	LT7	LT9	Description
PS00010	0.000368			ASX_HYDROXYL
PF12662	0.000247			cEGF
IPR026823	0.000247			cEGF
SM00038		2.5E-05		COLFI
PF01410		2.5E-05		COLFI
PF01391		0.001874	4.19E-06	Collagen
IPR008160		0.001874	4.19E-06	Collagen
IPR011029			0.00017	DEATH-like_dom
SM00181	2.1E-06	7.05E-05	1.5E-06	EGF
PF00008			9.42E-06	EGF
PS00022	8.89E-05	5.77E-05	4.19E-06	EGF
PS01186	9.45E-06	4.21E-06	7.02E-06	EGF
PS50026	0.000011		0.000209	EGF
PF07645	5.51E-05	0.000249		EGF
SM00179	2.47E-05	0.001874		EGF
PS01187	5.01E-05			EGF
IPR018097	4.01E-05			EGF_Ca-bd_CS
IPR001881	2.9E-05	0.002076		EGF-like_Ca-bd_dom
IPR013032	2.16E-06	1.02E-05	4.79E-08	EGF-like_CS
IPR000742	9.45E-06	4.12E-05	7.07E-04	EGF-like_dom
IPR000152	0.000121			EGF-type_Asp/Asn_hydroxyl_site
PD002078		2.5E-05		Fib_collagen_C
IPR000885		2.5E-05		Fib_collagen_C
SM00060			0.000799	FN3
PS50853			0.004262	FN3
IPR003961			0.000334	FN3_dom

IPR009030	9.37E-05	0.005339	1.59E-05	Growth_fac_rcpt_
PS00856	0.02688			GUANYLATE_KINASE_1
PS50052	0.02688			GUANYLATE_KINASE_2
IPR013783			4.15E-05	Ig-like_fold
IPR032695			0.0002	Integrin_dom
IPR011009	0.04356			Kinase-like_dom
IPR001791			8.36E-05	Laminin_G
2.60.40.10			4.19E-06	NA*
1.10.533.10			0.000225	NA
PS51461		2.5E-05		NC1_FIB
PF07714	0.04519			Pkinase_Tyr
IPR000719	0.00585			Prot_kinase_dom
PS00107	0.001612		0.00355	PROTEIN_KINASE_ATP
IPR017441	0.01304			Protein_kinase_ATP_BS
PS50011	0.006695			PROTEIN_KINASE_DOM
PS00109	0.01171			PROTEIN_KINASE_TYR
PS50001			0.005654	SH2
IPR008266	0.01171			Tyr_kinase_AS
IPR020635	0.004333			Tyr_kinase_cat_dom
SM00219	0.004333			TyrKc
PS01208	0.04908			VWFC_1

*NA = Not available

Table S4. Seed genes for the identification of a PDAC disease module with DADA¹³. Genes are ordered following their genetic location; genes annotation shown as well.

Gene	Chromosome	Description
HDAC1	1	histone deacetylase 1
GLI2	2	GLI family zinc finger 2
TGFA	2	transforming growth factor alpha
IHH	2	Indian hedgehog signaling molecule
AREG	4	amphiregulin
HHIP	4	hedgehog interacting protein
EGF	4	epidermal growth factor
MAML3	4	mastermind like transcriptional coactivator 3
APC	5	APC regulator of WNT signaling pathway
HDAC2	6	histone deacetylase 2
CCND3	6	cyclin D3
EGFR	7	epidermal growth factor receptor
RAC1	7	Rac family small GTPase 1
SHH	7	sonic hedgehog signaling molecule
MYC	8	MYC proto-oncogene@ bHLH transcription factor
DKK4	8	dickkopf WNT signaling pathway inhibitor 4
DKK1	10	dickkopf WNT signaling pathway inhibitor 1
FRAT1	10	FRAT regulator of WNT signaling pathway 1
CCND1	11	cyclin D1
ATM	11	ATM serine/threonine kinase
MMP7	11	matrix metalloproteinase 7
FOSL1	11	FOS like 1@ AP-1 transcription factor subunit
INHBE	12	inhibin subunit beta E
GLI1	12	GLI family zinc finger 1
WIF1	12	WNT inhibitory factor 1
IFNG	12	interferon gamma
ZIC2	13	Zic family member 2

JAG2	14	jagged canonical Notch ligand 2
SMAD3	15	SMAD family member 3
DLL4	15	delta like canonical Notch ligand 4
MAPK3	16	mitogen-activated protein kinase 3
SMAD4	18	SMAD family member 4
SMAD2	18	SMAD family member 2
SMAD7	18	SMAD family member 7
TGFB1	19	transforming growth factor beta 1
BMP2	20	bone morphogenetic protein 2

Table S5. Overlap between individual perturbation expression profile (PEEP) for LT PDAC survivors with top 1% gene list from NetICS and DADA. NetICS= Network-based Integration of Multi-omics Data²⁵; DADA= Degree-Aware Algorithms for Network Based Disease Gene Prioritization¹³.

Identified in NetICS		
Patient ID	Nr* of PEEP genes in top 1% NetICS list	Top 5 genes
L1	1	TNNI3
L2	143	TNNI3.IRS4.ACTN2.MAPK14.PLCG1.
L3	5	FOSL1.MYL2.FCGR3A.USF2.NOSTRIN
L4	38	ACTN2.PRKCA.ITGA8.AP2A1.NRAS
L5	20	TJP2.ITGA10.SDC4.TYROBP.SPI1
L6	26	NPY.MPDZ.CNTN6.ITGB4.TNNC1
L7	119	TNNI3.HLA-DQA1.APOA4.IRS4.NPY.
L8	18	GNAI3. ITGA10. KLRC3. SPI1.TNNC1
L9	336	ACTB. ACTG1.NFKB1.ITGB1.CTNNB1
Identified in DADA		
Patient ID	Nr of PEEP genes in top 1% DADA list	Top 5 genes
L2	8	GLI2.RAC1.CUL1.ESR1.FBLN1.UBC.WNT11
L3	1	FOSL1
L4	3	TGFA.EGF.CDK2
L5	2	UBC.CDON
L7	9	ESR1.GLI2.WNT11.GLI3.HDAC1.CDK2.UBC.CDON
L9	12	GLI2.GLI3.UBC.RAC1.ACTB.AREG.DKK1.GRB2.MMP9.TGFB1.EGF.FOSL1

*Nr=Number

Table S6. Overlap between top 1% genes from NetICS /DADA and clinical modules, as defined in section 2.1 and 2.2. WGCNA = Weighted correlation network analysis¹⁰; NetICS= Network-based Integration of Multi-omics Data²⁵; DADA= Degree-Aware Algorithms for Disease Gene Prioritization¹³ .

Category	Module No.	Gene Symbol	Chromosome	Gene Description
Common among NetICS AND WGCNA	M30	CDH1	16	cadherin 1
	M30	CREB3L3	19	cAMP responsive element binding protein 3 like 3
	M30	NGFR	17	nerve growth factor receptor
	M30	CDON	11	cell adhesion associated oncogene regulated
	M30	KIF3B	20	kinesin family member 3B
	M30	MET	7	MET proto-oncogen receptor tyrosine kinase
	M30	PIP5K1B	9	phosphatidylinositol-4-phosphate 5-kinase type 1 beta
	M30	PRKAB1	12	protein kinase AMP-activated non-catalytic subunit beta 1
	M30	SPTBN1	2	spectrin beta non-erythrocytic 1
	M30	TJP2	9	tight junction protein 2
	M30	LDLR	19	low density lipoprotein receptor
	M30	TUBGCP2	10	tubulin gamma complex associated protein 2
	M30	RAN	12	RAN member RAS oncogene family
	M30	EPHB2	1	EPH receptor B2
	M30	ANK2	4	ankyrin 2
	M30	UBC	12	ubiquitin C
	M30	DYNLL2	17	dynein light chain LC8-type 2
	M9	TAC1	7	tachykinin precursor 1
	M9	TEAD3	6	TEA domain transcription factor 3
	M9	CLDN11	3	claudin 11
	M9	FLT4	5	fms related tyrosine kinase 4
	M9	COL11A1	1	collagen type XI alpha 1 chain
	M9	HDAC7	12	histone deacetylase 7
	M9	FGFR2	10	fibroblast growth factor receptor 2

M9	CLTCL1	22	clathrin heavy chain like 1
M9	ACTN1	14	actinin alpha 1
M9	GLI2	2	GLI family zinc finger 2
M9	NOTCH3	19	notch receptor 3
M9	RASAL2	1	RAS protein activator like 2
M9	ACTB	7	actin beta
M9	IL4R	16	interleukin 4 receptor
M9	COL16A1	1	collagen type XVI alpha 1 chain
M9	MMP2	16	matrix metalloproteinase 2
M9	NOS1	12	nitric oxide synthase 1
M9	LAMB1	7	laminin subunit beta 1
M9	MYL9	20	myosin light chain 9
M9	FZD3	8	frizzled class receptor 3
M9	TGFB1	19	transforming growth factor beta 1
M9	USF2	19	upstream transcription factor 2 c-fos interacting
M9	SCN1B	19	sodium voltage-gated channel beta subunit 1
M9	GLI3	7	GLI family zinc finger 3
M9	ACTA2	10	actin alpha 2 smooth muscle
M9	COL1A1	17	collagen type I alpha 1 chain
M9	MAP2K6	17	mitogen-activated protein kinase kinase 6
M9	COL12A1	6	collagen type XII alpha 1 chain
M9	LAMA4	6	laminin subunit alpha 4
M9	RASGRF2	5	Ras protein specific guanine nucleotide releasing factor 2
M9	PDGFRB	5	platelet derived growth factor receptor beta
M9	COL7A1	3	collagen type VII alpha 1 chain
M9	GNAI2	3	G protein subunit alpha i2
M9	CBLB	3	Cbl proto-oncogene B
M9	GNB4	3	G protein subunit beta 4
M9	FN1	2	fibronectin 1
M9	TGFB3	14	transforming growth factor beta 3
M9	NPY	7	neuropeptide Y

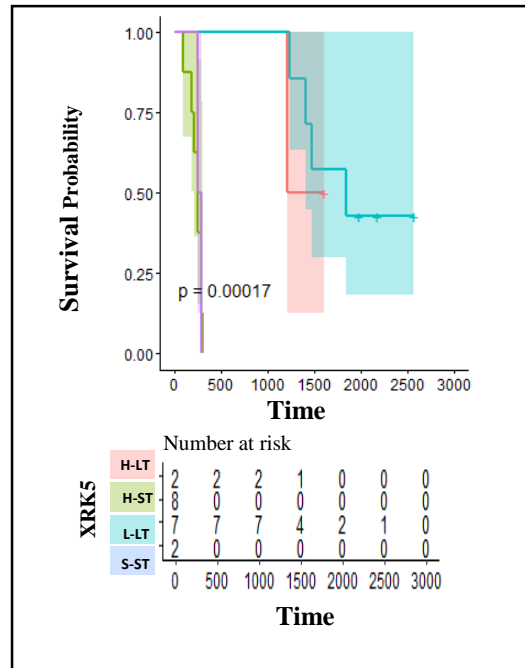
M9	CALD1	7	caldesmon 1
M9	LRP1	12	LDL receptor related protein 1
M9	COL10A1	6	collagen type X alpha 1 chain
M9	PLCG1	20	phospholipase C gamma 1
M9	C3	19	complement C3
M9	NUP214	9	nucleoporin 214
M9	SMO	7	smoothened frizzled class receptor
M9	TNNI3	19	troponin I3 cardiac type
M9	COL5A1	9	collagen type V alpha 1 chain
M9	RAF1	3	Raf-1 proto-oncogene serine/threonine kinase
M9	CHRM3	1	cholinergic receptor muscarinic 3
M9	IRS4	X	insulin receptor substrate 4
M9	COL4A2	13	collagen type IV alpha 2 chain
M9	LAMC1	1	laminin subunit gamma 1
M9	GABBR2	9	gamma-aminobutyric acid type B receptor subunit 2
M9	TUBB2A	6	tubulin beta 2A class IIa
M9	ITGA11	15	integrin subunit alpha 11
M9	LEF1	4	lymphoid enhancer binding factor 1
M9	COL2A1	12	collagen type II alpha 1 chain
M9	ACVRL1	12	activin A receptor like type 1
M9	CDH13	16	cadherin 13
M9	GNAL	18	G protein subunit alpha L
M9	COL6A1	21	collagen type VI alpha 1 chain
M9	COL6A2	21	collagen type VI alpha 2 chain
M9	HSPG2	1	heparan sulfate proteoglycan 2
M9	TPM3	1	tropomyosin 3
M9	COL8A1	3	collagen type VIII alpha 1 chain
M9	SERPINH1	11	serpin family H member 1
M9	MMP3	11	matrix metalloproteinase 3
M9	ADAM17	2	ADAM metalloproteinase domain 17
M9	GJA1	6	gap junction protein alpha 1

	M9	ASAP1	8	ArfGAP with SH3 domain ankyrin repeat and PH domain 1
	M9	MMP14	14	matrix metalloproteinase 14
	M9	FZD1	7	frizzled class receptor 1
	M9	SHC1	1	SHC adaptor protein 1
	M9	ITGA5	12	integrin subunit alpha 5
	M9	KCNJ3	2	potassium inwardly rectifying channel subfamily J member 3
	M9	COL6A3	2	collagen type VI alpha 3 chain
	M9	SPTA1	1	spectrin alpha erythrocytic 1
	M9	COL1A2	7	collagen type I alpha 2 chain
	M9	TUBA1A	12	tubulin alpha 1a
	M9	MLST8	16	MTOR associated protein LST8 homolog
	M9	RPSA	3	ribosomal protein SA
	M9	CTNNB1	3	catenin beta 1
	M9	COL3A1	2	collagen type III alpha 1 chain
	M9	AXIN2	17	axin 2
	M9	LIMS1	2	LIM zinc finger domain containing 1
	M9	PCDH7	4	protocadherin 7
	M9	CD14	5	CD14 molecule
	M9	FGG	4	fibrinogen gamma chain
	M9	COL8A2	1	collagen type VIII alpha 2 chain
	M9	LAMB2	3	laminin subunit beta 2
	M9	CEBPB	20	CCAAT enhancer binding protein beta
	M9	TUBB6	18	tubulin beta 6 class V
	M9	HLA-DQB1	CHR_HSCHR6_MHC_QBL_CTG 1	major histocompatibility complex class II DQ beta 1
	M9	FZD2	17	frizzled class receptor 2
	M9	KCNH7	2	potassium voltage-gated channel subfamily H member 7
	M9	P4HB	17	prolyl 4-hydroxylase subunit beta
	M9	MYL6B	12	myosin light chain 6B
	M9	HLA-DQA1	CHR_HSCHR6_MHC_QBL_CTG 1	major histocompatibility complex class II DQ alpha 1
	M9	FLNA	X	filamin A

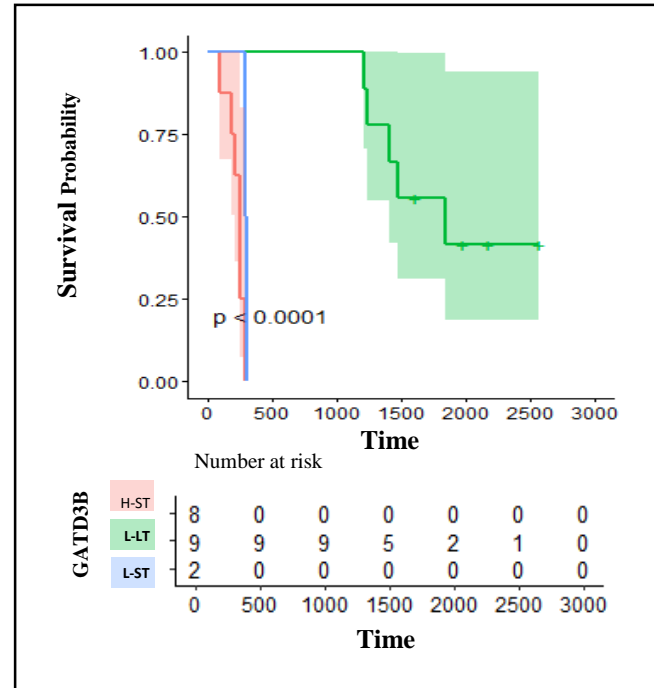
	M9	MAP3K5	6	mitogen-activated protein kinase kinase kinase 5
	M9	ITGBL1	13	integrin subunit beta like 1
	M9	COL5A2	2	collagen type V alpha 2 chain
	M9	NOTCH4	CHR_HSCHR6_MHC_QBL_CTG 1	notch receptor 4
	M9	ARPC4	3	actin related protein 2/3 complex subunit 4
	M9	UBE2V1	20	ubiquitin conjugating enzyme E2 V1
	M7	YWHAB	20	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein beta
	M7	NRAS	1	NRAS proto-oncogene GTPase
	M7	HSP90AA1	14	heat shock protein 90 alpha family class A member 1
	M7	HMGB1	13	high mobility group box 1
	M15	BSG	19	basigin (Ok blood group)
Common among DADA and WGCNA	M30	CDON	11	cell adhesion associated oncogene regulated
	M30	WNT11	11	Wnt family member 11
	M30	DKK4	8	dickkopf WNT signaling pathway inhibitor 4
	M30	BMP2	20	bone morphogenetic protein 2
	M9	DKK3	11	dickkopf WNT signaling pathway inhibitor 3
	M9	GLI2	2	GLI family zinc finger 2
	M9	SMAD7	18	SMAD family member 7
	M9	TGFB1	19	transforming growth factor beta 1
	M9	GLI3	7	GLI family zinc finger 3
	M9	GLI1	12	GLI family zinc finger 1
	M9	DLL4	15	delta like canonical Notch ligand 4
	M9	MYC	8	MYC proto-oncogene bHLH transcription factor
	M9	MMP7	11	matrix metalloproteinase 7
	M9	SMAD4	18	SMAD family member 4
	M9	WNT7A	3	Wnt family member 7A
	M34	WNT5A	3	Wnt family member 5A

Figure S1. Kaplan-Meier plots for significant survival genes differentially expressed between long-term (LT: alive or death ≥ 36 months) and short-term (ST: death ≥ 3 months and < 12 months) survival. A) Gene XRK5, p-value = 0.00017; B) Gene GATD3B, p-value < 0.0001 ; C) Gene CYP27C1, p-value < 0.0001) and D) Gene miR-765, p-value < 0.0001 . Here, H and L represents the high and low expression groups identified (appendix 3).

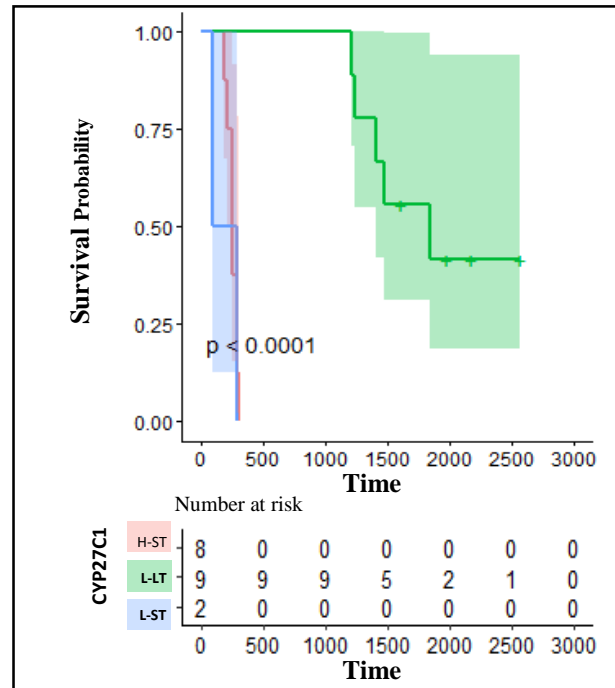
A)



B)



c)



D)

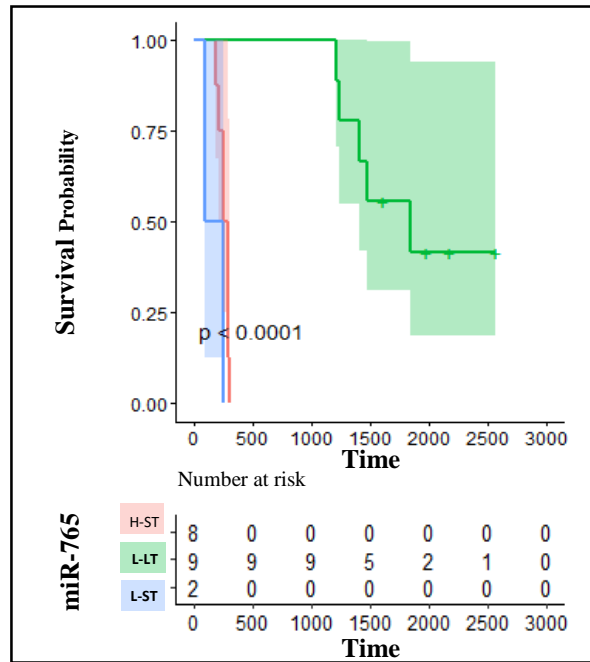


Figure S2. RNA-Seq supplemented with RT-PCR experiments: A) Correlation between RNA-Seq and RT-PCR expression values; B) Violin plot of TSPAN8 (gene ID: ENSG00000127324) by ST/LT PDAC survival; C) Same as B) but for REG4 (gene id: ENSG00000134193).

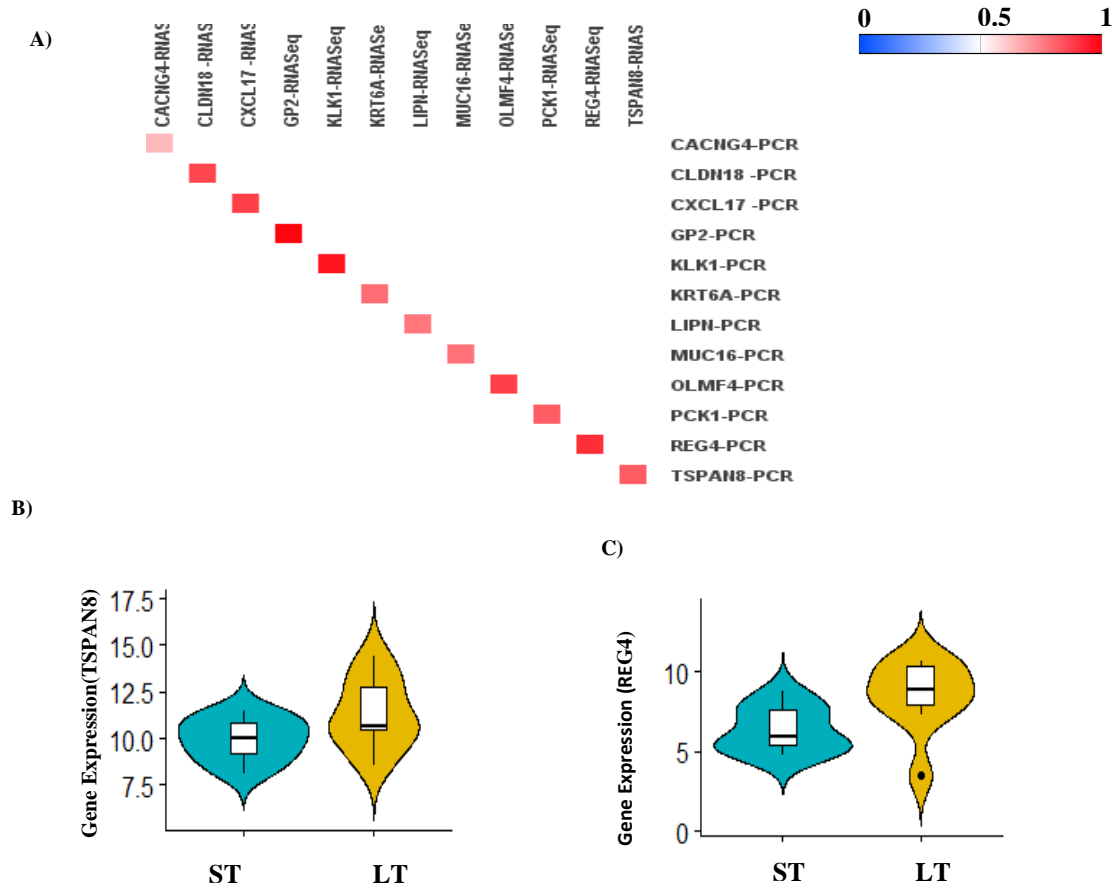


Figure S3. Clinical profile of WGCNA modules. Correlation patterns for WGCNA derived modules with significant association to clinical measurements (multiple testing adjusted p-value <0.05) are shown via corrplot²⁶ (section 2.1). The more extreme the association (+1/-1) the deeper the color (dark blue/dark red). The sizes of the circles are proportional to the correlation coefficients. WGCNA = Co-Expression Network Analysis. Adjusted p values are not indicated in plot. Complete detail of significant modules is given in section: Survival group heterogeneity

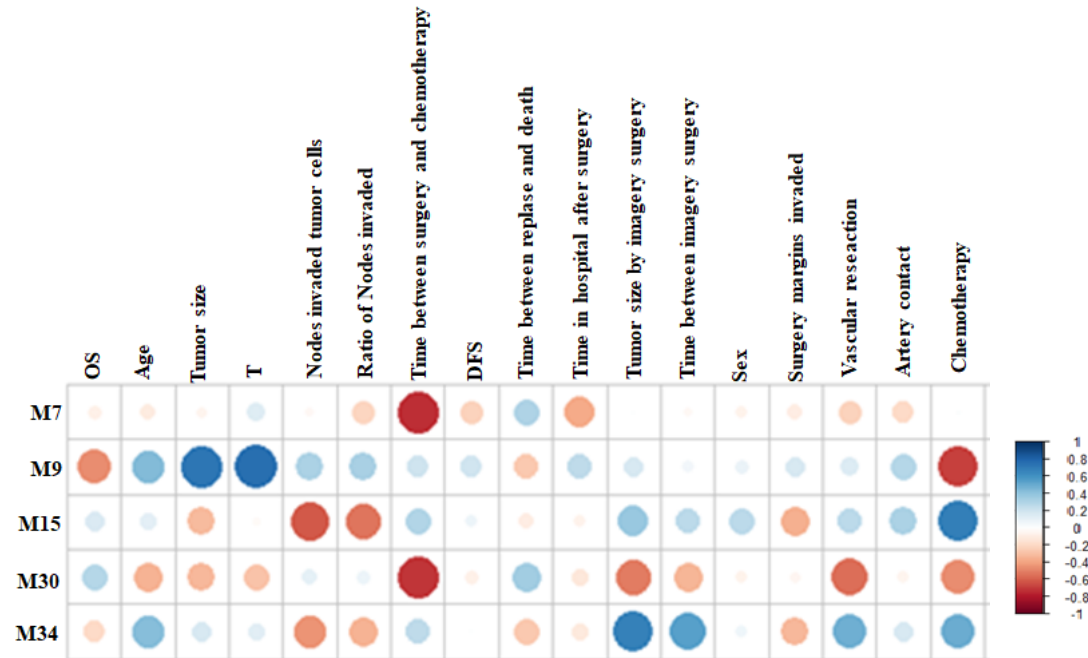
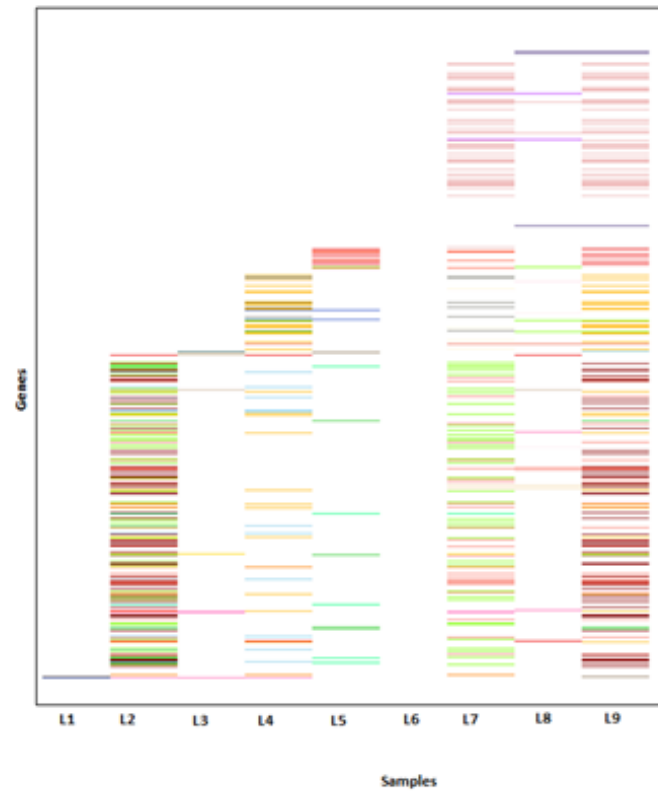


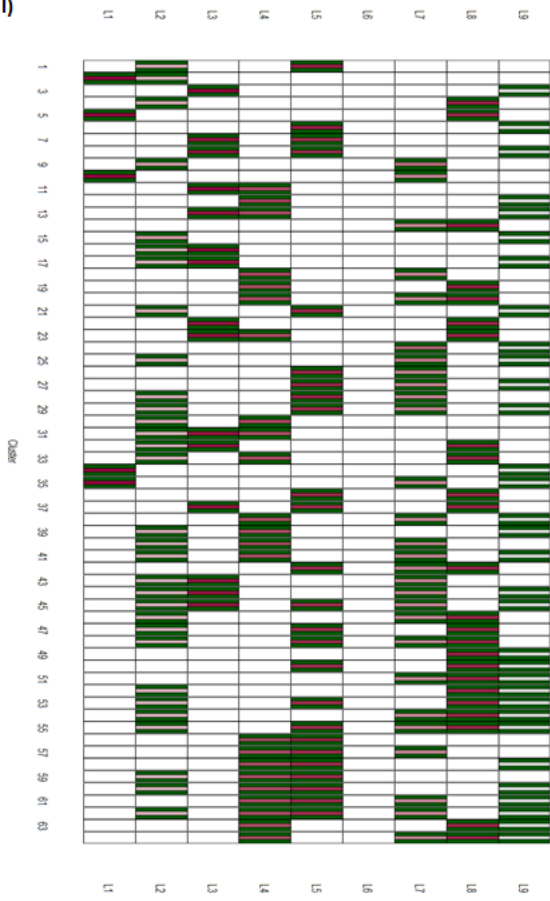
Figure S4. Bicluster analysis of perturbed genes in LT PDAC survivors: Input matrix is logical with genes (not) perturbed in an individual indicated by 1 (0). A) Sixty-four biclusters (BC) obtained from such a logical matrix (section 2.2) indicated as heatmap. Each color in heatmap represent cluster from 1 to 64. B) Advanced interpretation of identified biclusters via three different approaches. (I) **bicluster membership graph** based on BC cluster x LTS. (II) **Heatmap based on Jaccard similarity index** computed for the identified 64 biclusters ranging from 0 (no concordance) to 1 (perfect concordance). (III) **Hierarchical tree** constructed for the identified biclusters (appendix pp 4). C) Functional annotation of 64 biclusters with clusterProfiler⁶.

A)

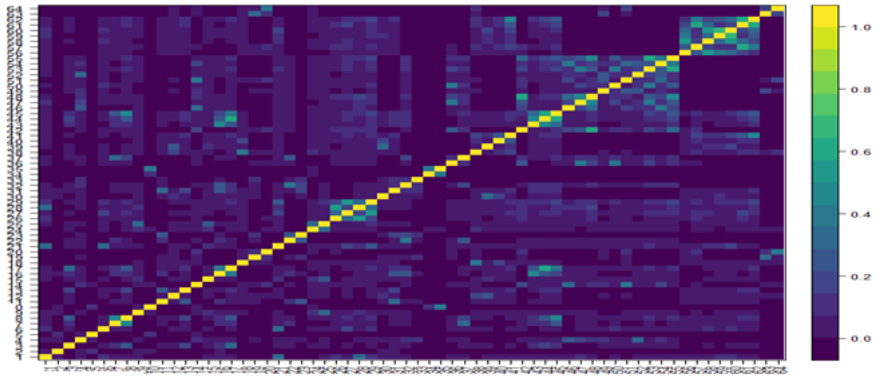


B)

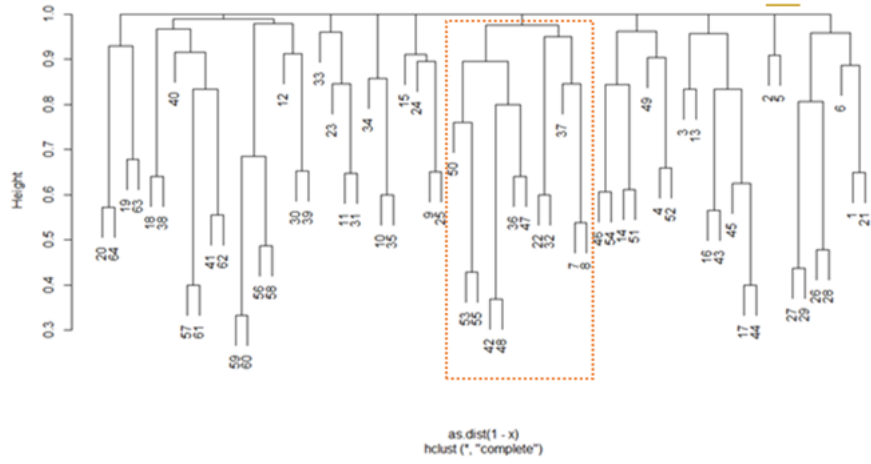
i)



ii)



iii)



C)

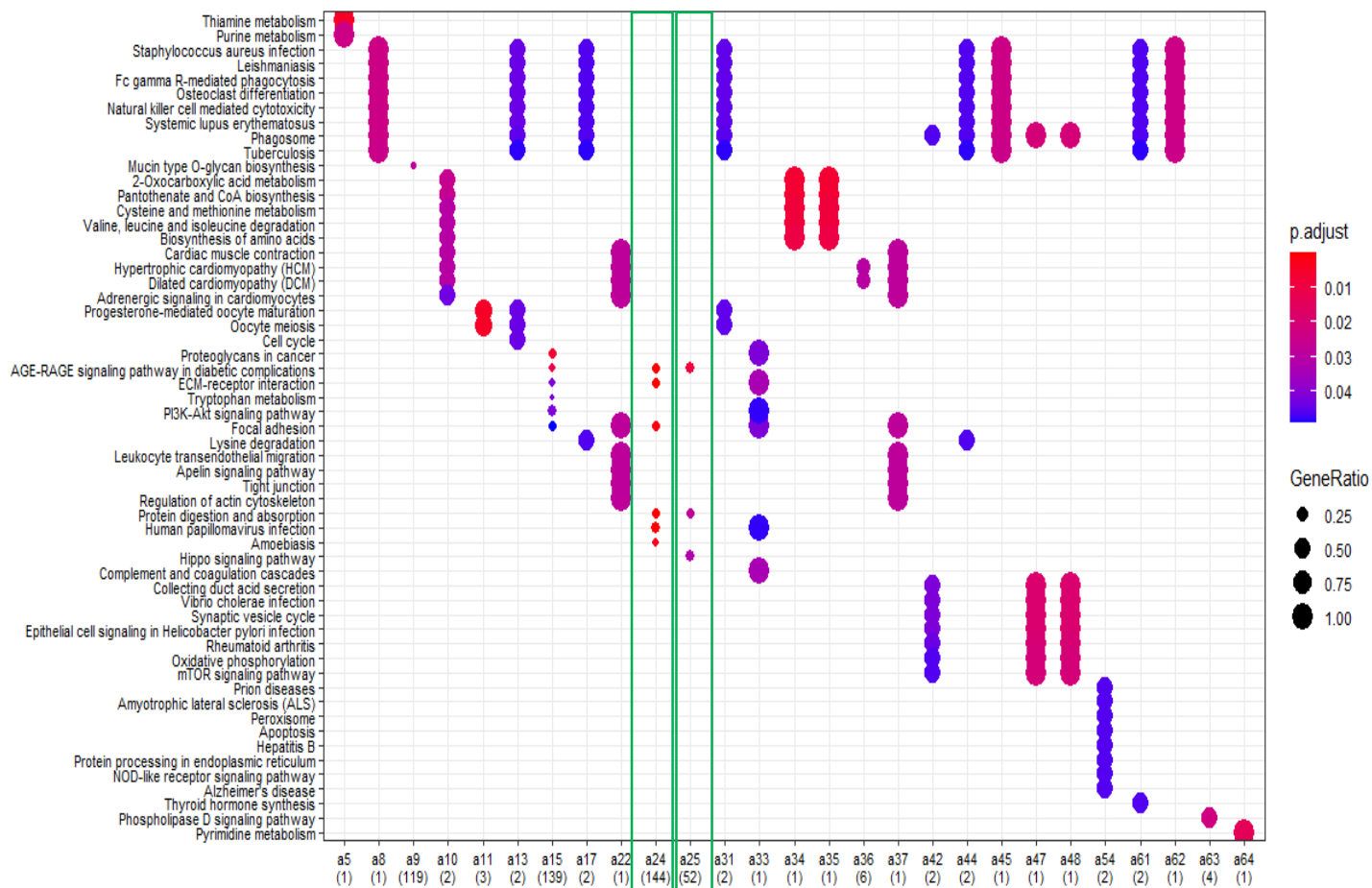


Figure S5. Bicluster analysis of enriched pathways in LT PDAC survivors: Input matrix is logical with enriched pathways (not) in an individual indicated by 1 (0). A) Three biclusters (BC) obtained from such a logical matrix (section 2.2). B) Advanced interpretation of identified biclusters via three different approaches. (I) **biclusters as lines of parallel coordinate graph** which indicates cluster specific detailed information. (II) **bicluster membership graph** based on BC cluster x LTS. (III) **Heatmap based on Jaccard similarity index** computed for the identified three biclusters ranging from 0 (no concordance) to 1 (perfect concordance). **Hierarchical tree** constructed for the identified biclusters (appendix pp 4).

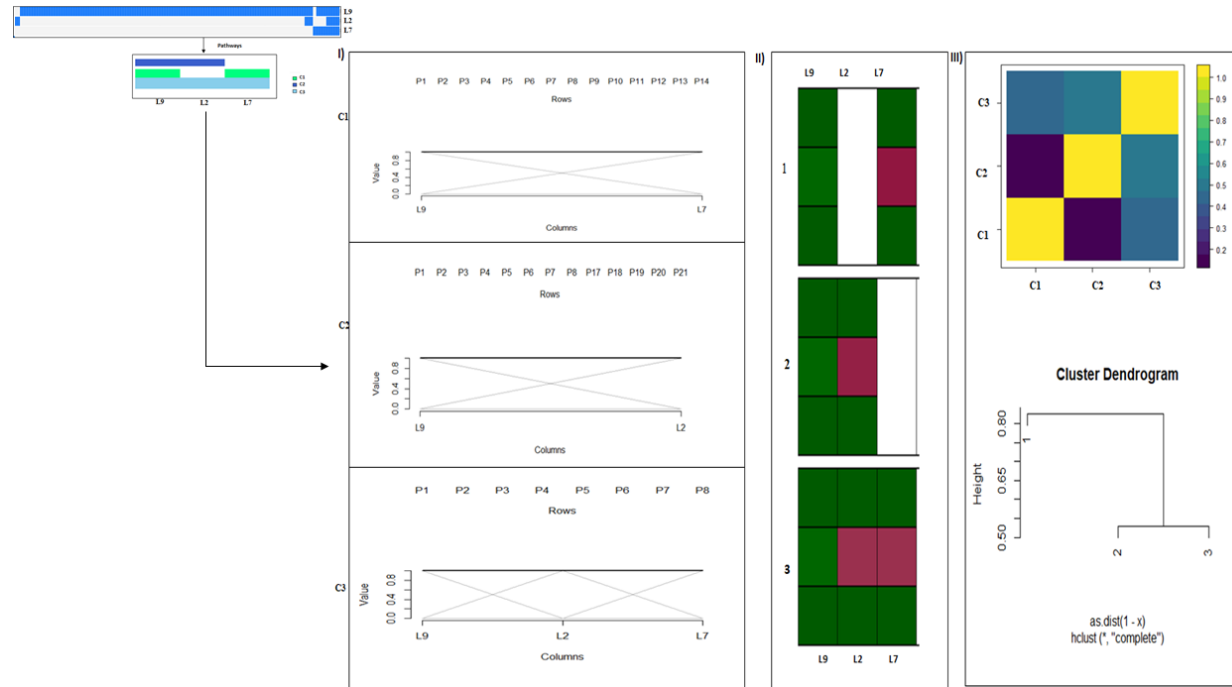


Figure S6. Bicluster analysis of enriched protein domains in LT PDAC survivors: Input matrix is logical with enriched protein domains (not) in an individual indicated by 1 (0). A) Four biclusters (BC) obtained from such a logical matrix (section 2.2). B) Advanced interpretation of identified biclusters via three different approaches. (I) **biclusters as lines of parallel coordinate graph** which indicates cluster specific detailed information. (II) **bicluster membership graph** based on BC cluster x LTS. (III) **Heatmap based on Jaccard similarity index** computed for the identified 4 biclusters ranging from 0 (no concordance) to 1 (perfect concordance). **Hierarchical tree** constructed for the identified biclusters (appendix pp 4).

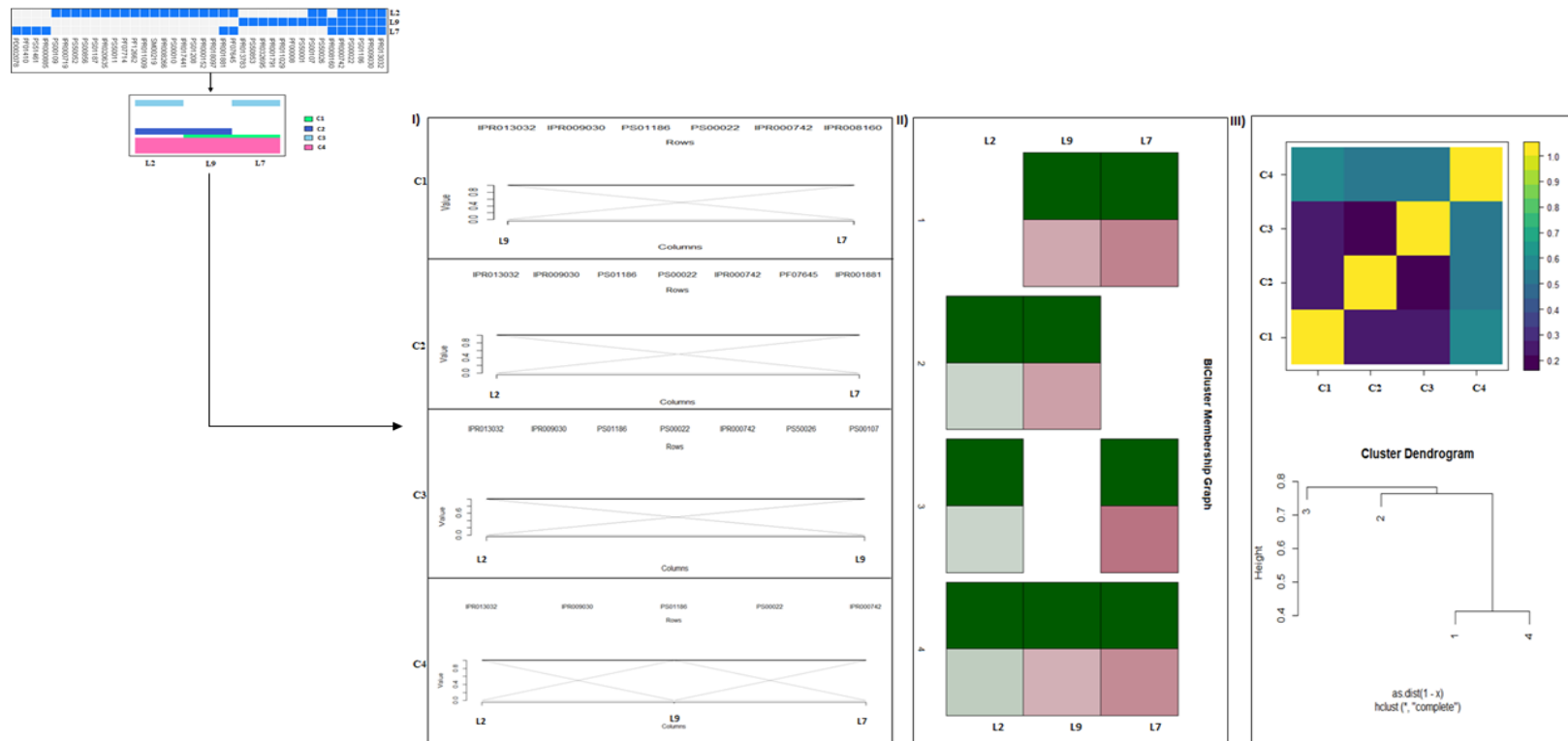


Figure S7. Superposition of PEEP induced perturbed genes to PDAC disease module derived via DADA. Individual-specific perturbed genes as identified by individual profiling with PEPPER¹⁵ (section 2.2) are highlighted with the same color per individual. LT2, LT3, LT4, LT5, LT7, and LT9 perturbed genes are indicated in red, green, light green, yellow, orange, and dark green, respectively. LT1, LT6 and LT8 specific perturbed genes showed no overlap in PDAC disease module derived via DADA (not shown in figure).

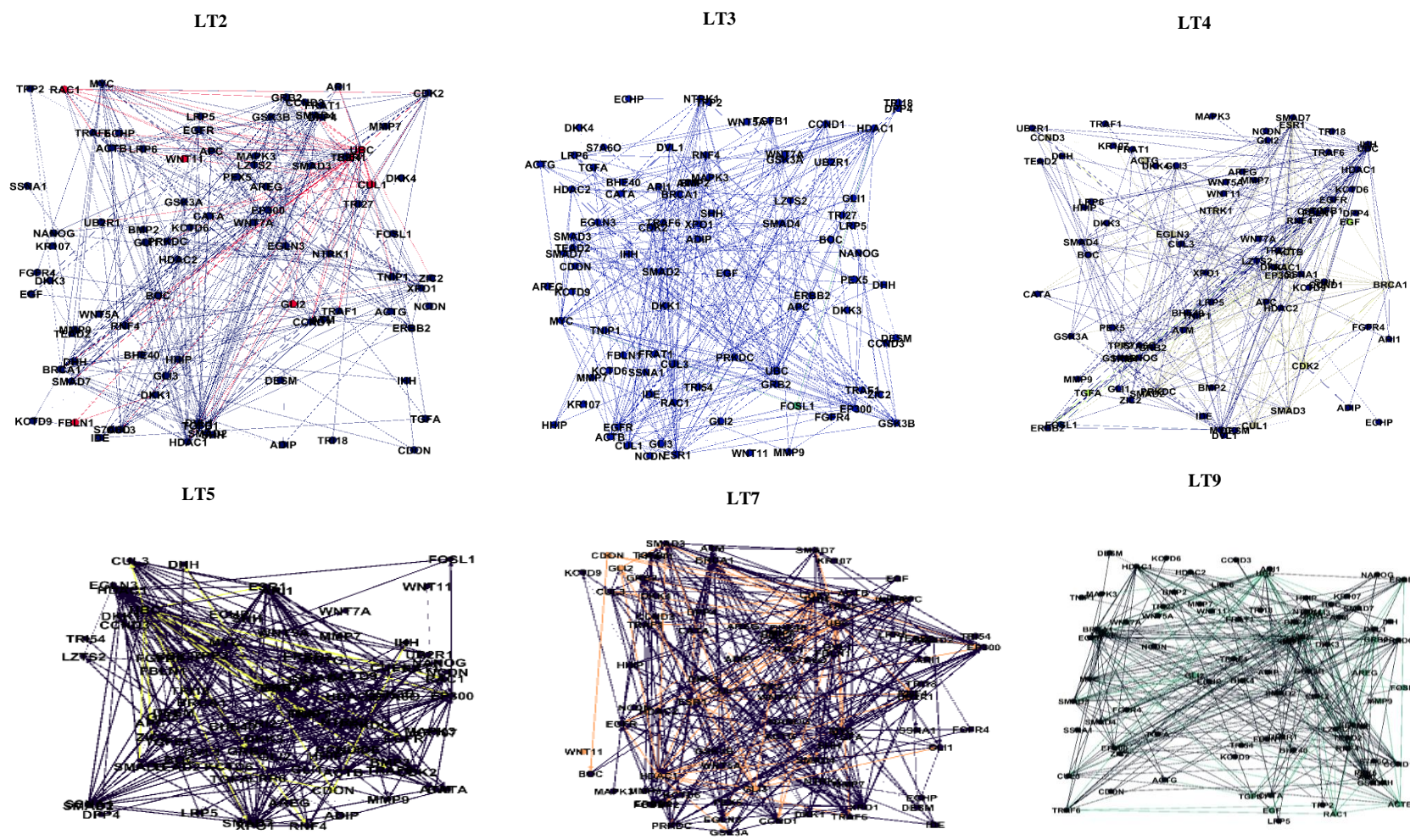
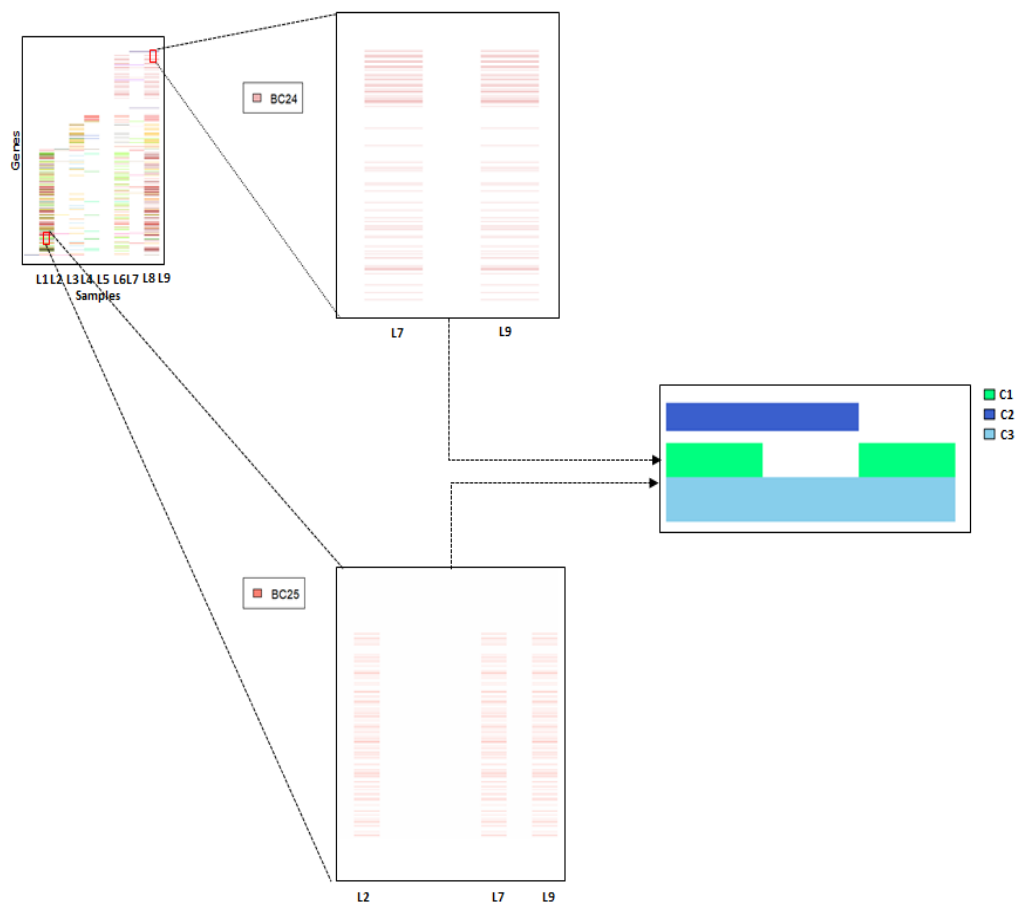


Figure S8. Aggregation of multi-level analyses. Genes present in groups BC24 and BC25 (Figure S4) were shown to be involved in pathway sets C2 and C3 (Figure S5) as indicated by the arrow (appendix 35).



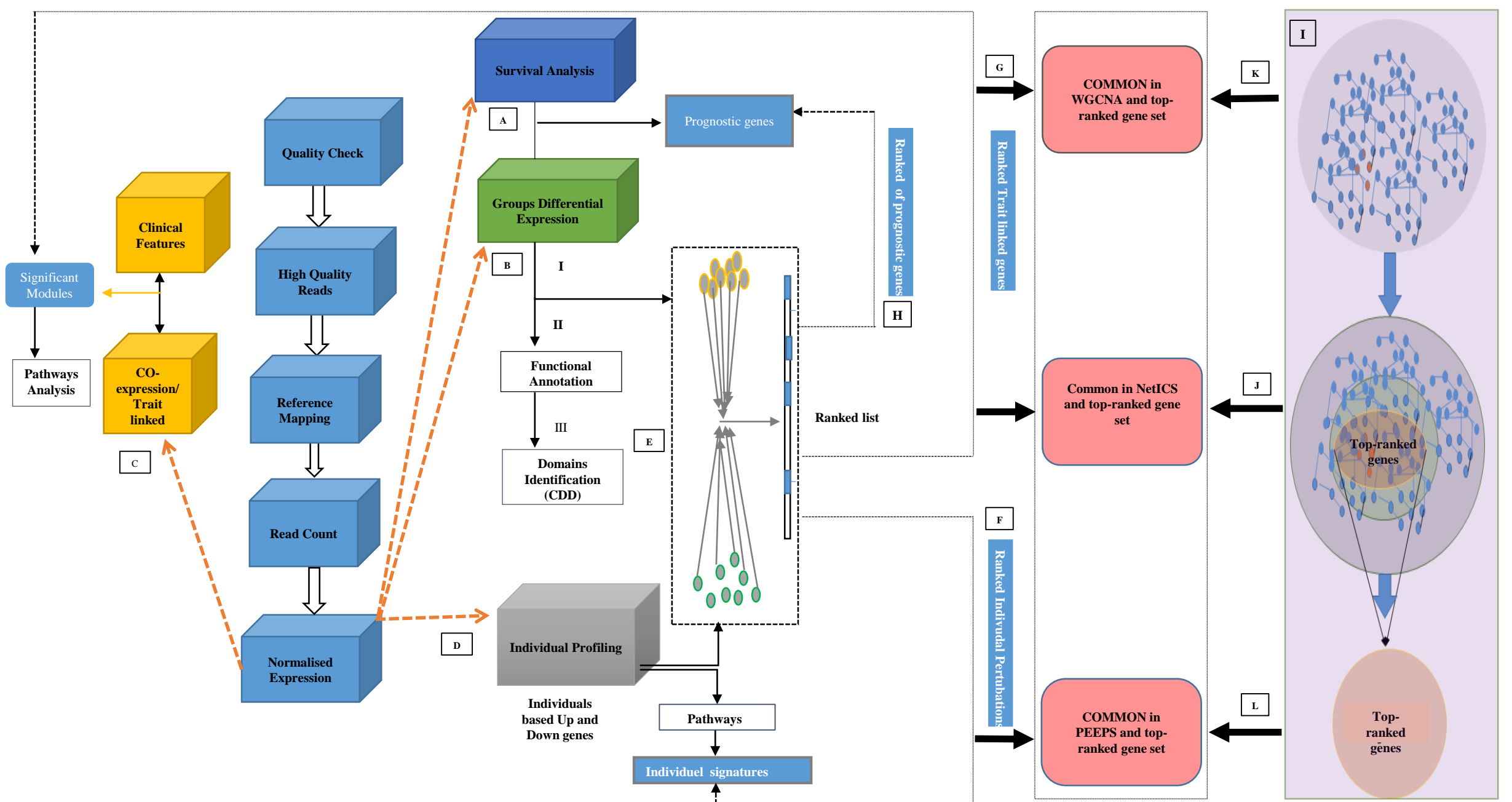


Figure 1: Flexible and interpretable omics integrative framework for RNA-seq data collected on two groups of patients, exemplified on PDAC ST/LT survival. RNA-seq quality-controlled data are inputted for A) Survival analysis; B) Group-based differential analysis via DESeq2⁶⁰; C) Weighted gene co-expression network analysis WGCNA²¹; D) Individual-based differential analysis (appendix pp 2-5); E) Genes are ranked based on the integration of individual and group-based differentially expressed genes via NetICS⁶¹; F-H) NetICS specific top 1% ranked genes are traced back in multiple previous analyses (A through E); I) DADA⁵ analysis starting from disease genes; J-L) DADA specific top 1% ranked genes are traced back in previous analyses (A through E).

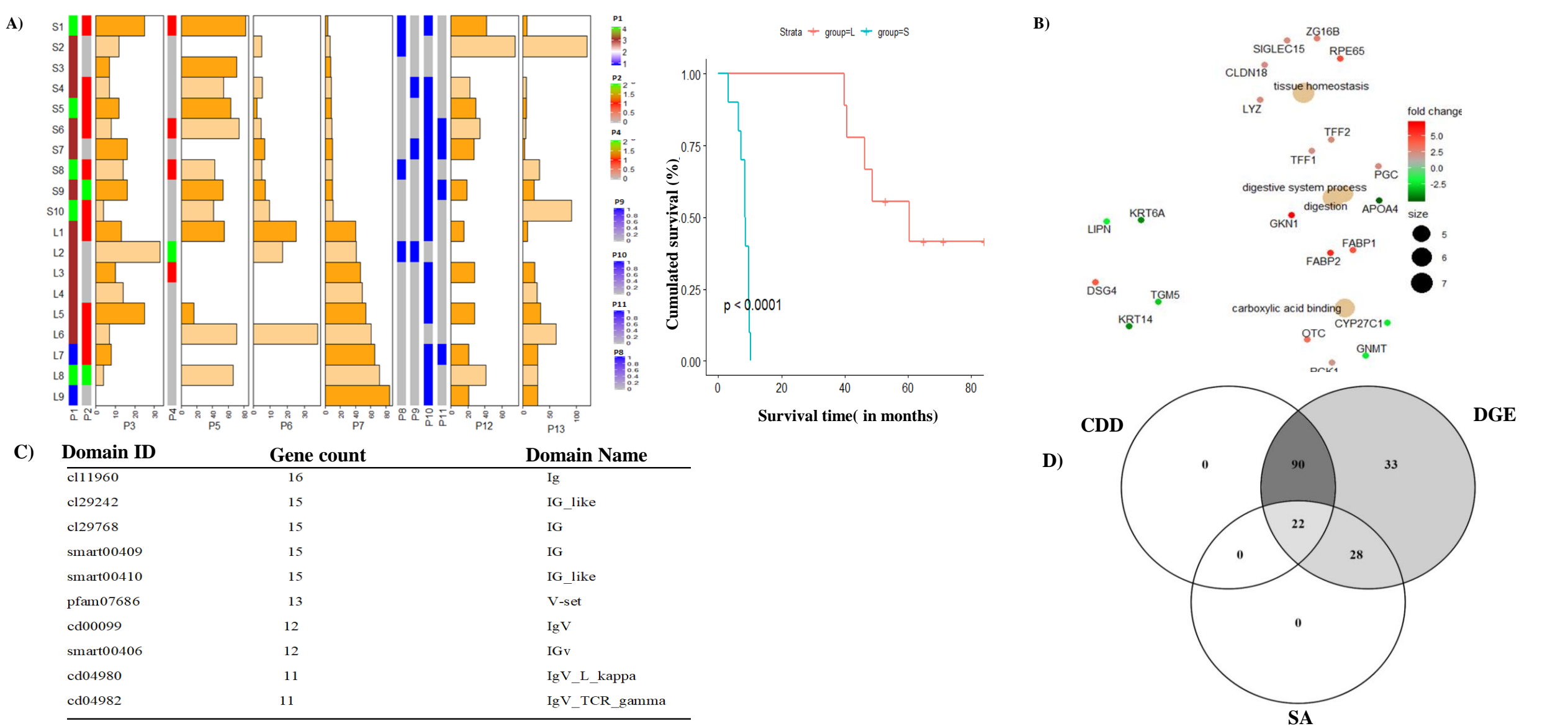


Figure 2: Overall Kaplan–Meier survival analysis of the ST and LT PDAC cohorts: A) Patient characteristic data for a selection of PDAC relevant traits are shown as mixed bar and heat map plot. P1 to P13 refer to patient specific clinical traits analyzed in this study (selective data has been shown in plot; full details given in appendix pp 9-11). P1 indicates Tumor stage (from 1 to 4). P3, P5, P6, P7, P12 and P13 indicate the frequency of number of nodes analyzed, time between surgery and chemotherapy (in days), disease free survival, OS (in months), tumor size by imagery (in mm) and Time between imagery and surgery, respectively. Remaining P2, P4, P8, P9, P10, P11, refers to status of N stage, surgical margin invaded by tumor cells, vascular resection, re-hospitalization after surgery, vascular resection, and artery contact, respectively. Here 0 and 1 indicate no and yes, reps. P7 clinical trait denotes overall survival and was used for the development of the Kaplan-Meier survival curves for short-term (ST) and long-term (LT) PDAC Survivors (STS: S1 to S10; LTS: L1 to L9); B) Identification of significant gene ontology of associated up and down-regulated DEGs and their relevant functions. Up and down-regulated genes are highlighted with red and green dots, respectively. The size of data points increases with increased significance (uncorrected for multiple testing – see appendix pp 3); C) Top-ranked conserved domains in differentially expressed gene sets; D) Venn-diagram showing the number of identified genes that are common to or different in multiple first-line analysis strategies (CDD: conserved domain database analysis, DGE: differential gene expression analysis, SA: survival analysis (appendix pp 3)).

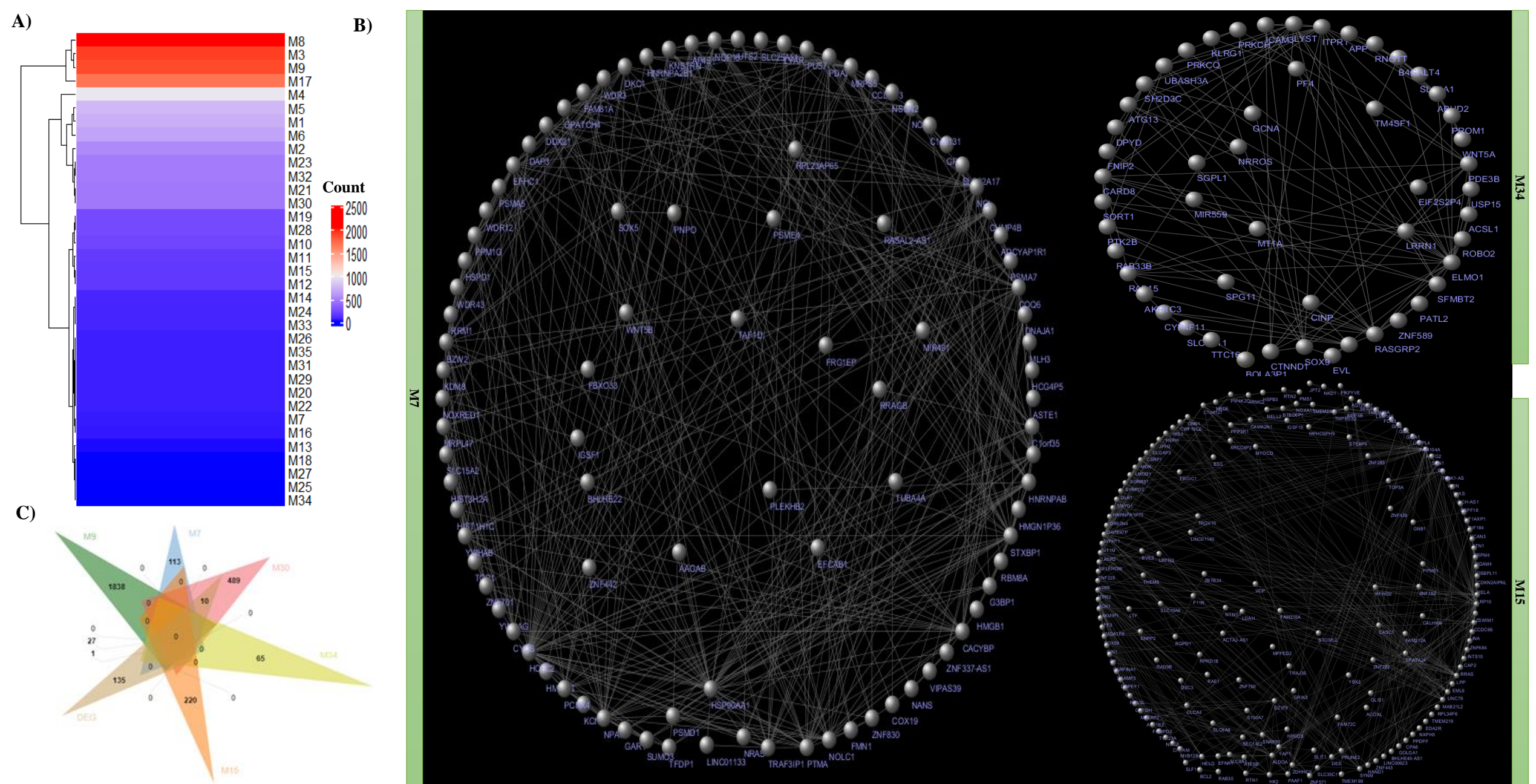
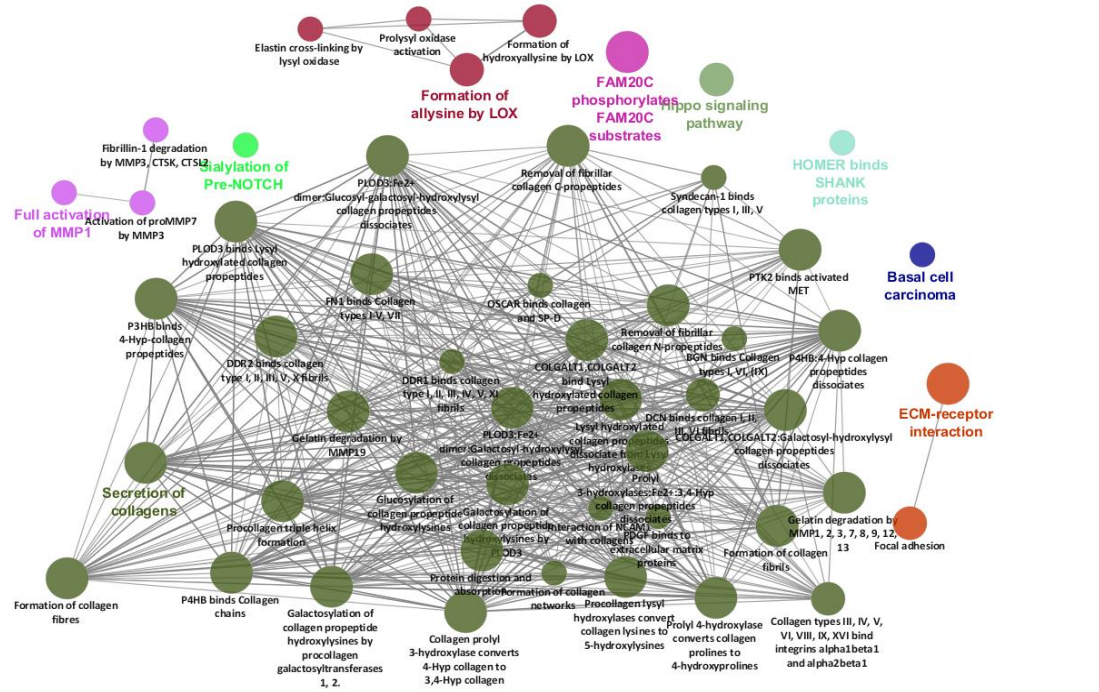
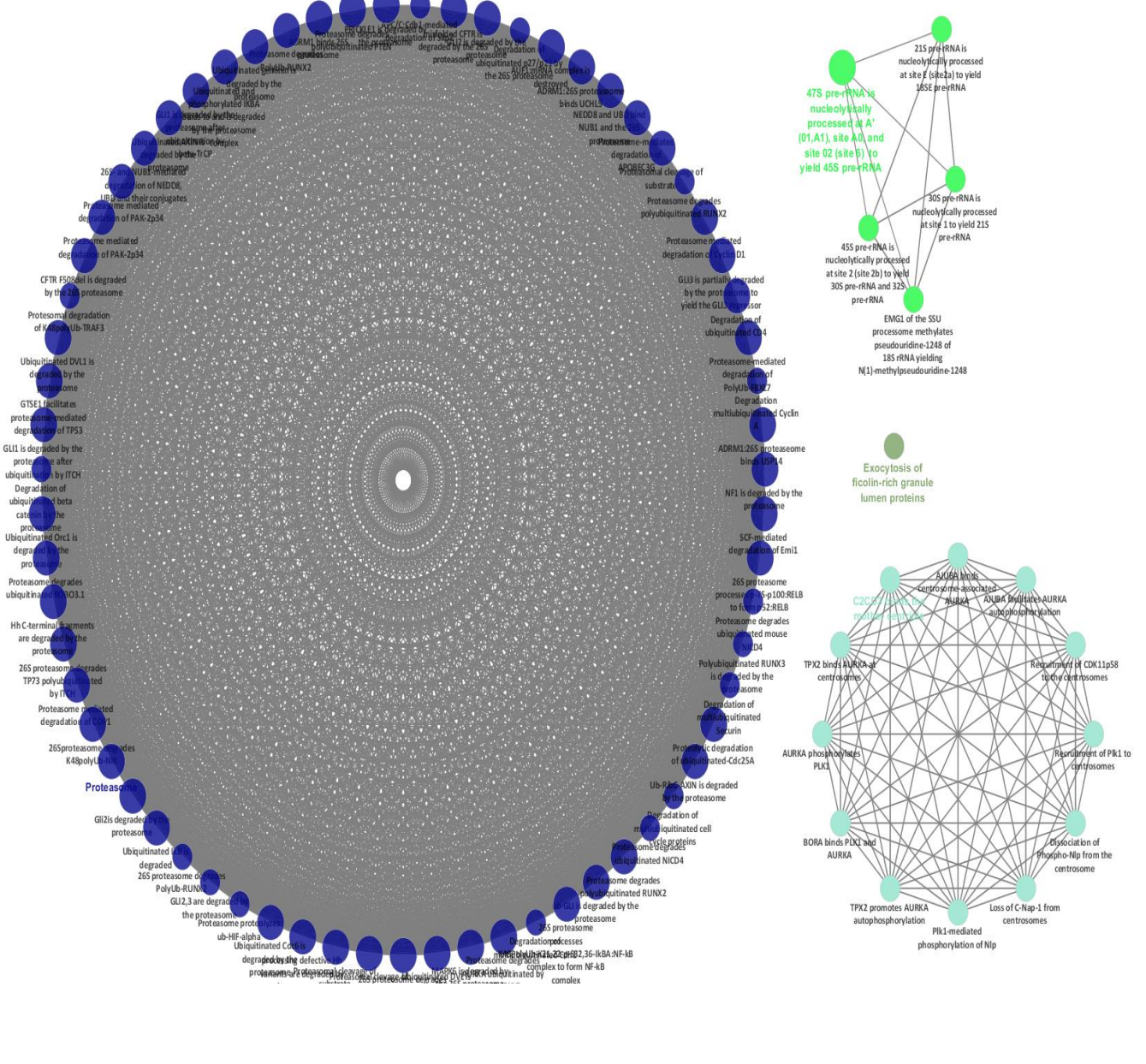


Figure 3: Clinical relevance of gene co-expression modules: A) Heatmap indicating the number of genes involved in each WGCNA-derived gene module; B) Network topology of three modules (M7, M15, M34), where nodes are genes and connections among nodes represent gene-gene interactions. In each network, the gene names are indicated in the circular layout as derived from Cytoscape.⁶² C) Venn diagram indicating the common genes between the identified significant DEGs and the five previously identified clinically relevant modules.

A)



B)



C)

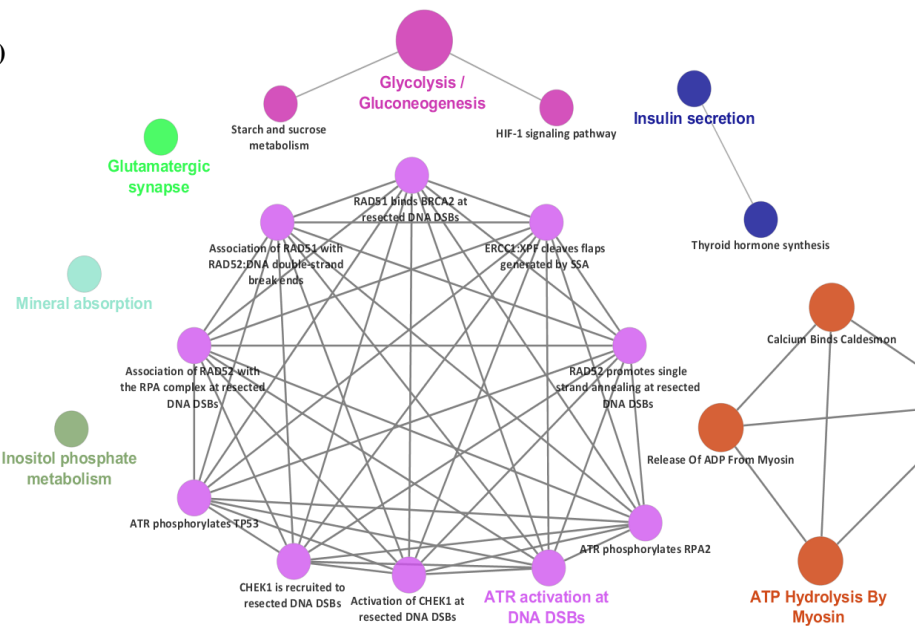


Figure 4: Functional follow-up of clinically relevant gene expression modules: A) Ten groups for module M9 comprising 33 significantly linked pathways; B) Three groups identified in the M7 modules; C) Depiction of the five groups identified in M15; For A-C, redundant groups with >50% overlap were merged. Each node in the network represents an enriched term; the size of each node follows the extent of enrichment significance. Connection among different nodes are based on kappa scores (≥ 0.4), as available from ClueGO.

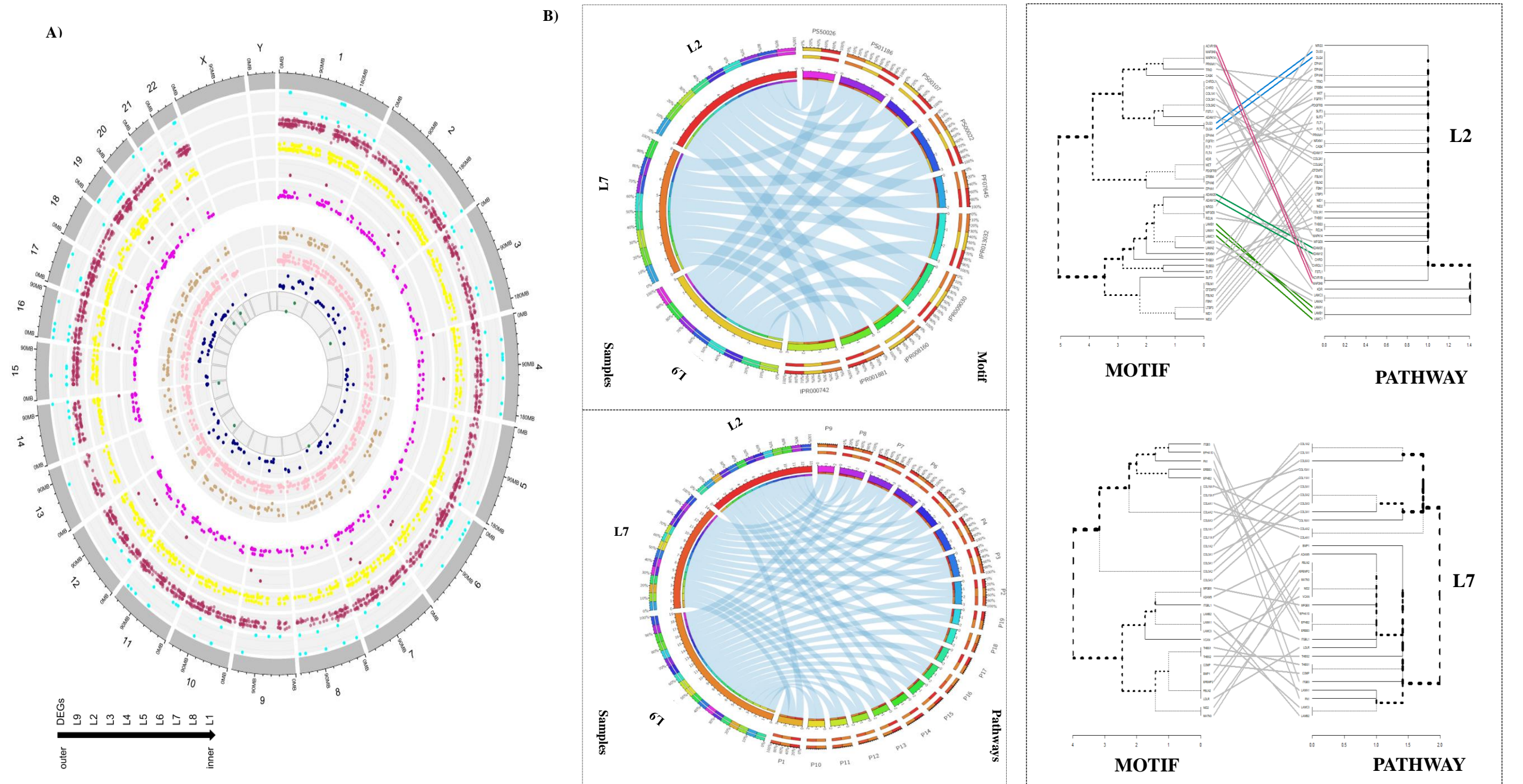


Figure 5: Genomic distributions of differentially expressed genes (DEGs) and PEEPs related to PDAC survivors using Circos plots and functional profiles of perturbation data: A) first outermost circle labeled with numbers represent chromosomes (same colors); the outermost track represents DEGs (up-regulated and down-regulated DEGs as scattered points); the nine innermost circles refer to the z-score for each LT survivor (LTS: ranging from LT1 to LT9) as scattered points. We have indicated perturbed genes only for chromosome 1 to 22 via track 2 to track 10 (outer to inner); B) Enriched KEGG pathways (P1 to P19 (out of 193)) and motifs common to at least 2 out of 9 LT individuals, shown via Circos Table Viewer (appendix pp 12-20). Each link refers to an LT survivor and a significantly enriched pathway (adjusted p-value < 0.05)/enriched motif based on the perturbed gene set found in that individual (data for LT2, LT7 and LT9 are shown). Uniquely enriched pathways across LT survivors are given in appendix pp 12-18; C) Visual comparison of two dendrograms developed from genes linked to enriched pathway and motif profiles. Similar sub-trees are connected with lines of the same color, while tree branches leading to distinct sub-trees are indicated with dashed lines.

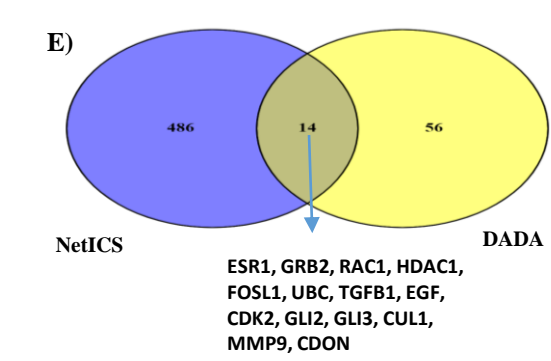
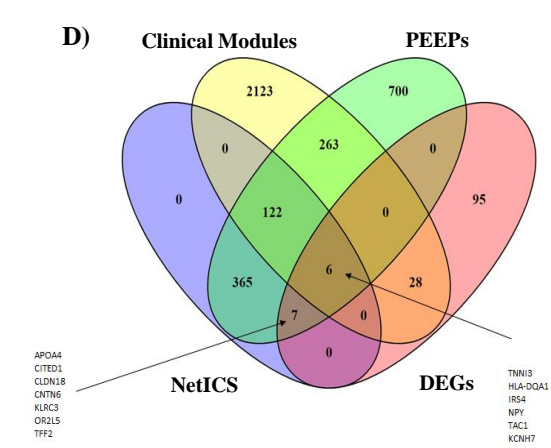
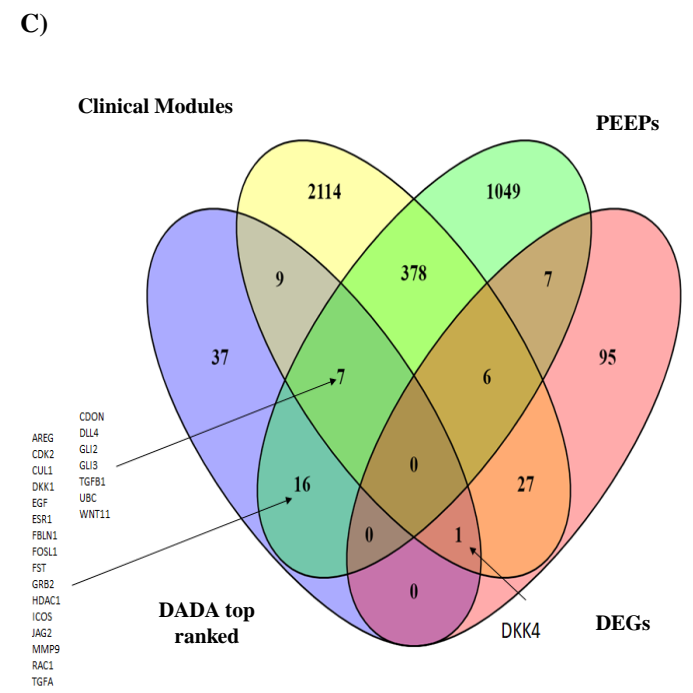
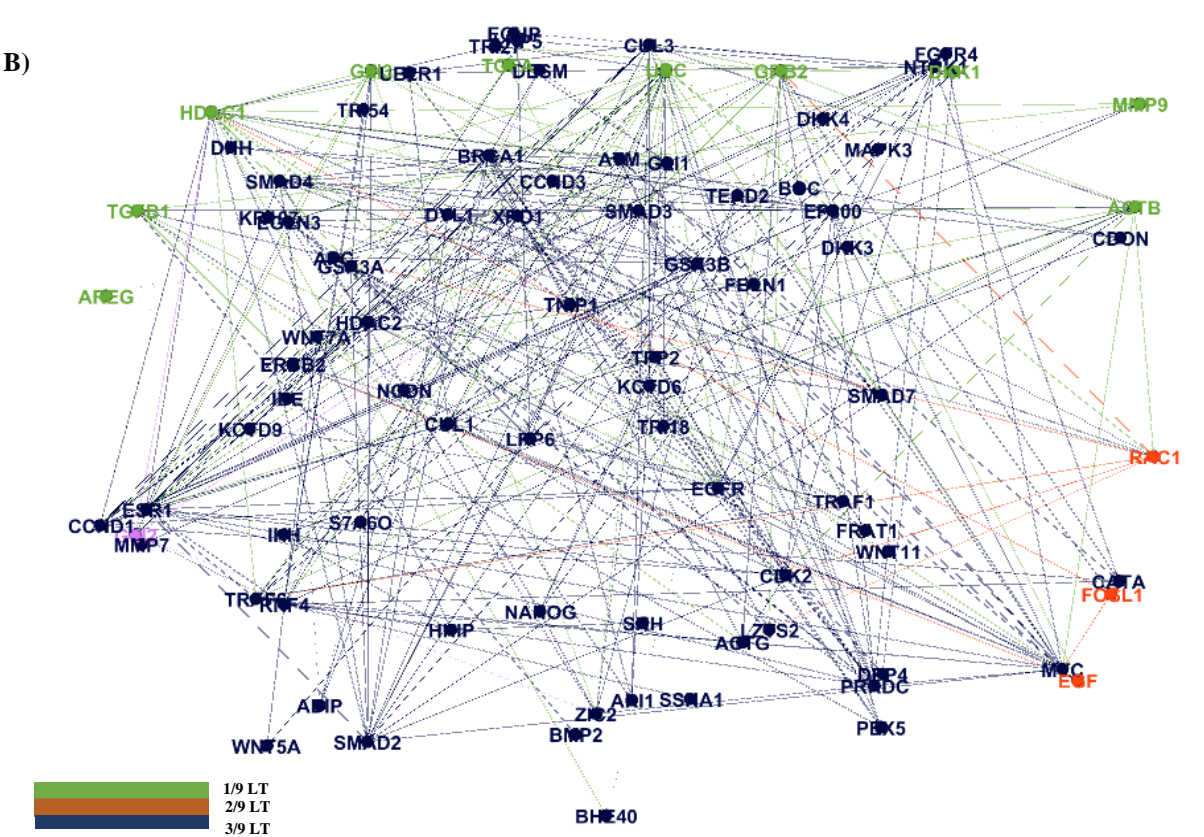
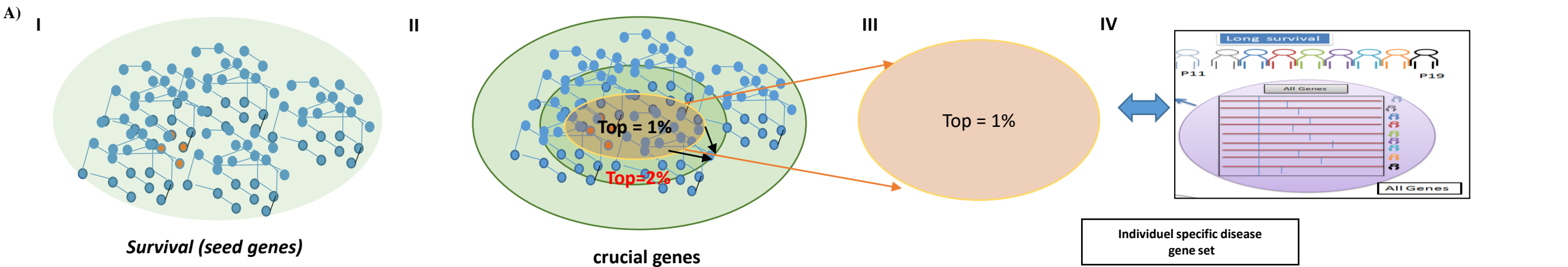


Figure 6: Exploitation of gene connectivity for LT PDAC survivor gene prioritization: A) DADA-oriented multi-step disease module identification: PDAC seed gene selection (I), restriction to top 1% of ranked genes (II-III) and intersection of retained gene list with individual perturbation gene expression profiles for LT survivors (IV); B) DADA-derived top-ranked genes found in at least one, two, or three LT survivors, indicated in green, orange and pink, respectively; C) Common genes to DADA and other gene prioritization approaches: DEGs, clinically relevant WGCNA gene modules, and PEEPs; D) Same as C) but with NetICS instead of DADA; E) Venn diagram showing the overlap between genes prioritized via NetICS and DADA. Common genes to top 1% NetICS individual gene lists and top 1% DADA genes are highlighted via arrows in C) and D).