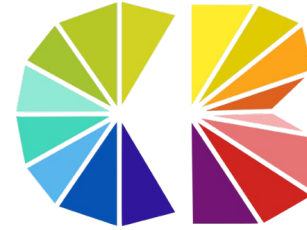




**Unidad Mixta de
Imagen Biomédica**

Fundació
Fisabio

CIBER
VALENCIA BIOMEDICAL
RESEARCH FOUNDATION



**Computational
Biomedicine**
Laboratory

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

Irene Pérez Díez
2025

PhD supervisors:
Francisco García García
María de la Iglesia Vayá



**UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA**

Índice

- I. Introducción
- II. Justificación y Objetivos
- III. Metodología Computacional Central
- IV. Identificación de Firmas Funcionales Basadas en el Sexo en NSCLC
- V. Identificación de Firmas Transcripcionales para estratificar pacientes en PDAC
- VI. Discusión General y Conclusiones

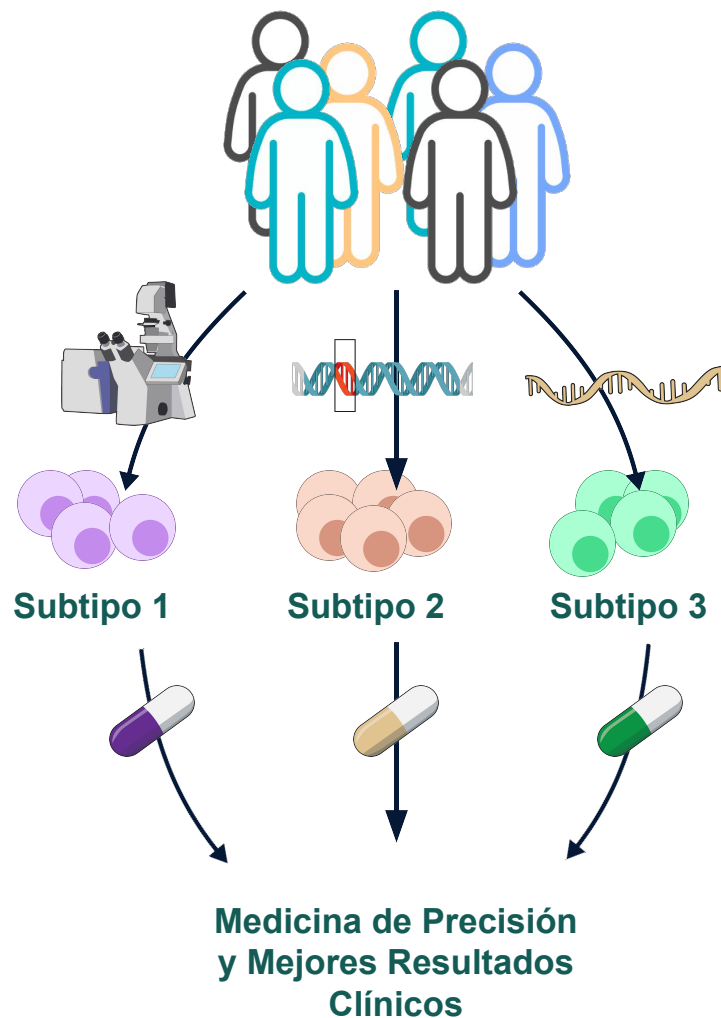
Parte I

Introducción

El reto de la Heterogeneidad en el Cáncer

El cáncer es una de las **principales causas de muerte** a nivel mundial, y presenta una **alta heterogeneidad** que dificulta la aplicación de terapias uniformemente efectivas. El **subtipado y la estratificación** de los pacientes es crucial.

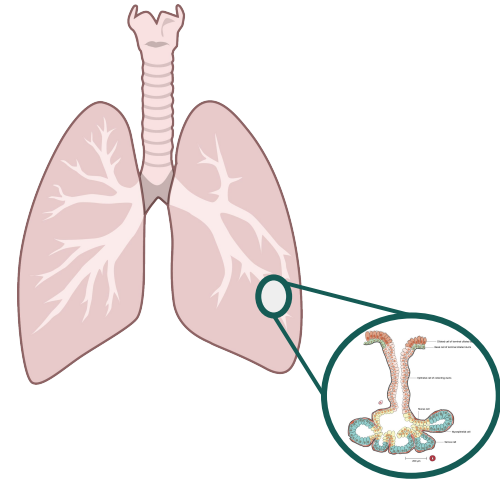
Variables clinicopatológicas relevantes para la enfermedad, como el **sexo del paciente**, suelen ser **ignoradas** en los estudios. Considerar estos factores en el subtipado es fundamental para el desarrollo de estrategias personalizadas.



Enfoque de la Investigación: Adenocarcinoma de Pulmón

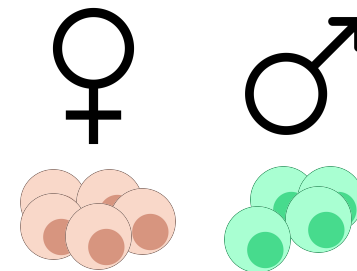
Adenocarcinoma de Pulmón (LUAD):

- > 50% de cáncer de pulmón de células no pequeñas (NSCLC).
- ~26% tasa de supervivencia a 5 años.
- Frecuente en no fumadores.

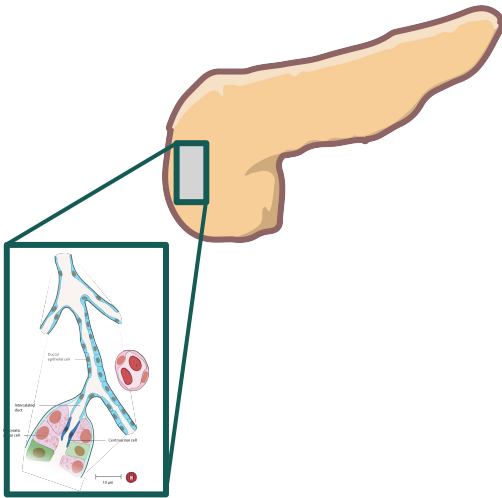


Desafío clave - Diferencias en base al sexo:

- Mayor incidencia y supervivencia en mujeres.
- Diferencias en supervivencia más pronunciadas.
- Desconocimiento de los mecanismos moleculares subyacentes.



Enfoque de la Investigación: Adenocarcinoma Ductal de Páncreas

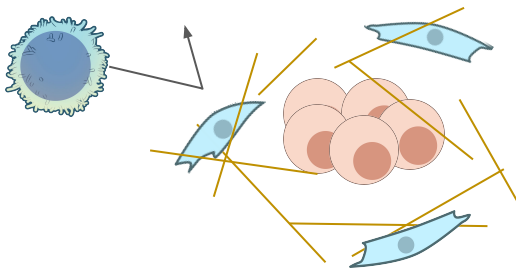


Adenocarcinoma Ductal de Páncreas (PDAC):

- >80% de los casos de cáncer de páncreas.
- ~12% tasa de supervivencia a los 5 años.

Desafío Clave - El Microambiente Tumoral (TME):

- Desmoplasia Extensa
- Inmunosupresor ("Frío")



Parte II

Justificación y Objetivos

Justificación: Uso de datos públicos para profundizar la comprensión de la Heterogeneidad del Cáncer

Oportunidad: Disponibilidad de datos públicos transcriptómicos en los repositorios, que podrían permitir caracterizar subtipos de cáncer e identificar biomarcadores.

Desafío: Dificultades de integración y limitaciones de los datos.

Enfoque Propuesto: Metaanálisis de Datos Transcriptómicos Públicos:

- **Estrategia:** Reanálisis e integración de los resultados
- **Beneficios:**
 - Superar limitaciones individuales
 - Incrementar el poder estadístico

Validación y estudios de caso: Adenocarcinoma de Pulmón (LUAD) y Adenocarcinoma Ductal Pancreático (PDAC).

- **Objetivo:** Demostrar la versatilidad y robustez del enfoque.

Objetivos

El **objetivo principal** fue desarrollar e implementar un enfoque computacional para caracterizar subtipos de cáncer a partir de datos públicos.

Los objetivos específicos:

1. Evaluar el **potencial de combinar datos publicados** para proporcionar a los investigadores una comprensión más profunda de los subtipos de cáncer.
2. Caracterizar a nivel transcriptómico las **diferencias funcionales en LUAD entre hombres y mujeres**.
3. Estudiar el **panorama transcriptómico del PDAC**.
4. **Identificar subtipos de PDAC y biomarcadores potenciales** basados en la supervivencia del paciente.

Parte III

Metodología Computacional Central

Metodología



Revisión Sistemática

Búsqueda sistemática de estudios publicados con datos disponibles.

Selección basada en el diseño del estudio y las variables del paciente.

Análisis Individuales

Preprocesamiento y estandarización de datos.

Evaluación de la normalización.

Análisis de expresión diferencial (y análisis funcional en LUAD).

Metaanálisis

Integración de los resultados del análisis utilizando el modelo de efectos aleatorios de DerSimonian y Laird.

Análisis adicionales

En PDAC, análisis de sobrerepresentación y análisis de supervivencia.

Desarrollo de herramientas web para mejorar la accesibilidad de los resultados.

Preprocesamiento Individual

Preprocesamiento:

- **Estandarización** de variables clínicas.
- **Normalización** de RNA-seq - media recortada de valores M (**TMM**).
- **Estandarización de la normalización** - transformación \log_2 .
- Anotación génica uniforme - IDs de **Entrez**.
- Eliminación de genes duplicados - mediana.

Análisis exploratorio:

- Diagramas de caja de expresión.
- Agrupamiento no supervisado.
- Análisis de Componentes Principales (PCA).

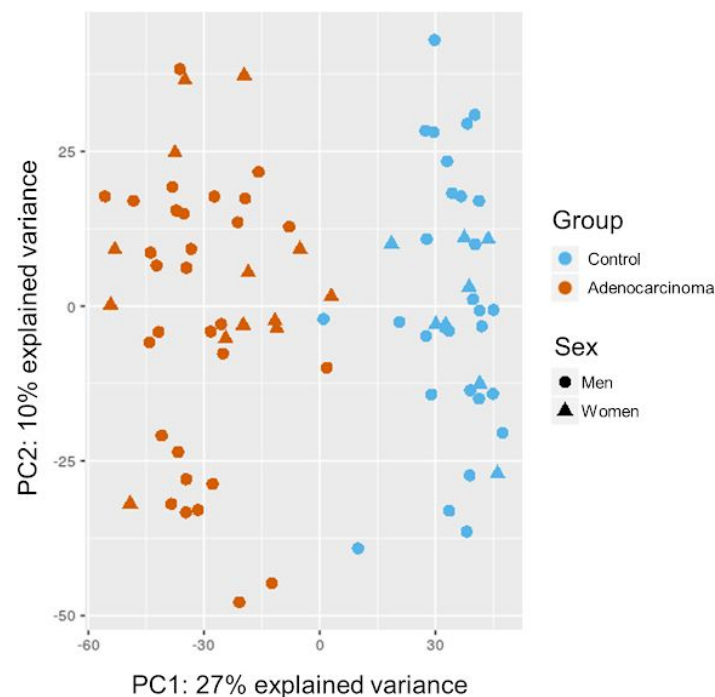


Gráfico de dispersión donde se muestran las dos primeras componentes de un PCA

Análisis de Expresión Diferencial - LUAD

Objetivo: Identificar genes donde la diferencia entre el tejido LUAD y el tejido Control dependa del sexo del paciente (es decir, identificar efectos de la enfermedad específicos del sexo)



Metodología: *Linear Models for Microarray and RNA-Seq Data (limma)*.

Ampliamente utilizada, muy versátil: diseños complejos y simples, implementación para microarrays y RNA-seq.

Contraste Matemático

$(LUAD.M - Control.M) - (LUAD.H - Control.H)$

Diseño de Contraste

$Expresión \sim Enfermedad * Sexo + Tabaco + Emparejamiento$

Corrección por comparaciones múltiples:
False Discovery Rate (FDR)

Resultados

Lista ordenada de genes con estadísticas que evalúan su implicación en la interacción entre el estado de adenocarcinoma de pulmón y el sexo del paciente.

Análisis de Expresión Diferencial - PDAC

Objetivo: Identificar genes expresados diferencialmente entre tejido PDAC y tejido no tumoral.



Metodología: *Linear Models for Microarray and RNA-Seq Data (limma)*.

Ampliamente utilizada, muy versátil: diseños complejos y simples, implementación para microarrays y RNA-seq.

Contraste Matemático

PDAC - Control

Diseño de Contraste

Expresión ~ Enfermedad + Emparejamiento

Corrección por comparaciones múltiples:
False Discovery Rate (FDR)

Resultados

Lista ordenada de genes con estadísticas que evalúan su alteración en PDAC.

Enriquecimiento Funcional

Objetivo: Identificar rutas y funciones asociadas con genes de interés

Adenocarcinoma de Pulmón

Metodología: *Gene Set Enrichment Analysis* (GSEA) usando *Multi-Dimensional Gene Set Analysis (mdgsa)*.

Enfoque en el enriquecimiento de rutas y funciones, en lugar de en los genes.

Adenocarcinoma Ductal de Páncreas

Methodology: Análisis de Sobrerrepresentación (ORA) usando *Cluster Profiler (clusterProfiler)*.

Enfoque en los genes, en lugar de en las funciones o rutas.

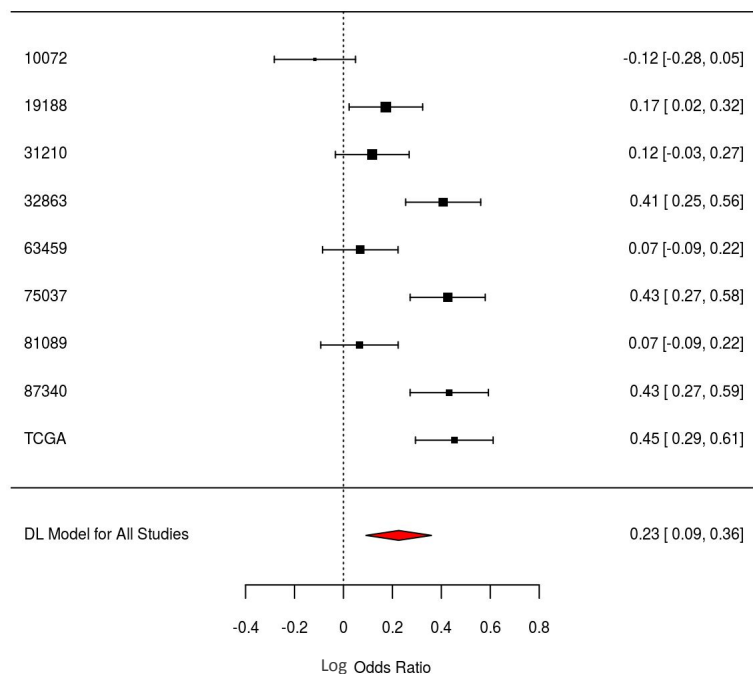
Resultado

Lista ordenada de rutas o funciones con estadísticas que evalúan su implicación en el contexto biológico estudiado.

Metaanálisis

Objetivo: Sintetizar los resultados de análisis previos a través de múltiples estudios independientes

Metodología: Modelo de efectos aleatorios de DerSimonian y Laird (DL) usando metafor.



Validación:

- Diagramas de bosque y diagramas de embudo: magnitud del efecto y precisión.
- Análisis de sensibilidad (*leave-one-out* validación cruzada).

Gráfico de bosque obtenido en el estudio de LUAD, correspondiente al Proceso Biológico de GO: GO:0002285 - activación de linfocitos implicada en la respuesta inmune.





Part IV

Identificación de Firmas Funcionales Basadas en el Sexo en NSCLC

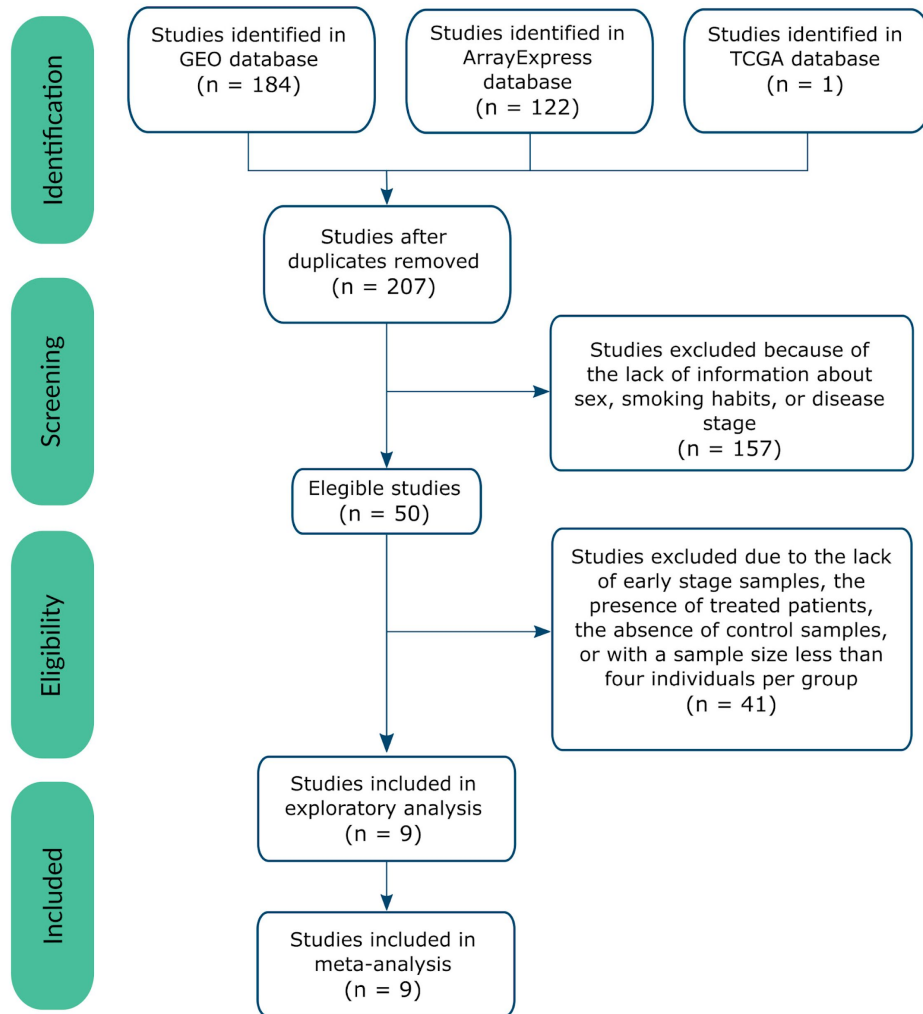


Systematic Review

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

Irene Pérez-Díez ^{1,2}, Marta R. Hidalgo ¹, Pablo Malmierca-Merlo ^{1,3}, Zoraida Andreu ¹, Sergio Romera-Giner ^{1,3}, Rosa Farràs ⁴, María de la Iglesia-Vayá ², Mariano Provencio ⁵, Atocha Romero ⁵ and Francisco García-García ^{1,6,*}

Resultados de la Selección de Estudios



Enfoque en estadio temprano.

Cohorte final:

9 estudios transcriptómicos.

Población de estudio:

- **1.366** muestras en estadio temprano:
 - 997 Casos (Tumor).
 - 369 Controles (Tejido Adyacente No Tumoral).
- Sexo: 56% Mujeres, 44% Hombres.
- Mediana de edad: 65.5 años.

Flujo de información a través de las diferentes fases de la revisión sistemática, siguiendo las directrices de la Declaración PRISMA.

Resultados del Análisis Individual

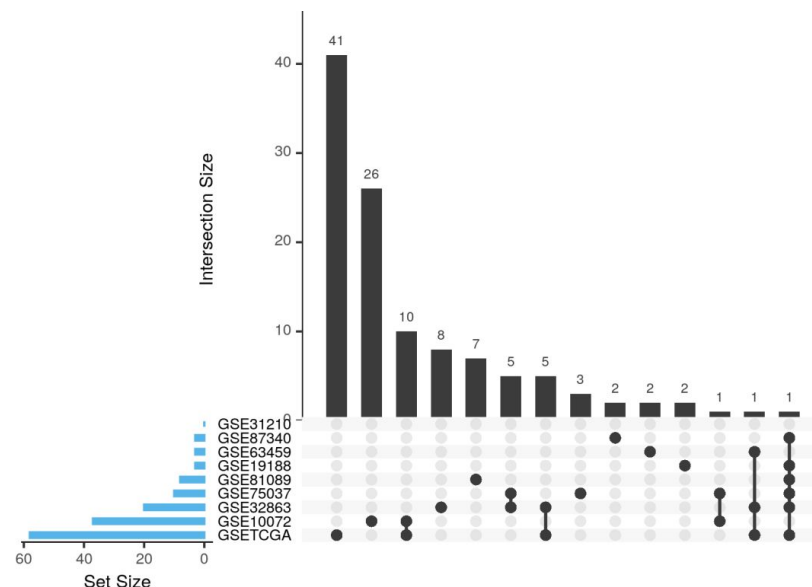
Control de calidad: Se identificaron y eliminaron **3 muestras anómalas**.

Expresión diferencial:

- Pocos genes significativos por estudio
- **Sin genes significativos comunes** entre los estudios.

Enriquecimiento funcional:

- Términos **GO** y rutas **KEGG**.
- La mayoría de los resultados significativos fueron **exclusivos de estudios individuales**.



Intersección de las funciones significativas entre estudios.

El análisis demuestra que los **resultados son inconsistentes** entre estudios, reforzando la idea de que una correcta **integración es necesaria**.

Resultados del Metaanálisis Funcional

Realizamos un metaanálisis por cada uno de los **8.672** términos GO y rutas KEGG detectados en al menos dos estudios.

Resultados significativos (FDR < 0.05): 129 términos o rutas significativas:

- 106 Procesos Biológicos.
- 3 Funciones Moleculares.
- 20 Rutas KEGG.

Pudimos asociar las funciones significativas a **21** grupos funcionales amplios (curados manualmente).

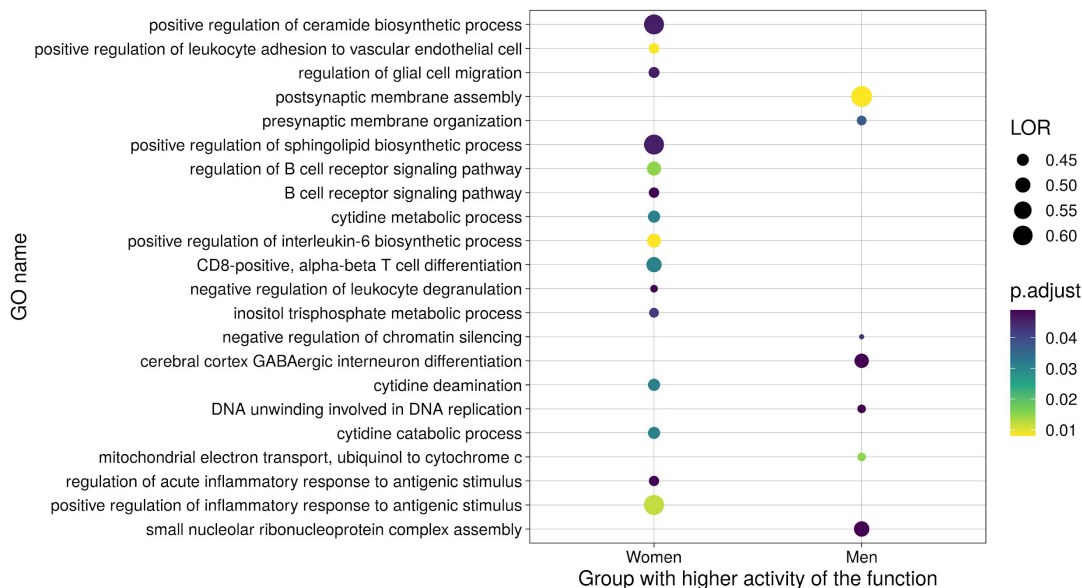
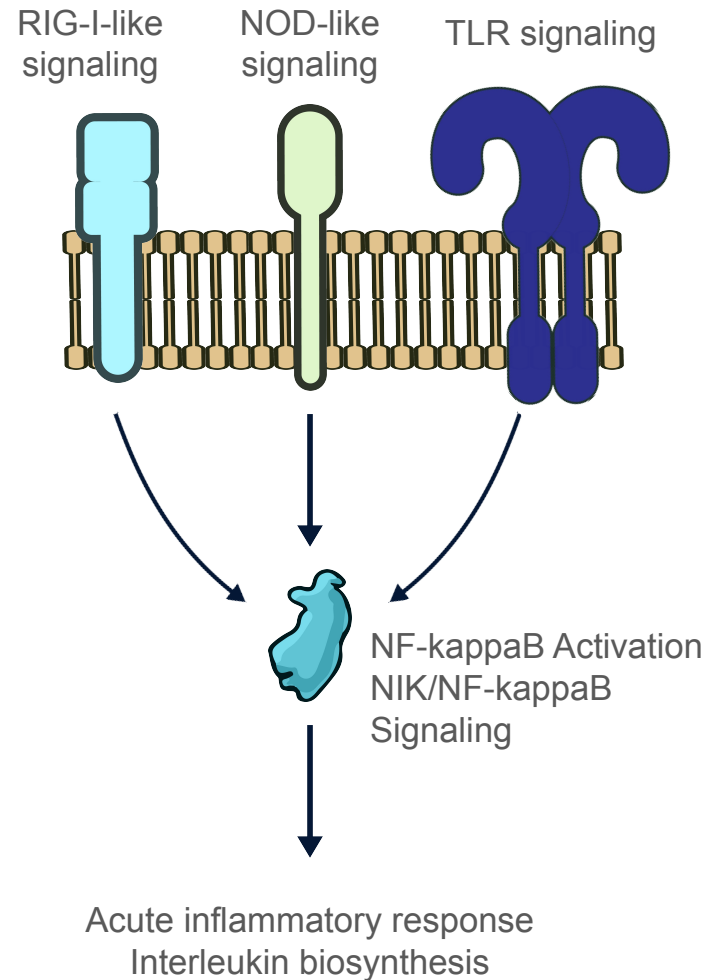
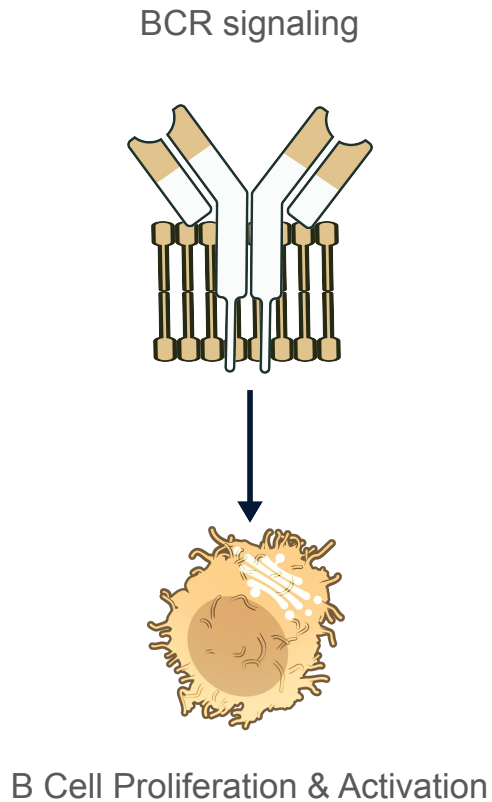


Diagrama de puntos resumen de los resultados referentes a procesos biológicos.

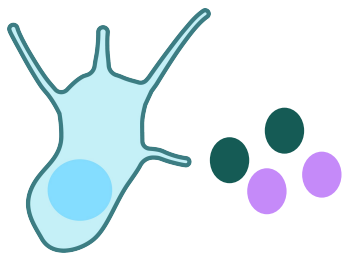
Funciones clave detectadas

~44% - Sistema inmune - Sobreactivación en mujeres



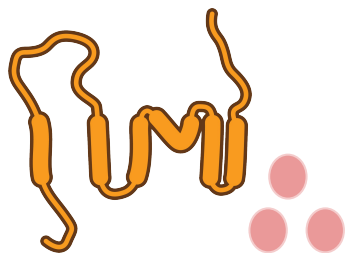
Funciones clave detectadas

Mujeres



Respuesta inmune

- Activación CD8+ T.
- Perfil de interleucinas complejo (IL-2, IL-10 vs IL-6, IL-8, IL-17)



Metabolismo

- Señalización purinérgica
- Ceramidas/esfingolípidos

Hombres



Daño y reparación del ADN

- Reparación del ADN
- Protección de telómeros
- Desenrollamiento ADN



Estrés oxidativo:

- Ciclo celular
- Procesos mitocondriales



Confirmamos diferencias en el sistema inmune y **revelamos alteraciones no descritas.**

Implicaciones potenciales en tratamiento: respuesta a inmunoterapia, inhibidores PARP.

Página web - resultados accesibles

- Analysis Summary
- Exploratory Analysis
- Differential Expression
- Gene Set Analysis
- Meta-Analysis
- Methods
- R Package

Authors

Irene Pérez Díez

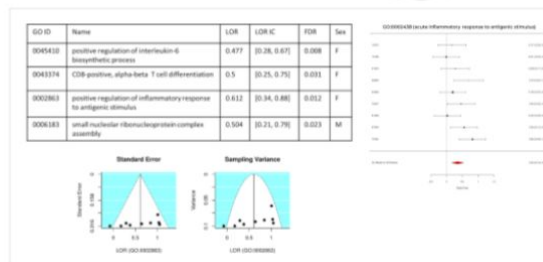
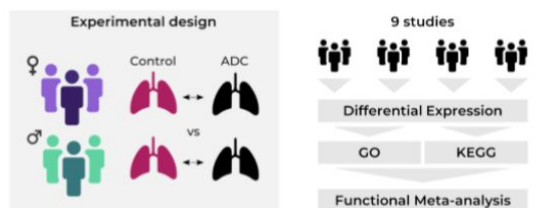
Marta R. Hidalgo

Pablo M. Merlo

Zoraida Andreu Martinez

Sergio Romera-Giner

Functional signatures in non-small-cell lung cancer: a systematic review and meta-analysis of sex differences in transcriptomic studies



Contexto y contribución metodológica

Algunos estudios han evaluado las diferencias en sexo en LUAD, con ciertas **limitaciones que abordamos** con nuestro estudio:

- Comparamos explícitamente pacientes hombres y mujeres, incluyendo muestras control.
- Aplicamos un metaanálisis funcional para integrar de forma robusta 9 estudios.
- Realizamos un enfoque amplio, sin centrarnos en un aspecto concreto de la biología

Limitación: Falta de detalles clínicos relevantes (estado mutacional), falta de estandarización en la información aportada, infrarrepresentación de estudios con mujeres - resalta la **necesidad de aplicar los principios FAIR** en investigación biomédica.

Fortalezas:

- Poder estadístico incrementado por el metaanálisis.
- Descripción funcional robusta, a pesar de la heterogeneidad.
- Centrado en estadio temprano.

Parte V

Identificación de Firmas Transcripcionales para estratificar pacientes en PDAC

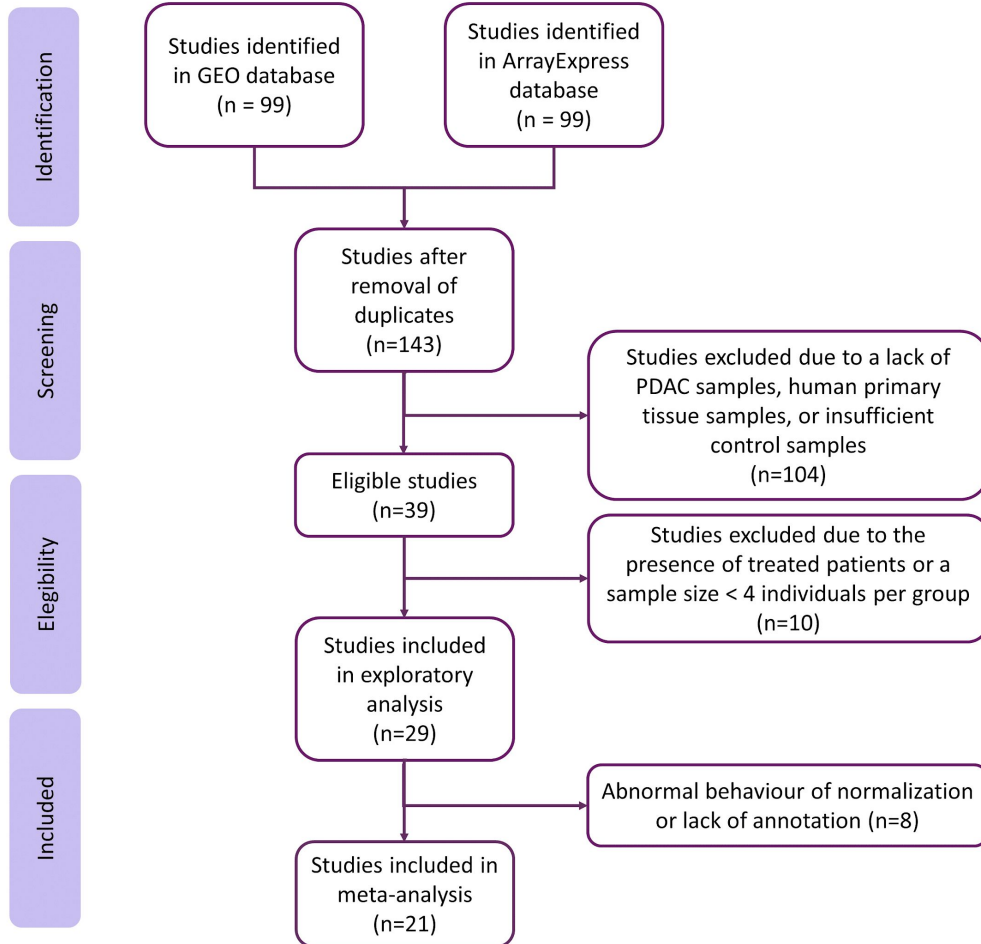


Systematic Review

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

Irene Pérez-Díez ^{1,2,3,†} , Zoraida Andreu ^{3,†}, Marta R. Hidalgo ^{1,3} , Carla Perpiñá-Clérigues ^{1,3,4} , Lucía Fantín ¹, Antonio Fernandez-Serra ^{3,5} , María de la Iglesia-Vaya ^{2,3} , José A. Lopez-Guerrero ^{3,5,6,*}  and Francisco García-García ^{1,3,*} 

Resultados de la Selección de Estudios



Cohorte final:

21 estudios transcriptómicos

Población de estudio:

- **922** muestras:
 - 602 Casos (Tejido tumoral).
 - 320 Controles (Tejido adyacente no tumoral).
- Sexo: 40% Mujeres, 60% Hombres (7 estudios).
- Mediana de edad: 64.5 años (6 estudios).

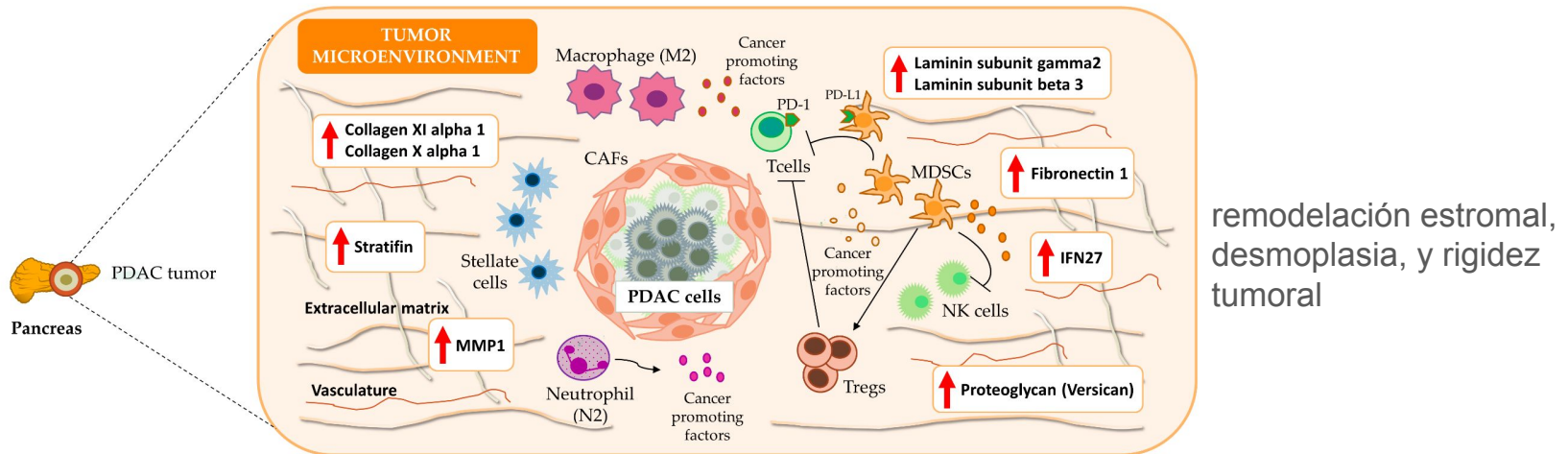
Flujo de información a través de las diferentes fases de la revisión sistemática, siguiendo las directrices de la Declaración PRISMA.

Metaanálisis de la expresión génica

Realizamos un metaanálisis de **24.365** genes (≥ 2 estudios).

Identificamos **1.153** genes expresados diferencialmente:

- FDR < 0.05.
- Valor absoluto del log2 de la magnitud de cambio ($|\log_2FC|$) > 0.6.
- Evaluado en al menos 11 estudios (>50%).



Visión general del microambiente del PDAC, según los resultados del metaanálisis.

Interpretación funcional de los resultados

Para mejorar la interpretabilidad, realizamos un **Análisis de Sobrerrepresentación** (ORA) en los 1.153 genes significativos utilizando términos GO BP con un umbral de FDR < 0.05.

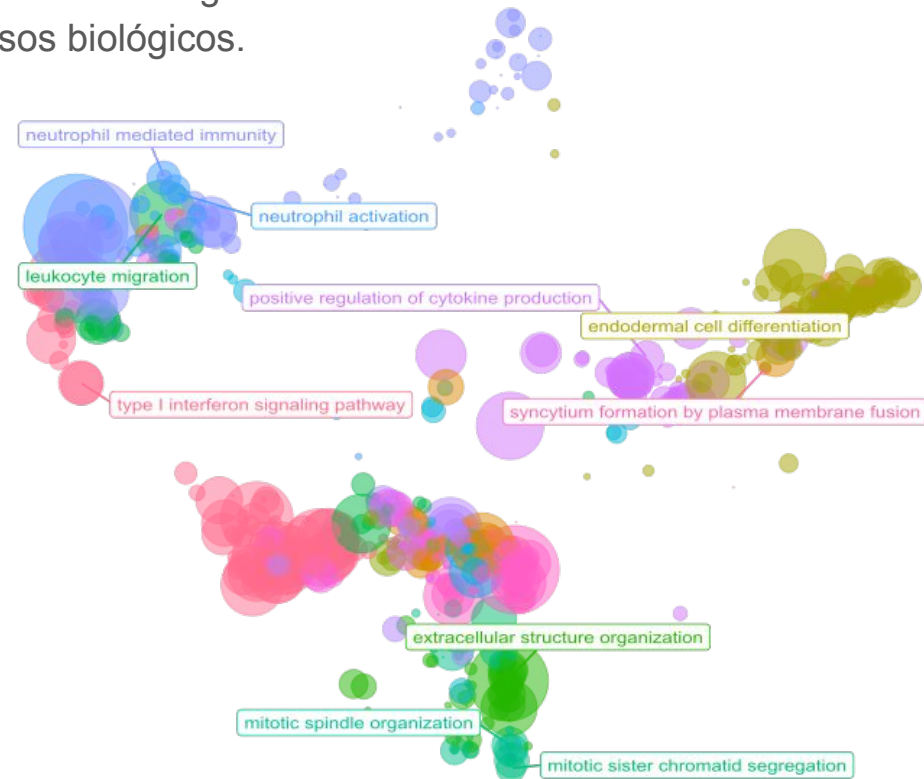
- **Genes sobreexpresados** (en PDAC): **546** procesos biológicos.
- **Genes infraexpresados** (en PDAC): **40** procesos biológicos.

TME:

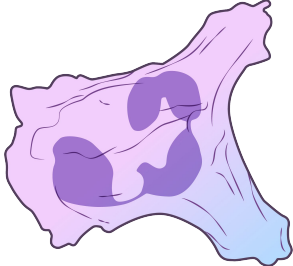
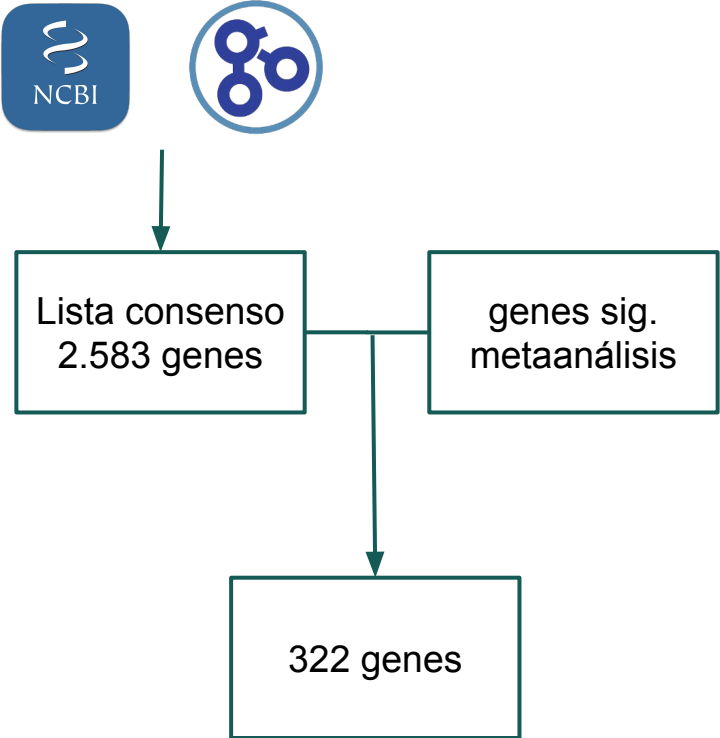
- Procesos del sistema inmunitario.
- Adhesión celular.
- Remodelación y degradación matriz extracelular (ECM).

Metástasis:

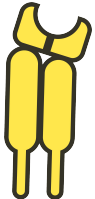
- Vascularización.
- Proliferación y migración celular.
- Metabolismo del colágeno.
- Transición epitelio-mesenquimal (EMT).



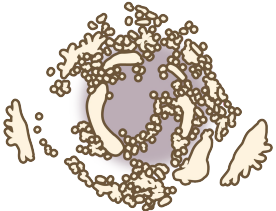
Enfoque en el Sistema Inmunitario



Señalización por interferón



Presentación de antígenos MHC-II



Regulación negativa de la muerte celular



Caracterización del Microambiente Tumoral

Procesos y genes clave:

- **Componentes de la ECM y remodelación:** Alta expresión de colágenos (COL11A1, COL10A1), Periostina (POSTN), Fibronectina (FN1), Versicano (VCAN).
 - Densificación del estroma, rigidez, barrera a infiltración inmunitaria, quimioresistencia
- **Adhesión e invasión celular:** Alta expresión de Stratifina (SFN), MMP1, Catepsina E (CTSE), Lamininas (LAMB3, LAMC2), CEACAMs.
 - Degradación ECM, migración celular, invasión, resistencia a terapias.
- **Biomarcadores potenciales:** Alta expresión de of Cistatina SN (CST1 - marcador proliferación) y Sulfatasa 1 (SULF1 - potencial de liberación de fármacos).
- **Enzimas de digestión pancreáticos:** Baja expresión de Quimiotripsina y Lipasas.
 - Reducción de la apoptosis e incremento de migración.

Nuestro enfoque destaca **procesos biológicos clave** que potencialmente impulsan la progresión del PDAC, la resistencia a la terapia y la metástasis.

Identificación de Firmas Pronósticas

Evaluamos el valor pronóstico de firmas de expresión génica usando una cohorte independiente de pacientes: **The Cancer Genome Atlas** (TCGA). Aplicamos análisis de **Kaplan-Meier** con prueba de **log-rank**.

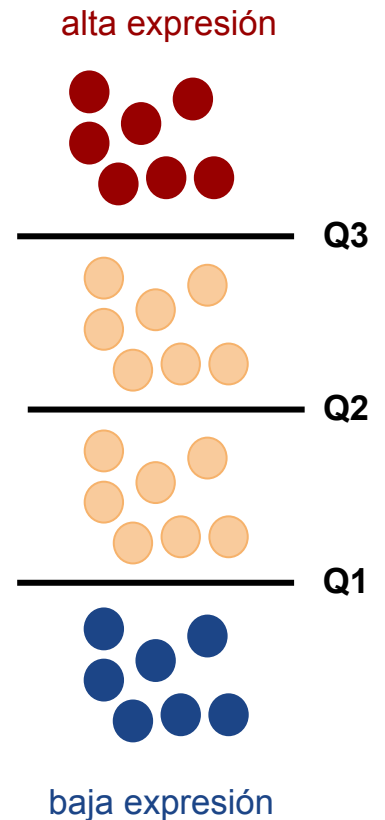
1. Análisis de supervivencia gen a gen:

- Gen altamente expresado si Z-score > Q3.

2. Análisis en base a puntuaciones de riesgo:

- Selección de genes
- Asignamos puntuación de riesgo a muestra en base al número de genes altamente expresados.
- Agrupación de pacientes en “Alto riesgo” o bajo riesgo si su puntuación de riesgo > mediana.

3. Análisis Multivariante: Riesgos Proporcionales de Cox.



Firma Pronóstica Inmunitaria

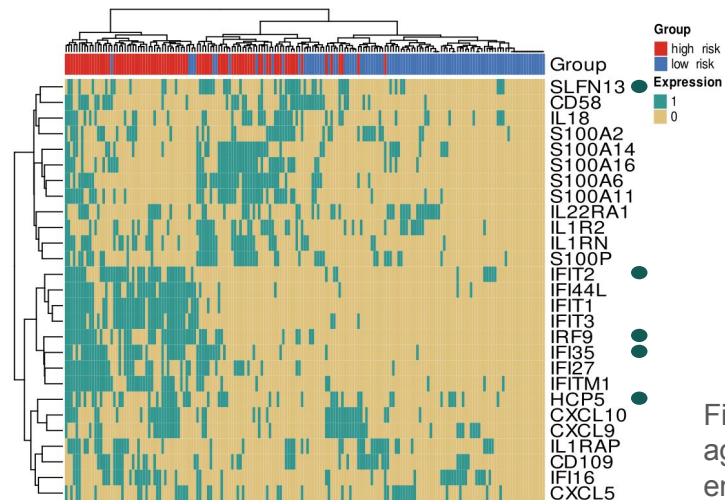
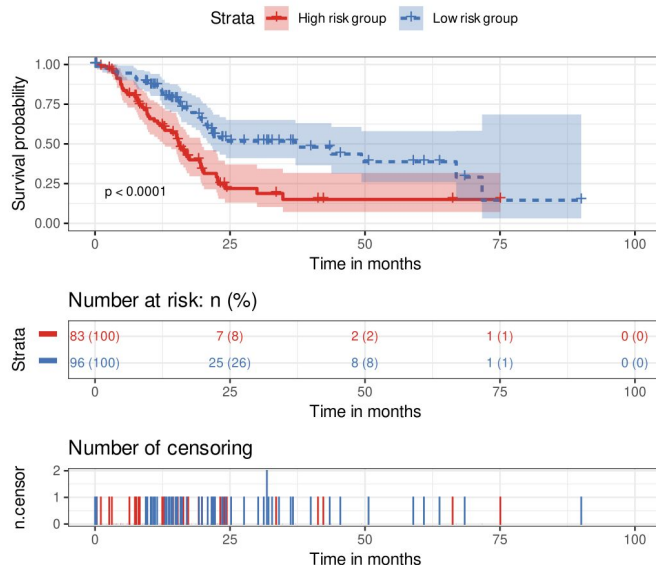
Análisis de supervivencia sobre 70 genes significativos clave del sistema inmunitario.

Análisis gen a gen: identificamos **28 genes** con asociación significativa.

- Alta expresión correlacionada consistentemente con **menor tasa de supervivencia**.
- Nuevos descubrimientos: 5 genes** identificados por primera vez en PDAC (HCP5, SLFN13, IRF9, IFIT2, IFI35).

Puntuación de riesgo (28 genes):

- El grupo de alto riesgo mostró **menor supervivencia** ($p < 0.0001$).
- Pronóstico independiente** según el modelo de Cox (HR = 2.36, $p < 0.05$).



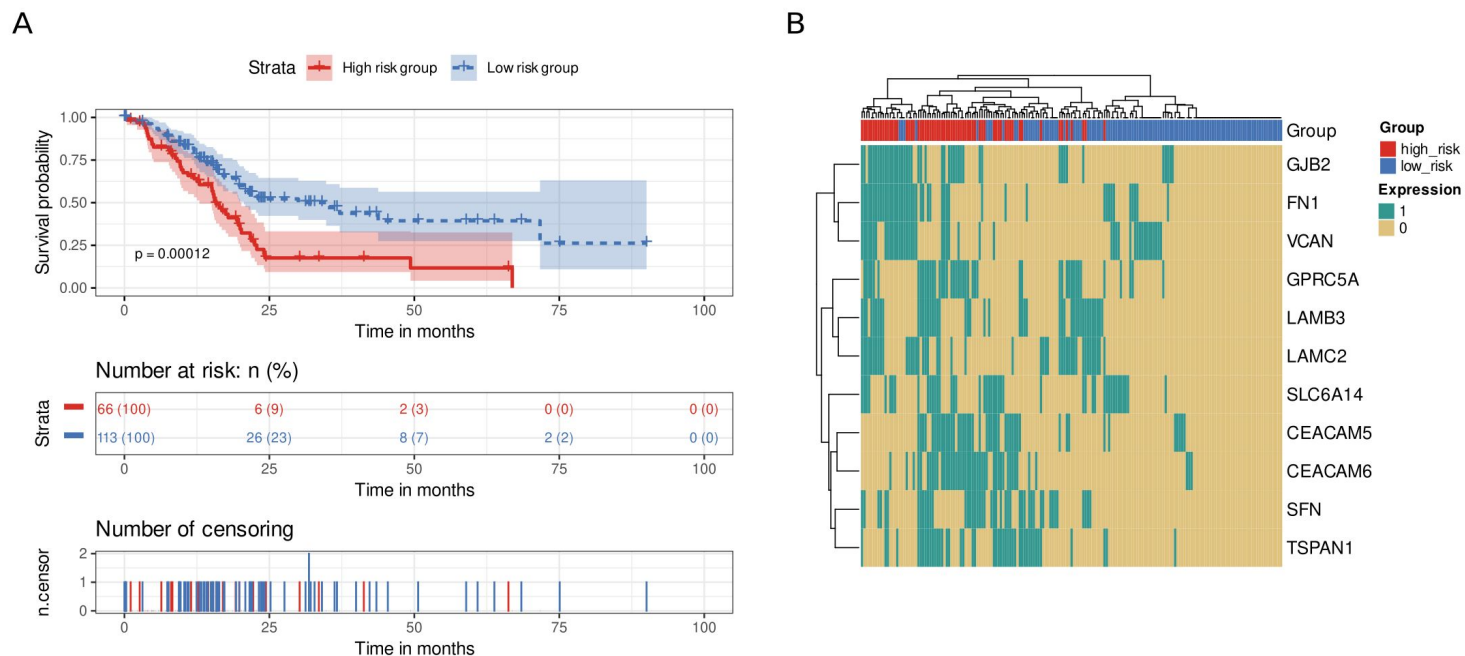
Firma génica para la agrupación de pacientes en alto o bajo riesgo.

Firma Pronóstica Estromal

Análisis de supervivencia sobre los genes significativos relacionados con la remodelación de ECM.

Individual Gene Analysis: identificamos **11 genes** con asociación significativa.

Puntuación de riesgo (11 genes): El grupo de alto riesgo mostró menor supervivencia ($p = 0.00012$).



Firma génica para la agrupación de pacientes en alto o bajo riesgo.

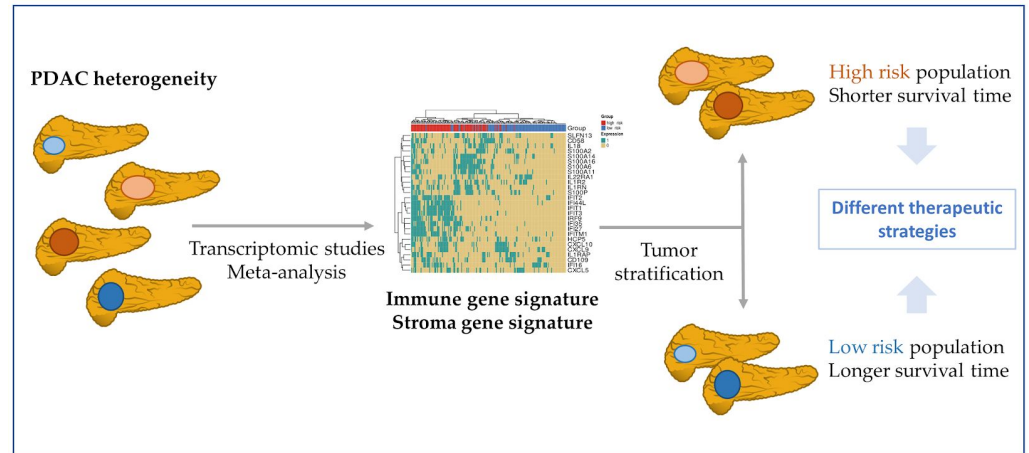
Firmas Pronósticas

Firmas pronósticas:

1. Firma inmunitaria (28 genes)

- Marcadores pronósticos nuevos (HCP5, SLFN13, IRF9, IFIT2, IFI35).

2. Firma estromal (11 genes)



Potencial utilidad clínica:

- Permiten estratificar pacientes en grupos de **Alto Riesgo y Bajo Riesgo** en relación a la supervivencia.
- Puede ayudar a distinguir tumores "**Fríos**" o "**Calientes**", guiando potencialmente decisiones terapéuticas.
- **Independencia de subtipos previos**: Mínima superposición con los genes incluidos en los subtipos de Bailey y Moffitt.
- **Viabilidad**: Numero manejable de genes, fácilmente trasladable a la clínica

Página web - resultados accesibles

MetaPDAC

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Unidad de Bioinformática y Bioestadística

Differential Expression
Gene Meta-analysis
ORA - Biological Process
Survival analysis

Differential expression analysis

Differential gene expression analyses were performed in R using limma (v.3.48.3), and a paired sample design was implemented in those datasets where applicable. Differentially-expressed genes were identified using p-values with Benjamini-Hochberg correction for a false discovery rate (FDR) at a significance level of 0.05.

Stats **Vulcano plot** Summary

Inputs

Select a dataset
GSE101448

Select logFC threshold
0,6

Select FDR threshold
0,05

Download the data

Results table

Show 10 entries

Search:

	logFC	se.coef	P.Value	adj.P.Val	geneNames
1	-4.63742323993859	0.159126702273134	5.56889904737696e-30	1.18016108612012e-25	pancreas associated transcription factor 1a
2	-5.53279902766447	0.207597128893606	2.14409847367455e-28	2.27188674270556e-24	serpin family I member 2

Contexto y contribución metodológica

Trabajos previos de integración analizaron de forma conjunta los datos de expresión. Nuestra aproximación maneja de forma mucho más robusta la heterogeneidad entre estudios, comparado con la unión de los datos crudos o normalizados.

Puntos fuertes:

- Caracterización exhaustiva del perfil transcripcional en PDAC.
- Identificación de firmas génicas con valor pronóstico independiente.
- Identificación de potenciales dianas terapéuticas

Limitaciones:

- Heterogeneidad de los datos: Abordado parcialmente por el modelo de efectos aleatorios.
- Falta de información clínica
- Se necesitan estudios adicionales para validar clínicamente las firmas pronósticas y dianas terapéuticas.

Parte VI

Discusión General y Conclusiones

Metodología Computacional

Marco Computacional Común:

Hemos aplicado una metodología consistente basada en el **metaanálisis** de **datos públicos transcriptómicos** para afrontar diferentes preguntas biológicas relacionadas con dos tipos de cáncer:

- **Adenocarcinoma de pulmón:** Investigar las diferencias funcionales en base al sexo.
- **Pancreatic Ductal Adenocarcinoma:** Caracterizar el TME e identificar firmas pronósticas.

Valor y versatilidad demostrados:

- Potencia estadística aumentada sobre los estudios individuales.
- Identificado señales biológicas robustas en medio de la heterogeneidad.
- Generar nuevos conocimientos e hipótesis a partir de datos existentes.

Destaca el valor de **reutilizar datos públicos (principios FAIR)** para avanzar en la investigación del cáncer, especialmente para problemas complejos como la caracterización de subtipos.

Fortalezas

1. **Metodología Robusta:** Mayor poder estadístico y fiabilidad.
2. **Manejo de la Heterogeneidad:** Selección rigurosa y modelo de efectos aleatorios.
3. **Nuevos conocimientos biológicos:**
 - Mecanismos alterados entre hombres y mujeres en LUAD.
 - Genes con impacto en supervivencia.
4. **Hallazgos relevantes a nivel clínico:**
 - Firmas pronósticas.
 - Relevancia del sexo en estudios de LUAD.
5. **Principios FAIR y Accesibilidad:** Páginas web interactivas para promover la transparencia y la reusabilidad.
6. **Validación por pares:** Resultados publicados en Q1.

Limitaciones

1. **Dependencia en la calidad de los datos:** La calidad del metaanálisis es dependiente de la calidad de los estudios individuales.
2. **Falta de información clínica:** Falta de metadatos detallados en la información aportada por los estudios.
3. **Falta de estandarización:** Diferencias en la anotación clínica y la anotación génica.
4. **Heterogeneidad:** Heterogeneidad técnica y biológica entre los estudios.

Conclusiones

Objetivo 1: Evaluar el **potencial de combinar datos publicados** para proporcionar a los investigadores una comprensión más profunda de los subtipos de cáncer.

1. Los datos publicados pueden ser **reutilizados eficazmente para avanzar el conocimiento científico** y profundizar en la comprensión de los subtipos de cáncer. Esto resalta la **importancia de la colaboración y la compartición de datos** en la comunidad científica.

2. El metaanálisis es la metodología estadística adecuada para este objetivo, ya que es robusto y mejora la potencia estadística, enmascarando posibles deficiencias de algunos de los datos publicados. El **metaanálisis puede utilizarse eficazmente para aprovechar los datos publicados** e impulsar un paso adelante la investigación sobre el cáncer.

Conclusiones

Objetivo 2: Caracterizar a nivel transcriptómico las **diferencias funcionales en LUAD entre hombres y mujeres**.

3. Los pacientes con adenocarcinoma de pulmón muestran **diferencias funcionales transcriptómicas en base al sexo del paciente**. Esto sugiere la importancia de considerar el sexo como un **factor biológico relevante en la investigación y el tratamiento del adenocarcinoma de pulmón**, lo que podría conducir a terapias más personalizadas y eficaces.

4. Los mecanismos moleculares subyacentes a las diferencias transcriptómicas basadas en el sexo en el adenocarcinoma de pulmón están **particularmente relacionados con el sistema inmunitario, la señalización purinérgica y el metabolismo lipídico**. Este conocimiento abre nuevas vías para la investigación de biomarcadores y terapias dirigidas específicamente a estas rutas biológicas.

Conclusiones

Objetivo 3: Estudiar el **panorama transcriptómico del PDAC**.

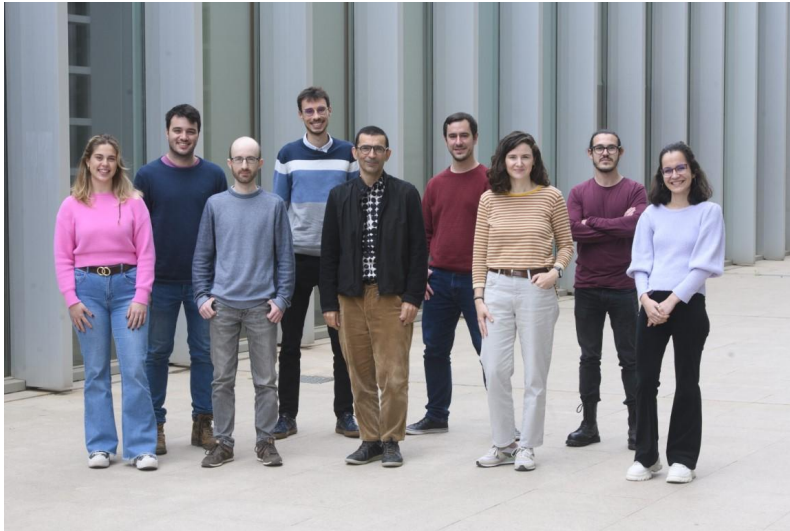
5. El **panorama transcriptómico del adenocarcinoma ductal pancreático** está fuertemente ligado a los componentes de la matriz extracelular y al sistema inmunitario. Esta asociación sugiere un **ambiente inmunosupresor y un estroma desmoplásico**, lo que puede influir en la respuesta al tratamiento y la progresión de la enfermedad.

Objetivo 4: **Identificar subtipos de PDAC y biomarcadores potenciales** basados en la supervivencia del paciente.

6. Basándonos en el panorama transcriptómico del adenocarcinoma ductal pancreático, identificamos **veintiocho genes inmunes y once genes relacionados con la matriz extracelular que impactaron la supervivencia del paciente**. Estos genes podrían utilizarse para estratificar pacientes y guiar la práctica médica.

Agradecimientos

Laboratorio de Biomedicina Computacional



Unidad Mixta de Imagen Biomédica



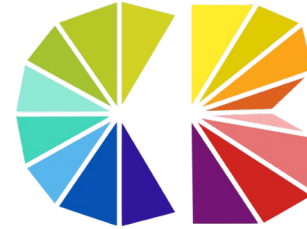
Dr. Zoraida Andreu
(Instituto Valenciano de
Oncología)



**Unidad Mixta de
Imagen Biomédica**

Fundació
Fisabio

CIBER
VALÈNCIA BIOMÈDICA
RESEARCH FOUNDATION



**Computational
Biomedicine**
Laboratory

Gracias

“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



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Erratas

- 4.3.4 - Pag. 60: Reference 152
- 4.3.4 - Pag. 59: Table 4.1
- 4.3.4 - Pag. 59: Table 4.2
- 4.3.3.2 - Pag 51: Tables 4.1 and 4.2

Table A.1 - Características clinicopatológicas de los estudios de NSCLC

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

Figure A.3 - COX model

Hazard Ratio for variables of interest

