Sejal MSc Thesis V2

by Sejal Msc

Submission date: 23-May-2026 04:10PM (UTC+0530)
Submission ID: 2967648849
File name: Sejal_DTU_Thesis_Report.docx (460.12K)
Word count: 7588
Character count: 50445

Chapter 1

INTRODUCTION

This chapter establishes the biological and diagnostic background of the study. It introduces pancreatic ductal carcinoma, pancreatic pseudocyst and the need for network-based biomarker discovery.

1.1 Background of Pancreatic Disease

The pancreas is a complex organ with endocrine and exocrine functions. Disease processes involving the pancreas may present with overlapping symptoms, imaging features and inflammatory responses. Among these disorders, pancreatic ductal carcinoma and pancreatic pseudocyst represent two clinically distinct but diagnostically challenging entities. PDAC is malignant, aggressive and frequently diagnosed late, whereas pancreatic pseudocyst is a benign inflammatory complication. The difficulty arises because both conditions may present with abdominal pain, altered pancreatic morphology and radiological abnormalities. A molecular framework that clarifies shared and divergent regulatory mechanisms can therefore support improved diagnostic interpretation.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

1.2 Pancreatic Ductal Carcinoma

Pancreatic ductal carcinoma accounts for the major proportion of pancreatic cancer cases and is associated with poor survival. The final research paper emphasizes that PDAC is silent in nature, and late-stage diagnosis contributes directly to poor survival rates. Its biology is shaped by uncontrolled proliferation, resistance to apoptosis, inflammatory signaling,

angiogenesis, stromal interactions and metabolic changes. Because early diagnosis remains difficult, molecular biomarkers that reflect disease biology are important for distinguishing malignant transformation from benign pancreatic changes.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

1.3 Pancreatic Pseudocyst and Diagnostic Ambiguity

Pancreatic pseudocyst is a localized fluid collection usually associated with pancreatitis. It is benign, but it can mimic PDAC in clinical and radiological settings. The research paper identifies this overlap as a central diagnostic problem. When inflammation, fluid accumulation and tissue remodeling occur, imaging alone may not always provide adequate separation between pseudocyst and malignant pancreatic lesions. This creates the need for biomarkers that can be interpreted together with clinical and imaging data.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

1.4 Need for Network-Based Biomarker Discovery

Traditional single-gene biomarker strategies may fail to capture the complexity of pancreatic disease. A regulatory network approach evaluates genes, miRNAs and transcription factors as interconnected elements rather than isolated molecules. Such an approach is valuable because disease phenotypes often emerge from coordinated dysregulation across multiple molecular layers. Network analysis can identify hub regulators that influence several targets and may therefore provide stronger diagnostic value than isolated biomarkers.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-

cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

1.5 Aim of the Thesis

The aim of this thesis is to expand the final research paper into a full-length study that explains how common regulatory biomarkers between PDAC and pancreatic pseudocyst can be identified using systems biology and network analysis. The thesis focuses on common genes retrieved from CTD, miRNA-gene interaction networks, transcription factor-gene networks, pancreas tissue-specific interaction networks and interpretation of regulatory hubs.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

1.6 Objectives

The specific objectives are to retrieve disease-associated genes for PDAC and pancreatic pseudocyst, identify common genes between the two conditions, construct regulatory interaction networks, identify hub miRNAs and TFs using degree centrality, interpret their functional relevance, and propose a diagnostic biomarker panel for future validation.

The present work follows a systems-level approach because pancreatic pathologies are rarely explained by a single molecular alteration. Instead, they involve layered changes in cell-cycle genes, inflammatory mediators, angiogenic regulators and apoptosis-associated molecules. By organizing these molecules into interpretable networks, the study provides a structured route for biomarker prioritization.

Chapter 2

LITERATURE REVIEW

This chapter reviews the conceptual foundation required to interpret the work. It describes pancreatic cancer biology, pseudocyst-related ambiguity, miRNA regulation, transcription factor control and systems-biology approaches.

2.1 Biological Features of PDAC

PDAC arises from ductal epithelial components of the pancreas and progresses through complex molecular events. Its aggressiveness is linked with early invasion, stromal reaction, immune evasion, poor vascular delivery of therapeutics and delayed clinical symptoms. At the molecular level, PDAC progression involves dysregulation of genes controlling proliferation, apoptosis, inflammation and angiogenesis. The regulatory hubs identified in the research paper, including MYC, CCND1, VEGFA and RELA, reflect these biological themes.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.2 Pancreatic Pseudocyst as a Benign Inflammatory Condition

Pancreatic pseudocyst typically develops after pancreatitis and reflects localized inflammatory injury. Although it is not malignant, it can share clinical or radiological resemblance with PDAC. This resemblance can lead to diagnostic uncertainty, especially when a pancreatic lesion is detected with nonspecific symptoms. Molecular biomarkers that distinguish inflammatory remodeling from malignant transformation may therefore improve diagnostic confidence.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.3 Role of Common Disease-Associated Genes

The final research paper identified 31 overlapping genes between PDAC and pancreatic pseudocyst. These common genes are relevant because they indicate shared biological processes in the pancreatic microenvironment. Some genes are involved in cell-cycle control, such as CCND1, CCND2 and CDK6. Others are linked with apoptosis, such as FAS and BCL2L1. VEGFA relates to angiogenesis, whereas RELA connects inflammatory signaling with survival pathways.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.4 microRNAs in Disease Regulation

microRNAs are short non-coding RNAs that regulate gene expression post-transcriptionally. They typically bind complementary sequences in target transcripts and reduce translation or promote mRNA degradation. In cancer, miRNAs may behave as tumor suppressors or oncogenic regulators. In the present study, hsa-miR-34a-5p, hsa-miR-16-5p and hsa-miR-335-5p were identified as central hub miRNAs, suggesting a broad regulatory influence over shared pancreatic disease genes.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.5 hsa-miR-34a-5p

hsa-miR-34a-5p is commonly interpreted as a tumor-suppressive miRNA. In the research paper, it showed the highest miRNA hub status with degree 11. Its regulatory interpretation includes suppression of genes associated with cell-cycle progression and proliferation, including CCND1, CDK6 and MYC. Loss or reduction of such regulation may permit oncogenic activity and contribute to malignant transformation.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.6 hsa-miR-16-5p

hsa-miR-16-5p was identified as a prominent miRNA hub with degree 9. Its reported targets include cyclin family genes such as CCND1 and CCND2. This places miR-16-5p in the context of G1/S cell-cycle control. Abnormal regulation of miR-16-mediated pathways may be relevant to both inflammatory remodeling and neoplastic progression in pancreatic disease.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.7 hsa-miR-335-5p

hsa-miR-335-5p also emerged as a hub miRNA with degree 9. The research paper relates this miRNA to genes involved in stemness and apoptosis, including POU5F1 and BCL2L1. Its role may therefore connect cell survival, plasticity and aggressive behavior in pancreatic pathology.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.8 Transcription Factor Hubs

Transcription factors coordinate the expression of large gene programs. In the research paper, MYC, CCND1, VEGFA, RELA and CDKN1B were highlighted as major regulatory hubs in the TF-gene network. Although not all are classical TFs in the same sense, they serve as high-connectivity regulatory nodes within the constructed network. Their degree values suggest substantial interaction density and biological relevance.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.9 Systems Biology and NetworkAnalyst

Systems biology provides a method for understanding complex biological systems through interactions and network organization. NetworkAnalyst supports the creation of regulatory networks from gene lists and integrates curated interaction information. In this thesis, NetworkAnalyst is used as the core platform to construct miRNA-gene, TF-gene and pancreas tissue-specific interaction networks.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

2.10 Hub-Based Biomarker Discovery

Hub nodes are molecules with many interactions in a network. They may be especially important because changes in their activity can affect several pathways simultaneously. Biomarker discovery based on hubs is useful because high-degree regulators can represent central points of disease control. However, hub status alone does not prove diagnostic performance; it must be followed by experimental and clinical validation.

This literature background supports the rationale for the current thesis: PDAC and pancreatic pseudocyst are biologically different but may share molecular signals. A network-driven strategy can use those shared signals to identify regulatory patterns that are more informative than single molecules considered independently.

Chapter 3

MATERIALS AND METHODS

This chapter explains the methodology followed in the study, including data retrieval, common gene identification, regulatory network construction, tissue-specific network integration and hub analysis.

3.1 Study Design

The study was designed as a computational systems-biology analysis. It used disease-associated gene information and regulatory network construction to identify shared biomarker candidates between PDAC and pancreatic pseudocyst. The workflow began with data retrieval, followed by common gene identification, construction of regulatory networks, hub selection and biological interpretation.

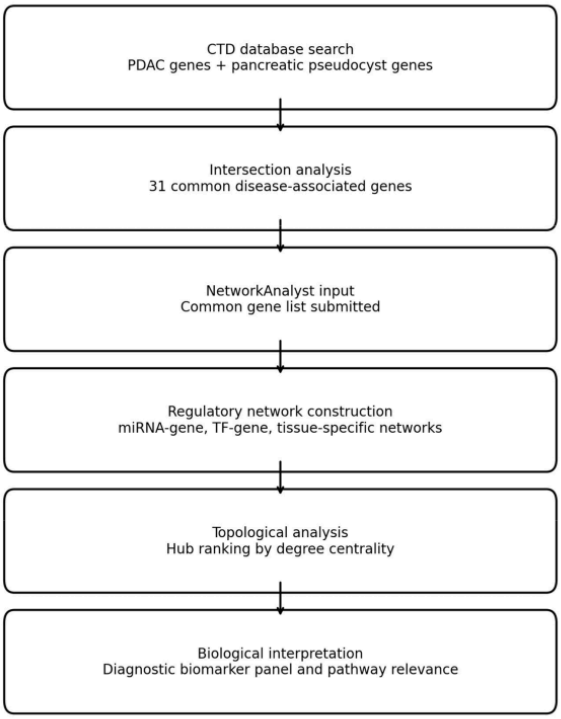

Figure 3. Methodological workflow of the regulatory network-driven biomarker discovery study.

3.2 Data Retrieval from CTD

Disease-associated genes for PDAC and pancreatic pseudocyst were retrieved from the Comparative Toxicogenomics Database. CTD was selected because it is a curated repository of disease-gene and molecular interaction information. Separate gene sets were obtained for each condition and prepared for intersection analysis.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.3 Common Gene Identification

The PDAC-associated gene set and pancreatic pseudocyst-associated gene set were compared to identify overlapping genes. The research paper reported 31 common genes. These overlapping genes were considered shared molecular candidates and were selected for downstream regulatory network construction.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.4 miRNA-Gene Network Construction

The common gene list was submitted to NetworkAnalyst to construct the miRNA-gene regulatory network. This network connected genes with predicted or curated miRNA regulators. Nodes represented genes and miRNAs, while edges represented regulatory interactions. Network degree was calculated to determine highly connected miRNA hubs.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.5 Transcription Factor-Gene Network Construction

A TF-gene regulatory network was generated using NetworkAnalyst. The purpose was to identify transcriptional regulators and high-connectivity gene hubs within the shared disease network. Degree centrality was used to rank the most connected nodes.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.6 Pancreas Tissue-Specific Network

To improve biological relevance, pancreas tissue-specific interaction information was integrated. Tissue-specific networks help determine whether identified regulators remain meaningful in the pancreatic microenvironment. This step supports contextual interpretation rather than relying only on general regulatory networks.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.7 Hub Selection Criteria

Hub nodes were selected using degree centrality. Degree centrality measures the number of connections linked to a node. The research paper selected the top three miRNA hubs and top TF or regulatory hubs for interpretation. The key miRNA hubs were hsa-miR-34a-5p, hsa-miR-16-5p and hsa-miR-335-5p. The key TF/regulatory hubs were MYC, CCND1, VEGFA, RELA and CDKN1B.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.8 Functional Interpretation

Functional interpretation was based on the known roles of hub genes and miRNAs in proliferation, apoptosis, inflammation and angiogenesis. The biological direction of miRNA-mediated regulation was interpreted using literature-supported relationships described in the research paper.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.9 Thesis Expansion Method

The final research paper was treated as the primary source. The Kaushiki thesis sample was used as a guide for chapter-wise content development, front matter, tables and academic writing style. The DTU thesis template was used as the formatting reference for title page,

declaration, certificate, content sequence, chapter headings, 12-point Times New Roman font and 1.5 line spacing.

This step contributes to the overall objective of prioritizing biomarkers that are both computationally central and biologically interpretable. The method emphasizes reproducibility, structured data use and transparent hub selection.

3.10 Tools and Data Summary

Tool/Resource	Purpose	Output Used in Thesis
CTD	Retrieval of disease-associated genes for PDAC and pancreatic pseudocyst	Common disease gene list
NetworkAnalyst	Construction of miRNA-gene, TF-gene and tissue-specific networks	Network maps, hub degree values
Degree centrality	Topological ranking of nodes	Hub miRNAs and TF/regulatory hubs
Literature interpretation	Biological explanation of hubs	Diagnostic relevance and validation needs

Chapter 4

RESULTS

This chapter presents the results obtained from the regulatory network-driven analysis. It interprets common genes, miRNA hubs, transcription factor hubs and pancreas tissue-specific network findings.

4.1 Identification of Common Genes

Intersection analysis of disease-associated genes retrieved from CTD resulted in the identification of 31 common genes between PDAC and pancreatic pseudocyst. The presence of these common genes indicates that malignant and benign inflammatory pancreatic conditions can share molecular pathways, especially those associated with inflammation, cell-cycle activation, apoptosis and angiogenesis.

Gene	Biological Group	Relevance
CCND1	Cell-cycle progression	Supports proliferation and G1/S transition
CCND2	Cell-cycle progression	Cyclin-mediated control of cell division
CDK6	Cell-cycle kinase	Regulates G1 phase progression
FAS	Apoptosis	Death receptor signaling
BCL2L1	Apoptosis resistance	Cell survival and anti-apoptotic control
CCL2	Inflammation	Chemokine-mediated inflammatory response
MYC	Transcriptional regulation/proliferation	Central regulator of growth and metabolism
VEGFA	Angiogenesis	Tumor vascularization and hypoxia adaptation
RELA	Inflammatory survival signaling	NF-kB pathway component
CDKN1B	Cell-cycle control	Regulator of cell-cycle progression

The table lists representative genes explicitly discussed in the research paper. These genes summarize the biological themes that emerged from the 31-gene overlap: cell-cycle control, apoptosis, inflammation and angiogenesis.

4.2 miRNA-Gene Regulatory Network

The miRNA-gene regulatory network identified hsa-miR-34a-5p, hsa-miR-16-5p and hsa-miR-335-5p as the most prominent miRNA hubs. hsa-miR-34a-5p showed degree 11, while hsa-miR-16-5p and hsa-miR-335-5p each showed degree 9. These hub miRNAs are mainly tumor suppressive and may repress genes involved in proliferation, apoptosis resistance, inflammation and angiogenesis.

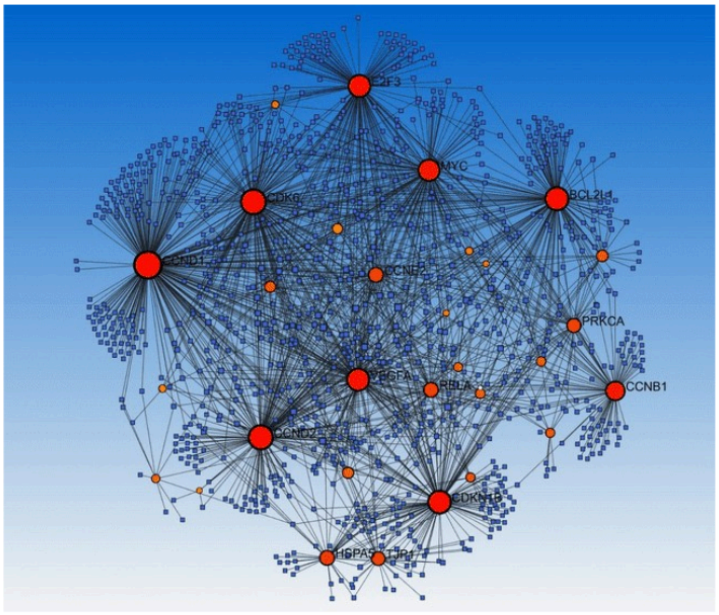

Figure 1. miRNA-gene interaction network from the final research paper.

Hub miRNA	Degree	Major Interpretation
hsa-miR-34a-5p	11	Tumor suppressor; regulates CCND1, CDK6 and MYC-related proliferation
hsa-miR-16-5p	9	Targets cyclin-related genes and supports cell-cycle suppression

hsa-miR-335-5p	9	Associated with stemness, apoptosis and tumor plasticity regulation

4.3 Transcription Factor-Gene Regulatory Network

The transcription factor-gene regulatory network identified MYC, CCND1, VEGFA, RELA and CDKN1B as major regulatory hubs. MYC showed the highest degree value of 79, followed by CCND1 with 64, VEGFA with 58, RELA with 35 and CDKN1B with 25. These hubs represent proliferative, angiogenic, inflammatory and cell-cycle regulatory themes that are relevant to both PDAC and pseudocyst-associated molecular changes.

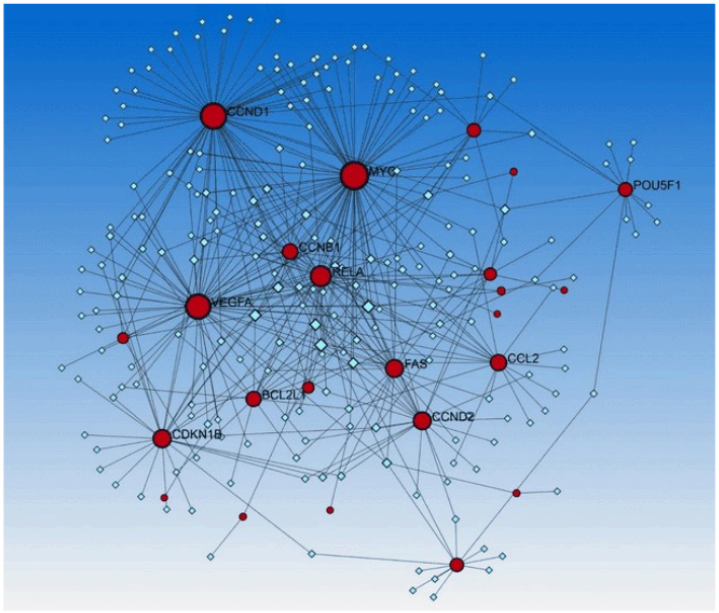

Figure 2. Transcription factor-gene regulatory network from the final research paper.

TF/Regulatory Hub	Degree	Functional Meaning
MYC	79	Global regulator of proliferation, metabolism and ribosomal biogenesis

CCND1	64	Cell-cycle progression and uncontrolled proliferation
VEGFA	58	Angiogenesis and vascular support
RELA	35	NF-kB-mediated inflammatory survival signaling
CDKN1B	25	Cell-cycle regulation and tumorigenesis-related control

4.4 Pancreas Tissue-Specific Integrated Network

The pancreas tissue-specific network reinforced the contextual relevance of the selected miRNAs and TF/regulatory hubs. Tissue-specific analysis is important because regulatory interactions may vary across organs. The persistence of these hubs in the pancreatic context supports their relevance to pancreatic disease biology and strengthens their candidacy for future biomarker validation.

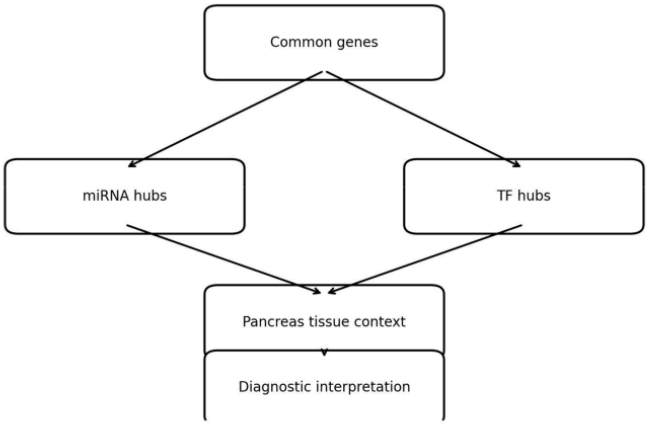

Figure 4. Integrated interpretation scheme linking common genes, miRNA hubs, TF hubs and pancreas tissue context.

4.5 Proposed Biomarker Panel

Based on the combined network evidence, the thesis proposes a candidate biomarker panel consisting of tumor-suppressive miRNAs and oncogenic regulatory hubs. The diagnostic

value of such a panel would not come from a single molecule alone but from the combined expression pattern across multiple regulatory layers.

Category	Candidate Biomarkers	Rationale
miRNA hubs	hsa-miR-34a-5p, hsa-miR-16-5p, hsa-miR-335-5p	Tumor-suppressive regulators controlling cell-cycle, apoptosis and plasticity genes
TF/regulatory hubs	MYC, CCND1, VEGFA, RELA, CDKN1B	High-degree regulators linked with proliferation, angiogenesis, inflammation and cell-cycle control
Combined panel	miRNA hubs + TF/regulatory hubs	Potential to distinguish malignant and benign inflammatory pancreatic processes after validation

Chapter 5

DISCUSSION

This chapter discusses the biological significance, diagnostic implications, strengths and limitations of the findings in relation to PDAC and pancreatic pseudocyst.

5.1 Interpretation of Shared Genes

The identification of 31 common genes suggests convergence between PDAC and pancreatic pseudocyst at the molecular level. This does not imply that pseudocyst is malignant; rather, it indicates that inflammation, tissue injury and cancer can activate overlapping biological processes. The shared genes reflect pathways that are important in both benign inflammatory remodeling and malignant progression.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.2 Biological Significance of miRNA Hubs

The three major miRNA hubs identified in the study are mostly tumor-suppressive in interpretation. Reduced activity of these miRNAs could allow increased expression of oncogenic targets such as MYC, CCND1 and CDK6. hsa-miR-34a-5p is particularly important because it showed the highest degree among miRNA hubs and is linked with multiple cell-cycle regulators.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.3 Biological Significance of TF/Regulatory Hubs

The TF/regulatory hubs provide a second layer of interpretation. MYC and CCND1 indicate proliferative pressure, VEGFA indicates angiogenic activity, RELA reflects inflammation-associated survival signaling and CDKN1B represents cell-cycle regulation. Together, these hubs form a regulatory pattern that connects inflammation with tumor progression.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.4 Diagnostic Implications

A combined biomarker panel may be more useful than a single molecule because PDAC and pseudocyst share some biological signals. A panel can capture a multi-dimensional pattern: tumor-suppressive miRNA reduction, oncogenic hub activation, inflammatory signaling and angiogenesis. Such a panel could eventually support clinical differentiation when imaging and symptoms are ambiguous.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.5 Strengths of the Study

The major strength of the study is its integrative design. It uses curated disease-gene data, multiple regulatory networks, degree centrality and tissue-specific context. This makes the biomarker selection process systematic and interpretable. The approach also reflects current systems-biology thinking by treating disease biology as a network rather than a list of isolated genes.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.6 Limitations

The study is computational and depends on database-derived interactions. It does not measure gene or miRNA expression in patient samples and does not test diagnostic performance through ROC analysis, sensitivity, specificity or independent clinical cohorts. Network degree also does not automatically confirm causal importance. Therefore, the proposed biomarkers require experimental and clinical validation.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.7 Validation Strategy

Future validation should include qRT-PCR analysis of miRNAs and mRNA targets in PDAC tissue, pseudocyst-associated samples and matched controls. Circulating miRNA analysis in plasma or serum could be explored for non-invasive diagnosis. Statistical modeling should evaluate whether the combined panel improves discrimination over individual biomarkers.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.8 Clinical Translation

For clinical translation, the biomarker panel would need to be robust, reproducible, cost-effective and interpretable alongside imaging data. It should be tested in well-defined patient cohorts with confirmed diagnosis. The ultimate goal would be to reduce diagnostic uncertainty and support timely clinical decisions.

The interpretation must remain cautious because shared network centrality is a hypothesis-generating result. Nevertheless, it is valuable because it narrows the number of candidates and identifies biologically meaningful regulators for future focused validation.

5.9 Limitations and Validation Requirements

Limitation	Impact	Suggested Solution
Database dependency	May include context-independent interactions	Validate in pancreatic tissue and disease-specific cohorts
No experimental expression data	Cannot confirm upregulation or downregulation in samples	Perform qRT-PCR or RNA sequencing
No clinical performance metrics	Diagnostic utility remains unknown	Use ROC analysis, sensitivity and specificity testing
Hub degree is not causality	High connectivity may not equal functional dominance	Use functional assays and knockdown/overexpression studies

Chapter 6

CONCLUSION AND FUTURE SCOPE

This chapter concludes the thesis and presents future directions for experimental validation and clinical translation.

6.1 Conclusion

This thesis presents a regulatory network-driven biomarker discovery study for pancreatic ductal carcinoma and pancreatic pseudocyst. The final research paper was expanded into a full thesis format following DTU-style front matter, chapter organization and academic formatting. The study identified 31 common disease-associated genes between PDAC and pancreatic pseudocyst using CTD-based retrieval and intersection analysis. These common genes represent shared biological processes including cell-cycle control, apoptosis, inflammation and angiogenesis.

NetworkAnalyst-based regulatory network construction identified hsa-miR-34a-5p, hsa-miR-16-5p and hsa-miR-335-5p as major miRNA hubs. The transcription factor-gene network identified MYC, CCND1, VEGFA, RELA and CDKN1B as important high-degree hubs. The pancreas tissue-specific network supported the contextual relevance of these regulators in pancreatic disease biology. Together, these findings suggest that a combined panel of tumor-suppressive miRNAs and oncogenic regulatory hubs may support future diagnostic differentiation of PDAC and pancreatic pseudocyst.

The thesis concludes that a systems-biology approach is useful for biomarker prioritization because it captures regulatory relationships rather than isolated molecules. However, the findings remain computational and should be treated as hypotheses for future validation. Experimental confirmation and clinical testing are necessary before the proposed biomarkers can be used in diagnostic practice.

6.2 Future Scope

Future work should validate the proposed biomarkers in patient-derived pancreatic tissue, pseudocyst fluid, serum and plasma samples. Expression levels of hsa-miR-34a-5p, hsa-miR-16-5p, hsa-miR-335-5p, MYC, CCND1, VEGFA, RELA and CDKN1B should be

measured using qRT-PCR, RNA sequencing or other molecular assays. Larger independent cohorts should be used to evaluate diagnostic accuracy.

Machine learning models may be developed using the combined biomarker panel to classify PDAC and pancreatic pseudocyst. Such models should be evaluated with sensitivity, specificity, area under the ROC curve and external validation. Additional layers such as proteomics, methylation data and single-cell sequencing can be integrated to improve biological resolution. Ultimately, the proposed network-based panel may contribute to improved diagnostic workflows when combined with clinical examination and imaging.

APPENDIX A: COMMON GENE AND HUB SUMMARY

This appendix summarizes the main molecular entities described in the thesis. The complete computational workflow should be validated with expression data and patient cohorts before any clinical use.

Entity Type	Molecules	Interpretation
Common genes	31 overlapping genes from CTD	Shared molecular background between PDAC and pancreatic pseudocyst
miRNA hubs	hsa-miR-34a-5p, hsa-miR-16-5p, hsa-miR-335-5p	Tumor-suppressive regulatory candidates
TF/regulatory hubs	MYC, CCND1, VEGFA, RELA, CDKN1B	High-connectivity nodes involved in proliferation, angiogenesis, inflammation and cell-cycle control

APPENDIX B: SAMPLE VALIDATION PLAN

• Collect confirmed PDAC, pancreatic pseudocyst and control samples under ethical approval.

• Measure expression of candidate miRNAs and hub genes using qRT-PCR or sequencing.

• Evaluate individual and combined diagnostic performance using statistical models.

• Validate the model in an independent cohort.

• Compare biomarker results with imaging and clinical parameters.

APPENDIX C: DECLARATION OF COMPUTATIONAL NATURE

The present thesis is based on computational and database-guided analysis. It does not report newly generated wet-lab data. The biomarker panel proposed here is intended for future experimental validation.

APPENDIX D: EXTENDED BACKGROUND NOTES

D.1 Importance of early diagnostic discrimination

Early differentiation between malignant and benign pancreatic lesions is clinically important because the management pathway is very different. PDAC requires urgent oncological evaluation, staging and treatment planning, while pancreatic pseudocyst is usually managed according to size, symptoms, infection risk and pancreatitis-related complications. When the two entities overlap clinically, a molecular panel may help provide an additional layer of evidence. The proposed network-based panel should therefore be viewed as a decision-support concept rather than a standalone diagnostic replacement.

For a future clinical workflow, these markers should be tested together with radiological and biochemical indicators. The integration of molecular biomarkers with imaging can improve interpretability and may reduce uncertainty in borderline cases. This is especially important for pancreatic diseases where delayed diagnosis can strongly influence outcome.

D.2 Inflammatory overlap in pancreatic lesions

Inflammation can create molecular signals that resemble cancer-associated pathways. Cytokine signaling, immune-cell recruitment, hypoxia, tissue repair and extracellular matrix remodeling can all influence gene expression. This explains why a benign inflammatory pseudocyst may share certain genes with PDAC. The identification of shared genes should not be interpreted as malignant conversion; instead, it reveals that both conditions can activate common biological processes in the pancreatic microenvironment.

For a future clinical workflow, these markers should be tested together with radiological and biochemical indicators. The integration of molecular biomarkers with imaging can improve interpretability and may reduce uncertainty in borderline cases. This is especially important for pancreatic diseases where delayed diagnosis can strongly influence outcome.

D.3 Value of miRNA biomarkers

miRNAs are attractive biomarker candidates because they can regulate multiple target genes and may be detectable in tissue or circulation. A single miRNA can influence a pathway-level response, and a small set of miRNAs can summarize complex regulatory states. In the

present thesis, hsa-miR-34a-5p, hsa-miR-16-5p and hsa-miR-335-5p were prioritized because they showed high centrality in the miRNA-gene network and had biologically interpretable roles.

For a future clinical workflow, these markers should be tested together with radiological and biochemical indicators. The integration of molecular biomarkers with imaging can improve interpretability and may reduce uncertainty in borderline cases. This is especially important for pancreatic diseases where delayed diagnosis can strongly influence outcome.

D.4 Need for combined panels

A combined panel is likely to be more informative than any one marker. PDAC and pseudocyst may share inflammatory and remodeling signatures, so a robust diagnostic panel must capture multiple dimensions of disease biology. The proposed panel combines tumor-suppressive miRNA hubs with high-degree regulatory genes. Such a panel can potentially represent proliferation, apoptosis, inflammation and angiogenesis simultaneously.

For a future clinical workflow, these markers should be tested together with radiological and biochemical indicators. The integration of molecular biomarkers with imaging can improve interpretability and may reduce uncertainty in borderline cases. This is especially important for pancreatic diseases where delayed diagnosis can strongly influence outcome.

APPENDIX E: CANDIDATE BIOMARKER DETAIL SHEETS

hsa-miR-34a-5p

This miRNA was the highest-degree miRNA hub in the regulatory network. It is interpreted as a tumor-suppressive regulator and is associated with inhibition of cell-cycle progression. In the research paper, it was linked to regulation of CCND1, CDK6 and MYC. Its diagnostic relevance arises from its ability to represent a broad anti-proliferative regulatory program.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

hsa-miR-16-5p

This miRNA was identified as a high-degree hub with a degree value of 9. It is associated with control of cyclin-related genes and G1/S transition. In the context of PDAC and pseudocyst, altered miR-16-5p activity may indicate abnormal cell-cycle activation shared by inflammatory and neoplastic pancreatic changes.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

hsa-miR-335-5p

This miRNA was also identified as a high-degree hub with a degree value of 9. It is linked with genes associated with stemness, apoptosis and cell plasticity. Its inclusion in the panel is important because tumor progression is not only about proliferation but also about survival, invasiveness and adaptive cellular states.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

MYC

MYC showed the highest degree among the major regulatory hubs. It is a global regulator of proliferation, metabolism and biosynthetic activity. Its high degree indicates broad network connectivity. In diagnostic interpretation, MYC can represent the proliferative and metabolic dimension of malignant biology.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

CCND1

CCND1 showed a high degree value and is directly associated with cell-cycle progression. It is particularly relevant because it is also discussed as a target of tumor-suppressive miRNAs. The relationship between CCND1 and miRNA hubs supports the value of studying regulatory pairs rather than isolated genes.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

VEGFA

VEGFA is a major angiogenic regulator. Its presence as a hub indicates the importance of vascular and hypoxia-associated processes in pancreatic disease biology. In PDAC, angiogenic signaling can support tumor growth, while inflammatory lesions may also involve vascular remodeling.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

RELA

RELA is a component of NF-kB signaling and connects inflammation with survival pathways. Its inclusion is biologically meaningful because pseudocyst is inflammatory while PDAC can exploit inflammatory signaling for progression. RELA may therefore represent a bridge between benign inflammation and malignant survival signaling.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

CDKN1B

CDKN1B is involved in cell-cycle regulation and showed hub status in the TF-gene network. Its interpretation depends on context, but its presence in the network supports the importance of cell-cycle control in shared pancreatic disease signatures.

Recommended validation: measure expression in confirmed PDAC samples, pseudocyst-associated samples and non-diseased controls. Evaluate whether the candidate improves classification alone and in combination with other markers. Compare the molecular result with imaging, histopathology and clinical diagnosis wherever available.

APPENDIX F: PROPOSED EXPERIMENTAL VALIDATION PROTOCOL

Step 1: Cohort definition

Define three groups: confirmed PDAC, confirmed pancreatic pseudocyst and appropriate controls. Inclusion and exclusion criteria should be documented clearly. Clinical diagnosis should be verified using histopathology, imaging, biochemical tests and clinician assessment.

Step 2: Sample collection

Collect tissue, pseudocyst fluid, serum or plasma depending on the validation design. Samples should be handled under standardized conditions to reduce technical variation. RNA preservation protocols should be followed strictly because miRNA and mRNA quality can affect downstream results.

Step 3: RNA isolation and quality control

Extract total RNA including the small RNA fraction. Assess concentration, purity and integrity using suitable instruments. Samples with poor purity or degraded RNA should be excluded or processed separately.

Step 4: Expression measurement

Measure miRNA expression using qRT-PCR, small RNA sequencing or digital PCR. Measure target gene expression using qRT-PCR or RNA sequencing. Internal controls should be selected carefully and tested for stability across sample groups.

Step 5: Statistical analysis

Compare expression levels between PDAC, pseudocyst and control groups. Use appropriate tests depending on distribution and sample size. Evaluate diagnostic performance using ROC curves, sensitivity, specificity, positive predictive value and negative predictive value.

Step 6: Panel modeling

Develop a combined biomarker score using logistic regression, random forest, support vector machine or another supervised learning method. The model should be trained on one dataset and tested on an independent dataset to reduce overfitting.

Step 7: Biological confirmation

Functional validation may include knockdown or overexpression studies in pancreatic cell models. Such experiments can test whether altering miRNA hubs affects the expression of MYC, CCND1, VEGFA, RELA or other network targets.

Validation Stage	Expected Output	Decision Point
Expression screening	Differential expression values	Select robust candidates
Panel development	Combined diagnostic score	Test whether panel outperforms individual markers
Independent validation	External performance metrics	Confirm reproducibility
Functional testing	Evidence of regulatory effect	Support biological mechanism

APPENDIX G: SAMPLE DATA COLLECTION FORMAT

The following sample format can be used during future validation studies. It is included to support reproducibility and planning. Actual patient-related data must be collected only after ethical approval and informed consent.

Field	Description
Sample ID	Unique anonymized identifier
Clinical group	PDAC / pancreatic pseudocyst / control
Age and sex	Basic demographic variables
Clinical diagnosis method	Histopathology, imaging, biochemical markers or clinician-confirmed diagnosis
Sample type	Tissue, cyst fluid, serum or plasma
RNA quality metric	Purity and integrity value
miRNA expression	Ct value or normalized expression
Gene expression	Ct value or normalized expression
Final classification	Confirmed diagnostic category

G.1 Data security

All data should be anonymized before analysis. Patient identifiers should not be stored in the analysis file. Access to raw clinical information should be restricted to authorized researchers only.

G.2 Data quality

Missing values, outliers and batch effects should be documented. Technical replicates can improve reliability. Normalization controls must be selected carefully for both miRNA and mRNA assays.

G.3 Reporting

Results should be reported with confidence intervals and validation metrics. Negative findings should also be reported because they help refine biomarker panels and prevent overinterpretation.

APPENDIX H: EXTENDED GLOSSARY

Term	Meaning
Biomarker	A measurable biological feature that can indicate a normal process, disease process or response to intervention.
Degree centrality	A network measure representing the number of connections associated with a node.
Hub node	A highly connected network element that may influence several other molecules.
Regulatory network	A graph showing relationships between regulators such as miRNAs or transcription factors and target genes.
Systems biology	An approach that studies biological components as interacting systems rather than isolated parts.
Diagnostic panel	A group of biomarkers interpreted together to improve diagnostic classification.
Tissue-specific network	A regulatory or interaction network filtered or contextualized for a particular tissue environment.
Validation cohort	An independent group of samples used to test whether findings are reproducible.

This glossary is intended to help readers interpret the network-based terminology used throughout the thesis. The definitions are simplified for academic report use and should be understood in the context of computational biomarker discovery.

APPENDIX I: EXTENDED DISCUSSION ON NETWORK INTERPRETATION

I.1 Why network context matters

A gene list provides useful information, but it does not explain how genes influence each other or how regulatory molecules coordinate disease behavior. A network representation adds structure by connecting genes with upstream regulators and related pathways. In this thesis, network context helped prioritize molecules that occupy central positions across shared PDAC and pseudocyst biology. This is important because diseases involving inflammation and cancer may not be separated by single markers alone.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.2 Interpretation of shared versus disease-specific markers

The study focused on common genes because the diagnostic problem arises from overlap. A shared gene does not automatically become a discriminatory biomarker. Instead, diagnostic value may appear when shared genes are interpreted through regulatory direction, network centrality and expression pattern. For example, a tumor-suppressive miRNA and an oncogenic hub may create a regulatory balance that differs between pseudocyst and PDAC. Therefore, future validation should examine both absolute expression and relationships among markers.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.3 Role of cell-cycle signaling

Cell-cycle regulators appear repeatedly in the study. CCND1, CCND2, CDK6 and CDKN1B are linked with control of cellular proliferation. In a benign inflammatory lesion, cell-cycle signals may reflect repair and regeneration. In PDAC, similar signals may reflect uncontrolled growth. This difference in biological meaning is why a panel needs clinical context and tissue-specific interpretation.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.4 Role of apoptosis and cell survival

Apoptosis-related genes such as FAS and BCL2L1 represent another shared biological dimension. In cancer, evasion of apoptosis supports survival of abnormal cells. In inflammatory pancreatic disease, apoptosis and survival pathways can also be activated during tissue damage and repair. The network-based approach helps connect these genes with miRNAs and transcriptional regulators, allowing more nuanced interpretation.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.5 Role of angiogenesis

VEGFA emerged as a high-degree hub in the TF-gene network. Angiogenesis is important for tumor growth, but vascular remodeling may also occur during inflammation. The diagnostic question is not whether VEGFA is present, but how strongly it contributes to a broader molecular pattern. Combining VEGFA with miRNA hubs and proliferation markers may improve interpretation.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.6 Role of inflammatory signaling

RELA connects the study to NF-kB-related inflammatory survival pathways. This is particularly relevant because pseudocyst is an inflammatory condition, while PDAC can exploit inflammatory signaling for tumor progression. RELA may therefore be a bridge marker that requires careful contextual analysis. Its degree centrality supports biological relevance, but clinical specificity must be tested experimentally.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.7 Importance of tissue-specific integration

Tissue-specific network integration reduces the risk of overinterpreting generic interactions that may not be relevant to the pancreas. Regulatory databases often include interactions from many tissues and experimental systems. By adding pancreas-specific context, the analysis becomes more aligned with the biological environment of PDAC and pseudocyst. This supports stronger hypothesis generation for future validation.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

I.8 From computational discovery to clinical use

Computational biomarker discovery is an early stage in translational research. It can identify candidates, reveal mechanisms and reduce the number of molecules requiring wet-lab testing. However, clinical use requires rigorous validation, reproducibility, standardization and regulatory acceptance. The proposed panel should therefore be considered a candidate framework that guides future laboratory and clinical studies.

In practical terms, this means that the thesis findings should be used to design experiments rather than to make direct diagnostic decisions. The strength of the work is its ability to generate a biologically coherent and testable panel from a small research-paper dataset.

APPENDIX J: CHAPTER-WISE EXTENDED SUMMARY

J.1 Introduction summary

The introduction explains the clinical challenge created by the similarity between PDAC and pancreatic pseudocyst. It also establishes why regulatory network analysis is suitable for identifying candidate biomarkers in complex pancreatic conditions.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

J.2 Literature review summary

The literature review describes the key biological layers relevant to the study: pancreatic cancer biology, inflammatory pseudocyst formation, miRNA-mediated regulation, transcription factor hubs and systems-biology biomarker discovery.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

J.3 Methods summary

The methods chapter describes retrieval of disease-associated genes, intersection analysis to identify common genes, NetworkAnalyst-based construction of regulatory networks and hub prioritization using degree centrality.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

J.4 Results summary

The results chapter reports 31 common genes, three major miRNA hubs and five major TF/regulatory hubs. It also explains the pancreas tissue-specific network and proposes a combined biomarker panel.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

J.5 Discussion summary

The discussion chapter interprets the biological relevance of shared genes and regulatory hubs. It emphasizes that the findings are computational and require validation before clinical application.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

J.6 Conclusion summary

The conclusion states that a network-driven approach can identify biologically meaningful biomarker candidates for differentiating PDAC and pancreatic pseudocyst, but future experimental validation is essential.

This section was included to help readers follow the logical progression from clinical problem to computational analysis, results, interpretation and future work.

APPENDIX K: QUALITY CHECKLIST FOR FUTURE WORK

Data source check

Confirm that disease-associated genes are downloaded from the intended database version and that search terms are documented.

Gene list check

Remove duplicate gene symbols and standardize nomenclature before intersection analysis.

Network construction check

Record NetworkAnalyst settings, selected databases and network filters so that the analysis can be reproduced.

Hub analysis check

Report degree centrality values and clarify why selected cutoffs or top-ranked nodes were chosen.

Biological interpretation check

Support each proposed biomarker with literature and avoid claiming clinical validity before experimental testing.

Validation check

Use independent cohorts and appropriate statistical measures such as sensitivity, specificity and ROC-AUC.

Reporting check

Clearly distinguish computational predictions from experimentally validated findings.

This checklist summarizes the quality-control logic that should accompany the proposed work in future studies. It helps ensure that the computational analysis can be repeated and that conclusions remain proportional to the evidence available.

Sejal MSc Thesis V2

ORIGINALITY REPORT

4%	2%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Delhi Technological University Student Paper	2%
2	Zhuoyuan Xin, Zhao Yang, Jianting Xu, Chaoying Li, Tong Shao, Guoqing Wang, Chong Li. "Gene expression profiling and construction of a putative gene regulatory network of bladder cancer tumor-initiating cells", Oncotarget, 2017 Publication	<1%
3	tudr.thapar.edu:8080 Internet Source	<1%
4	Hecker, Michael, Madhan Thamilarasan, Dirk Koczan, Ina Schröder, Kristin Flechtner, Sherry Freiesleben, Georg Füllen, Hans-Jürgen Thiesen, and Uwe Zettl. "MicroRNA Expression Changes during Interferon-Beta Treatment in the Peripheral Blood of Multiple Sclerosis Patients", International Journal of Molecular Sciences, 2013. Publication	<1%

5	newsmanager.commpartners.com Internet Source	<1%
6	geneglobe.qiagen.com Internet Source	<1%
7	www.mdpi.com Internet Source	<1%
8	worldwidescience.org Internet Source	<1%
9	www.coursehero.com Internet Source	<1%
10	Hsei-Wei Wang. "Pediatric primary central nervous system germ cell tumors of different prognosis groups show characteristic miRNome traits and chromosome copy number variations", BMC Genomics, 2010 Publication	<1%
11	Umesh Chaudhari, Lotta Pohjolainen, Heikki Ruskoaho, Virpi Talman. "Genome-wide profiling of miRNA-gene regulatory networks in mouse postnatal heart development—implications for cardiac regeneration", Frontiers in Cardiovascular Medicine, 2023 Publication	<1%
12	Min Li. "MicroRNAs: Control and Loss of Control in Human Physiology and Disease", World Journal of Surgery, 04/2009	<1%

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		