UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VALENCIA BIOMEDICAL RESEARCH FOUNDATION
CENTRO DE INVESTIGACIÓN PRÍNCIPE FELIPE

PhD in

Biotechnology

"Development and application of novel computational approaches for the characterization of cancer subtypes"

PhD Candidate:

Irene Pérez Díez

PhD Supervisors:

Francisco García García

María de la Iglesia Vayá

UPV Supervisor:

Máximo Ibo Galindo Orozco

April 2025

Development and application of novel computational approaches for the characterization of cancer subtypes

Keywords:

cancer subtypes, meta-analysis, transcriptomics, lung adenocarcinoma, pancreatic ductal adenocarcinoma, cancer heterogeneity, biomarkers, molecular profile, functional profile, prognosis

Abstract:

Cancer remains a global health crisis, demanding further research to understand its molecular basis. Even within the same cancer type, inter-patient variability is an obstacle to disease understanding and therapy development. Cancer subtyping, therefore, becomes essential to address cancer heterogeneity. One promising approach to unravel this heterogeneity is through cancer subtyping based on the transcriptomic landscape of patients. Through gene expression profiling, researchers can identify groups of patients that may represent distinct cancer subtypes with unique biological characteristics, response to therapies and clinical outcomes. However, this approach has its major challenge in requiring large sample sizes, which are crucial for the identification of meaningful subtypes.

This is where transcriptomics data meta-analysis emerges as a powerful statistical tool. By systematically collecting and re-analyzing data from public repositories, researchers can integrate sample sizes across multiple studies, overcoming the limitations of individual datasets. This approach allows for the identification of subtle yet critical differences in gene expression patterns that might be missed in smaller cohorts.

In this thesis, we explored cancer heterogeneity in two distinct contexts: lung adenocarcinoma and pancreatic ductal adenocarcinoma. We used an in-silico approach, leveraging published data through meta-analysis, overcoming limitations of individual studies, and driving novel discoveries. We have identified sex-specific transcriptomic differences in lung adenocarcinoma, particularly in the immune sys-

tem, purinergic signaling, and lipid metabolism pathways. We have also characterized the transcriptomic landscape of pancreatic ductal adenocarcinoma and its links to patient survival, revealing two prognostic gene signatures associated with the immune system and the extracellular matrix.

Desenvolupament i aplicació de nous enfocaments computacionals per a la caracterització de subtipus de càncer

Paraules clau:

subtipus de càncer, metaanàlisi, transcriptòmica, adenocarcinoma de pulmó, adenocarcinoma ductal pancreàtic, heterogeneïtat del càncer, biomarcadors, perfil molecular, perfil funcional, pronòstic

Resum:

El càncer continua sent una crisi sanitària a nivell mundial, i es requereix més recerca per a comprendre les seves bases moleculars. Fins i tot dins d'un mateix tipus de càncer, la variabi-litat entre pacients és un obstacle per a la comprensió de la malaltia i el desenvolupament de teràpies. Per tant, la subtipificació dels diferents tipus de càncer resulta essencial per a abordar la seva heterogeneïtat. Un enfocament prometedor per a desentranyar aquesta heterogeneïtat és la subtipificació del càncer basat en el context transcriptòmic dels pacients. Mitjançant perfils d'expressió gènica, els investigadors poden identificar grups de pacients que poden representar diferents subtipus de càncer amb característiques biològiques, resposta a teràpies i resultats clínics únics. No obstant això, aquest enfocament representa un gran repte, ja que requereix una gran grandària mostral, crucial per a la identificació de subtipus significatius.

És ací on el metaanàlisi de dades transcriptòmiques emergeix com una potent eina estadística. Mitjançant la recopilació sistemàtica i la reanàlisi de dades de repositoris públics, els investigadors poden integrar grandàries mostrals de múltiples estudis, superant les limitacions dels conjunts de dades individuals. Aquest enfocament permet identificar diferències subtils però crítiques en els patrons d'expressió gènica que podrien passar desapercebudes en cohorts més petites.

En aquesta tesi, explorem l'heterogeneïtat del càncer en dos contextos diferents: l'adenocar-cinoma de pulmó i l'adenocarcinoma ductal pancreàtic. Hem utilitzat un enfocament in-silico, aprofitant les dades publicades mitjançant una estratègia de

metaanàlisi, superant les limitacions dels estudis individuals i impulsant nous descobriments. Hem identificat diferències transcriptòmiques específiques del sexe en l'adenocarcinoma de pulmó, en particular en les rutes del sistema immunitari, la senyalització purinèrgica i el metabolisme lipídic. També hem caracteritzat el paisatge transcriptómico de l'adenocarcinoma ductal pancreàtic i els seus vincles amb la supervivència dels pacients, revelant dues signatures genètiques pronòstiques associades amb el sistema immunitari i la matriu extracelul·lar.

Desarrollo y aplicación de nuevos enfoques computacionales para la caracterización de subtipos de cáncer

Palabras clave:

subtipos de cáncer, metaanálisis, transcriptómica, adenocarcinoma de pulmón, adenocarcinoma ductal pancreático, heterogeneidad del cáncer, biomarcadores, perfil molecular, perfil funcional, pronóstico

Resumen:

El cáncer sigue siendo una crisis sanitaria a nivel mundial, y se requiere más investigación para comprender sus bases moleculares. Incluso dentro de un mismo tipo de cáncer, la variabilidad entre pacientes es un obstáculo para la comprensión de la enfermedad y el desarrollo de terapias. Por tanto, la subtipificación de los distintos tipos de cáncer resulta esencial para abordar su he-terogeneidad. Un enfoque prometedor para desentrañar esta heterogeneidad es la subtipificación del cáncer basado en el contexto transcriptómico de los pacientes. Mediante perfiles de expresión génica, los investigadores pueden identificar grupos de pacientes que pueden representar distintos subtipos de cáncer con características biológicas, respuesta a terapias y resultados clínicos únicos. Sin embargo, este enfoque representa un gran reto, ya que requiere un gran tamaño muestral, crucial para la identificación de subtipos significativos.

Es aquí donde el metaanálisis de datos transcriptómicos emerge como una potente herra-mienta estadística. Mediante la recopilación sistemática y el reanálisis de datos de repositorios públicos, los investigadores pueden integrar tamaños muestrales de múltiples estudios, superando las limitaciones de los conjuntos de datos individuales. Este enfoque permite identificar diferencias sutiles pero críticas en los patrones de expresión génica que podrían pasar desapercibidas en cohortes más pequeñas.

En esta tesis, exploramos la heterogeneidad del cáncer en dos contextos distintos: el adenocarcinoma de pulmón y el adenocarcinoma ductal pancreático. Hemos

utilizado un enfoque in-silico, aprovechando los datos publicados mediante una estrategia de metaanálisis, superando las limitaciones de los estudios individuales e impulsando nuevos descubrimientos. Hemos identificado diferencias transcriptómicas específicas del sexo en el adenocarcinoma de pulmón, en particular en las rutas del sistema inmunitario, la señalización purinérgica y el metabolismo lipídico. También hemos caracterizado el paisaje transcriptómico del adenocarcinoma ductal pancreático y sus vínculos con la supervivencia de los pacientes, revelando dos firmas genéticas pronósticas asociadas con el sistema inmunitario y la matriz extracelular.

"Magicians and scientists are, on the face of it, poles apart. Certainly, a group of people who often dress strangely, live in a world of their own, speak a specialized language and frequently make statements that appear to be in flagrant breach of common sense have nothing in common with a group of people who often dress strangely, speak a specialized language, live in ... er ..."

Terry Pratchett

The Science of Discworld, 1999

Contents

List of Figures

List of Tables

Glossary

A | B | C | E | F | G | H | K | L | M | N | O | P | R | S | T | Z

A

Antigen-presenting Cell A type of immune cell that boosts immune responses by showing antigens on its surface to other cells of the immune syste 16, 20

ArrayExpress ArrayExpress is a database of functional genomics experiments 6

B

Binary Logarithm of Fold Change Binary logarithm of a measure of the magnitude of transformation between initial and final values. 30, 32, 34

BP Biological Process 15–17, 25

C

CC Cellular Component 15, 25

Cluster of Differentiation Protocol used for the identification and investigation of cell surface molecules providing targets for immunophenotyping of cells. 38, 42

Confidence Interval The confidence interval estimates the interval of probable values of the studied population 16, 18

E

Extracellular Matrix A large network of proteins and other molecules that surround, support, and give structure to cells and tissues in the body 28, 32, 34, 35, 39–41

F

FAIR principles The FAIR principles are a set of guidelines intended to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets 5, 6, 18

False Discovery Rate False discovery rate is a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons 16, 25, 30, 32, 34

G

Gene Expression Omnibus The Gene Expression Omnibus is a public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomic data submitted by the scientific community 5, 6, 10, 20, 22, 28, 29, 31, 32

Gene Ontology The Gene Ontology is a knowledgbase that provides a framework and set of concepts for describing the functions of gene products from all organisms 5, 14–17, 21, 25, 31, 34, 37, 38

Genomic Data Commons Genomic Data Commons is a database created to promote precision medicine in oncology. It supports the import and standardization of genomic

and clinical data from cancer research programs. 6

GSEA Gene Set Enrichment Analysis 19, 25

H

Human Leukocyte Antigens A type of molecule found on the surface of most cells in the body. HLAs play an important part in the body's immune response to foreign substances. 38, 41, 42

K

Kyoto Encyclopedia of Genes and Genomes The Kyoto Encyclopedia of Genes and Genomes is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances, conceived as a computer representation of the biological system. 5, 14–17, 25

L

Log Odds Ratio Natural logarithm of the Odds Ratio, a statistic that quantifies the strength of the association between two events. 16–18, 25

LUAD Lung Adenocarcinoma 2, 3, 7

M

MF Molecular Function 15, 16, 25

N

NSCLC Non-Small Cell Lung Carcinoma 2, 10, 18, 20–23

O

Over-Representation Analysis Statistical method that determines whether genes from pre-defined sets are present more than would be expected in a subset of data. 30, 31, 34, 37, 38, 40

P

PCA Principal Component Analysis 11, 23, 24, 30

PDAC Pancreatic Ductal Adenocarcinoma 3, 7, 27–29, 31, 32, 34, 35, 39–43

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses 22, 29, 33

R

Reactome Reactome is an open-source, open access, manually curated and peer-reviewed biological pathway database. 5

RNA sequencing RNA-Seq is an RNA analysis technique based on Next Generation Sequencing that allows the identification, measurement, and comparison of gene expression in a target transcriptome. 4, 23, 29–31

S

Standard Error The standard error is a measure of the dispersion of the sampling distribution 16, 25

T

The Cancer Genome Atlas The Cancer Genome Atlas is a comprehensive, collaborative effort led by the National Institutes of Health to map the genomic changes associated with specific types of tumors to improve the prevention, diagnosis and treatment of cancer 6, 10, 11, 19, 29, 31, 38, 42

TMM Trimmed Mean of M-values 23

Tumor Microenvironment The normal cells, molecules, and blood vessels that surround and feed a tumor cell. A tumor can change its microenvironment, and the microenvironment can affect how a tumor grows and spreads 3

Z

Z-Score The number of standard deviations by which the value of a raw score is above or below the mean value of what is being observed or measured. 31

Chapter 1

Introduction

1.1 Overview of cancer subtyping

Cancer is a complex and heterogeneous disease comprising a wide range of malignancies characterized by different biological properties and clinical behaviors. Globally, cancer is a leading cause of premature death [1], with incidence and mortality rates varying across different cancer types [1, 2]. Even within the same cancer type, cancer subtypes can exhibit different origin mechanisms and disease progression, significantly impacting prognosis and survival. Cancer subtyping plays, therefore, a fundamental role in categorizing tumors into subgroups based on specific molecular, genetic, or clinical features. This approach is essential for understanding the underlying mechanisms of tumorigenesis and the variations in treatment response and patient outcomes.

Cancer subtypes are distinct subclasses or clusters within a specific cancer type characterized by unique molecular or clinical signatures. Historically, several strategies have been employed to classify and subtype cancers, mainly:

- Histological and Morphological analysis
- Tumor location and origin
- Molecular and Genetic Markers
- Gene Expression Profiling
- Immunophenotyping
- Clinical behavior and response to treatment
- Staging systems

High-throughput technologies can be used to subtype based on gene expression profiling, performing comprehensive gene expression analysis of tumors [3]. It enables the identification of differentially expressed genes that distinguish subtypes, shedding light on dysregulated biological processes and molecular pathways. Parker and colleagues [4] developed the PAM50 assay, one of the earliest applications of transcriptomics for classifying cancer subtypes based on gene expression.

The authors used the microarray technology to measure the expression of 50 genes and successfully classified breast cancer into four different subtypes: luminal A, luminal B, HER2-enriched, and basal-like types. The PAM50 assay paved the way for further gene expression-based tests, highlighting the potential of transcriptomics for improved breast cancer subtyping.

Cancer subtyping offers several advantages in clinical practice and research, enhancing our understanding of disease biology. Following the PAM50 example, by classifying breast cancer into different intrinsic subtypes, practitioners could better assess the risk of recurrence and survival rates for patients [4]. Diagnosis by subtypes added significant prognostic and predictive information for patients and helped assess the likelihood of efficacy from neoadjuvant chemotherapy. Consequently, PAM50 became a foundational tool for breast cancer subtyping, providing valuable prognostic and predictive information to guide treatment decisions and improve patient outcomes. Furthermore, understanding the biology of each PAM50 subtype allowed researchers to develop novel targeted therapies for specific subtypes. Thus, it is crucial in developing personalized medicine approaches, as it allows the identification of biomarkers that can predict treatment responses and guide therapeutic decisions [5, 6]. It can also lead to the discovery of novel therapeutic targets by identifying subtype-specific vulnerabilities and dependencies on specific pathways [7].

1.1.1 Sex as a biological variable in cancer subtyping

Historically, medical research has been centered on male physiology, with few studies considering this variable in their experimental design and analyses [8, 9]. Sex is an essential modifier of disease via genetic, epigenetic, and hormonal regulations [10], and cancer is not an exception. Generally speaking, non-reproductive cancers are more frequently developed by males, who also exhibit shorter survival times, even after adjusting for risk factors such as smoking or dietary habits. Sex differences in cancer are thought to be related to genetic differences, the incomplete inactivation of the X chromosome in female individuals, the presence of Y

chromosome-encoded oncogenes, and the chromatin remodeling effects of in-utero testicular testosterone in male cells [8]. These mechanisms influence metabolism, growth regulation, angiogenesis, and immunity, which are hallmarks of cancer [11].

For instance, significant sex differences in pathways related to glycolysis, fatty acid, and bile acid metabolism have been found in some non-reproductive cancers [12]. Glucose intake, an essential part of cellular proliferative growth, has a different impact on men and women, with practices such as intermittent fasting correlated with an increased incidence of liver [13] and colon [14] cancer in males but not in females. Linked to lipid metabolism, males with obesity have disproportionately high rates of colon and hematological cancers [15]. The excess of adipose tissue is frequently associated with chronic low-grade inflammation, which can promote cancer by driving DNA damage. In contrast, acute inflammatory responses can be positive against cancer, with females having a more significant benefit from this effect overall [12].

Furthermore, sex can also influence response to treatment. Sex-specific analysis to perform molecular subtyping has been proposed to help tailor treatment to patients with glioblastomas and colon cancer [16, 17]. Integrating sex into the treatment research design will positively impact other cancer types and is especially promising for discovering novel cancer immunotherapies [18]. Sex perspective should be included in further cancer research, as cancer studies are enhanced by sex-specific targeting.

1.2 Lung adenocarcinoma

Lung cancer represents a significant public health challenge, accounting for the highest cancer incidence and mortality worldwide [2]. Lung adenocarcinoma (LUAD), a subtype of non-small cell lung carcinoma (NSCLC), accounts for more than 50% of all lung cancer diagnoses, with an increasing frequency over time [19, 20]. It shows a five-year survival rate of 26.3%, although this rate fluctuates, influenced by race, sex, and tumor stage [21]. Furthermore, it is the most common subtype diagnosed

in people who have never smoked [22], with incidence rates in non-smokers of 14.4 to 20.8 per 100,000 person-year in women and 4.8 to 13.7 per 100,000 person-year in men [23].

LUAD usually evolves from the mucosal glands, and it is characterized by complex genomic landscape, with numerous underlying genetic and epigenetic mechanisms, characterizes LUAD. While mutations in TP53 and LRP1B genes are common in NSCLC, the major disrupted signaling pathways in LUAD are RAS-MEK-ERK and PIK3CA-MTOR, involving high rates of mutations in KRAS, EGFR, MET, and BRAF genes. These are clinically relevant, as they are potentially targetable [24].

Several studies have studied the epidemiological differences in LUAD between males and females. Females exhibit a higher predominance of LUAD and higher survival rates [20, 25, 26]. Furthermore, LUAD shows a more pronounced survival rate difference between male and female lung adenocarcinoma patients when compared to other tumor types [25]. Overall, the evidence suggests that sex has a significant influence on LUAD.

1.3 Pancreatic ductal adenocarcinoma

Pancreatic cancer originates from any cellular types that compose the pancreatic gland. It can affect any organ region, although it frequently originates from the exocrine component's ducts and is predominantly observed in the pancreatic head. Pancreatic Ductal Adenocarcinoma (PDAC) is the prevailing subtype of pancreatic cancer, accounting for over 80% of all diagnosed pancreatic neoplasms [2]. PDAC is characterized by aggressive behavior, late diagnosis, and limited treatment options [27]. This subtype has the lowest cancer survival rate (12%), and it is has become the third leading cause of cancer-related deaths by 2023 [2].

PDAC arises from a complex interplay of ambient factors and genetic and epigenetic alterations contributing to disease progression. Multiple key signaling pathways are implicated in PDAC development and progression, including KRAS, TP53, CDKN2A, and SMAD4 genes. These alterations disrupt cellular homeostasis, pro-

mote tumor growth, and confer resistance to conventional therapies [28]. Furthermore, PDAC is characterized by an extensive desmoplastic reaction involving activating stromal cells, extracellular matrix deposition, and altered tumor microenvironment (TME). This fibrotic response creates a barrier to drug delivery and contributes to the disease's aggressive nature [29, 30].

1.4 Transcriptomics

Transcriptomics holds immense significance in modern biology and bioinformatics. Genomics studies the static genetic composition of an organism, while transcriptomics attempts to decode gene expression and delves into the dynamic spectrum of messenger RNA molecules, known as the transcriptome. This transcriptome forms a bridge between genotype and phenotype, which makes it a crucial intermediary between the information encoded in the genes and the functionality of the proteins.

Transcriptomics techniques allow researchers to study gene expression patterns, revealing which genes are actively transcribed and how their expression levels change in response to environmental stimuli, developmental stages, and other factors such as pathological states. This ability to monitor gene expression has transformed our understanding of the molecular mechanisms underlying biological phenomena.

1.4.1 High throughput technologies

Since the late 1990s, a series of technological innovations have transformed transcriptomics and made it a widespread discipline. The contemporary transcriptomics workhorse is RNA sequencing (RNA-seq), which uses high-throughput sequencing to capture the whole transcriptome of an organism. However, despite their decline since 2014, microarrays have been the main driving force in transcriptomics for most of its history [31]. They make up a substantial portion of the data available in public repositories for research use.

Microarrays were designed as a set of probes (short nucleotide oligomers) arrayed on a solid substrate. Transcript abundance is measured by fluorescence intensity based on the hybridization of the transcript to the probes, as each transcript is fluorescently labeled [32]. This process allows researchers to perform relative abundance analyses that have driven the discovery of therapeutic targets [33], disease characterization [34, 35] or the development of prognostic and diagnostic classifiers [36, 37].

In contrast to microarrays, RNA-seq captures and quantifies the transcripts present in an RNA extract and allows researchers to explore the whole transcriptome rather than relying on a limited number of predefined probes. The generated nucleotide sequences are aligned to a reference genome and converted into read counts that can be used to model gene expression levels accurately. Just as microarrays, RNA-seq technology has been widely applied in disease research, improving our knowledge [38] and enabling the identification of novel therapeutic targets [39] and the prediction of patient responses to treatment [40].

1.4.2 Analysis strategies

Since different transcriptomic datasets have unique properties, the preprocessing of their data involves distinct methods tailored to the research objective and the technology used. Both technologies, microarray, and RNA-seq, will provide a matrix with the quantification of gene expression after passing through quality control and reads or probes annotation stages. This so-called counts matrix is the starting point of transcriptomic analyses. Once the transcript counts are available, the biological questions to be answered by the research will drive the analysis to be conducted.

Differential expression analysis is the quintessential analysis in transcriptomics. Counts are normalized, modeled, and statistically analyzed to study differential gene expression, resulting in pair-wise tests between the compared groups and the probability estimates for the computed differences. This approach has proven to be helpful in a broad range of fields, including biomarker detection [41] and disease characterization [42].

Another essential analysis in the transcriptomics toolset is survival analysis, in which gene expression data is integrated with clinical information to identify potential prognostic and predictive biomarkers. Researchers can detect gene signatures associated with longer or shorter survival times by analyzing gene expression profiles across patient groups with different survival outcomes [41]. This knowledge can be used to develop new diagnostic assays, treatment strategies, and personalized medicine approaches.

These methodologies provide researchers with lists of genes or transcripts of interest, generally ranked according to statistical criteria. These lists can enrich their value if combined with functional annotations from biological databases, such as The Gene Ontology (GO), the Kyoto Encyclopedia of Genes and Genomes (KEGG), or Reactome. Functional enrichment analysis is the method that uses statistical approaches to identify significantly enriched or depleted groups of genes associated with different phenotypes in order better to understand the underlying biological processes or signaling pathways.

1.4.3 Data repositories

Vast amounts of transcriptomics data are generated yearly, and sharing them with the scientific community in an open, systematic and standardized way is crucial for advancing research. By making data publicly available, researchers can build upon previous work, identify patterns otherwise unnoticed, and accelerate discovery. Transcriptomics data repositories play a pivotal role in facilitating this sharing, adhering to quality control and data management standards, and ensuring the application of the FAIR principles [43].

Gene Expression Omnibus (GEO) was developed by the National Center for Biotechnology Information in 2002 [44]. Since then, it has been the leading public repository of high-throughput gene expression data [45], allowing researchers to access raw and processed data with experimental descriptions and sample metadata. Tens of thousands of publications[1] have reused and reanalyzed data published in this

[1] https://www.ncbi.nlm.nih.gov/geo/info/citations.html

repository to advance science.

The European counterpart, developed in 2003 by the European Bioinformatics Institute, is ArrayExpress [46]. This repository provides added value to researchers as a team of data curators ensures the quality and reliability of all submitted data. Furthermore, ArrayExpress acts as a mirror of a subset of curated GEO datasets, improving data sharing and interoperability.

In cancer research, the data repository by excellence is Genomic Data Commons (GDC), launched in 2016 by the National Cancer Institute. GDC integrates and harmonizes genomic, transcriptomic, and clinical data from multiple research projects, including The Cancer Genome Atlas (TCGA) [47]. This project started in 2006 as a collaboration between the National Cancer Institute and the National Human Genome Research Institute, collecting data from more than 126000 patients, spanning 33 different types of cancer.

1.4.4 Meta-analysis

Systematic reviews and meta-analyses, when combined, provide a comprehensive and reliable understanding of a specific research question. The first step of the process is the systematic review, a method for comprehensively gathering and critically analyzing all relevant research on a specific topic. It is based on systematically searching different databases and screening studies to critically assess their quality and methodology [48, 49]. These studies are screened and evaluated based on predefined inclusion and exclusion criteria. Lastly, a meta-analysis summarizes and analyzes the data retrieved from the selected studies. Meta-analysis is a statistical methodology that allows researchers to combine the results of multiple studies addressing the same research question. The result is a more precise and reliable estimate of a given phenomenon than single studies [48]. Meta-analysis relies on the quality of the underlying studies. Therefore, the systematic review paves the way by comprehensively identifying and evaluating studies, while the meta-analysis quantitatively synthesizes their work.

As discussed in the previous section, several scientific organizations have cre-

ated repositories to systematically store transcriptomic studies and make them available to the scientific community, constituting a relevant source of information for transcriptomic research. Even though the cost of high-throughput technologies has significantly decreased with time, this has been a restriction for generalizing its use. Therefore, a considerable number of studies include an adjusted or reduced sample size, which may limit their detection power. Meta-analysis represents a valuable approach to fill this gap, integrating published data and significantly improving the detection power in transcriptomics research.

Chapter 2

Justification and Objectives

2.1 Justification

Cancer remains one of the leading causes of mortality worldwide [2]. Further research in this area is required, as the development of effective treatment benefits from understanding underlying molecular mechanisms and identifying cancer variability between patients. The field of bioinformatics has made remarkable progress in analyzing large-scale genomic data to characterize cancer subtypes and explore potential biomarkers. Many datasets have been generated and, due to the increasing application of FAIR principles to publication requirements [43], are made available to other researchers online. However, despite these advancements, challenges still need to be addressed, particularly in integrating and interpreting diverse datasets from multiple sources.

The small sample size of most of the experiments and their confinement to a specific scenario represent limiting factors in the evaluation of these transcriptomic studies. By reanalyzing data from different sources, the experiments proposed in this work can overcome sample size limitations, increase statistical power, and identify biomarkers that may have been overlooked in individual studies. This approach not only complements the findings of the original data contributors but also provides a fresh perspective into methodologies for cancer subtyping and biomarker discovery.

Lung adenocarcinoma and pancreatic ductal adenocarcinoma were consensually chosen with our collaborators to illustrate the methodology's broad applicability. This decision was based on the clinical burden these diseases represent. Both LUAD and PDAC are characterized by inter-patient heterogeneity, thus being relevant to cancer subtyping strategies. Additionally, these two cancer types offer contrasting challenges: LUAD shows sex disparities in its incidence and survival rates, while PDAC is characterized by a profoundly immunosuppressive microenvironment. By applying the proposed approach in these distinct contexts, we wanted to prove its versatility and robustness in deciphering intricate transcriptomic patterns and yielding clinically relevant insights despite being based on already published data.

Overall, the proposed computational approach overcomes the challenges of integrating data from different sources by using meta-analysis techniques to integrate datasets on a results level rather than on a raw data level. Moreover, this research encourages to further extend the use of generated data. This utilization of published datasets holds significant value, especially when considering the difficulties associated with data collection and data scarcity in certain contexts. This approach empowers scientists with technical expertise but limited data access to contribute significantly to the field and expands research possibilities. Thus, this thesis offers a framework to propel the in-depth understanding of cancer subtypes, even when primary data collection may be challenging or limited.

2.2 Objectives

The main objective was to develop and implement computational approaches to characterize cancer subtypes.

The specific objectives were addressed in the specific scientific publications and were:

- To assess the potential of combining published data to give researchers a deeper understanding of cancer subtypes
- To characterize the transcriptomic functional differences in LUAD between men and women
- To study the transcriptomic landscape of PDAC
- To identify PDAC subtypes and potential biomarkers based on patient survival

These objectives will contribute to advancing the understanding of cancer heterogeneity and the identification of potential personalized treatment strategies.

Chapter 3

Identification of sex-based functional signatures in lung cancer

3.1 Overview

Differences in epidemiological and clinical patterns of lung adenocarcinoma have been described between male and female patients. Our research aimed to assess the molecular mechanisms underlying those differences through functional profiling and meta-analysis of lung adenocarcinoma expression datasets. In this chapter, we display the research work we performed in this regard, culminating in a scientific publication.

3.2 Reference and contribution of the candidate

Pérez-Díez, I.; Hidalgo, M.R.; Malmierca-Merlo, P.; Andreu, Z.; Romera-Giner, S.; Farràs, R.; de la Iglesia-Vayá, M.; Provencio, M.; Romero, A.; García-García, F. Functional Signatures in Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Sex-Based Differences in Transcriptomic Studies. Cancers 2021, 13, 143. DOI: 10.3390/cancers13010143. PMID: 33526761.

The candidate participated in study design, software development, formal analysis, investigation, data curation, data visualization and writing of the manuscript.

3.3 Functional Signatures in Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Sex-Based Differences in Transcriptomic Studies

3.3.1 Introduction

Lung cancer is the most frequently diagnosed cancer and the leading cause of cancer-related death worldwide, representing 18.4% of all cancer deaths [19]. Exposure to

tobacco, domestic biomass fuels, asbestos, and radon represent the most relevant lung cancer risk factors [19, 20]; however, as has become evident from studies of other cancer types, sex-based differences (sexual dimorphisms) may also have significant relevance in lung cancer [19, 50, 51]. Lung cancer exhibits sex-based disparities in clinical characteristics and outcomes, with better survival observed in women [20, 25, 52]. While lung cancer incidence worldwide is higher in men, there exists an increasing trend in women that cannot be solely explained by tobacco consumption [19, 53]. Furthermore, studies have reported sex-dependent differences in estrogen receptors and their impact on lung cancer [54–56]; however, conflicting results have attributed lung cancer susceptibility in women to genetic variants, hormonal factors, molecular abnormalities, and oncogenic viruses [20, 26, 57, 58]. Adenocarcinoma represents the most frequent non-small cell lung cancer (NSCLC) subtype in both sexes [59], with a higher predominance in women compared to men (41% of cases in women versus 34% in men) [20, 25, 26]. Interestingly, Wheatley-Price *et al.* demonstrated a more pronounced survival rate difference between male and female lung adenocarcinoma patients when compared to other tumor types [25]. The molecular causes underlying such sex-biased patterns remain largely unknown, as limited efforts have been made for lung adenocarcinoma, with few studies considering this differential perspective [60–63]. These limitations can be partially addressed through meta-analysis, a robust methodology that combines information from related but independent studies to derive results with increased statistical power and precision [64, 65]. As current treatment strategies do not cure most lung cancer patients, and invasive diagnostic techniques (e.g., via biopsies and bronchoscopies) often induce discomfort in patients, meta-analyses that improve our understanding of sex-specific molecular mechanisms in lung adenocarcinoma may facilitate the discovery of non-invasive prognostic and diagnostic biomarkers. We performed a meta-analysis based on functional profiles of transcriptomic studies to explore the molecular mechanisms underlying sex-based differences in early-stage lung adenocarcinoma. We carried out exhaustive review and selection steps to guarantee the homogeneity of the selected studies and the subsequent comparison and integration of the data in the meta-analysis with an appreciation of this strategy's specific lim-

itations. After the systematic review, we retrieved and analyzed nine studies from GEO [44] and TCGA [47], and then combined the results in a random-effects meta-analysis. This approach allowed the identification of functional alterations caused by lung adenocarcinoma in both male and female patients, comparing tumor samples with adjacent non-tumor tissue. In this study, we identified immune responses, purinergic signaling, and lipid metabolism as the main biological processes that display differences between male and female lung adenocarcinoma patients, with the acute immune response increased in female patients. Overall, our findings provide evidence that sex-based differences influence cancer biology and may impact response to treatment. Furthermore, underlying sex-based differences may contribute to the discovery of sex-specific prognostic and diagnostic biomarkers and the improvement of personalized therapies.

3.3.2 Results

We organized our findings into three sections: the first describes the studies evaluated and selected in the systematic review; the second section reports on the results of the bioinformatic analysis of each of these selected studies as follows: (i) exploratory analysis, (ii) differential expression, and (iii) functional characterization; while the third section presents the results of the differential functional profiling by sex.

3.3.2.1 Study Search and Selection

The systematic review identified 207 non-duplicated studies, of which 48.8% included both male and female patients (Figure A.1). We applied inclusion and exclusion criteria (see Figure 3.1) to select a set of homogeneous and comparable studies to ensure the reliability of the subsequent analyses. To ensure study homogeneity (and in the hope of contributing to the early diagnosis of disease), we focused on those studies of early-stage disease and selected nine transcriptomic studies for further analysis (Table 3.1). The selected studies represented a population of 1366 early-stage samples (369 controls and 997 cases), of which 44% were from men, and

56% from women (Figure 3.2), with a median age of 65.54 years old. Table 3.1, Figure 3.2, and Table A.1 contain further information regarding the selected studies and the clinicopathological characteristics of the study population.

Table 3.1 – Studies selected after the systematic review.

Study	Platform	Publication
GSE10072	Affymetrix Human Genome U133A Array	[66]
GSE19188	Affymetrix Human Genome U133 Plus 2.0 Array	[67]
GSE31210	Affymetrix Human Genome U133 Plus 2.0 Array	[33, 68]
GSE32863	Illumina HumanWG-6 v3.0 Expression BeadChip	[34]
GSE63459	Illumina HumanRef-8 v3.0 Expression BeadChip	[36]
GSE75037	Illumina HumanWG-6 v3.0 Expression BeadChip	[37]
GSE81089	Illumina HiSeq 2500	[69]
GSE87340	Illumina HiSeq 2000	[70]
TCGA	Illumina HiSeq 2000	[47]

3.3.2.2 Individual Analysis

As the normalized data derives from different platforms, we performed exploratory and processing steps for the data set to ensure the comparability and integration of subsequent analyses. The exploratory analysis found a lack of abnormal behavior except for three samples in the principal component analysis (PCA) and unsupervised clustering; therefore, we excluded the GSM47570 and GSM47578 samples in study GSE19188, and the GMS773784 sample in study GSE31210 from further analysis. The differential expression results for each study demonstrated a large number of differentially expressed genes when comparing female lung adenocarcinoma samples to female control samples and male adenocarcinoma samples to male control samples (see Table A.2). However, the evaluation of sex-based differences in lung adenocarcinoma patients provided a small number of significantly affected genes (see Table A.3), with no intersecting genes. We performed an individual functional enrichment analysis of GO terms and KEGG pathways to identify the possible implications of these sex-specific differentially-expressed genes in pathways relevant to lung adenocarcinoma. The identified pathways revealed a diversity of significant results among datasets, which we have summarized in Table 3.2. When analyzing intersections, UpSet plots (Figure 3.3, equivalent to a Venn diagram) illustrate the

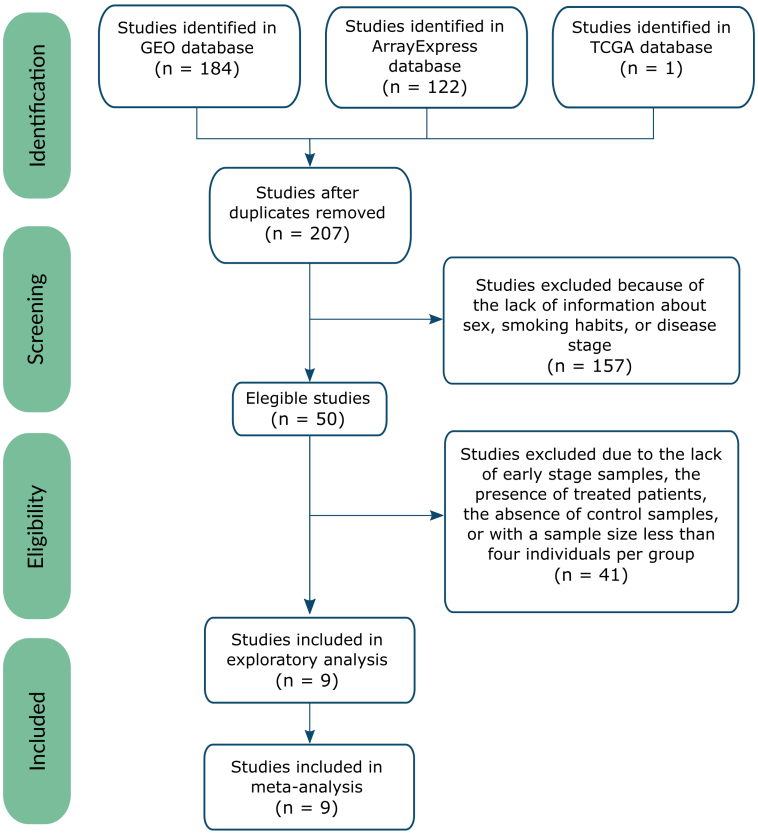


Figure 3.1 – Flow of information through the different phases of the systematic review, following PRISMA Statement guidelines [48].

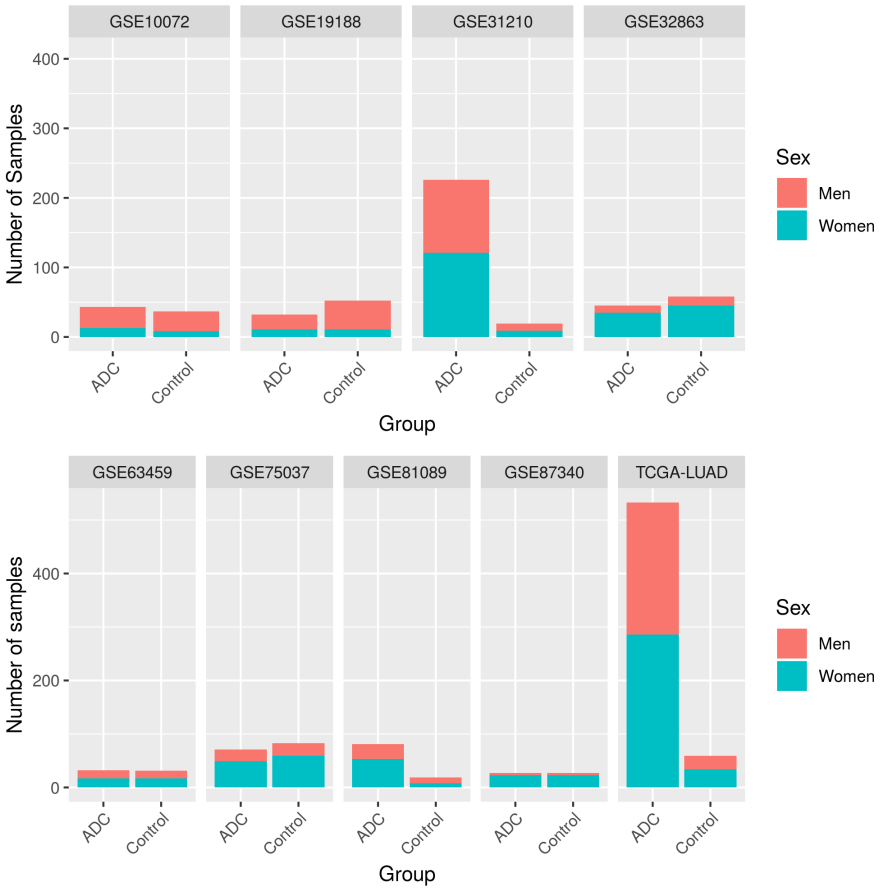


Figure 3.2 – Number of samples per study, divided by sex and experimental group (ADC: lung adenocarcinoma samples).

degree of intersection between studies, demonstrating that most significant results are exclusive of each study. This data highlights the need for integrated strategies, such as meta-analyses, to increase the statistical power of any findings.

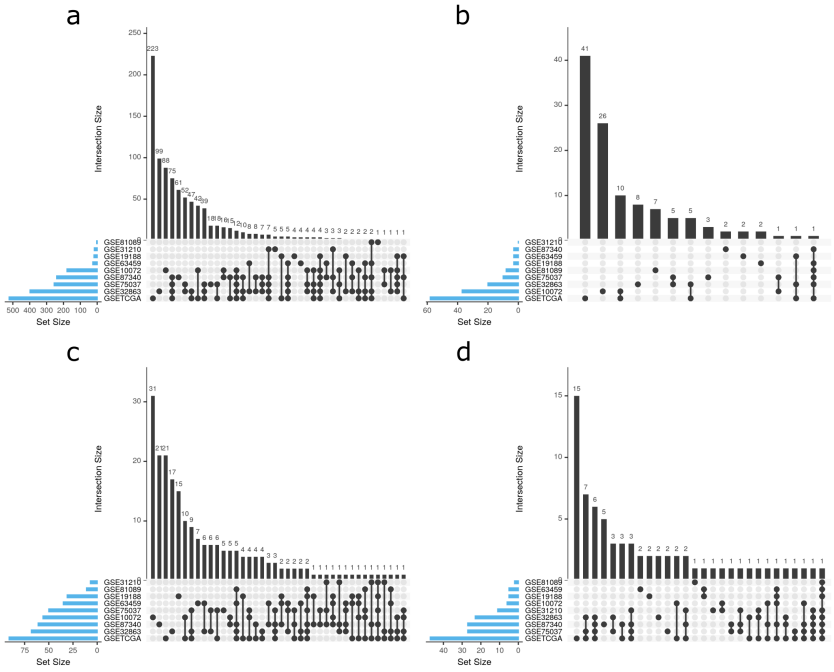


Figure 3.3 – The intersection of significant functions between studies. UpSet plots for (a) Gene Ontology (GO) biological process, (b) GO molecular functions, (c) GO cellular components, and (d) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. UpSet plots detailing the number of common elements among GO terms in our functional enrichment analysis. Horizontal bars indicate the number of significant elements in each study. The vertical bars indicate the common elements in the sets, indicated with dots under each bar. The single points represent the number of unique elements in each group.

Table 3.2 – Summary of functional enrichment analysis results by GO functions (BP: Biological Process, MF: Molecular Functions, CC: Cellular Component), and KEGG pathways. "Up" terms are overrepresented in female lung adenocarcinoma patients, while "Down" terms are overrepresented in male lung adenocarcinoma patients.

Study	Significant GO BP			Significant GO MF			Significant GO CC			Significant KEGG		
	Up	Down	Total	Up	Down	Total	Up	Down	Total	Up	Down	Total
GSE10072	26	153	179	29	8	37	16	40	56	1	5	6
GSE19188	7	12	19	0	3	3	11	20	31	1	4	5
GSE31210	21	2	23	0	0	0	4	0	4	8	2	10
GSE32863	428	51	479	28	5	33	40	41	81	27	6	33
GSE63459	0	26	26	0	3	3	8	27	35	1	4	5
GSE75037	245	35	280	14	4	18	14	21	35	22	5	27
GSE81089	2	1	3	7	1	8	7	4	11	1	1	2
GSE87340	178	62	240	3	0	3	48	13	61	26	1	27
TCGA	294	228	522	28	30	58	21	70	91	30	17	47

3.3.2.3 Meta-analysis

We performed a functional meta-analysis for each of the 8672 GO functions and KEGG pathways, including every term found in at least two studies. Results with a false discovery rate (FDR) of < 0.05 included 106 GO biological processes (BP), 3 GO molecular functions (MF), and 20 KEGG pathways, which were associated with 21 wider functional groups. We rejected potential bias on the significant results after the inspection of funnel plots; furthermore, sensitivity analyses failed to indicate alterations in the results due to the inclusion of any study. The results for the 129 significant GO terms and KEGG pathways are further detailed in Table A.4, including FDR, the log odds ratio LOR, and its 95% confidence interval CI, and the standard error SE of the LOR. In addition, Table A.5 details the number of functions involved in each of the genes of this functional signature, and Figure A.2 displays the prognostic description of this set of significant functions.

3.3.2.3.1 Upregulated Functions

We discovered that 43.88% of detected functions related to the immune response (Table A.4 and Figure 3.4), which all displayed upregulation in female lung adenocarcinoma patients. This finding agrees with studies that report more robust innate and adaptive immune responses in women who, for example, present with more efficient antigen-presenting cells (APC) than males [71, 72]. The results provided evidence for the positive regulation of an acute inflammatory response, with CD8+ alpha-beta T cell differentiation and activation, B cell proliferation and activation, and an increase of interleukin (IL) biosynthesis, including IL-2, 6, 8, 10, and 17. Several immune-related signaling pathways also displayed differences between female and male lung adenocarcinoma patients. We discovered the upregulation of the MyD88-independent toll-like receptor signaling pathway, NIK/NF-kappa β signaling pathway, FC-epsilon receptor signaling pathway, B-cell receptor signaling pathway, Toll-like receptor signaling pathway, NOD-like receptor signaling pathway, and RIG-I-like receptor signaling pathways in female lung adenocarcinoma patients. Overall, these findings suggest the relevance of sex-based differences in

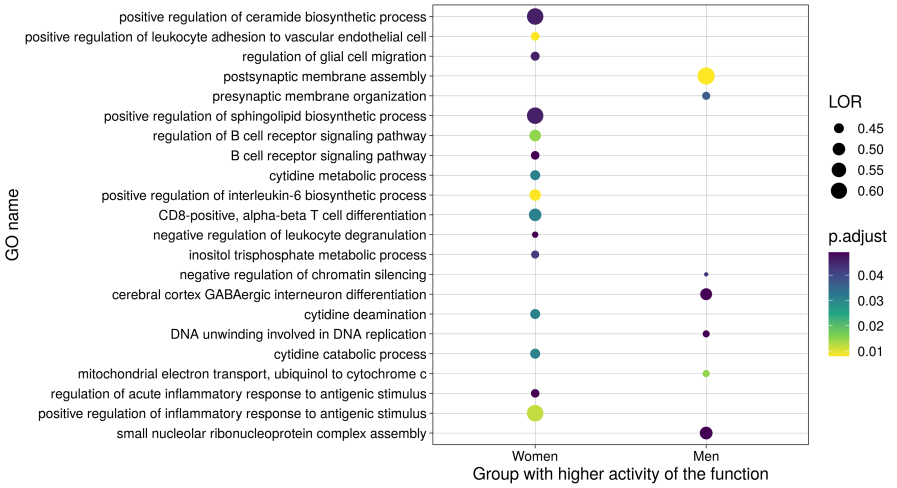


Figure 3.4 – Summary dot plot of GO BP meta-analysis results. Only those significant terms with a LOR over 0.4 are shown.

immune responses to lung cancer, which may represent a significant contributor to the marked differences observed during disease progression in male and female patients. Furthermore, such findings may help to define novel therapeutic targets and may have important implications for immunotherapy.

We also uncovered evident sex-based differences in cell metabolism. Studies have shown the significant upregulation of lipid metabolism (ceramide and sphingolipid biosynthetic processes) in females and the higher utilization of carbohydrates by males [73, 74]. Furthermore, lipid metabolism and signaling are widely accepted as major players in cancer biology [75]. "Metabolism- Nucleic acids metabolism and signaling" was the second most abundant functional group upregulated in female lung adenocarcinoma patients, comprising 8.63% of the altered functions. These GO terms and KEGG pathways are mainly related to purinergic signaling through G protein-coupled receptors and cytidine metabolism. Other functional groups upregulated in female lung adenocarcinoma patients include cell migration and homeostasis. Overall, further explorations of sex differences in metabolic pathways may provide new perspectives for treatment approaches with sex-specific effects.

3.3.2.3.2 Downregulated Functions

It should be noted that 23.26% of the significant functions exhibited lower activity in female compared to male lung adenocarcinoma patients. Downregulated functional groups include those related to cell cycle progression, cell junctions, DNA repair and telomere protection, mitochondrial processes, neural development, post-translational changes, post-transcriptional changes, protein degradation, and transcription regulation (Table A.4 and Figure 3.4). Taken together, the downregulated pathways in female patients suggest the existence of lower levels of oxidative stress compared to male patients [76], which may contribute to the existence of a less permissive tumorigenic environment.

3.3.2.4 Metafun-NSCLC Web Tool

The Metafun-NSCLC web tool (https://bioinfo.cipf.es/metafun-nsclc) contains information related to the nine studies and 1329 samples involved in this study. For each study, this resource includes fold-changes of genes and LOR of functions and pathways that users can explore to identify profiles of interest. We carried out a total of 8672 meta-analyses. For each of the 129 significant functions and pathways, Metafun-NSCLC depicts the global activation level for all studies and each study's specific contribution using statistical indicators (LOR, CI, and p-value) and graphical representations by function as forest and funnel plots. This open resource hopes to contribute to data sharing between researchers, the elaboration of innovative studies, and the discovery of new findings. Here, we also highlight the importance of including/reporting sex-related data in the results of clinical studies given their general importance in tumor risk, treatment response, and outcomes in lung adenocarcinoma and other cancers/disorders. The integration of sex-based differences in this manner has the potential to significantly impact the cancer biology field.

3.3.3 Discussion

Despite the profuse evidence for the influence of sex on rates and patterns of metastasis, the expression of prognostic biomarkers, and therapeutic responses in several cancer types [77, 78], sex-based differences have not been consistently considered when studying cancer, designing therapies, or constructing clinical trials. Cases of NSCLC, including adenocarcinoma, exhibit differences in incidence, prevalence, and severity in female and male patients [19, 20, 25, 79]. Elucidating the molecular basis for this sex-based differential impact will have clinical relevance, as this information can guide/improve both diagnosis and treatment.

Biomedical research generally underrepresents female patients, with sex-based differences rarely considered [80, 81]. Our systematic review of transcriptomic studies revealed that only 48.8% of lung adenocarcinoma-related datasets considered both sexes, a figure similar (49%) to that reported by Woitowich *et al.* [81]. Sex-based differences impact disease biomarkers, drug responses, and treatment [80], and, therefore, sex must represent a critical component of experimental design. Added to this problem, we faced a lack of standardization among studies and detailed clinical information (i.e., mutations, smoking status, stages) when searching for suitable datasets. The consideration of Findable, Accessible, Interoperable, and Reusable FAIR data principles [43], a requisite for quality science, would ensure that generated data can be of further use throughout the scientific community.

To the best of our knowledge, only four studies have attempted to address the functional alterations caused by lung adenocarcinoma in both male and female patients—those by Shi *et al.* [62], which considered female patients, Araujo *et al.* [60], Yuan *et al.* [61], and Li *et al.* [63]. Shi *et al.* [62] integrated samples from two datasets for a differential expression analysis followed by a functional enrichment analysis, whereas Araujo *et al.* [60] independently processed six datasets and jointly analyzed their results. Yuan *et al.* [61] compared male and female patients with various cancer types (including lung adenocarcinoma) using the TCGA dataset, but did not include control samples in the statistical comparison. Li *et al.* [63] evaluated the differences in lung adenocarcinoma by focusing only on metabolic pathways. By

including both male and female patients and controls in our gene expression comparison, in contrast with the analysis performed by Shi *et al.* [62] and Yuan *et al.* [61], we have effectively unveiled sex-based differences occurring in lung adenocarcinoma. Of note, 88% of the results reported by Yuan *et al.* [61] relate to the sex chromosome, but not necessarily due to cancer. Shi *et al.* [62] described the consequences of cancer development in female patients, but the authors failed to compare said effects with male patients. Although Araujo *et al.* [60] do not describe the statistical comparisons performed (and do not perform a statistical integration such as meta-analysis), the authors describe the results obtained for each dataset. Our study addressed sex-based differences in male and female lung adenocarcinoma patients through meta-analysis to address previous limitations and improve on those approaches employed by others. Despite certain supposed limitations to our approach (the presence of studies with different sample sizes and types of platforms), meta-analyses can integrate selected studies by eliminating inconsistency in individual studies, thereby increasing the statistical power, and highlighting robust disease-associated functions. We performed a meta-analysis using a random-effects model on Gene Set Enrichment Analysis (GSEA) results independently obtained from each study to evaluate the functions differentially altered between male and female lung adenocarcinoma patients. To make results comparable and reduce biases in the type of analysis used, we applied the same bioinformatics strategy from normalized expression matrices to GSEA results. This robust methodology integrates groups of data and provides results with higher statistical power and precision [64, 65] and reveals findings that cannot be obtained through the intersection or addition of results in individual studies. Nevertheless, we selected only samples from early-stage lung adenocarcinoma patients to reduce variability and included smoking status in the differential expression linear model. While the inclusion of the mutational status of relevant genes to lung adenocarcinoma (e.g., EGFR) could have revealed important insight, the lack of this information in the majority of the studies and the resultant limited sample size hampered this aim. We would support the inclusion of this type of data as a requirement for the publication of new data to repositories, such as the GEO.

The immune system plays a crucial role in the development of cancer [82], and several studies have reported sex-based differences in immune responses (reviewed by Klein and Flanagan [72]). Tumor cells evade the immune system using different strategies, including the modulation of antigen-presentation and the suppression of regulatory T cells. Therefore, sex differences in APCs and their downstream effector cells, among other components, may contribute to the sexual disparity observed in various aspects of cancer development and may significantly impact antitumor immunity and immunotherapy. Adult females generally present more robust innate and adaptive immune responses than males, as evidenced by increased phagocytic activity of neutrophils and macrophages, more efficient APCs, and differences in lymphocyte subsets (B cells, CD4+T cells, CD8+T cells) and cytokine production. Accordingly, our results demonstrate an enrichment of immune response-related terms in female lung adenocarcinoma patients, which agrees with the findings of Araujo *et al.* [60]. The analyzed functions suggest the positive regulation of CD8+ alpha-beta T cell activation and differentiation in female lung adenocarcinoma patients, which play an essential role in antitumor immunity [83, 84]. Furthermore, Ye *et al.*[84] discovered a more abundant population of effector memory CD8+ T cells in female lung adenocarcinoma patients, which agrees with our results. A previous study described CD8+ lymphocyte levels as a prognostic biomarker in NSCLC [85], and specifically in lung adenocarcinoma [86], with a correlation between higher levels of CD8+ lymphocytes with higher survival rates and lower disease recurrence. Elevated levels of active CD8+ T cells in female lung adenocarcinoma patients could form part of the molecular mechanism underlying higher survival rates when compared to male lung adenocarcinoma patients. Activation of the Notch signaling pathway decreases CD8+ T lymphocyte activity in lung adenocarcinoma [87]; therefore, the downregulation of the Notch signaling pathway discovered in female lung adenocarcinoma patients could explain higher CD8+ T activity when compared to male lung adenocarcinoma patients.

Concerning the immune response, we also detected differences that supported the increased production of IL-2, which is known to stimulate T cell proliferation and the production of effector T cells, thereby amplifying the lymphocytic response

[88]. Higher levels of IL-2 could also relate to increased activity of CD8+ T cells in female lung adenocarcinoma patients. Increased levels of IL-10 are also supported in female lung adenocarcinoma patients and, although IL-10 has anti-inflammatory and anti-immune activities [89, 90], studies have suggested a dual role in cancer. In advanced lung adenocarcinoma, high expression of IL-10 receptor 1 correlates with worse prognosis [90], while IL-10 expression by T-regulatory cells inhibits apoptosis through Programmed death-ligand 1 inhibition [89]. Nevertheless, IL-10 correlates with better prognosis when expressed by CD8+ T cells in early-stage NSCLC [91], and it seems to activate the antitumor control of CD8+ T cells [92]. IL-2 and IL-10 could display increased activity in early-stage female patients, alongside a higher population of active CD8+ T cells than males, conferring women a survival advantage.

We also detected the positive regulation of IL-6 biosynthesis in female lung adenocarcinoma patients, with increased IL-6 levels correlating with worse prognosis in NSCLC patients in previous studies [93, 94]. Network analysis in non-smoking female lung adenocarcinoma patients described IL-6 as one of the pathology's central nodes [62], and these findings agree with our results, which provide evidence of the critical role of IL-6 in tumor progression in female lung adenocarcinoma patients. IL-8 and IL-17 exhibit increased production and biosynthesis in female lung adenocarcinoma patients, with said interleukins known to influence tumor growth and metastasis and correlate with worse prognosis [95–97].

Although altered immune responses can positively and negatively influence tumor progression, our findings have detected GO terms that point to an elevated acute immune response in female compared to male lung adenocarcinoma patients. Of note, sex-based immunological differences in lung adenocarcinoma might have an impact on immunotherapy response. Different studies have addressed the role of sex in immunotherapy [50, 84, 98] and established improved survival for female NSCLC patients. The discovered molecular pathways differentially activated between male and female lung adenocarcinoma patients may underlie phenotypic differences regarding immunotherapy response. Sex-based differences in metabolism occur under physiological conditions and in the presence of cancer. Here, we de-

tected an upregulation of purinergic signaling and nucleic acid metabolism in female lung adenocarcinoma patients, a finding not described in previous lung cancer studies. An NSCLC-based study described an antitumor effect of the P2X4 receptor [99], which also exhibits sexual dimorphism in murine brain microglia [100]. Other P2 and A2 receptors play a role in NSCLC [99], but evidence of sex-based differences in receptor expression in human NSCLC patients has yet to be reported. Purinergic signaling and the role of purinergic receptors may also have relevance to innate and adaptive responses in different inflammatory and neurodegenerative diseases and several cancer types, including pancreatic ductal carcinoma, hepatocellular, hepatobiliary carcinoma cells, and breast cancer [101–105]. Of note, studies have linked the upregulation of purinergic signaling with poor prognosis in pancreatic ductal adenocarcinoma [104]. We also discovered significant differences in lipid metabolism, with the positive regulation of ceramide and sphingolipid biosynthetic processes upregulated in female lung adenocarcinoma patients. The presence of lipids can promote tumorigenesis, while higher adipose tissue levels are associated with poorer outcomes in several cancers [75]. Thus, exploring the differential roles of purinergic signaling and lipid metabolism between male and female lung adenocarcinoma patients may represent an interesting proposition to improve sex-specific risk-stratification of patients, prevention, diagnosis, and treatment. DNA damage and repair-related genes also presented sex-based differences in lung adenocarcinoma patients. In general, males present with a higher level of DNA damage, and females present with a lower DNA repair capacity [106], and, in agreement, we detected DNA repair and telomere protection as a downregulated functional group in female lung adenocarcinoma patients (with both mechanisms involved in tumor growth prevention [107]). Furthermore, we discovered the upregulation of DNA repair and an increase of DNA unwinding in male lung adenocarcinoma patients. DNA unwinding has emerged as a new target in cancer therapy with a primary focus on helicase inhibitors [108]. Besides, regarding DNA repair, the evaluation of poly (ADP-ribose) polymerase inhibitors in NSCLC cell lines has suggested potential therapeutic activity [109, 110], and there may be value in exploring both treatment approaches in NSCLC patients, especially male patients. With these results in mind,

we propose future studies focused on DNA repair and lipid/purinergic metabolism in female lung adenocarcinoma patients and the immune response in male lung adenocarcinoma patients in the hope of developing enhanced therapeutic strategies. Our study has characterized functional differences between the sexes in lung adenocarcinoma, shedding light on the functional basis behind this pathology in male and female patients. While our meta-analysis confirmed the conclusions of other studies, we also report previously undescribed alterations in biological processes that may broaden this field of study. Further knowledge regarding how those factors related to the functional mechanisms, described above, differentially impact male and female lung adenocarcinoma patients, and may improve our understanding of the disease and improve treatment and diagnosis through biomarker identification.

3.3.4 Materials and methods

Bioinformatics and statistical analysis employed R software v.3.5.3 [111]. Table A.6 details R packages and versions.

3.3.4.1 Study Search and Selection

Publicly available datasets were collected from GEO [44], ArrayExpress [112], and TCGA [47]. A systematic search of studies published in the period 2004–2018 was conducted in 2019 following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [48]. Two researchers involved in the study carried out the literature search, and the consistency of the review and selection procedures used was evaluated and confirmed. Several keywords were employed in the search, including lung adenocarcinoma, non-small-cell lung carcinoma, Homo sapiens, and excluding cell lines. Eleven variables were considered for each study, including the clinical characteristics of the patients (e.g., sex and smoking habit) and experimental design (e.g., sample size and sample extraction source). The final inclusion criteria were:

- Sex, disease stage, and smoking habit variables registered;

- RNA extracted directly from human lung biopsies;
- Both normal and lung adenocarcinoma samples available;
- Patients who had not undergone treatment before biopsy;
- Sample size of > 3 for case and control groups in both sexes.

Finally, normalized gene expression data of six array NSCLC datasets (GSE10072, GSE19188, GSE31210, GSE32863, GSE63459, and GSE75037) and counts matrix of three RNA-seq NSCLC datasets (GSE81089, GSE87340, and TCGA-LUAD) were retrieved.

3.3.4.2 Individual Transcriptomics Analysis

Individual transcriptomics analysis consisted of three steps: pre-processing, differential expression analysis, and functional enrichment analysis Figure 3.5.

Data pre-processing included the standardization of the nomenclature of the clinical variables included in each study, normalization of RNA-seq counts matrix, and exploratory analysis. RNA-seq counts were pre-processed with the **edgeR** [113] R package using the trimmed mean of m-values (TMM) method [114]. We assessed the normalization methods performed by the original authors for each dataset, and log2 transformed the matrices when necessary. Annotation from probe set to Entrez identifiers from the National Center for Biotechnology Information [115] database and gene symbol was carried out with the **biomaRt** [116] R package. When dealing with duplicated probe-to-Entrez mappings, the median of their expression values was calculated. The exploratory analysis included unsupervised clustering and PCA to detect patterns of expression between samples and genes and the presence of batch effects in each study Figure 3.5. Differential expression analyses were performed using the **limma** [117] R package. To detect differentially expressed genes in male and female lung adenocarcinoma patients, the following contrast was applied:

$$(ADC.W - Control.W) - (ADC.M - Control.M)$$

where ADC.W, Control.W, ADC.M and Control.M correspond to lung adenocarcinoma affected women, control women, lung adenocarcinoma affected men, and

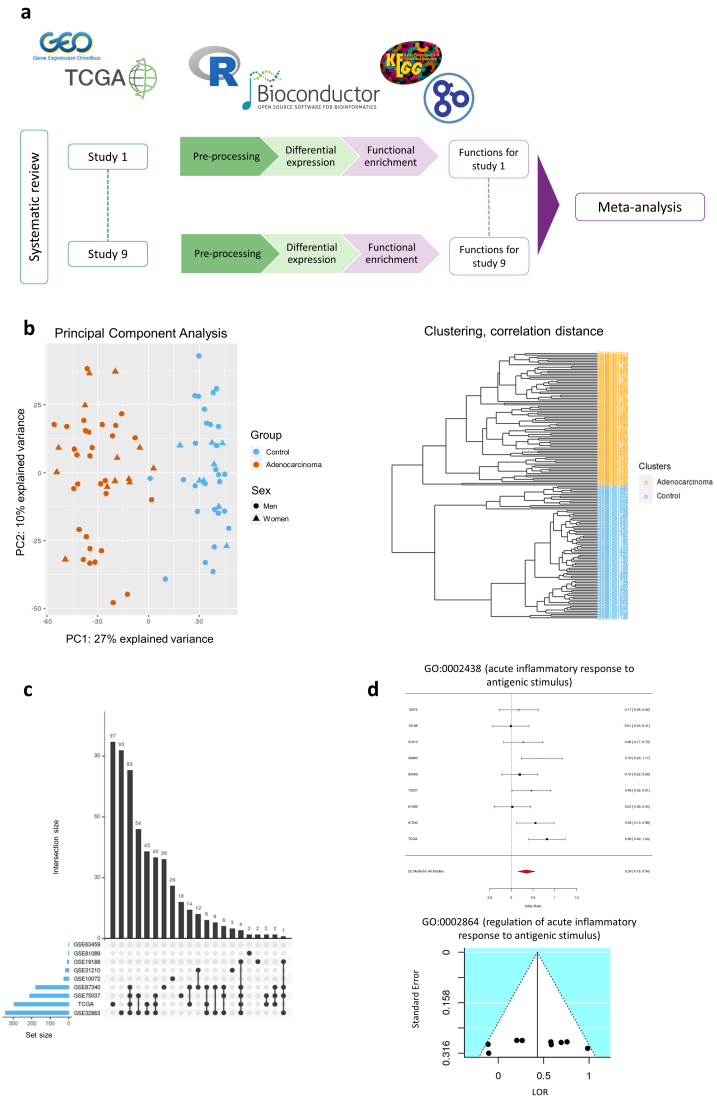


Figure 3.5 – Workflow and analysis design. (a) Summary of the analysis design followed in this work, (b) PCA plot and hierarchical clustering analysis as an example of exploratory analyses (to explore possible batch effects and to assure expected data behavior) performed at the pre-processing stage to assess for the integrity of the data, (c) example of UpSet plot as an intersection analysis for functional enrichment analysis results, and (d) examples of forest and funnel plots to assess meta-analysis results.

control men, respectively. Paired samples design was implemented, and tobacco consumption was included as a batch effect on the limma linear model to reduce its impact on data. p-values were calculated and corrected for FDR [118]. This comparison allows the detection of genes and functions altered by the disease and that have higher or lower activity in women when compared to men. Significant functions and genes were considered when FDR < 0.05. Functional enrichment analyses were performed using the GSEA implemented in the **mdgsa** [119] R package. p-values were, again, corrected for FDR. For functional annotation, two functional databases were used: the KEGG Pathway database [120] and GO [121]. GO terms were analyzed and propagated independently for each GO ontology: BPs, MFs, and cellular components (CC). Those annotations excessively specific or generic were filtered out, keeping functions with blocks of annotations between 10 and 500. Intersections within groups were analyzed with UpSet plots [122] Figure 3.5.

3.3.4.3 Functional Meta-Analysis

Functional GSEA results were integrated into a functional meta-analysis [123] implemented with **mdgsa** and **metafor** [124] R packages. Meta-analysis was applied under the DerSimonian and Laird random-effects model [125], taking into account individual study heterogeneity. This model considers the variability of individual studies by increasing the weights of studies with less variability when computing meta-analysis results. Thus, the most robust functions between studies are highlighted. A total of 6467 GO BP terms, 785 GO MF terms, 1207 GO CC terms, and 213 KEGG pathways were evaluated. p-values, FDR corrected p-values, LOR, and 95% CIs of the LOR were calculated for each evaluated function. Functions and pathways with FDR < 0.05 were considered significant, and both funnel and forest plots were computed for each Figure 3.5. These representations were checked to assess for possible biased results, where LOR represents the effect size of a function, and the SE of the LOR serves as a study precision measure [126]. Sensitivity analysis (leave-one-out cross-validation [124]) was conducted for each significant function to verify possible alterations in the results due to the inclusion of any study.

3.3.4.4 Metafun-NSCLC Web Tool

All data and results generated in the different steps of the meta-analysis are available in the Metafun-NSCLC web tool (https://bioinfo.cipf.es/metafun-nsclc), which is freely accessible to any user and allows the confirmation of the results described in this manuscript and the exploration of other results of interest. The front-end was developed using the Angular Framework. All graphics used in this web resource have been implemented with Plot.ly except for the exploratory analysis cluster plot, which was generated with **ggplot2** [127]. This easy-to-use resource is divided into five sections: (1) Summary of analysis results in each phase. Then, for each of the studies, the detailed results of the (2) exploratory analysis, (3) differential expression, and (4) functional profiling. The user can interact with the web tool through graphics and tables and search for specific information for a gene or function. Finally, Section (5) provides several indicators for the significant functions identified in the meta-analysis that inform whether they are more active in men or women.

3.3.5 Conclusions

Sex-based molecular differences may influence the incidence and outcome of lung adenocarcinoma and, therefore, may have important clinical implications. We identified immune responses, purinergic signaling, and lipid-related processes as the main biological processes altered between male and female lung adenocarcinoma patients by a meta-analysis of transcriptomic datasets. Said processes exhibit increased activity in female lung adenocarcinoma patients, whereas other processes (such as DNA repair) are more active in male lung adenocarcinoma. Although further studies are required to verify and fully explore these findings, our results provide new clues to understand the molecular mechanisms of sex-based differences in lung adenocarcinoma patients and new perspectives regarding the identification of biomarkers and therapeutic targets.

Chapter 4

Identification of transcriptional signatures to stratify pancreatic ductal adenocarcinoma patients

4.1 Overview

In this chapter, we explore the transcriptomic landscape of Pancreatic Ductal Adenocarcinoma and analyze whether genes from the most significantly altered pathways are related to patient survival. This combines two traditional approaches to cancer subtyping: gene expression profiling and clinical behavior. We propose gene signatures that can subtype "hot" and "cold" immune tumors, to help drive patient treatment based on how the tumor is expected to respond to immunotherapy.

4.2 Reference and contribution of the candidate

Pérez-Díez, I.; Andreu, Z.; Hidalgo, M.R.; Perpiñá-Clérigues, C.; Fantín, L.; Fernandez-Serra, A.; de la Iglesia-Vaya, M.; Lopez-Guerrero, J.A.; García-García, F. A Comprehensive Transcriptional Signature in Pancreatic Ductal Adenocarcinoma Reveals New Insights into the Immune and Desmoplastic Microenvironments. Cancers 2023, 15, 2887. DOI: 10.3390/cancers15112887. PMID: 37296850.

The candidate participated in data curation, formal analysis, investigation, methodology, software development, validation, visualization and writing of the manuscript.

4.3 A Comprehensive Transcriptional Signature in Pancreatic Ductal Adenocarcinoma Reveals New Insights into the Immune and Desmoplastic Microenvironments

4.3.1 Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common type of pancreatic cancer, representing over 80% of all diagnosed pancreatic neoplasms. This highly

lethal cancer has a poor prognosis, with a median survival rate of fewer than six months, although its five-year survival rate increased to 12% in recent years [2]. While it is currently the third leading cause of cancer-related deaths worldwide [2], the yearly increase in its incidence may make PDAC the second leading cause of cancer-related deaths by 2030 [2]. The absence of reliable biomarkers for effective screening and early diagnosis at the pre-symptomatic stages when treatments function most effectively represents a primary reason why most PDAC cases remain incurable. Currently, most patients present locally advanced (30–35%) or metastatic (50–55%) PDAC at diagnosis [128]. In advanced-stage PDAC patients, curative surgery remains impossible, and systemic therapeutic options (including immunotherapy) remain limited and ineffective [129]. Among the solid tumors, PDAC represents an immunologically "cold" tumor characterized by sparse T cell infiltration [30, 130]; in contrast, immunologically "hot" tumors (such as melanoma) suffer from a high neoantigen load and immune cell infiltration [131]. PDAC tumors possess distinctive features such as an extracellular matrix (ECM) composition and a fibrotic stroma, which make it highly desmoplastic and significantly influence immune responses [132]. PDAC cells strongly interact with the surrounding microenvironment, which includes components such as immune cells, cytokines, metabolites, fibroblasts, and hyaluronan. These interactions create a highly fibrotic and active organized stroma (desmoplastic stroma) and an immunosuppressive environment that makes PDAC invasive and highly resistant to immunotherapy [29, 30]; therefore, the characterization of the stroma and tumor immune microenvironment in PDAC patients represents a critical step in developing more effective therapeutic strategies. In the last few years, several investigations have focused on studying gene expression in PDAC to better understand the molecular composition of this devastating cancer and identify different molecular subtypes of pancreatic cancer that improve the stratification of patients for clinical strategies [133, 134]. Bailey and colleagues defined four molecular subtypes of pancreatic cancer: squamous, pancreatic progenitor, immunogenic and aberrantly differentiated endocrine exocrine (ADEX) [134], while Moffitt's group identified two stromal subtypes that were defined as "normal" and "activated" [133]. Nevertheless, in clinical practice, it is

difficult to perform this broad molecular test on each patient. Therefore, and despite these new insights in pancreatic cancer, the diagnostic and prognostic outcomes of PDAC patients are extremely poor compared to those of other types of cancers. Additionally, new studies need to be conducted to understand the extreme complexity of PDAC and find simpler genetic signatures that can be incorporated into clinical practice and improve the clinical setting for PDAC patients and families. We aimed to understand the stroma and tumor immune microenvironments of PDAC patients by retrieving and analyzing transcriptomic data from 21 different studies (representing a population of 922 samples; 320 controls and 602 cases) from the Gene Expression Omnibus GEO and ArrayExpress data repositories. Through meta-analysis, we identified a series of gene signatures with survival prognostic value that may play a significant role in therapeutic decision making for PDAC patients, including five genes not previously related to PDAC survival. We also provide a friendly user web tool with detailed and interactive visualization of our comprehensive meta-analysis results.

4.3.2 Materials and Methods

For all bioinformatics and statistical analyses, we employed R software v.4.1.3 [111] (Table A.7 details the R packages and versions).

4.3.2.1 Study Search and Selection

Publicly available datasets were collected from GEO [44] and ArrayExpress databases [112]. Data available in the Cancer Genome Atlas (TCGA) [47] were excluded from the original search with the purpose of using this dataset as an external cohort for survival analysis. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [48], a systematic search of published studies was conducted in 2021 (period: 2002–2021). Three researchers in the study conducted the literature search, and the consistency of the review and selection procedures used was evaluated and confirmed. A broad search was performed using the MeSH (Medical Subject Headings) thesaurus keyword "pancreatic cancer", after

which stringent filters were applied. The final inclusion criteria were:

- Normal and PDAC samples available.
- RNA extracted directly from human pancreas biopsies.
- Patients had not undergone treatment before biopsy.
- Sample size > 4 for PDAC and control groups.

Finally, normalized gene expression from twenty-seven microarray studies (GSE86436, GSE71989, GSE62452, GSE62165, GSE60979, GSE56560, GSE55643, GSE46234, GSE43795, GSE43288, GSE41368, GSE32676, GSE28735, GSE27890, GSE22780, GSE19650, GSE18670, GSE16515, GSE15471, GSE1542, GSE11838, GSE102238, GSE101448, E-MTAB-3365, E-MTAB-1791, E-MEXP-950, and E-EMBL-6) and the count matrices of two RNA-seq (GSE119794 and GSE136569) datasets were retrieved for further analysis.

4.3.2.2 Study Search and Selection

Datasets were individually analyzed in two steps: preprocessing and differential expression analysis.

The nomenclature of clinical variables included in each study was standardized for data preprocessing, and then, exploratory analysis was performed. Prior to exploratory analysis, RNA-seq raw count matrices were normalized using the trimmed mean of m values from the **edgeR** package [113, 114]. The normalization method performed by the original authors for each microarray dataset was assessed, and the matrices were log2 transformed when necessary. Exploratory analysis included expression boxplots, unsupervised clustering, and principal component analysis (PCA) to detect patterns of expression between samples and genes and the presence of batch effects in each study. Differential gene expression analyses were performed in R using **limma** [117], and a paired sample design was implemented in those datasets where applicable. Differentially expressed genes were identified using p values with Benjamini-Hochberg correction [118] for a false discovery rate (FDR) at a significance level of 0.05.

4.3.2.3 Gene Expression Meta-Analysis

Gene expression analysis results were integrated into a meta-analysis using the DerSimonian & Laird random effects model [125], considering individual study heterogeneity. This model considers the variability in individual studies by increasing the weights of studies with less variability when meta-analysis results are computed. A total of 24,365 genes were evaluated. p values, FDR-corrected p values, the binary logarithm of Fold Change (log2FC), and 95% confidence intervals of log2FC were calculated for each evaluated gene, and both funnel and forest plots were computed for each gene. These representations were assessed for possible biased results, where log2FC represents the effect size of a function, and the standard error of the log2FC serves as a study precision measure [126]. Genes were considered significant when FDR < 0.05, absolute log2FC > 0.6, and were measured in at least eleven studies. Sensitivity analysis (leave-one-out cross-validation [124]) was conducted for each significant gene to verify alterations in the results, owing to the inclusion of any study.

Statistically significant results from the gene expression meta-analysis were functionally enriched by over-representation analysis (ORA) using **clusterProfiler** [135, 136] and **ReactomePA** [137]. Gene Ontology terms [121, 138] and Reactome pathway [139] enrichments were performed following this approach. Only those functions and pathways with more than ten differentially expressed genes found in the gene set were considered. Functional enrichment was explored and visualized with the **rrvgo** package [140].

4.3.2.4 Web tool

To make the data and results of our research widely accessible, a web tool was developed using the **shiny** package in R. The tool was developed in a user-friendly manner, allowing users to navigate and interact with the data. Users can then select different variables and parameters to visualize the data in numerous ways. The tool also includes interactive plots and tables to display the analysis results. The web tool is hosted on a secure server and is regularly maintained to ensure stability

and performance. The source code for the tool is also publicly available and can be accessed through our GitHub repository: https://github.com/ipediez/ShinyReport (accessed on 1 May 2023).

4.3.2.5 Survival Analysis

RNA-seq expression data and metadata from patients in the Pancreatic adenocarcinoma TCGA cohort were downloaded from cBioPortal [141]. Z-score of RNA-seq expression were used for survival analysis. For each analyzed gene, samples were divided into two groups based on their expression levels. Samples with expression Z-score below the lower quartile were classified as having low levels of expression, whereas samples exceeding the upper quartile were classified as having high levels of expression. Forty-five samples with high levels of expression and forty-five samples with low levels of expression were included for survival analysis. Gene-wise Kaplan–Meier survival analysis compared the low-level and high-level expression groups. This method estimates the probability of survival over time based on the expression levels of the gene of interest. The log-rank test was used to compare the survival curves between distinct groups of samples. For risk-score-based survival, genes were tagged as highly expressed for a given sample when the expression levels were above the upper quartile. Then, samples were clustered into "high-risk" and "low-risk" groups based on the number of highly expressed genes. The cutoff was set as the median of the highly expressed genes in each sample. Furthermore, a proportional hazard model using Cox regression was implemented to study the impact of clinicopathological variables on survival and evaluate the contribution of the risk score in a multivariate model.

4.3.3 Results

We performed a systematic review and differential gene expression analysis of PDAC transcriptomic studies from GEO [44] and ArrayExpress [112] databases to explore the stroma and immune environments in PDAC patients. We then integrated the results of each differential gene expression analysis into a meta-analysis. The bio-

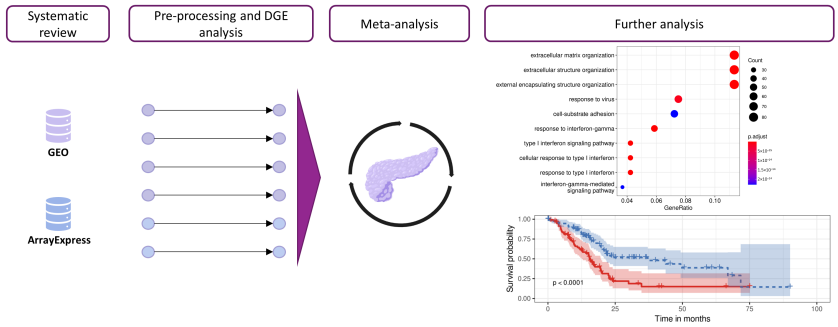


Figure 4.1 – Workflow and analysis design. Relevant studies from GEO and ArrayExpress databases were retrieved, and data exploration and preprocessing were then performed. After differential gene expression analysis, the results from different studies were integrated into a gene meta-analysis. Functional profiling methodologies were applied to explore the biological implications of the results.

logical context of the meta-analysis results was explored via functional enrichment using an ORA of GO terms and pathways (Figure 4.1). Finally, we conducted survival analysis to explore the impact of specific candidate genes on patients' outcomes.

4.3.3.1 Systematic Review

The systematic review identified 143 non-duplicated studies. Then, we excluded studies with samples from patients under cancer treatment and studies where the sample size was less than four in the PDAC or the control group, resulting in a subset of twenty-nine studies (Figure 4.2). We discarded eight studies after exploratory analysis, giving a final set of twenty-one homogeneous and comparable studies for further analysis. The selected studies included 922 samples (320 controls and 602 cases). Although most studies did not include relevant sample metadata, we assessed the clinical characteristics when they were available. Table A.8 and Table A.9 contain further information regarding the selected studies and clinicopathologic characteristics of the study population.

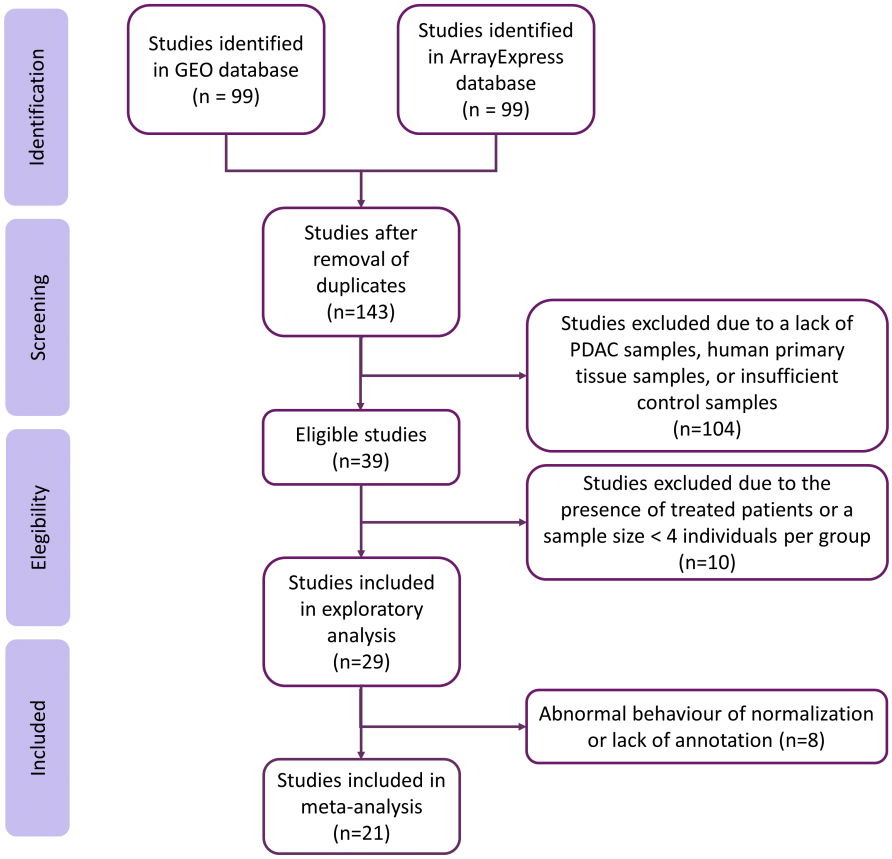


Figure 4.2 – Flow of information through the distinct phases of the systematic review, following PRISMA Statement guideline.

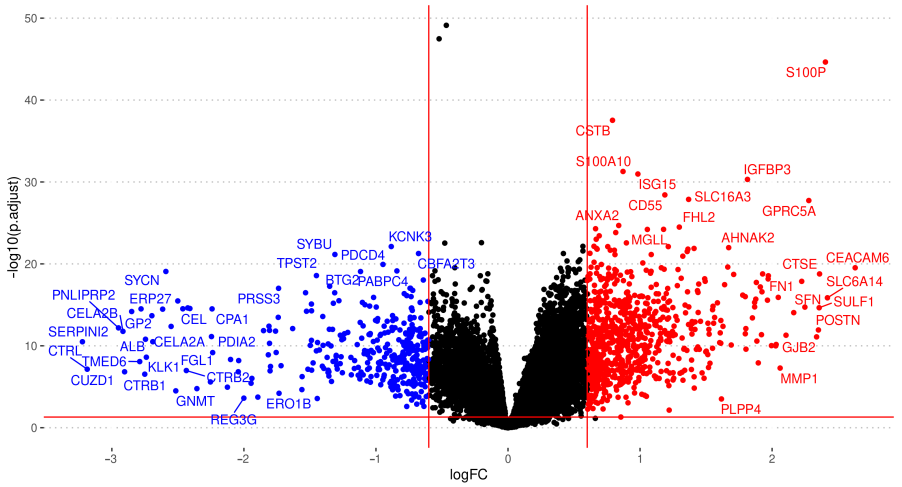


Figure 4.3 – Volcano plot summarizing the gene expression meta-analysis. Significantly over-expressed genes are shown in red, and significantly under-expressed genes are shown in blue (FDR < 0.05; absolute log2FC > 0.6). Genes that do not show significant differential expression are represented in black. Only genes found in at least eleven studies are shown.

4.3.3.2 Integration of Differential Expression Profiles

Exploratory analysis found abnormal normalization or a lack of annotation in eight studies, which we excluded from further analysis (listed in Table A.9). Then, we performed the independent differential gene expression analysis of each study and meta-analysis for 24,365 genes evaluated in the different datasets, including every gene found in at least two studies. We considered results with an FDR < 0.05, an absolute log2FC > 0.6, and those evaluated in at least eleven studies to be significant; overall, 1153 genes met these criteria (Figure 4.3; further details are given in Table A.10).

We noted the presence of genes encoding ECM components (e.g., collagens, fibronectin, laminin, and stratifin), proteoglycans (e.g., versican), cell adhesion molecules, integrins, matrix metallopeptidases, and additional peptidases and enzymes that impact mechano-contractility, epithelial tension, and the stiffness of the tumor stroma, which can promote tumor progression and resistance to therapy (Figure 4.4). **??** displays the twenty genes with the highest and lowest log2FC values from the meta-analysis; these genes mainly play roles in ECM remodeling, desmoplasia, metabolism,

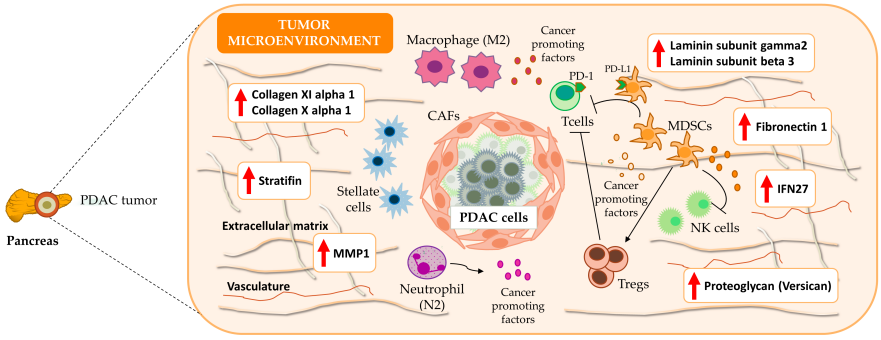


Figure 4.4 – Overview of PDAC microenvironment. Meta-analysis results indicated an overexpression of several ECM components, e.g., stratifin, fibronectin 1, different laminin subtypes (gamma2 and beta3), collagens, and proteoglycans that characterize the dense and desmoplastic stroma of PDAC tumors. Additionally, the results highlight the presence of immune components such as IFN27, which contribute to an increase in the number of M2 macrophages and a decrease in the number of CD8+ T cells. Therefore, the desmoplastic stroma and the immune system favor immune tolerance and poor prognosis in PDAC. The red upward-pointing arrows denote genes exhibiting significant overexpression in the conducted meta-analysis. IFN27: interferon alpha inducible protein; MMP1: matrix metallopeptidase 1; NK cells: natural killer cells; T cells: T effector lymphocytes; Tregs: T regulatory lymphocytes T.

and the immune system. Table A.10 reports a complete list of significantly affected genes.

We performed ORA using GO biological process terms to identify the possible implications of 1153 significantly differentially expressed genes in the PDAC samples. We considered only those biological processes with at least ten associated genes and an adjusted p value under 0.05. We found 546 over-represented biological processes among the over-expressed genes and 40 biological processes over-represented among the under-expressed genes (Table A.11). ORA revealed the enrichment of terms related to the tumor microenvironment (Figure 4.5), with GO terms related to the immune system, cell adhesion, and ECM remodeling/degradation. Of note, additional over-represented functions were related to metastasis (vascularization, cell migration, collagen, mesenchymal transition, cell proliferation, and peptidyl modifications) [132, 142].

Table 4.1 – Top twenty genes up-regulated in PDAC patients

Gene Symbol	Gene Name	Expression Level	Function
CEACAM6	CEA cell adhesion molecule 6	UP	ECM remodeling
SLC6A14	Solute carrier family 6 member 14	UP	ECM remodeling
S100P	S100 calcium-binding protein P	UP	ECM remodeling
CTSE	Cathepsin E	UP	ECM remodeling
SULF1	Sulfatase 1	UP	ECM remodeling
POSTN	Periostin	UP	ECM remodeling
GJB2	Gap junction protein beta 2	UP	ECM remodeling
GPRC5A	G protein-coupled receptor class C group 5 member A	UP	ECM remodeling
SFN	Stratifin	UP	ECM remodeling
FN1	Fibronectin 1	UP	ECM remodeling
LAMC2	Laminin subunit gamma 2	UP	ECM remodeling
CEACAM5	CEA cell adhesion molecule 5	UP	ECM remodeling
MMP1	Matrix metallopeptidase 1	UP	ECM remodeling
COL11A1	Collagen type XI alpha 1 chain	UP	ECM remodeling
TSPAN1	Tetraspanin 1	UP	ECM remodeling
IFI27	Interferon alpha inducible Protein 27	UP	Immune System
CST1	Cystatin SN	UP	Epithelial -mesenchymal transition
LAMB3	Laminin subunit beta 3	UP	ECM remodeling
COL10A1	Collagen type X alpha 1 chain	UP	ECM remodeling
VCAN	Versican	UP	ECM remodeling

Table 4.2 – Top twenty genes down-regulated in PDAC patients

Gene Symbol	Gene Name	Expression Level	Function
CTRB2	Chymotrypsinogen B2	DOWN	ECM remodeling
PLA2G1B	Phospholipase A2 group IB	DOWN	Metabolism
CTRC	Chymotrypsin C	DOWN	ECM remodeling
GNMT	Glycine N-methyltransferase	DOWN	Metabolism
AQP8	Aquaporin 8	DOWN	H2O2 transport
SYCN	Syncolin	DOWN	Exocytosis
CPA2	Carboxypeptidase A2	DOWN	Metabolism
CELA2A	Chymotrypsin-like elastase 2A	DOWN	ECM remodeling
GP2	Glycoprotein 2	DOWN	Metabolism
KLK1	Kallikrein 1	DOWN	Serine protease
ALB	Albumin	DOWN	Oncotic pressure
CTRB1	Chymotrypsinogen B1	DOWN	ECM remodeling
ERP27	Endoplasmic reticulum protein 27	DOWN	Lipid and protein synthesis
TMED6	Transmembrane p24 trafficking protein 6	DOWN	Insulin secretion
PNLIPRP1	Pancreatic lipase-related protein 1	DOWN	Metabolism
CUZD1	CUB and zona pellucida like domain 1	DOWN	ECM remodeling and Immune System
CELA2B	Chymotrypsin-like elastase 2B	DOWN	ECM remodeling
PNLIPRP2	Pancreatic lipase-related protein 2	DOWN	Metabolism
CTRL	Chymotrypsin-like	DOWN	ECM remodeling
SERPINI2	Serpin family I member 2	DOWN	Protease inhibitor

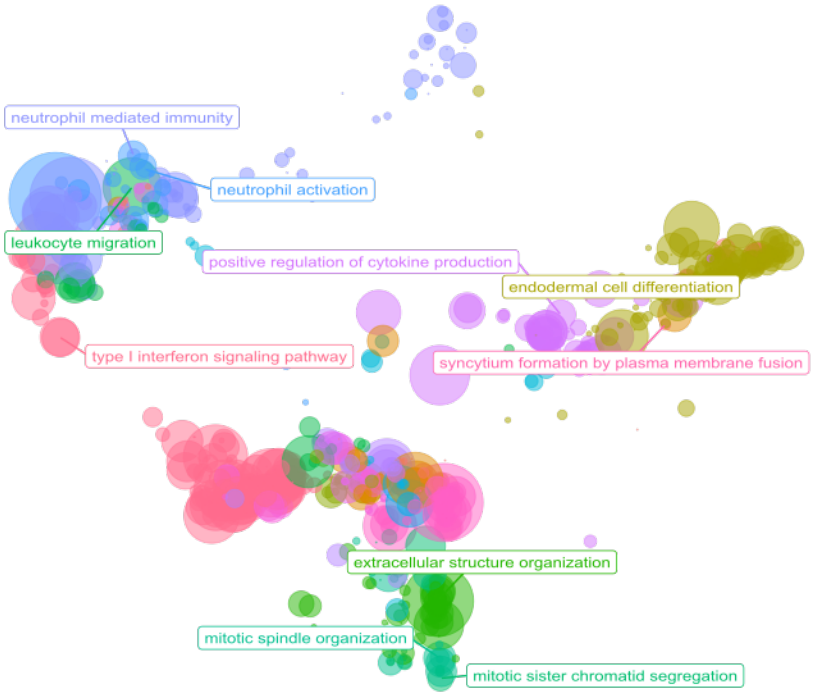


Figure 4.5 – Scatter plot of ORA results. The scatterplot reports the GO biological process representative terms after redundancy reduction in a two-dimensional space derived from the semantic similarities between GO terms. The dot size represents the number of biological processes related to a GO term. The parent terms of the main clusters are labeled.

4.3.3.3 Interactive Tool for Results Visualization

The web tool contains comprehensive information regarding the data and results of the meta-analysis of gene expression. The application includes tables and plots for the differential expression results of twenty-one datasets included in the study and meta-analysis results. Statistical indicators, such as the log odds ratio, confidence intervals, and adjusted p values, are provided to estimate each study's global expression and specific contribution. The web tool is available online: https://bioinfo.cipf.es/MetaPDAC/ (accessed on 1 January 2023).

4.3.3.4 Immune System: A Functional Overview in PDAC

To focus our analysis on the tumor immune microenvironment, we extracted a consensus list of genes related to the immune system and inflammation from NCBI and GO databases (mainly framed in the categories of HLA, interleukin, CI, interferon, chemokine, and S100 genes, Table A.12). Considering an FDR threshold of 0.05 and an absolute fold change greater than 0.6, we discovered the significant differential expression of 322 immune genes in our meta-analysis results. To explore the functional involvement of these results, we performed ORA on this group of genes using GO biological process terms and Reactome pathways. We considered significant functional terms with at least ten associated genes and an adjusted p value < 0.05. We discovered the over-representation of thirty-three GO terms and twenty-seven pathways among the over-expressed immune-related genes and none when under-expressed genes were analyzed. The enriched terms suggest the increased activity of neutrophil-related immune response, the negative regulation of cell killing, interferon signaling, and an antigen presentation via major histocompatibility complex II.

4.3.3.5 Immune and Stromal Survival Signatures Impact PDAC Prognosis

We explored the 322 differentially expressed immune-related genes and identified a set of 70 genes of particular interest in our experimental research (Table 4.3). We performed survival analysis using the pancreatic adenocarcinoma TCGA cohort for each of these genes and found statistically significant differences in twenty-eight genes (IFI27, IL1R2, IL1RN, IL1RAP, IL18, IL22RA1, HCP5, SLFN13, CD58, CD109, IFI44L, IFI16, IFITM1, IFIT1, IFIT3, IRF9, IFIT2, IFI35, CXCL10, CXCL5, CXCL9, S100P, S100A6, S100A2, S100A16, S100A11, S100A14, and S100A10), which shared a pattern: a higher expression in patients associated with a lower rate of survival. As far as we are aware, this is the first time that HCP5, SLFN13, IRF9, IFIT2, and IFI35 have been related to prognosis value in PDAC patients (Figure A.3).

We analyzed genes that displayed statistical significance as a “signature”, dividing the samples into high-risk and low-risk groups based on the number of highly

Table 4.3 – Subset of immune-related genes. Genes in bold possess statistically significant differences according to survival analysis.

Functional Group	Genes
HLA	HLA-F, HLA-DRB5, HLA-B, HLA-A, HCP5, HLA-DRA, HLA-DPA1, HLA-DQB1, HLA-DQA1, HLA-DMB, HLA-DRB1, HLA-G, HLA-DPB1, SLFN12, SLFN13, SLFN11
Interleukin	IL1R2, IL1RN, IL1RAP, IL7R, IL2RG, IRAK3, IL18, LIF, IL22RA1
CD	CD58, CD109, CD52, CD53, CD74, CD14, CCDC80, CCDC141, CCDC69, DCDC2, PDCD4
Interferon	IFI27, IFI44L, IFI6, STING1, IFI16, IFITM1, ISG20, IFIT1, IFIT3, IFITM2, IRF9, IFIT2, IFNGR2, IFITM3, IFI35
Chemokine	CCL20, CCL18, CXCL10, CXCL5, CXCL8, CXCR4, CKLF, CXCL9, CXCL3, CXCL14, CXCL12
S100	S100P, S100A6, S100A2, S100A16, S100A11, S100A4, S100A14, S100A10

expressed genes (above the upper quartile). We set the median (six highly expressed genes) as the cutoff value to divide the samples into groups. Interestingly, patients in the high-risk group possessed shorter survival times than those in the low-risk group did (p value < 0.0001, Figure 4.6A). Furthermore, we studied the effect of this signature in a multivariate Cox model including age, alcoholic history, the presence of chronic pancreatitis, diabetes diagnostic, tumor grade, and the American Joint Committee on Cancer classification of a metastatic tumor and a residual tumor as covariates. The proposed signature was the only variable with p value < 0.05 and showed a hazard ratio of 2.36 (Supplementary Figure S4). We then analyzed the co-occurrence of highly expressed genes in the samples, finding two main co-occurrence groups that related to high-risk patients: i) the interferon gene family (IFN genes) and ii) the S100 and IL genes (S100A14, S100A16, S100A6, S100A11, IL1R2, IL1RN, and S100P) (Figure 4.6B).

To explore how a desmoplastic environment can affect patients' survival, we employed an homologous approach using genes related to ECM remodeling (Table 4.1 and Table 4.2). We discovered eleven genes whose survival analysis showed statistically significant differences (CEACAM5, CEACAM6, FN1, GJB2, GPRC5A, LAMB3, LAMC2, SFN, SLC6A14, TSPAN1, and VCAN). Again, we divided the sam-

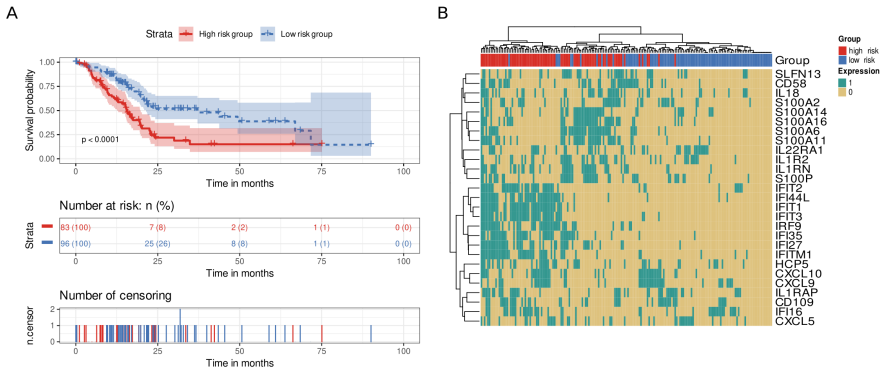


Figure 4.6 – A twenty-eight gene signature clustered patients into high-risk or low-risk groups based on the number of highly expressed signature genes in their transcriptomic profile. Patients with at least six highly expressed genes were classified as having a high risk, whereas those with five or fewer were classified as having a low risk. (A) Kaplan–Meier curve. Patients from the high-risk group (red) had shorter survival times than patients from the low-risk group did (blue). Below, the number of still alive patients and percentage in each group at 0, 25, 50, 75, and 100 months, and the censored events. (B) Heatmap demonstrating the patterns of high expression between genes and samples. Gene expression was coded as 1 for a sample above the upper quartile.

ples into high-risk and low-risk groups using the median of the number of highly expressed genes as the cutoff value (median = 3). Patients with high levels of expression in three or more genes from the signature presented lower survival times than those with fewer highly expressed genes did (p value = 0.00012, Figure 4.7A). Of note, we distinguished a cluster of co-occurrence of patients with high levels of GJB2, FN1, and VCAN at the same time (Figure 4.7B).

Finally, we performed comparative analysis between the immune and stromal survival signatures identified in our work and other signatures generated in previous works for patient stratification [133, 134]. These results provided insight into the level of intersection between this group of signatures (Table A.13).

4.3.4 Discussion

Using comprehensive meta-analysis, we explored the immune environment and desmoplastic stroma of PDAC tumors to contribute to a deeper understanding of tumorigenesis and the design of effective therapeutic strategies, such as immunother-

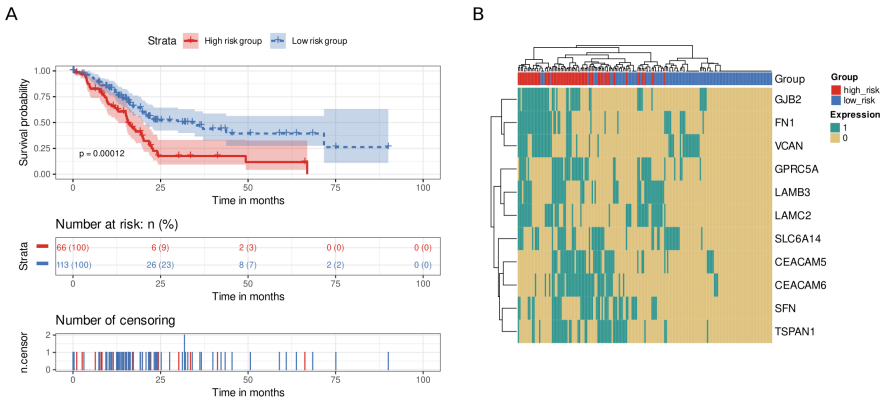


Figure 4.7 – Survival analysis of ECM remodeling genes. An eleven-gene signature clustered patients into high-risk or low-risk groups based on the number of highly expressed signature genes in their transcriptomic profile. Patients with at least three highly expressed genes were classified as having a high risk, whereas those with five or fewer were classified as having a low risk. (A) Kaplan–Meier curve. Patients from the high-risk group (red) had shorter survival times than patients from the low-risk group did (blue). Below, the number of still alive patients and percentage in each group at 0, 25, 50, 75, and 100 months, and the censored events. (B) Heatmap demonstrating the patterns of high expression between genes and samples. Gene expression was coded as 1 for a sample above the upper quartile.

apies. ECM components from the desmoplastic stroma tightly interact with the immune environment and contribute to immune evasion by modulating immune cell infiltration, thus influencing cell proliferation, tumor progression, and overall survival [143, 144]. The meta-analysis and ORA results characterized differences in the gene-expression landscape of PDAC tumors and identified more than 1000 dysregulated genes, most of them with immune system- and desmoplasia-related roles. We discovered thirty-nine genes (twenty-eight immune-related genes and eleven stroma-related genes) that impact PDAC patients' survival. Among the top forty dysregulated genes (**??**), we observed the upregulation of collagens (COL11A1 and COL10A1), which influence immune infiltration and chemoresistance and confer a poor prognosis [145–147]. PDAC patients also presented with upregulated periostin expression, which has been linked to a shorter overall survival [148], and cystatin SN, which contributes to pancreatic cancer cell proliferation and may represent a potential biomarker for the early detection of pancreatic cancer [149]. Stratifin and matrix metallopeptidase 1 also appeared to be upregulated in PDAC patients; stratifin stimulates matrix metallopeptidase 1 expression in fibroblasts, contributing to

remodel ECM [150]. The increased expression of fibronectin in the PDAC stroma has also been reported. The observed upregulation of cathepsin E and sulfatase 1 expression in the PDAC microenvironment might also benefit the development of therapeutic strategies with polymer drug conjugates since they may contribute to drug release [**mohamed_cysteine_2006** , 151, 153].
The analysis of the top forty dysregulated genes also provided evidence for the downregulation of genes coding for proteolytic enzymes released by the pancreas (e.g., chymotrypsin, chymotrypsinogen, lipases, and phospholipases). Pancreatic cancer cells express around 20% of chymotrypsin C normal cells expression, with this enzyme participating in cancer cell apoptosis and migration [154]. A recent report suggested that a combination of trypsinogen and chymotrypsinogen displayed an anti-tumorigenic potential [155].

Focusing on the immune environment, PDAC tumors develop a wide range of mechanisms to evade the immune system (e.g., a low level of expression of HLA antigens, immunosuppressive signals that inhibit natural killer and T cell functions, and the presence of immunosuppressive cells). This creates an immunotolerant environment in which the immune system of PDAC patients does not robustly recognize and target cancer cells [156]. We explored the expression of seventy genes of particular interest, including those from the HLA, interleukin, CI, interferon, chemokine, and S100 categories. The survival analysis of these genes in the pancreatic adenocarcinoma TCGA cohort identified a twenty-eight immune-related gene signature with a prognostic value that was used to cluster PDAC patients into high-risk and low-risk groups.

The proposed signature possessed significance in univariate and multivariate Cox models with clinicopathological variables, significantly adding statistical power to the survival analysis. This signature could aid in the stratification of patients (Figure 4.8) who could benefit from immunotherapeutic strategies, given that it could contribute to distinguishing “cold” PDAC tumors (characterized by the low presence of T cells (CD8+) and natural killer cells, high presence of immunosuppressive cell populations, and poor prognoses and responses to immunotherapy) from “hot tumors” (with an opposite profile) [147, 157]. We uncovered two high

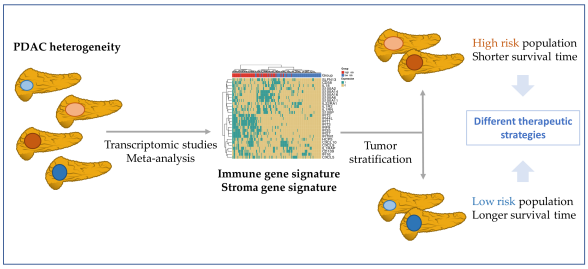


Figure 4.8 – Patient stratification based on PDAC molecular features. Meta-analysis from transcriptomic studies allows a better understanding of the PDAC environment. In this study, the found gene signatures might contribute to the stratification of PDAC patients. In a first step, the immune or the stroma gene signatures can divide patients into high- and low-risk populations. After, with a focus on the immune signature co-occurrence, patients could be divided into those with more S100/IL genes and those with more IFN expressed genes. The knowledge about these molecular features of PDAC tumors may guide the design of more effective therapeutic strategies.

gene-expression co-occurrence patterns, one composed of IFN genes and the other of S100/IL genes. The IFN signaling pathways participate in PDAC development, while the over-expression of S100 genes blocks the infiltration and cytotoxic activity of CD8+ T cells, and the low level of expression of IL1RN and IL1R2 has been associated with increase survival in PDAC patients [158–160].

To the best of our knowledge, this is the first report of data suggesting a link between the HCP5, SLFN13, IRF9, IFIT2, and IFI35 immune genes and PDAC prognosis, presenting the discriminatory power of clustering PDAC patients. The remaining genes of the immune gene signature have been individually associated with PDAC or other cancers, with data suggesting that their overexpression could impact patients' diagnosis, prognosis, and response to treatment [161–166]; however, we report that a joint gene expression signature of these genes impacts PDAC patients' survival.

Focusing on the PDAC stroma, the altered genes include several types of collagens, fibronectins, and proteolytic enzymes, such as metalloproteases and pepti-

dases (**??** and Table A.10), which significantly contribute to ECM composition and stromal remodeling and support desmoplasia and immunosuppression [167]. The survival analysis of significantly dysregulated stromal gene expression from the meta-analysis of the pancreatic adenocarcinoma TCGA cohort revealed a gene signature with prognostic capacity that clustered PDAC patients into high-risk and low-risk groups. We observed a co-occurrence pattern in high-risk patients, indicating a subgroup of PDAC patients with a high level of expression of GJB2, FN1, and VCAN genes. These results indicate stromal heterogeneity in PDAC [168] and the need to characterize it to stratify patients (Figure 4.8).

With respect to other dysregulated genes, the upregulation of CEACAM5 and CEACAM6 represents an early event in pancreatic carcinogenesis, with these genes being candidates for immunotherapies [169–171]. Furthermore, laminins LAMBC2 and LAMB3 support cancer progression and resistance to gemcitabine—one of the main chemotherapeutics used in PDAC patients [172, 173]. In general, the association of the stroma signature with a poor prognosis is consistent with the one described in previous studies for each gene: CEACAM5 [174], CEACAM6 [175] , FN1 [176], GJB2 [177], GPRC5A [178], LAMB3 [179, 180], LAMC2 [179, 180], SFN [181], SLC6A14 [182], TSPAN1 [183], and VCAN [176].

With respect to other similar approaches, we are aware of two additional studies in which expression datasets were integrated to explore the nature of the PDAC in depth: one by Gooneskere and colleagues, who integrated six PDAC and three other pancreatic carcinomas datasets [184], and one by Irigoyen and colleagues, who integrated two peripheral blood datasets [185]. Both approaches integrate different datasets at the gene level to increase the number of samples and perform unique differential gene expression analysis. In contrast, our approach analyzed each dataset independently, and then integrated the results, evaluating their robustness. From the experimental design point of view, both studies differ greatly from ours, since Grooneskere et al.'s [184] one is not specifically focused on PDAC , and Irigoyen et al.'s [185] one does not analyze pancreatic tissue. From a methodological point of view, our study contributes to a more profound and robust analysis of the PDAC expression landscape by integrating data after differential gene expression analysis

had been performed, thus avoiding the necessity to control heterogeneity among studies and retaining the full potential of biological differences.

Other molecular studies based on whole transcriptome and genomic analyses of pancreatic tumors have found specific gene signatures that identify different molecular subtypes [133, 134]. However, the aim of this study was not to identify molecular subtypes, such as in the cited works. The immune signature or the stromal signature presented in this work establishes patient survival groups (high-risk group and low-risk group), which could help practitioners to decide if the patient could benefit from immunotherapy, for example, or not. Intersection analysis indicated that there is hardly any overlap between the gene signatures found in our study and the signatures described by Bailey et al. [134] or Moffitt et al. [133] (as shown in the supplemental analysis (Table A.13). Therefore, the proposed gene signatures show subtype-independent survival value and display a reasonable number of genes for them to be translated to clinics. Nevertheless, more and deeper studies are needed for this purpose. Additionally, the works by Moffitt et al. [133] and Bailey et al. [134] are enormously rich and provide comprehensive molecular stratification to facilitate personalized treatment and the identification of therapeutic targets. Unfortunately, extensive molecular analyzes are difficult to translate to clinical practice for individual patients.

A potential limitation of our study has been the relative heterogeneity among the sample sizes and sequencing platforms used. The meta-analysis methodology, which integrates data groups and provides results with higher statistical power and precision [64, 65], addresses this issue by independently comparing each study and combining the results. A lack of clinical and/or molecular information in most studies, such as survival time, stage condition, or molecular pattern, represents an additional limitation. We employed TCGA data for survival analysis, but additional analyses should integrate other covariates of interest in the study.

Finally, we provided an interactive web tool that allows users to explore our results, facilitating the accessibility, transparency, and reusability of our research. Overall, the web tool provides a detailed and interactive visualization of the meta-analysis results, allowing users to further explore and understand the gene expres-

sion patterns identified in the studies. Other functionalities include the capability to customize and filter the data to further investigate specific aspects of the analysis in more detail. In this manner, we aim to align our research with the FAIR principles to share our data in a way that can be of further use to the scientific community who studies this aggressive and lethal tumor.

4.3.5 Conclusions

Therapeutic strategies to overcome the immune microenvironment and the desmoplastic stroma barriers remain limited and generally unsuccessful. This study performs a comprehensive transcriptional meta-analysis of the molecular PDAC environment. The results highlight the relevance of the interaction between the immune system and stroma, revealing an impact on patients' survival. The identified gene signatures provide new insights into the potential therapeutic targets for this deadly disease that can help to stratify its heterogeneity. Future studies are needed to explore the benefits of targeting the immune and stromal microenvironments as a treatment strategy for PDAC.

Chapter 5

General discussion

Cancer is a leading cause of premature death worldwide [186] and, according to the World Health Organization, is expected to be the cause of more than 15 million deaths per year by 2040 [187]. One of the most challenging aspects of cancer is its heterogeneity, as each type of cancer has its genetic profile. As a result, treatments are not effective for all types of cancer [188]. In addition, there is a broad biological diversity in patients that makes the development of the disease and its possible treatment differ between individuals. Thus, for example, inherited mutations can cause greater susceptibility to one type of cancer than another, or differences in the immune system, especially between men and women, can influence the effectiveness of treatment [10]. This diversity of cancer subtypes and variability among patients are some of the reasons why cancer treatment is not always successful. Personalized medicine approaches and tailored treatment can maximize effectiveness and reduce adverse patient effects. This scenario is where cancer subtyping enters the equation, offering researchers the knowledge of different molecular landscapes of the disease and physicians the ability to predict a patient's prognosis and response to treatment more accurately.

Initially, we researched lung adenocarcinoma, a type of cancer in which epidemiological differences between men and women have been described [20, 25, 52]. We proved that the main affected functional domains were the immune system, purinergic signaling, and lipid metabolism. We then explored the molecular mechanisms underlying these differences. Interestingly, our results pointed to an elevated acute immune response in female LUAD patients, aligning with previous studies on female LUAD patients [84] or LUAD/NSCLC patients [85, 86, 91]. The discovered molecular pathways differentially activated between males and females could underlie the phenotypic differences regarding immunotherapy response observed in the literature [50, 84, 98, 189].

Furthermore, we showed that sex can be a crucial variable when studying and treating LUAD, as it could be the case for other cancer types. The omission of sex consideration is a recurring deficiency in research design and reporting, with medical research centered historically on male physiology [80, 81]. Although there is evidence about the sex differences in disease prevalence, manifestation, and response

to treatment, sex-based biology and medicine are still viewed as a specialized area of interest rather than a central consideration in medical research. This outdated paradigm must change to reflect the growing body of scientific evidence and ensure that sex is appropriately incorporated into all aspects of medical research, from study design to data analysis and interpretation. Otherwise, we could be drawing the wrong conclusions from studies. Analyzing individuals of one sex may make results not extrapolate to the other, and mixing them without analyzing possible differences may mask relevant disease and treatment effects. Thus, the omission of sex as a relevant variable could prevent us from finding better treatments and could lead to poor medical decisions for half our population. Only by taking sex into account as a fundamental biological variable can we reach the potential for more effective and personalized treatments that truly address the unique needs of patients of both sexes.

Following our LUAD research, we delved into the transcriptomic landscape of pancreatic ductal adenocarcinoma. In this analysis, we aimed to understand the intricate relationship between the significantly altered genes and pathways and patient survival. Our findings revealed a substantial association between the altered genes and functions with the immune system and the extracellular matrix. Additionally, we formulated two distinct gene signatures, which were subsequently validated using an external cohort, allowing us to establish a clear link between these signatures and patient outcomes.

The meta-analysis strategy has allowed us to describe the transcriptomic scenario of PDAC in-depth, confirming previous discoveries and providing new insight not described before. By integrating individual studies, this approach allowed us to report two new gene signatures impacting patient survival. By describing these gene signatures, we have also reported a link between the HCP5, SLFN13, IRF9, IFIT2, and IFI35 immune genes and PDAC prognosis for the first time. Previous studies have associated the remaining genes shaping these prognostic signatures with poor prognosis [174–183], suggesting their overexpression could impact a patient's diagnosis, prognosis, and treatment response. Our research is the first to combine them in a joint gene expression signature impacting PDAC patients' survival. These findings

could be used to stratify patients, distinguishing between “hot” or “cold” tumors, and drive therapeutic decisions, as patient response to immunotherapy may be remarkably different for a patient depending on this.

Our research shows that the immune system and its interplay with the tumor microenvironment are essential in tumor progression and treatment response in the two cases studied. The status of the tumor ("hot" or "cold") leads to different cancer progression and prognosis [147, 157], and generalizing one of these statuses to all patients is not only inefficient but counterproductive. These differences can be related to clinical variables, as for LUAD. Sex as a variable is a driving force of immune response differences between LUAD patients, and this knowledge could be used to drive therapeutic decisions. In other cases, the differences are independent of known clinical variables, as we showed for PDAC. Individual PDAC patients show variability in how their tumor behaves and interacts with the immune system, leading to different survival times and potentially influencing tolerance or resistance to immunotherapy. Further studies could assess if the proposed gene signatures are related to immunotherapy response.

5.1 Strengths

This thesis’s major strength relies on using integration strategies of transcriptomic profiles based on a meta-analysis, which allow the generation of new and relevant biological knowledge in the described types of cancer. The applied methodology provides greater robustness and statistical power compared to individual study analyses. Meta-analysis studies a larger sample size, leading to more precise and reliable effect size estimates, detecting even small but real effects that the noise of individual studies might mask. Furthermore, while individual studies are susceptible to biases, meta-analysis can lessen the impact of these biases and explore and quantify the variability in effect sizes across studies. Meta-analysis finds a consensus pattern, allowing for more precise and objective conclusions than individual studies. Lastly, the work presented here has been extensively evaluated by experts in the field through the publication process, demonstrating the validity and significance of

the research.

5.2 Limitations

There are some limitations in the work presented in this thesis. First, most studies lack clinical and molecular information, such as survival time, stage condition, molecular pattern, or sex. Moreover, the systematic review performed for LUAD studies revealed that less than 50% of the datasets considered this variable in their study design. This lack of information represents a limitation, as variables potentially impacting the results cannot be considered. We also faced a lack of standardization among clinical information and gene annotation between studies. Some information is lost when transforming data from one codification to another, as gene annotation conversion is not always one-to-one. Lastly, a potential limitation of our studies has been the quality of the individual studies. Although the meta-analysis approach addresses the heterogeneity among sample sizes and sequencing platforms used by the individual analyses, its results are only as good as the quality of the studies it includes. We attempted to address this issue by applying quality control filters during the study selection phase of our work. Even though all the studies evaluated had been published through a peer review process, some were discarded after an exploratory analysis due to questionable data quality.

5.3 Future perspectives

Based on our findings, collaborators from the Instituto Valenciano de Oncologia will perform experimental validation to confirm the described PDAC gene signatures and their impact on treatment response. Once confirmed, this strategy could be applied to different cancer types of interest to draft further personalized medicine approaches.

Chapter 6

Conclusions

1. Published data can be effectively reused to develop scientific knowledge further and deepen the understanding of cancer subtypes. This highlights the importance of collaboration and data sharing in the scientific community.
2. Meta-analysis is the proper statistical methodology for this aim, as it is robust and improves the statistical power, masking possible deficiencies from some of the published data. Meta-analysis can be effectively used to leverage published data and push a step forward cancer research
3. Lung adenocarcinoma patients show sex-based transcriptomic functional differences. This suggests the importance of considering sex as a relevant biological factor in lung cancer research and treatment, which could lead to more personalized and effective therapies.
4. The molecular mechanisms underlying lung adenocarcinoma sex-based transcriptomic differences are particularly related to the immune system, purinergic signaling, and lipid metabolism. This insight opens new avenues for biomarker research and therapies specifically targeting these biological pathways.
5. The pancreatic ductal adenocarcinoma transcriptomic landscape is strongly linked to extracellular matrix components and the immune system. This association suggests an immunosuppressive environment and a desmoplastic stroma, which could influence treatment response and disease progression.
6. Based on the pancreatic ductal adenocarcinoma transcriptomic landscape, we identified twenty-eight immune genes and eleven extracellular matrix-related genes that impacted patient survival. These genes could be used to stratify patients and drive medical practice.

Bibliography

[1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," eng, *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, May 2021. DOI: 10.3322/caac.21660.

[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," en, *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023. DOI: 10.3322/caac.21763.

[3] C. M. Perou *et al.*, "Molecular portraits of human breast tumours," en, *Nature*, vol. 406, no. 6797, pp. 747–752, Aug. 2000. DOI: 10.1038/35021093.

[4] J. S. Parker *et al.*, "Supervised risk predictor of breast cancer based on intrinsic subtypes," *Journal of Clinical Oncology*, vol. 27, pp. 1160–1167, Mar. 2009. DOI: 10.1200/jco.2008.18.1370.

[5] P. S. Hammerman *et al.*, "Comprehensive genomic characterization of squamous cell lung cancers," en, *Nature*, vol. 489, no. 7417, pp. 519–525, Sep. 2012. DOI: 10.1038/nature11404.

[6] D. M. Muzny *et al.*, "Comprehensive molecular characterization of human colon and rectal cancer," en, *Nature*, vol. 487, no. 7407, pp. 330–337, Jul. 2012. DOI: 10.1038/nature11252.

[7] C. Curtis *et al.*, "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," en, *Nature*, vol. 486, no. 7403, pp. 346–352, Jun. 2012. DOI: 10.1038/nature10983.

[8] F. Mauvais-Jarvis *et al.*, "Sex and gender: Modifiers of health, disease, and medicine," *Lancet (London, England)*, vol. 396, no. 10250, pp. 565–582, 2020. DOI: 10.1016/S0140-6736(20)31561-0.

[9] A. Gogovor *et al.*, "Sex and gender considerations in reporting guidelines for health research: A systematic review," *Biology of Sex Differences*, vol. 12, no. 1, p. 62, Nov. 2021. DOI: 10.1186/s13293-021-00404-0.

[10] J. Grahovac, "The importance of sex as a biological variable in cancer research," en, *Oncology Insights*, no. 1, pp. 8–15, 2023, ISSN: 3009-3848.

[11] D. Hanahan, "Hallmarks of Cancer: New Dimensions," *Cancer Discovery*, vol. 12, no. 1, pp. 31–46, Jan. 2022. DOI: 10.1158/2159-8290.CD-21-1059.

[12] S. Haupt, F. Caramia, S. L. Klein, J. B. Rubin, and Y. Haupt, "Sex disparities matter in cancer development and therapy," *Nature Reviews Cancer*, vol. 21, pp. 393–407, Apr. 2021. DOI: 10.1038/s41568-021-00348-y.

[13] H. Han *et al.*, "Blood glucose concentration and risk of liver cancer: Systematic review and meta-analysis of prospective studies," *Oncotarget*, vol. 8, pp. 50 164–50 173, Apr. 2017. DOI: 10.18632/oncotarget.16816.

[14] A. Vulcan, J. Manjer, and B. Ohlsson, "High blood glucose levels are associated with higher risk of colon cancer in men: A cohort study," *BMC Cancer*, vol. 17, Dec. 2017. DOI: 10.1186/s12885-017-3874-4.

[15] A. M. Barberio *et al.*, "Central body fatness is a stronger predictor of cancer risk than overall body size," *Nature Communications*, vol. 10, Jan. 2019. DOI: 10.1038/s41467-018-08159-w.

[16] C. M. Lopes-Ramos *et al.*, "Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism," *Cancer Research*, vol. 78, no. 19, pp. 5538–5547, Oct. 2018. DOI: 10.1158/0008-5472.CAN-18-0454.

[17] W. Yang *et al.*, "Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data," *Science Translational Medicine*, vol. 11, no. 473, eaao5253, Jan. 2019. DOI: 10.1126/scitranslmed.aao5253.

[18] C. Tannenbaum, R. P. Ellis, F. Eyssel, J. Zou, and L. Schiebinger, "Sex and gender analysis improves science and engineering," en, *Nature*, vol. 575, no. 7781, pp. 137–146, Nov. 2019. DOI: 10.1038/s41586-019-1657-6.

[19] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," en, *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018. DOI: 10.3322/caac.21492.

[20] J. A. Barta, C. A. Powell, and J. P. Wisnivesky, "Global Epidemiology of Lung Cancer," en, *Annals of Global Health*, vol. 85, no. 1, p. 8, Jan. 2019. DOI: 10.5334/aogh.2419.

[21] *SEER Cancer Statistics Review, 1975-2017*, en. [Online]. Available: https://seer.cancer.gov/csr/1975_2017/index.html (visited on 04/17/2024).

[22] G. A. Rivera and H. Wakelee, "Lung Cancer in Never Smokers," en, in *Lung Cancer and Personalized Medicine: Current Knowledge and Therapies*, ser. Advances in Experimental Medicine and Biology, A. Ahmad and S. Gadgeel, Eds., Cham: Springer International Publishing, 2016, pp. 43–57, ISBN: 978-3-319-24223-1. DOI: 10.1007/978-3-319-24223-1_3.

[23] H. A. Wakelee *et al.*, "Lung cancer incidence in never smokers," *Journal of Clinical Oncology*, vol. 25, pp. 472–478, Feb. 2007. DOI: 10.1200/jco.2006.07.2983.

[24] D. I. Suster and M. Mino-Kenudson, "Molecular Pathology of Primary Non-small Cell Lung Cancer," *Archives of Medical Research*, Cancer: A Chronic Pandemic, vol. 51, no. 8, pp. 784–798, Nov. 2020. DOI: 10.1016/j.arcmed.2020.08.004.

[25] P. Wheatley-Price *et al.*, "The influence of sex and histology on outcomes in non-small-cell lung cancer: A pooled analysis of five randomized trials," English, *Annals of Oncology*, vol. 21, no. 10, pp. 2023–2028, Oct. 2010. DOI: 10.1093/annonc/mdq067.

[26] C. M. North and D. C. Christiani, "Women and Lung Cancer: What is New?" *Seminars in Thoracic and Cardiovascular Surgery*, vol. 25, no. 2, pp. 87–94, Jun. 2013. DOI: 10.1053/j.semtcvs.2013.05.002.

[27] D. P. Ryan, T. S. Hong, and N. Bardeesy, "Pancreatic Adenocarcinoma," *New England Journal of Medicine*, vol. 371, no. 11, pp. 1039–1049, Sep. 2014. DOI: 10.1056/NEJMra1404198.

[28] A. Maitra and R. H. Hruban, "Pancreatic Cancer," *Annual Review of Pathology: Mechanisms of Disease*, vol. 3, no. 1, pp. 157–188, 2008. DOI: 10.1146/annurev.pathmechdis.3.121806.154305.

[29] A. Di Federico *et al.*, "Immunotherapy in Pancreatic Cancer: Why Do We Keep Failing? A Focus on Tumor Immune Microenvironment, Predictive Biomarkers and Treatment Outcomes," eng, *Cancers*, vol. 14, no. 10, p. 2429, May 2022. DOI: 10.3390/cancers14102429.

[30] N. A. Ullman, P. R. Burchard, R. F. Dunne, and D. C. Linehan, "Immunologic Strategies in Pancreatic Cancer: Making Cold Tumors Hot," eng, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 40, no. 24, pp. 2789–2805, Aug. 2022. DOI: 10.1200/JCO.21.02616.

[31] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLoS Computational Biology*, vol. 13, no. 5, e1005457, May 2017. DOI: 10.1371/journal.pcbi.1005457.

[32] V. Romanov, S. N. Davidoff, A. R. Miles, D. W. Grainger, B. K. Gale, and B. D. Brooks, "A critical comparison of protein microarray fabrication technologies," en, *Analyst*, vol. 139, no. 6, pp. 1303–1326, Feb. 2014. DOI: 10.1039/C3AN01577G.

[33] H. Okayama *et al.*, "Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas," *Cancer Research*, vol. 72, no. 1, p. 100, Jan. 2012. DOI: 10.1158/0008-5472.CAN-11-1403.

[34] S. A. Selamat *et al.*, "Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression," eng, *Genome Research*, vol. 22, no. 7, pp. 1197–1211, Jul. 2012.

[35] S. Yang *et al.*, "A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2," *Cancer Research*, vol. 76, no. 13, pp. 3838–3850, Jun. 2016. DOI: 10.1158/0008-5472.CAN-15-2841.

[36] A. I. Robles *et al.*, "An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma Based on mRNA, microRNA, and DNA Methylation Biomarkers," *Journal of Thoracic Oncology*, vol. 10, no. 7, pp. 1037–1048, Jul. 2015. DOI: 10.1097/JTO.0000000000000560.

[37] L. Girard *et al.*, "An Expression Signature as an Aid to the Histologic Classification of Non–Small Cell Lung Cancer," *Clinical Cancer Research*, vol. 22, no. 19, p. 4880, Oct. 2016. DOI: 10.1158/1078-0432.CCR-15-2900.

[38] Y. Kong *et al.*, "circNFIB1 inhibits lymphangiogenesis and lymphatic metastasis via the miR-486-5p/PIK3R1/VEGF-C axis in pancreatic cancer," *Molecular Cancer*, vol. 19, no. 1, p. 82, May 2020. DOI: 10.1186/s12943-020-01205-6.

[39] J. Lin *et al.*, "Network-based integration of mRNA and miRNA profiles reveals new target genes involved in pancreatic cancer," eng, *Molecular Carcinogenesis*, vol. 58, no. 2, pp. 206–218, Feb. 2019. DOI: 10.1002/mc.22920.

[40] T. Goldmann *et al.*, "PD-L1 amplification is associated with an immune cell rich phenotype in squamous cell cancer of the lung," en, *Cancer Immunology, Immunotherapy*, vol. 70, no. 9, pp. 2577–2587, Sep. 2021. DOI: 10.1007/s00262-020-02825-z.

[41] I. Pérez-Díez *et al.*, "A Comprehensive Transcriptional Signature in Pancreatic Ductal Adenocarcinoma Reveals New Insights into the Immune and Desmoplastic Microenvironments," en, *Cancers*, vol. 15, no. 11, p. 2887, Jan. 2023. DOI: 10.3390/cancers15112887.

[42] I. Pérez-Díez *et al.*, "Functional Signatures in Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Sex-Based Differences in Transcriptomic Studies," *Cancers*, vol. 13, no. 1, 2021. DOI: 10.3390/cancers13010143.

[43] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160 018, Mar. 2016. DOI: 10.1038/sdata.2016.18.

[44] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," eng, *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, Jan. 2002. DOI: 10.1093/nar/30.1.207.

[45] E. Clough and T. Barrett, "The Gene Expression Omnibus Database," eng, *Methods in Molecular Biology (Clifton, N.J.)*, vol. 1418, pp. 93–110, 2016. DOI: 10.1007/978-1-4939-3578-9_5.

[46] A. Brazma *et al.*, "ArrayExpress—a public repository for microarray gene expression data at the EBI," *Nucleic Acids Research*, vol. 31, no. 1, pp. 68–71, Jan. 2003. DOI: 10.1093/nar/gkg091.

[47] Cancer Genome Atlas Research Network *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project," eng, *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, Oct. 2013. DOI: 10.1038/ng.2764.

[48] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," eng, *PLoS medicine*, vol. 6, no. 7, e1000097, Jul. 2009. DOI: 10.1371/journal.pmed.1000097.

[49] N. Shaheen *et al.*, "Appraising systematic reviews: A comprehensive guide to ensuring validity and reliability," *Frontiers in Research Metrics and Analytics*, vol. 8, p. 1 268 045, Dec. 2023. DOI: 10.3389/frma.2023.1268045.

[50] F. Conforti *et al.*, "Cancer immunotherapy efficacy and patients' sex: A systematic review and meta-analysis," *The Lancet Oncology*, vol. 19, no. 6, pp. 737–746, Jun. 2018. DOI: 10.1016/S1470-2045(18)30261-4.

[51] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," en, *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020. DOI: 10.3322/caac.21590.

[52] M. Barquín *et al.*, "Sex is a strong prognostic factor in stage IV non-small-cell lung cancer patients and should be considered in survival rate estimation," *Cancer Epidemiology*, vol. 67, p. 101 737, Aug. 2020. DOI: 10.1016/j.canep.2020.101737.

[53] A. Jemal *et al.*, "Higher Lung Cancer Incidence in Young Women Than Young Men in the United States," *New England Journal of Medicine*, vol. 378, no. 21, pp. 1999–2009, May 2018. DOI: 10.1056/NEJMoa1715907.

[54] M. J. Fasco, G. J. Hurteau, and S. D. Spivack, "Gender-dependent expression of alpha and beta estrogen receptors in human nontumor and tumor lung tissue," en, *Molecular and Cellular Endocrinology*, vol. 188, no. 1, pp. 125–140, Feb. 2002. DOI: 10.1016/S0303-7207(01)00750-X.

[55] L. P. Stabile *et al.*, "Human Non-Small Cell Lung Tumors and Cells Derived from Normal Lung Express Both Estrogen Receptor α and β and Show Biological Responses to Estrogen," en, *Cancer Research*, vol. 62, no. 7, pp. 2141–2150, Apr. 2002.

[56] H. Kawai *et al.*, "Estrogen Receptor α and β are Prognostic Factors in Non–Small Cell Lung Cancer," en, *Clinical Cancer Research*, vol. 11, no. 14, pp. 5084–5089, Jul. 2005. DOI: 10.1158/1078-0432.CCR-05-0200.

[57] C. Bain *et al.*, "Lung Cancer Rates in Men and Women With Comparable Histories of Smoking," *JNCI: Journal of the National Cancer Institute*, vol. 96, no. 11, pp. 826–834, Jun. 2004. DOI: 10.1093/jnci/djh143.

[58] S. Kligerman and C. White, "Epidemiology of Lung Cancer in Women: Risk Factors, Survival, and Screening," *American Journal of Roentgenology*, vol. 196, no. 2, pp. 287–295, Feb. 2011. DOI: 10.2214/AJR.10.5412.

[59] W. D. Travis *et al.*, "International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma," *Journal of Thoracic Oncology*, vol. 6, no. 2, pp. 244–285, Feb. 2011. DOI: 10.1097/JTO.0b013e318206a221.

[60] J. M. Araujo *et al.*, "Repeated observation of immune gene sets enrichment in women with non-small cell lung cancer," *Oncotarget*, vol. 7, no. 15, pp. 20 282–20 292, Mar. 2016. DOI: 10.18632/oncotarget.7943.

[61] Y. Yuan *et al.*, "Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients," English, *Cancer Cell*, vol. 29, no. 5, pp. 711–722, May 2016. DOI: 10.1016/j.ccell.2016.04.001.

[62] K. Shi, N. Li, M. Yang, and W. Li, "Identification of Key Genes and Pathways in Female Lung Cancer Patients Who Never Smoked by a Bioinformatics Analysis," en, *Journal of Cancer*, vol. 10, no. 1, pp. 51–60, 2019. DOI: 10.7150/jca.26908.

[63] Y. Li, C.-L. He, W.-X. Li, R.-X. Zhang, and Y. Duan, "Transcriptome analysis reveals gender-specific differences in overall metabolic response of male and female patients in lung adenocarcinoma," en, *PLOS ONE*, vol. 15, no. 4, e0230796, Apr. 2020. DOI: 10.1371/journal.pone.0230796.

[64] S. L. Normand, "Meta-analysis: Formulating, evaluating, combining, and reporting," eng, *Statistics in Medicine*, vol. 18, no. 3, pp. 321–359, Feb. 1999. DOI: 10.1002/(sici)1097-0258(19990215)18:3<321::aid-sim28>3.0.co;2-p.

[65] J. P. T. Higgins *et al.*, *Cochrane Handbook for Systematic Reviews of Interventions*, en. John Wiley & Sons, Sep. 2019, ISBN: 978-1-119-53661-1.

[66] M. T. Landi *et al.*, "Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival," *PLOS ONE*, vol. 3, no. 2, e1651, Feb. 2008. DOI: 10.1371/journal.pone.0001651.

[67] J. Hou *et al.*, "Gene Expression-Based Classification of Non-Small Cell Lung Carcinomas and Survival Prediction," *PLOS ONE*, vol. 5, no. 4, e10312, Apr. 2010. DOI: 10.1371/journal.pone.0010312.

[68] M. Yamauchi *et al.*, "Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma," *PLOS ONE*, vol. 7, no. 9, e43923, Sep. 2012. DOI: 10.1371/journal.pone.0043923.

[69] A. Mezheyeuski *et al.*, "Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals distinct immune profiles that classify lung cancer patients," *The Journal of Pathology*, vol. 244, no. 4, pp. 421–431, Apr. 2018. DOI: 10.1002/path.5026.

[70] Z. Sun *et al.*, "Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers," *BMC Medical Genomics*, vol. 7, no. 1, p. 486, Jun. 2014. DOI: 10.1186/1755-8794-7-32.

[71] M. B. Cook *et al.*, "Sex Disparities in Cancer Incidence by Period and Age," en, *Cancer Epidemiology and Prevention Biomarkers*, vol. 18, no. 4, pp. 1174–1182, Apr. 2009. DOI: 10.1158/1055-9965.EPI-08-1118.

[72] S. L. Klein and K. L. Flanagan, "Sex differences in immune responses," en, *Nature Reviews Immunology*, vol. 16, no. 10, pp. 626–638, Oct. 2016. DOI: 10.1038/nri.2016.90.

[73] M. J. Lukey, W. P. Katt, and R. A. Cerione, "Targeting amino acid metabolism for cancer therapy," en, *Drug Discovery Today*, vol. 22, no. 5, pp. 796–804, May 2017. DOI: 10.1016/j.drudis.2016.12.003.

[74] A. Poff, A. P. Koutnik, K. M. Egan, S. Sahebjam, D. D'Agostino, and N. B. Kumar, "Targeting the Warburg effect for cancer treatment: Ketogenic diets for management of glioma," en, *Seminars in Cancer Biology*, Current Vision on Target Enzymes for Cancer Therapy, vol. 56, pp. 135–148, Jun. 2019. DOI: 10.1016/j.semcancer.2017.12.011.

[75] L. M. Butler *et al.*, "Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention," en, *Advanced Drug Delivery Reviews*, Jul. 2020. DOI: 10.1016/j.addr.2020.07.013.

[76] P. Gaignard *et al.*, "Effect of Sex Differences on Brain Mitochondrial Function and Its Suppression by Ovariectomy and in Aged Mice," en, *Endocrinology*, vol. 156, no. 8, pp. 2893–2904, Aug. 2015. DOI: 10.1210/en.2014-1913.

[77] S. K. Pal and A. Hurria, "Impact of Age, Sex, and Comorbidity on Cancer Therapy and Disease Progression," *Journal of Clinical Oncology*, vol. 28, no. 26, pp. 4086–4093, Sep. 2010. DOI: 10.1200/JCO.2009.27.0579.

[78] L. Mervic, "Time Course and Pattern of Metastasis of Cutaneous Melanoma Differ between Men and Women," en, *PLOS ONE*, vol. 7, no. 3, e32955, Mar. 2012. DOI: 10.1371/journal.pone.0032955.

[79] S. Sun, J. H. Schiller, and A. F. Gazdar, "Lung cancer in never smokers — a different disease," en, *Nature Reviews Cancer*, vol. 7, no. 10, pp. 778–790, Oct. 2007. DOI: 10.1038/nrc2190.

[80] A. M. Kim, C. M. Tingen, and T. K. Woodruff, "Sex bias in trials and treatment must end," *Nature*, vol. 465, no. 7299, pp. 688–689, Jun. 2010. DOI: 10.1038/465688a.

[81] N. C. Woitowich, A. Beery, and T. Woodruff, "A 10-year follow-up study of sex inclusion in the biological sciences," *eLife*, vol. 9, C. Sugimoto, P. Rodgers, R. Shansky, and L. Schiebinger, Eds., e56344, Jun. 2020. DOI: 10.7554/eLife.56344.

[82] D. S. Vinay *et al.*, "Immune evasion in cancer: Mechanistic basis and therapeutic strategies," eng, *Seminars in Cancer Biology*, vol. 35 Suppl, S185–S198, Dec. 2015. DOI: 10.1016/j.semcancer.2015.03.004.

[83] N. K. Altorki *et al.*, "The lung microenvironment: An important regulator of tumour growth and metastasis," *Nature Reviews Cancer*, vol. 19, no. 1, pp. 9–31, Jan. 2019. DOI: 10.1038/s41568-018-0081-9.

[84] Y. Ye *et al.*, "Sex-associated molecular differences for cancer immunotherapy," *Nature Communications*, vol. 11, no. 1, p. 1779, Apr. 2020. DOI: 10.1038/s41467-020-15679-x.

[85] D.-Q. Zeng *et al.*, "Prognostic and predictive value of tumor-infiltrating lymphocytes for clinical therapeutic research in patients with non-small cell lung cancer," *Oncotarget*, vol. 7, no. 12, 2016. DOI: 10.18632/oncotarget.7282.

[86] S.-L. Ye, X.-Y. Li, K. Zhao, and T. Feng, "High expression of cd8 predicts favorable prognosis in patients with lung adenocarcinoma," *Medicine*, vol. 96, e6472, Apr. 2017. DOI: 10.1097/md.0000000000006472.

[87] S. Li, Z. Wang, and X.-J. Li, "Notch signaling pathway suppresses CD8+ T cells activity in patients with lung adenocarcinoma," en, *International Immunopharmacology*, vol. 63, pp. 129–136, Oct. 2018. DOI: 10.1016/j.intimp.2018.07.033.

[88] A. K. Abbas, E. Trotta, D. R. Simeonov, A. Marson, and J. A. Bluestone, "Revisiting IL-2: Biology and therapeutic prospects," *Science Immunology*, vol. 3, no. 25, eaat1482, Jul. 2018. DOI: 10.1126/sciimmunol.aat1482.

[89] J. M. Vahl *et al.*, "Interleukin-10-regulated tumour tolerance in non-small cell lung cancer," *British Journal of Cancer*, vol. 117, no. 11, pp. 1644–1655, Nov. 2017. DOI: 10.1038/bjc.2017.336.

[90] Y. Gao *et al.*, "IL-10 suppresses IFN-ɣ-mediated signaling in lung adenocarcinoma," *Clinical and Experimental Medicine*, vol. 20, no. 3, pp. 449–459, Aug. 2020. DOI: 10.1007/s10238-020-00626-3.

[91] D. Miotto *et al.*, "CD8+ T cells expressing IL-10 are associated with a favourable prognosis in lung cancer," *Lung Cancer*, vol. 69, no. 3, pp. 355–360, Sep. 2010. DOI: 10.1016/j.lungcan.2009.12.012.

[92] J. Emmerich *et al.*, "IL-10 Directly Activates and Expands Tumor-Resident CD8+ T Cells without De Novo Infiltration from Secondary Lymphoid Organs," *Cancer Research*, vol. 72, no. 14, p. 3570, Jul. 2012. DOI: 10.1158/0008-5472.CAN-12-0721.

[93] G.-S. Shang, L. Liu, and Y.-W. Qin, "IL-6 and TNF-α promote metastasis of lung cancer by inducing epithelial-mesenchymal transition," *Oncology Letters*, vol. 13, no. 6, pp. 4657–4660, Jun. 2017. DOI: 10.3892/ol.2017.6048.

[94] E. M. Silva *et al.*, "High systemic IL-6 is associated with worse prognosis in patients with non-small cell lung cancer," *PLOS ONE*, vol. 12, no. 7, e0181125, Jul. 2017. DOI: 10.1371/journal.pone.0181125.

[95] A. M. Lewis, S. Varghese, H. Xu, and H. R. Alexander, "Interleukin-1 and cancer progression: The emerging role of interleukin-1 receptor antagonist as a novel therapeutic agent in cancer treatment," *Journal of Translational Medicine*, vol. 4, no. 1, p. 48, Nov. 2006. DOI: 10.1186/1479-5876-4-48.

[96] F. Wu *et al.*, "The role of interleukin-17 in lung cancer," *Mediators of Inflammation*, vol. 2016, pp. 1–6, 2016. DOI: 10.1155/2016/8494079.

[97] A. Gottschlich, S. Endres, and S. Kobold, "Can we use interleukin-1β blockade for lung cancer treatment?" *Translational Lung Cancer Research*, vol. 7, S160–S164, 2018. DOI: 10.21037/tlcr.2018.03.15.

[98] C. Wang *et al.*, "Effect of sex on the efficacy of patients receiving immune checkpoint inhibitors in advanced non-small cell lung cancer," *Cancer Medicine*, vol. 8, no. 8, pp. 4023–4031, Jul. 2019. DOI: 10.1002/cam4.2280.

[99] D. Vijayan, M. J. Smyth, and M. W. L. Teng, "Purinergic Receptors: Novel Targets for Cancer Immunotherapy," in *Oncoimmunology: A Practical Guide for Cancer Immunotherapy*, L. Zitvogel and G. Kroemer, Eds., Cham: Springer International Publishing, 2018, pp. 115–141, ISBN: 978-3-319-62431-0. DOI: 10.1007/978-3-319-62431-0_7.

[100] J. M. Crain, M. Nikodemova, and J. J. Watters, "Expression of P2 nucleotide receptors varies with age and sex in murine brain microglia," *Journal of Neuroinflammation*, vol. 6, no. 1, p. 24, Aug. 2009. DOI: 10.1186/1742-2094-6-24.

[101] E. Tak *et al.*, "Upregulation of P2Y2 nucleotide receptor in human hepatocellular carcinoma cells," eng, *The Journal of International Medical Research*, vol. 44, no. 6, pp. 1234–1247, Dec. 2016. DOI: 10.1177/0300060516662135.

[102] L. E. B. Savio, P. de Andrade Mello, C. G. da Silva, and R. Coutinho-Silva, "The P2X7 Receptor in Inflammatory Diseases: Angel or Demon?" eng, *Frontiers in Pharmacology*, vol. 9, p. 52, 2018. DOI: 10.3389/fphar.2018.00052.

[103] A. Asif, M. Khalid, S. Manzoor, H. Ahmad, and A. U. Rehman, "Role of purinergic receptors in hepatobiliary carcinoma in Pakistani population: An approach towards proinflammatory role of P2X4 and P2X7 receptors," eng, *Purinergic Signalling*, vol. 15, no. 3, pp. 367–374, 2019. DOI: 10.1007/s11302-019-09675-0.

[104] L.-P. Hu *et al.*, "Targeting Purinergic Receptor P2Y2 Prevents the Growth of Pancreatic Ductal Adenocarcinoma by Inhibiting Cancer Cell Glycolysis," eng, *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 25, no. 4, pp. 1318–1330, 2019. DOI: 10.1158/1078-0432.CCR-18-2297.

[105] M. Xu *et al.*, "The lysosomal TRPML1 channel regulates triple negative breast cancer development by promoting mTORC1 and purinergic signaling pathways," eng, *Cell Calcium*, vol. 79, pp. 80–88, 2019. DOI: 10.1016/j.ceca.2019.02.010.

[106] Y. Zhu, X. Shao, X. Wang, L. Liu, and H. Liang, "Sex disparities in cancer," en, *Cancer Letters*, vol. 466, pp. 35–38, Dec. 2019. DOI: 10.1016/j.canlet.2019.08.017.

[107] J. W. Shay and W. E. Wright, "Senescence and immortalization: Role of telomeres and telomerase," en, *Carcinogenesis*, vol. 26, no. 5, pp. 867–874, May 2005. DOI: 10.1093/carcin/bgh296.

[108] A. Datta and R. M. Brosh, "New Insights Into DNA Helicases as Druggable Targets for Cancer Therapy," *Frontiers in Molecular Biosciences*, vol. 5, p. 59, 2018. DOI: 10.3389/fmolb.2018.00059.

[109] R. Abbotts *et al.*, "DNA methyltransferase inhibitors induce a BRCAness phenotype that sensitizes NSCLC to PARP inhibitor and ionizing radiation," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, p. 22 609, Nov. 2019. DOI: 10.1073/pnas.1903765116.

[110] J. Luo *et al.*, "Fluzoparib increases radiation sensitivity of non-small cell lung cancer (NSCLC) cells without BRCA1/2 mutation, a novel PARP1 inhibitor undergoing clinical trials," *Journal of Cancer Research and Clinical Oncology*, vol. 146, no. 3, pp. 721–737, Mar. 2020. DOI: 10.1007/s00432-019-03097-6.

[111] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/.

[112] A. Athar *et al.*, "ArrayExpress update - from bulk to single-cell expression data," eng, *Nucleic Acids Research*, vol. 47, no. D1, pp. D711–D715, Jan. 2019. DOI: 10.1093/nar/gky964.

[113] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data," eng, *Bioinformatics (Oxford, England)*, vol. 26, no. 1, pp. 139–140, Jan. 2010. DOI: 10.1093/bioinformatics/btp616.

[114] M. D. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, no. 3, R25, Mar. 2010. DOI: 10.1186/gb-2010-11-3-r25.

[115] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: Gene-centered information at NCBI," en, *Nucleic Acids Research*, vol. 33, no. suppl_1, pp. D54–D58, Jan. 2005. DOI: 10.1093/nar/gki031.

[116] S. Durinck *et al.*, "BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis," en, *Bioinformatics*, vol. 21, no. 16, pp. 3439–3440, Aug. 2005. DOI: 10.1093/bioinformatics/bti525.

[117] M. E. Ritchie *et al.*, "Limma powers differential expression analyses for RNA-sequencing and microarray studies," eng, *Nucleic Acids Research*, vol. 43, no. 7, e47, Apr. 2015. DOI: 10.1093/nar/gkv007.

[118] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," en, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

[119] D. Montaner and J. Dopazo, "Multidimensional Gene Set Analysis of Genomic Data," *PLOS ONE*, vol. 5, no. 4, e10348, Apr. 2010. DOI: 10.1371/journal.pone.0010348.

[120] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," en, *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, Jan. 2000. DOI: 10.1093/nar/28.1.27.

[121] M. Ashburner *et al.*, "Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium," eng, *Nature Genetics*, vol. 25, no. 1, pp. 25–29, May 2000. DOI: 10.1038/75556.

[122] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "UpSet: Visualization of Intersecting Sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1983–1992, Dec. 2014.

[123] F. García-García, "Methods of functional enrichment analysis in genomic studies," Ph.D. dissertation, Universitat de València, València, 2016.

[124] W. Viechtbauer, "Conducting Meta-Analyses in R with the metafor Package," en, *Journal of Statistical Software*, vol. 36, pp. 1–48, Aug. 2010. DOI: 10.18637/jss.v036.i03.

[125] R. DerSimonian and N. Laird, "Meta-analysis in clinical trials," eng, *Controlled Clinical Trials*, vol. 7, no. 3, pp. 177–188, Sep. 1986. DOI: 10.1016/0197-2456(86)90046-2.

[126] J. A. Sterne and M. Egger, "Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis," eng, *Journal of Clinical Epidemiology*, vol. 54, no. 10, pp. 1046–1055, Oct. 2001. DOI: 10.1016/s0895-4356(01)00377-8.

[127] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN: 978-3-319-24277-4. [Online]. Available: https://ggplot2.tidyverse.org.

[128] W. Park, A. Chawla, and E. M. O'Reilly, "Pancreatic Cancer: A Review," eng, *JAMA*, vol. 326, no. 9, pp. 851–862, Sep. 2021. DOI: 10.1001/jama.2021.13027.

[129] A. Henriksen, A. Dyhl-Polk, I. Chen, and D. Nielsen, "Checkpoint inhibitors in pancreatic cancer," eng, *Cancer Treatment Reviews*, vol. 78, pp. 17–30, Aug. 2019. DOI: 10.1016/j.ctrv.2019.06.005.

[130] F. N. Shamohammadi *et al.*, "Controversial role of γδ T cells in pancreatic cancer," eng, *International Immunopharmacology*, vol. 108, p. 108 895, Jul. 2022. DOI: 10.1016/j.intimp.2022.108895.

[131] C. Dong, D. Dang, X. Zhao, Y. Wang, Z. Wang, and C. Zhang, "Integrative Characterization of the Role of IL27 In Melanoma Using Bioinformatics Analysis," eng, *Frontiers in Immunology*, vol. 12, p. 713 001, 2021. DOI: 10.3389/fimmu.2021.713001.

[132] L. Ostios-Garcia, J. Villamayor, E. Garcia-Lorenzo, D. Vinal, and J. Feliu, "Understanding the immune response and the current landscape of immunotherapy in pancreatic cancer," eng, *World Journal of Gastroenterology*, vol. 27, no. 40, pp. 6775–6793, Oct. 2021. DOI: 10.3748/wjg.v27.i40.6775.

[133] R. A. Moffitt *et al.*, "Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma," en, *Nature Genetics*, vol. 47, no. 10, pp. 1168–1178, Oct. 2015. DOI: 10.1038/ng.3398.

[134] P. Bailey *et al.*, "Genomic analyses identify molecular subtypes of pancreatic cancer," en, *Nature*, vol. 531, no. 7592, pp. 47–52, Mar. 2016. DOI: 10.1038/nature16965.

[135] T. Wu *et al.*, "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data," English, *The Innovation*, vol. 2, no. 3, Aug. 2021. DOI: 10.1016/j.xinn.2021.100141.

[136] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: An R package for comparing biological themes among gene clusters," eng, *Omics: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, May 2012. DOI: 10.1089/omi.2011.0118.

[137] G. Yu and Q.-Y. He, "Reactomepa: An r/bioconductor package for reactome pathway analysis and visualization," *Molecular BioSystems*, vol. 12, pp. 477–479, 2016. DOI: 10.1039/c5mb00663e.

[138] Gene Ontology Consortium, "The Gene Ontology resource: Enriching a GOld mine," eng, *Nucleic Acids Research*, vol. 49, no. D1, pp. D325–D334, Jan. 2021. DOI: 10.1093/nar/gkaa1113.

[139] M. Gillespie *et al.*, "The reactome pathway knowledgebase 2022," eng, *Nucleic Acids Research*, vol. 50, no. D1, pp. D687–D692, Jan. 2022. DOI: 10.1093/nar/gkab1028.

[140] S. Sayols, "Rrvgo: A Bioconductor package to reduce and visualize Gene Ontology terms," *microPublication Biology*, vol. 2023, 2023. DOI: 10.17912/micropub.biology.000811.

[141] E. Cerami *et al.*, "The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data," eng, *Cancer Discovery*, vol. 2, no. 5, pp. 401–404, May 2012. DOI: 10.1158/2159-8290.CD-12-0095.

[142] E. Hessmann *et al.*, "Microenvironmental determinants of pancreatic cancer," *Physiological Reviews*, vol. 100, pp. 1707–1751, Oct. 2020. DOI: 10.1152/physrev.00042.2019.

[143] C. J. Whatcott *et al.*, "Desmoplasia in Primary Tumors and Metastatic Lesions of Pancreatic Cancer," eng, *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 21, no. 15, pp. 3561–3568, Aug. 2015. DOI: 10.1158/1078-0432.CCR-14-1051.

[144] X. Zhou, Y. Liu, M. Hu, M. Wang, X. Liu, and L. Huang, "Relaxin gene delivery modulates macrophages to resolve cancer fibrosis and synergizes with immune checkpoint blockade therapy," eng, *Science Advances*, vol. 7, no. 8, eabb6596, Feb. 2021. DOI: 10.1126/sciadv.abb6596.

[145] H. Wang, R. Ren, Z. Yang, J. Cai, S. Du, and X. Shen, "The COL11A1/Akt/CREB signaling axis enables mitochondrial-mediated apoptotic evasion to promote chemoresistance in pancreatic cancer cells through modulating BAX/BCL-2 function," eng, *Journal of Cancer*, vol. 12, no. 5, pp. 1406–1420, 2021. DOI: 10.7150/jca.47032.

[146] X. Zheng, X. Liu, H. Zheng, H. Wang, and D. Hong, "Integrated bioinformatics analysis identified COL11A1 as an immune infiltrates correlated prognosticator in pancreatic adenocarcinoma," eng, *International Immunopharmacology*, vol. 90, p. 106 982, Jan. 2021. DOI: 10.1016/j.intimp.2020.106982.

[147] Y.-T. Liu and Z.-J. Sun, "Turning cold tumors into hot tumors by improving T-cell infiltration," eng, *Theranostics*, vol. 11, no. 11, pp. 5365–5386, 2021. DOI: 10.7150/thno.58390.

[148] C. Neuzillet *et al.*, "Periostin- and podoplanin-positive cancer-associated fibroblast subtypes cooperate to shape the inflamed tumor microenvironment in aggressive pancreatic adenocarcinoma," eng, *The Journal of Pathology*, vol. 258, no. 4, pp. 408–425, Dec. 2022. DOI: 10.1002/path.6011.

[149] J. Jiang, H.-L. Liu, Z.-H. Liu, S.-W. Tan, and B. Wu, "Identification of cystatin SN as a novel biomarker for pancreatic cancer," eng, *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 36, no. 5, pp. 3903–3910, May 2015. DOI: 10.1007/s13277-014-3033-3.

[150] C. Chavez-Muñoz, J. Morse, R. Kilani, and A. Ghahary, "Primary human keratinocytes externalize stratifin protein via exosomes," eng, *Journal of Cellular Biochemistry*, vol. 104, no. 6, pp. 2165–2173, Aug. 2008. DOI: 10.1002/jcb.21774.

[151] I. M. Berquin and B. F. Sloane, "Cathepsin B expression in human tumors," eng, *Advances in Experimental Medicine and Biology*, vol. 389, pp. 281–294, 1996. DOI: 10.1007/978-1-4613-0335-0_35.

[152] M. M. Mohamed and B. F. Sloane, "Cysteine cathepsins: Multifunctional enzymes in cancer," eng, *Nature Reviews. Cancer*, vol. 6, no. 10, pp. 764–775, Oct. 2006. DOI: 10.1038/nrc1949.

[153] S. P. Atkinson, Z. Andreu, and M. J. Vicent, "Polymer Therapeutics: Biomarkers and New Approaches for Personalized Cancer Treatment," eng, *Journal of Personalized Medicine*, vol. 8, no. 1, p. 6, Jan. 2018. DOI: 10.3390/jpm8010006.

[154] H. Wang, W. Sha, Z. Liu, and C.-W. Chi, "Effect of chymotrypsin C and related proteins on pancreatic cancer cell migration," eng, *Acta Biochimica Et Biophysica Sinica*, vol. 43, no. 5, pp. 362–371, May 2011. DOI: 10.1093/abbs/gmr022.

[155] A. González-Titos, P. Hernández-Camarero, S. Barungi, J. A. Marchal, J. Kenyon, and M. Perán, "Trypsinogen and chymotrypsinogen: Potent anti-tumor agents," eng, *Expert Opinion on Biological Therapy*, vol. 21, no. 12, pp. 1609–1621, Dec. 2021. DOI: 10.1080/14712598.2021.1922666.

[156] A. Makkouk and G. J. Weiner, "Cancer immunotherapy and breaking immune tolerance: New approaches to an old challenge," eng, *Cancer Research*, vol. 75, no. 1, pp. 5–10, Jan. 2015. DOI: 10.1158/0008-5472.CAN-14-2538.

[157] S. J. S. Rubin, R. S. Sojwal, J. Gubatan, and S. Rogalla, "The Tumor Immune Microenvironment in Pancreatic Ductal Adenocarcinoma: Neither Hot nor Cold," eng, *Cancers*, vol. 14, no. 17, p. 4236, Aug. 2022. DOI: 10.3390/cancers14174236.

[158] M. Fujisawa *et al.*, "Involvement of the Interferon Signaling Pathways in Pancreatic Cancer Cells," eng, *Anticancer Research*, vol. 40, no. 8, pp. 4445–4455, Aug. 2020. DOI: 10.21873/anticanres.14449.

[159] H. Zhuang *et al.*, "Prognostic values and immune suppression of the S100A family in pancreatic cancer," eng, *Journal of Cellular and Molecular Medicine*, vol. 25, no. 6, pp. 3006–3018, Mar. 2021. DOI: 10.1111/jcmm.16343.

[160] K. M. Herremans *et al.*, "The interleukin-1 axis and the tumor immune microenvironment in pancreatic ductal adenocarcinoma," eng, *Neoplasia (New York, N.Y.)*, vol. 28, p. 100 789, Jun. 2022. DOI: 10.1016/j.neo.2022.100789.

[161] A. Rodolosse, E. Chalaux, T. Adell, H. Hagège, A. Skoudy, and F. X. Real, "PTF1alpha/p48 transcription factor couples proliferation and differentiation in the exocrine pancreas [corrected]," eng, *Gastroenterology*, vol. 127, no. 3, pp. 937–949, Sep. 2004. DOI: 10.1053/j.gastro.2004.06.058.

[162] H. Shen *et al.*, "PLZF inhibits proliferation and metastasis of gallbladder cancer by regulating IFIT2," eng, *Cell Death & Disease*, vol. 9, no. 2, p. 71, Jan. 2018. DOI: 10.1038/s41419-017-0107-3.

[163] Y. Liu *et al.*, "Long Noncoding RNA HCP5 Regulates Pancreatic Cancer Gemcitabine (GEM) Resistance By Sponging Hsa-miR-214-3p To Target HDGF," eng, *OncoTargets and Therapy*, vol. 12, pp. 8207–8216, 2019. DOI: 10.2147/OTT.S222703.

[164] B. Yuan, Q. Guan, T. Yan, X. Zhang, W. Xu, and J. Li, "LncRNA HCP5 Regulates Pancreatic Cancer Progression by miR-140-5p/CDK8 Axis," eng, *Cancer Biotherapy & Radiopharmaceuticals*, vol. 35, no. 9, pp. 711–719, Nov. 2020. DOI: 10.1089/cbr.2019.3294.

[165] Y. Hu, B. Wang, K. Yi, Q. Lei, G. Wang, and X. Xu, "IFI35 is involved in the regulation of the radiosensitivity of colorectal cancer cells," eng, *Cancer Cell International*, vol. 21, no. 1, p. 290, Jun. 2021. DOI: 10.1186/s12935-021-01997-7.

[166] J. Xu *et al.*, "Schlafen family is a prognostic biomarker and corresponds with immune infiltration in gastric cancer," eng, *Frontiers in Immunology*, vol. 13, p. 922 138, 2022. DOI: 10.3389/fimmu.2022.922138.

[167] W. J. Ho, E. M. Jaffee, and L. Zheng, "The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities," eng, *Nature Reviews. Clinical Oncology*, vol. 17, no. 9, pp. 527–540, Sep. 2020. DOI: 10.1038/s41571-020-0363-5.

[168] A. N. Hosein, R. A. Brekken, and A. Maitra, "Pancreatic cancer stroma: An update on therapeutic targeting strategies," eng, *Nature Reviews. Gastroenterology & Hepatology*, vol. 17, no. 8, pp. 487–505, Aug. 2020. DOI: 10.1038/s41575-020-0300-1.

[169] B. Rizeq, Z. Zakaria, and A. Ouhtit, "Towards understanding the mechanisms of actions of carcinoembryonic antigen-related cell adhesion molecule 6 in cancer progression," eng, *Cancer Science*, vol. 109, no. 1, pp. 33–42, Jan. 2018. DOI: 10.1111/cas.13437.

[170] J. Zińczuk *et al.*, "Expression of Chosen Carcinoembryonic-Related Cell Adhesion Molecules in Pancreatic Intraepithelial Neoplasia (PanIN) Associated with Chronic Pancreatitis and Pancreatic Ductal Adenocarcinoma (PDAC)," eng, *International Journal of Medical Sciences*, vol. 16, no. 4, pp. 583–592, 2019. DOI: 10.7150/ijms.32751.

[171] Z.-W. Han *et al.*, "The old CEACAMs find their new role in tumor immunotherapy," eng, *Investigational New Drugs*, vol. 38, no. 6, pp. 1888–1898, Dec. 2020. DOI: 10.1007/s10637-020-00955-w.

[172] H. Zhang *et al.*, "LAMB3 mediates apoptotic, proliferative, invasive, and metastatic behaviors in pancreatic cancer by regulating the PI3K/Akt signaling pathway," eng, *Cell Death & Disease*, vol. 10, no. 3, p. 230, Mar. 2019. DOI: 10.1038/s41419-019-1320-z.

[173] Y. Okada, N. Takahashi, T. Takayama, and A. Goel, "LAMC2 promotes cancer progression and gemcitabine resistance through modulation of EMT and ATP-binding cassette transporters in pancreatic ductal adenocarcinoma," eng, *Carcinogenesis*, vol. 42, no. 4, pp. 546–556, Apr. 2021. DOI: 10.1093/carcin/bgab011.

[174] Y. Lu *et al.*, "Identification of Critical Pathways and Potential Key Genes in Poorly Differentiated Pancreatic Adenocarcinoma," eng, *OncoTargets and Therapy*, vol. 14, pp. 711–723, 2021. DOI: 10.2147/OTT.S279287.

[175] B. Johnson and D. Mahadevan, "Emerging Role and Targeting of Carcinoembryonic Antigen-related Cell Adhesion Molecule 6 (CEACAM6) in Human Malignancies," eng, *Clinical Cancer Drugs*, vol. 2, no. 2, pp. 100–111, Feb. 2015. DOI: 10.2174/2212697X02666150602215823.

[176] X. Lei, G. Chen, J. Li, W. Wen, J. Gong, and J. Fu, "Comprehensive analysis of abnormal expression, prognostic value and oncogenic role of the hub gene FN1 in pancreatic ductal adenocarcinoma via bioinformatic analysis and in vitro experiments," eng, *PeerJ*, vol. 9, e12141, 2021. DOI: 10.7717/peerj.12141.

[177] H. Fukuhisa *et al.*, "Gene regulation by antitumor miR-130b-5p in pancreatic ductal adenocarcinoma: The clinical significance of oncogenic EPS8," eng, *Journal of Human Genetics*, vol. 64, no. 6, pp. 521–534, Jun. 2019. DOI: 10.1038/s10038-019-0584-6.

[178] E. Jahny *et al.*, "The G Protein-Coupled Receptor RAI3 Is an Independent Prognostic Factor for Pancreatic Cancer Survival and Regulates Proliferation via STAT3 Phosphorylation," eng, *PloS One*, vol. 12, no. 1, e0170390, 2017. DOI: 10.1371/journal.pone.0170390.

[179] S. Islam, T. Kitagawa, B. Baron, Y. Abiko, I. Chiba, and Y. Kuramitsu, "ITGA2, LAMB3, and LAMC2 may be the potential therapeutic targets in pancreatic ductal adenocarcinoma: An integrated bioinformatics analysis," eng, *Scientific Reports*, vol. 11, no. 1, p. 10 563, May 2021. DOI: 10.1038/s41598-021-90077-x.

[180] H. Lin *et al.*, "S100A10 Promotes Pancreatic Ductal Adenocarcinoma Cells Proliferation, Migration and Adhesion through JNK/LAMB3-LAMC2 Axis," eng, *Cancers*, vol. 15, no. 1, p. 202, Dec. 2022. DOI: 10.3390/cancers15010202.

[181] F. Robin *et al.*, "Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma," eng, *British Journal of Cancer*, vol. 123, no. 1, pp. 72–80, Jul. 2020. DOI: 10.1038/s41416-020-0863-1.

[182] B. K. Schniers *et al.*, "Deletion of Slc6a14 reduces cancer growth and metastatic spread and improves survival in KPC mouse model of spontaneous pancreatic cancer," eng, *The Biochemical Journal*, vol. 479, no. 5, pp. 719–730, Mar. 2022. DOI: 10.1042/BCJ20210855.

[183] C. Zhou *et al.*, "TSPAN1 promotes autophagy flux and mediates cooperation between WNT-CTNNB1 signaling and autophagy via the MIR454-FAM83A-TSPAN1 axis in pancreatic cancer," eng, *Autophagy*, vol. 17, no. 10, pp. 3175–3195, Oct. 2021. DOI: 10.1080/15548627.2020.1826689.

[184] N. C. W. Goonesekere, X. Wang, L. Ludwig, and C. Guda, "A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers," en, *PLoS ONE*, vol. 9, no. 4, J. D. Hoheisel, Ed., e93046, Apr. 2014. DOI: 10.1371/journal.pone.0093046.

[185] A. Irigoyen *et al.*, "Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers," eng, *PloS One*, vol. 13, no. 4, e0194844, 2018. DOI: 10.1371/journal.pone.0194844.

[186] A. C. Society, *The burden of cancer*, The Cancer Atlas, 2019. [Online]. Available: https://canceratlas.cancer.org/the-burden/the-burden-of-cancer/ (visited on 02/28/2024).

[187] W. H. Organization, *Cancer tomorrow*, World Health Organization, 2022. [Online]. Available: https://gco.iarc.fr/tomorrow/en/dataviz/isotype?years=2040&single_unit=500000&types=1 (visited on 02/28/2024).

[188] B. Marte, "Tumour heterogeneity," *Nature*, vol. 501, pp. 327–327, Sep. 2013. DOI: 10.1038/501327a.

[189] F. Conforti *et al.*, "Sex-Based Heterogeneity in Response to Lung Cancer Immunotherapy: A Systematic Review and Meta-Analysis," *JNCI: Journal of the National Cancer Institute*, vol. 111, no. 8, pp. 772–781, May 2019. DOI: 10.1093/jnci/djz094.

Appendix A

Appendix: Additional Figures and Tables

Note Tables too big to be rendered in this document are available at Zenodo [1].

[1] https://zenodo.org/doi/10.5281/zenodo.11032404

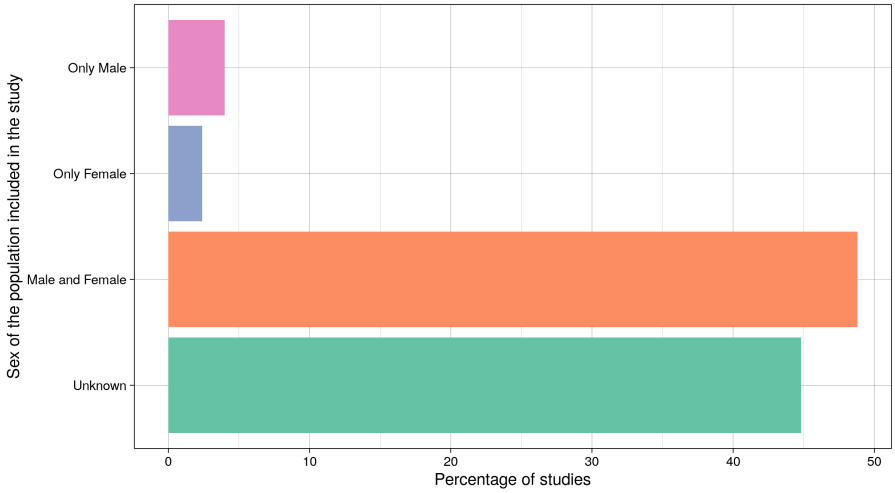
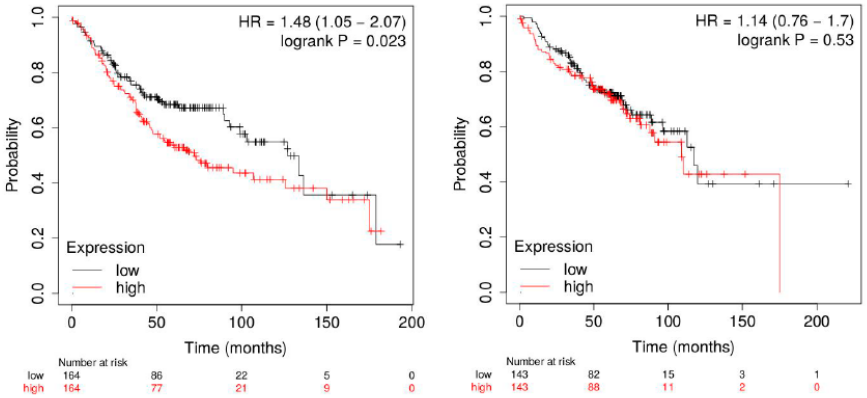
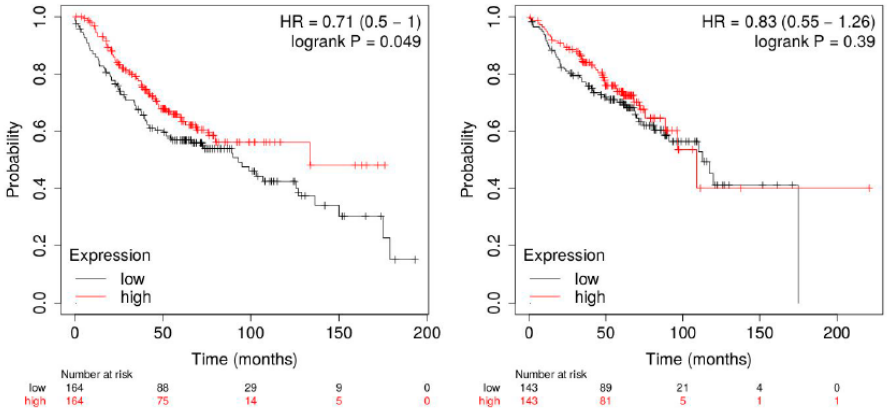


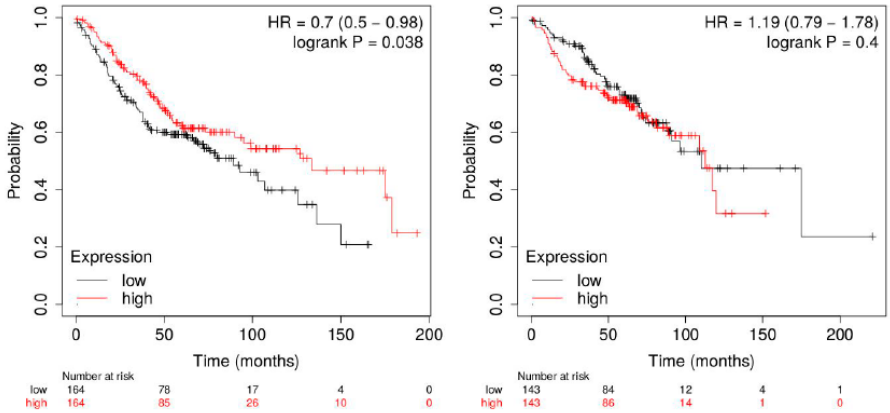
Figure A.1 – Information regarding sex distribution among reviewed studies.



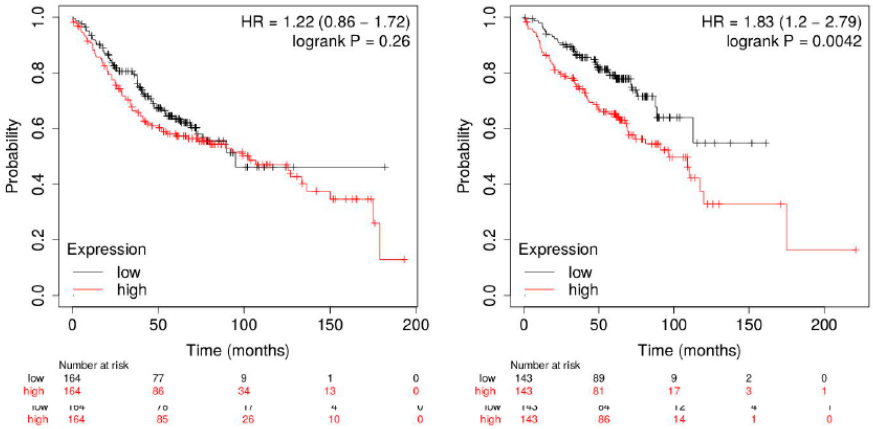
(a) Overall survival sex differences between low and high expression groups, in genes related to "negative regulation of leukocyte degranulation (GO:0043301)".



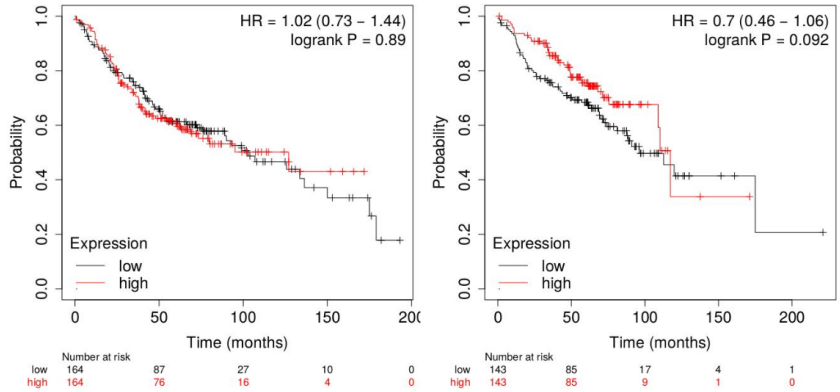
(b) Overall survival sex differences between low and high expression groups, in genes related to "inositol trisphosphate metabolic process (GO:0032957)".



(c) Overall survival sex differences between low and high expression groups, in genes related to "negative regulation of chromatin silencing (GO:0031936)".



(d) Overall survival sex differences between low and high expression groups, in genes related to "inner ear auditory receptor cell differentiation (GO:0000491)".



(e) Overall survival sex differences between low and high expression groups, in genes related to "positive regulation of leukocyte adhesion to vascular endothelial cells (GO:1904996)".

Figure A.2 – Prognostic effect of transcriptional pathways. Overall survival differences between low and high expression groups, in men (left) and women (right). Low and high groups include under or overexpressed genes respectively, related to each pathway. TCGA and GEO datasets containing expression levels and survival information were used from Kaplan-Meier Plotter web tool.

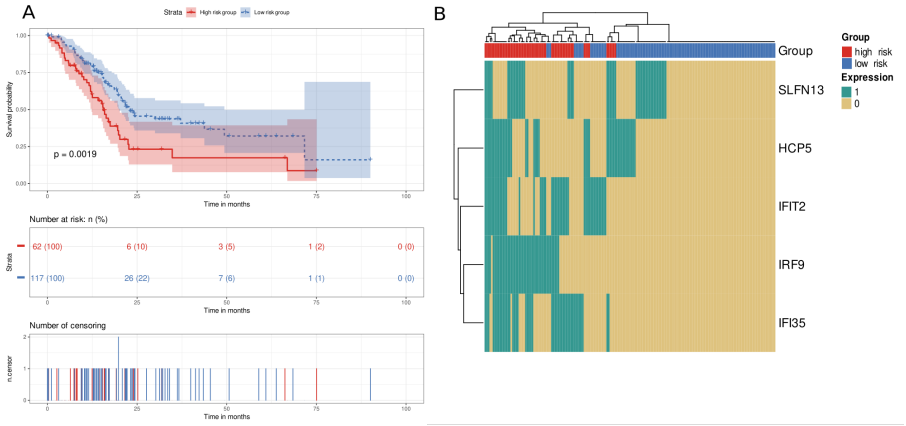


Figure A.3 – A five gene (HCP5, SLFN13, IRF9, IFIT2, and IFI35) signature clustered patients into high-risk or low-risk groups based on the number of highly expressed signature genes in their transcriptomic profile. Patients with at least two highly expressed genes were classified as having a high risk, whereas those with five or fewer were classified as having a low risk. (A) Kaplan–Meier curve. Patients from the high-risk group (red) had shorter survival times than patients from the low-risk group did (blue). Below, the number of still alive patients and percentage in each group at 0, 25, 50, 75, and 100 months, and the censored events. (B) Heatmap demonstrating the patterns of high expression between genes and samples. Gene expression was coded as 1 for a sample above the upper quartile.

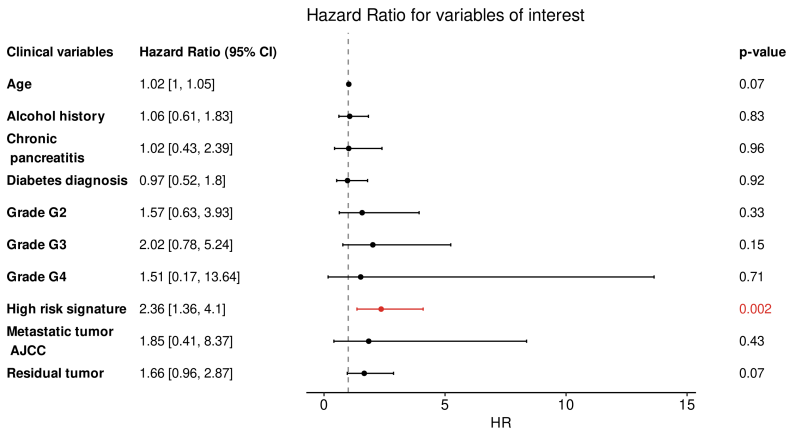


Figure A.4 – Hazard Ratio of variables of interest included in COX model. The proposed signature was the only variable with p value < 0.05.

Table A.1 – Distribution of the clinicopathological characteristics of each study population. Tumor stage percentages were computed over the total number of adenocarcinoma samples.

Study	Samples	Age (median + SD)	Sex (%)		Smoking Status (%)		Histology (%)		Stage (%) *		Mutations and Fusions (%)	
GSE10072	80	60 ± 6.68	Men Women	59 (73.75%) 21 (26.25%)	Non-smoker Smoker	19 (23.75%) 61 (76.25%)	Adenocarcinoma Control	43 (53.75%) 37 (46.25%)	IA IB IIA IIB	5 (1.63%) 17 (39.53%) 3 (6.98%) 18 (41.86%)	Not reported	
GSE19188	84	65.07 ± 10.39	Men Women	62 (73.81%) 22 (26.19%)	Non-smoker Smoker	50 (49.52%) 34 (40.47%)	Adenocarcinoma Control	32 (38.10%) 52 (61.90%)	IA IB IIB	10 (31.25%) 15 (46.88%) 7 (21.97%)	Not reported	
GSE31210	246	61 ± 8.08	Men Women	116 (47.15%) 130 (52.85%)	Non-smoker Smoker	123 (50%) 123 (50%)	Adenocarcinoma Control	226 (91.87%) 20 (8.13%)	IA IB IIA	114 (50.44%) 54 (23.89%) 58 (25.67%)	ALK EGFR KRAS Wild-type	11 (4.87%) 127 (56.19%) 20 (8.85%) 68 (30.09%)
GSE32863	90	72 ± 9.33	Men Women	20 (22.22%) 70 (77.78%)	Non-smoker Smoker	47 (52.22%) 43 (47.78%)	Adenocarcinoma Control	45 (50%) 45 (50 %)	IA IB IIA IIB	16 (35.56%) 18 (40.00%) 9 (20.00%) 2 (4.44%)	EGFR KRAS LKB1 KRAS+LKB1 Wild-type	12 (26.67%) 15 (33.33%) 4 (8.89%) 2 (4.44%) 12 (26.67%)
GSE63459	63	64 ± 11.37	Men Women	29 (46.03%) 34 (53.97%)	Non-smoker Smoker	8 (12.70%) 55 (87.30%)	Adenocarcinoma Control	32 (50.79%) 31 (49.21%)	I	32 (100%)	KRAS KRAS + tp53 tp53 Unknown Wild-type	1 (3.13%) 2 (6.25%) 10 (31.25%) 5 (15.63%) 14 (43.75%)
GSE75037	154	70 ± 9.73	Men Women	46 (29.87%) 108 (70.13%)	Non-smoker Smoker	55 (35.71%) 99 (64.29%)	Adenocarcinoma Control	71 (46.11%) 83 (53.89%)	IA IB IIA IIB	25 (35.21%) 26 (36.62%) 3 (4.23%) 17 (23.94%)	EGFR KRAS LKB1 KRAS+LKB1 Wild-type	16 (22.54%) 25 (35.21%) 6 (8.45%) 5 (7.04%) 19 (26.76%)
GSE81089	100	67.5 ± 7.37	Men Women	39 (39%) 61 (61%)	Non-smoker Smoker	8 (8%) 92 (92%)	Adenocarcinoma Control	81 (81%) 19 (19T)	IA IB IIA IIB	45 (53.32%) 18 (22.2%) 8 (9.88%) 11 (13.58%)	Not reported	
GSE87340	54	67 ± 12.44	Men Women	8 (14.81%) 46 (85.19%)	Non-smoker Smoker	54 (100%) 0 (0%)	Adenocarcinoma Control	27 (50%) 27 (50%)	IA IB I	10 (37.04%) 17 (62.96%) 5 (1.20%)	Not reported	
TCGA	458	65.79 ± 9.93	Men Women	209 (45.63%) 249 (54.37%)	Non-smoker Smoker	132 (28.82%) 326 (71.18%)	Adenocarcinoma Control	415 (90.61%) 43 (9.39%)	IA IB II IIA	136 (32.77%) 151 (36.39%) 1 (0.024%) 50 (12.05%)	Not reported	

Table A.2 – Summary of differential expression analysis results in individual studies. Two exploratory differential expression analyses were performed (ADC Women - Control Women, ADC.W -Control.W; ADC Men - Control Men, ADC.M - Control.M), together with the contrast of interest: (ADC.W - Control.W) - (ADC.M - Control.M). When performing the contrast of interest, "Up" terms are overrepresented in female lung adenocarcinoma patients, while "Down" terms are overrepresented in male lung adenocarcinoma patients.

Study		(ADC.W - ControlW) - (ADC.M - ControlM)	ADC.W - Control.W	ADC.M - Control-M
GSE10072	Up	0	1199	3182
	Down	0	1296	2688
GSE19188	Up	0	2348	5828
	Down	0	2111	3830
GSE31210	Up	0	3310	3560
	Down	0	2370	2543
GSE32863	Up	6	5243	2458
	Down	1	4507	2172
GSE63459	Up	0	2409	1561
	Down	0	2064	1611
GSE75037	Up	1	5779	4115
	Down	1	5117	3680
GSE81089	Up	3	2654	3564
	Down	0	3383	4262
GSE87340	Up	1	4887	1875
	Down	3	4958	1841
TCGA	Up	1	5861	5397
	Down	0	5569	5269

Table A.3 – Genes differentially expressed between male and female lung adenocarcinoma patients.

ENTREZ ID	Gene Name	Up / Down	logFC	adj.pval	Study
9086	Eukaryotic translation initiation factor 1A Y-linked	Up	0.606 1.514	0.014 $7.38 \cdot 10^{-7}$	GSE32863 GSE75037
146330	F-box and leucine rich repeat protein 16	Down	1.247 2.157	0.028 0.003	GSE32863 GSE75037
3394	interferon regulatory factor 8	Up	1.177	0.028	GSE32863
80301	pleckstrin homology domain containing O2	Up	0.986	0.028	GSE32863
3689	integrin subunit beta 2	Up	1.393	0.043	GSE32863
11309	solute carrier organic anion transporter family member 2B1	Up	1.123	0.049	GSE32863
83706	fermitin family member 3	Up	0.866	0.049	GSE32863
252948	testis-specific transcript, Y-linked 16	Up	1.655 2.983	$1.98 \cdot 10^{-7}$ $3.3 \cdot 10^{-14}$	GSE81089 TCGA
107987337	ZFY antisense RNA 1	Up	1.609	0.001	GSE81089
6736	sex determining region Y	Up	1.519	0.007	GSE81089
694	BTG anti-proliferation factor 1	Up	0.859	0.007	GSE87340
64582	G protein-coupled receptor 135	Down	2.213	0.007	GSE87340
22979	EFR3 homolog B	Down	1.799	0.03	GSE87340
54753	zinc finger protein 853	Down	1.581	0.04	GSE87340

Table A.4 – All significant GO terms and KEGG pathways in the functional meta-analysis. Available online at Zenodo .

Table A.5 – Genes annotated to significant GO terms and KEGG pathways in functional meta-analysis. Available online at Zenodo.

Table A.6 – Software and versions used in Pérez-Díez *et al.* [42]

Software / R package	Version
R	3.5.3
AnnotationDbi	1.44.0
Biobase	2.42.0
biomaRt	2.38.0
edgeR	3.24.3
GEOQuery	2.50.5
ggdendro	0.1-20
ggpubr	0.2
hgu133plus2.db	3.2.3
illuminaHumanv3.db	1.26.0
KEGG.db	3.2.3
limma	3.38.3
mdgsa	1.14.0
metafor	2.1-0
methods	3.5.3
org.Hs.eg.db	3.7.0
reshape	0.8.8
stats	3.5.3
SummarizedExperiment	1.12.0
TCGAbiolinks	2.10.5
tidyverse	1.2.1
UpSetR	1.3.3
utils	3.5.3

Table A.7 – Software version used in Pérez-Díez *et al.* [41]

Software / R package	Version
R	4.1.3
rrvgo	1.4.4
tidyverse	1.3.1
edgeR	3.34.1
ArrayExpress	1.52.0
affy	1.70.0
GEOquery	2.60.0
ggplot2	3.3.6
limma	3.48.3
org.Hs.eg.db	3.13.0
metafor	3.4.0
reshape	0.8.9
ggpubr	0.4.0
biomaRt	2.48.3
clusterProfiler	4.0.5
topGO	2.44
AnnotationDbi	1.54.1
rbioapi	0.7.6
GO.db	3.13.0
ReactomePA	1.36.0
survival	3.4.0
survminer	0.4.9
cBioPortalData	2.14.10

Table A.8 – PDAC dataset inclusion. List of datasets included in the study, alongside its expression profiling technology, number of samples, and inclusion/exclusion flag. Available online at Zenodo .

Table A.9 – PDAC datasets clinical characteristics. Summary table of clinicopathologic variables for all the included studies. Individual sub-tables, including all the clinicopathologic variables available for each study. Available online at Zenodo.

Table A.10 – Gene meta-analysis results. Summary statistics, name, symbol and ENSEMBL ID of genes with FDR adjusted p-value < 0.05. Available online at Zenodo.

Table A.11 – ORA Results. Summary statistics, name, and GO ID. Available online at Zenodo.

Table A.12 – NCBI and GO Immune system genes. Available online at Zenodo.

Table A.13 – Gene intersection between the defined gene signatures and other signatures in the literature. Available online at Zenodo.