

III Jornadas de Genómica Clínica

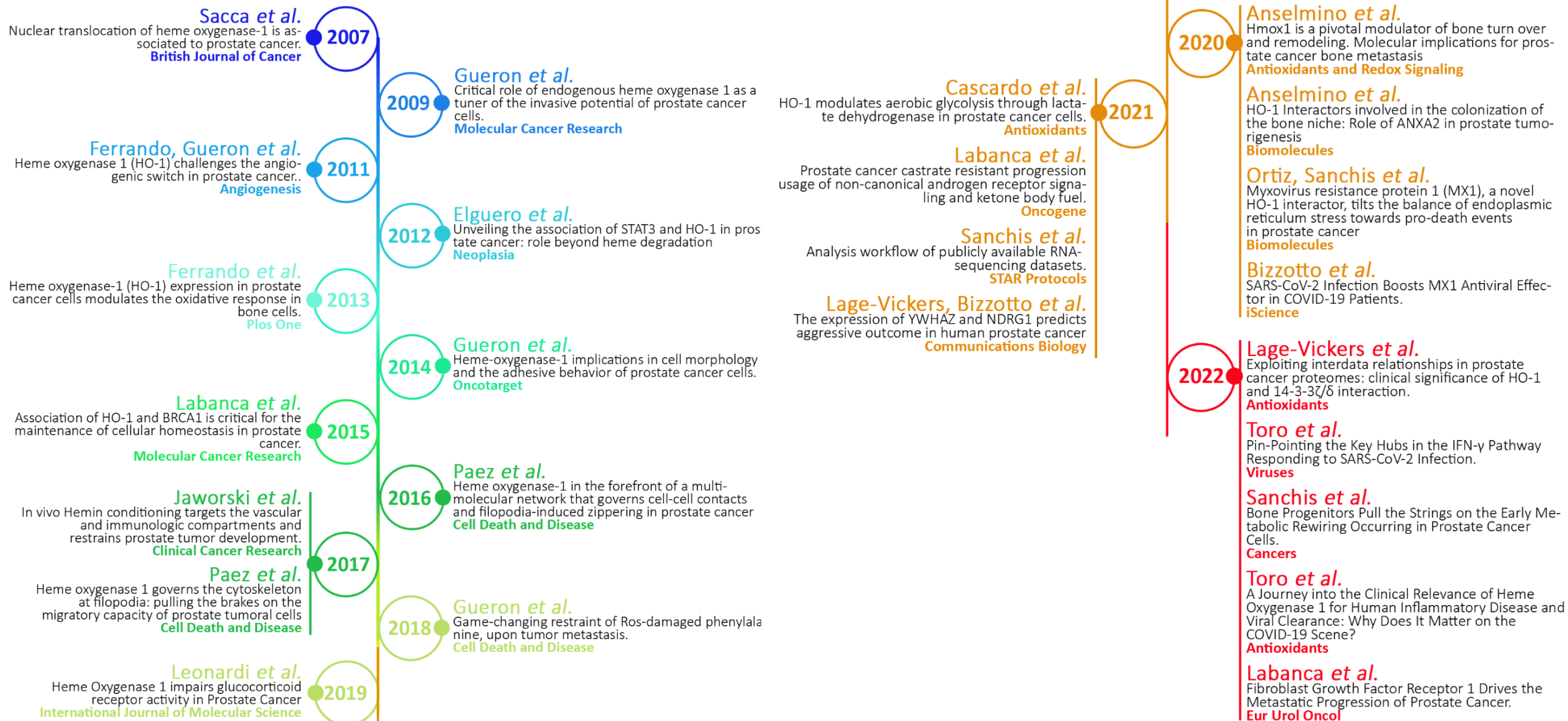
25/10/2024

Desarrollo de firmas transcriptómicas para estratificación de riesgo en cáncer

Dr. Juan Bizzotto

Laboratorio de Inflamación y Cáncer,
Depto. QB, FCEN-UBA / IQIBICEN-CONICET



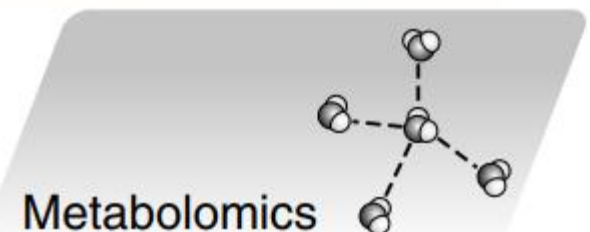
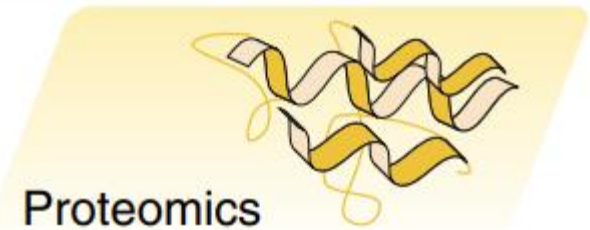
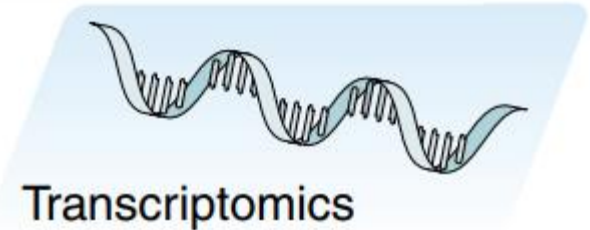


ómicas

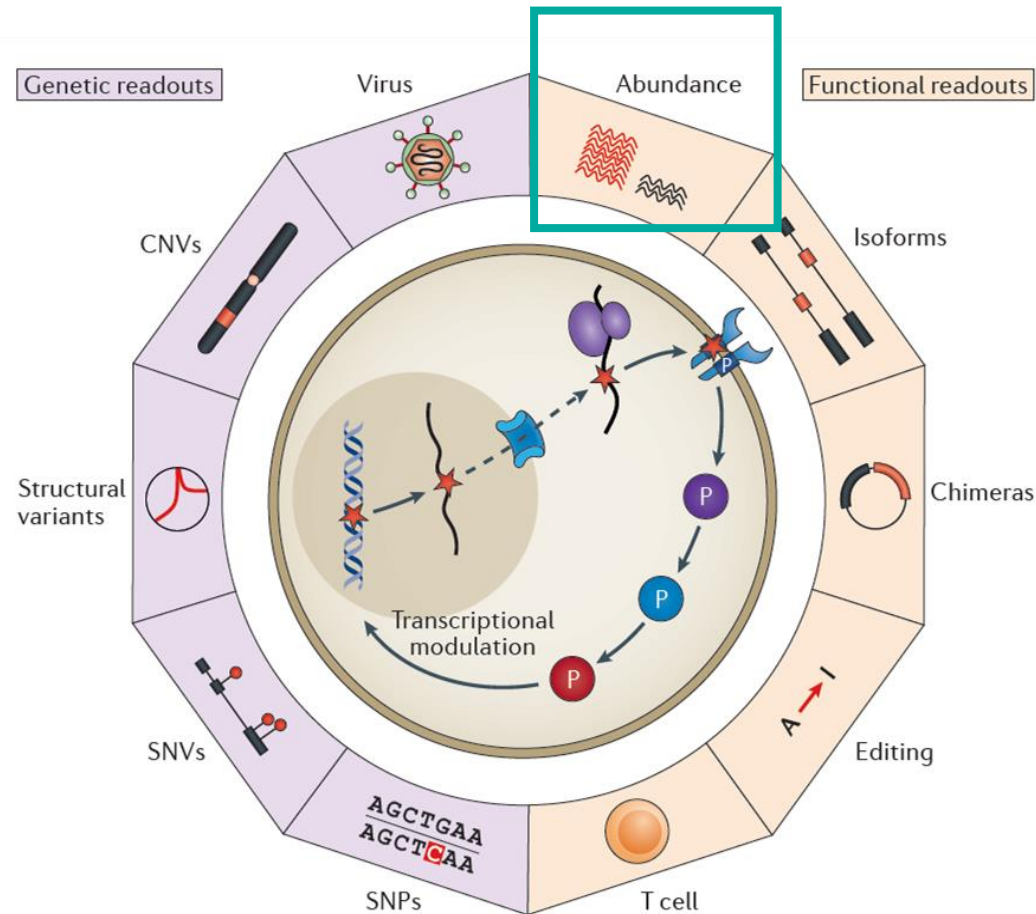
La "ómica" en biología se refiere a un enfoque **sistemático y exhaustivo** para estudiar diferentes aspectos moleculares de los sistemas biológicos.

Estos enfoques suelen implicar la **investigación a gran escala mediante tecnologías de alto rendimiento**, de diversas moléculas biológicas, como genes, proteínas, metabolitos y otros componentes celulares.

Molecular data



Transcriptómica permite estudiar múltiples capas de información



El transcriptoma es el conjunto de todas las moléculas de ARN (“transcriptos”) presentes en una célula o grupo de células en un **momento determinado**.

El estudio del transcriptoma permite caracterizar el fenotipo y el genotipo (que se expresa)

Abundancia de transcritos en muestras de pacientes

genes

pacientes

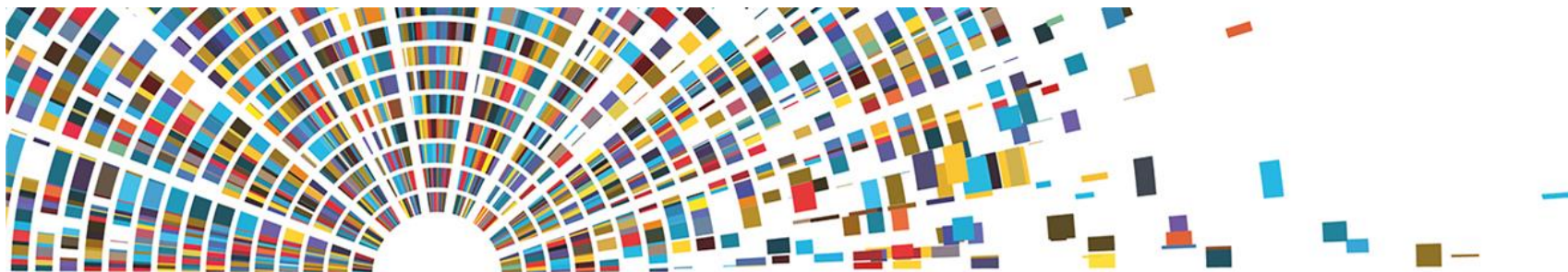
550 muestras

20 000 genes

	TCGA-XJ-A83F-01	TCGA-G9-6348-01	TCGA-CH-5766-01	TCGA-EJ-A65G-01	TCGA-G9-6354-01
ARHGEF10	9.3554	8.8729	8.5581	9.2085	9.0514
HIF3A	5.1517	5.9049	4.9716	6.7795	5.3511
RNF17	0	0.4008	0.7574	0	0
RNF10	12.4656	12.3538	12.295	11.9701	12.5973
RNF11	11.1274	11.5348	11.9867	11.3146	11.3622
RNF13	10.5783	10.5856	11.2172	11.3116	10.6387
GTF2IP1	12.6987	12.2242	12.3527	12.5196	12.4639
REM1	4.5629	5.1002	4.3699	3.3596	3.9809
sep-09	13.0502	12.4365	12.3659	11.108	12.7733
RTN4RL2	5.4558	6.1261	6.3596	5.9495	5.1578
C16orf13	10.8922	9.9575	9.158	10.1131	10.8401
C16orf11	0	0	0.4278	0	0
FGFR1OP2	7.9025	8.8094	8.0633	8.8149	7.8995
TSKS	0.6742	0.9713	1.7724	0.659	0.6497
ATRX	10.2488	10.3802	11.591	10.6535	9.9265
PMM2	10.2546	10.6489	10.7094	10.1118	10.997
LOC10027	5.3752	4.1152	4.9395	5.4353	4.032
ASS1	10.5984	11.1878	10.6394	8.697	10.619
NCBP1	9.3001	9.0393	9.9152	9.5014	9.2123
ZNF709	6.7892	6.3236	7.141	6.8893	7.3845
ZNF708	8.3151	7.3714	8.7473	8.2306	8.0741

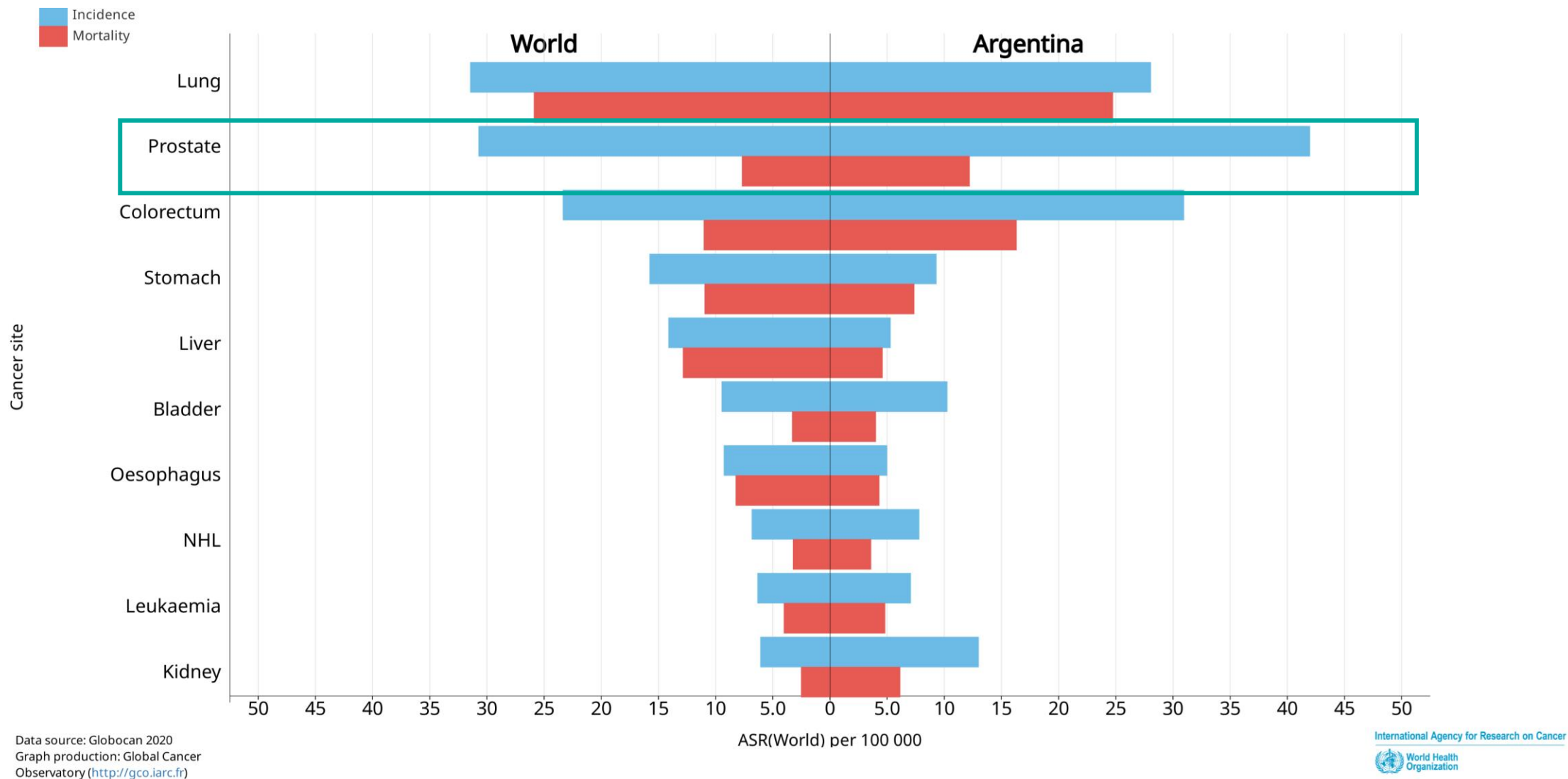
La (bio)informática es una pieza clave en el estudio de las *ómicas*

- Cantidad de datos masiva
- Nos permite ser globales (estudiar todo el genoma en simultáneo)
- Permite estudiar la naturaleza multifactorial de la enfermedad y la complejidad de la biología subyacente
- Herramientas para el análisis y visualización de datos
- La información genética es digital
- Subyace todas las herramientas y las aplicaciones



Desarrollo de una **firma pronóstica** en cáncer de próstata

Cáncer de próstata: incidencia y mortalidad



Tasas de incidencia y mortalidad estandarizadas por edad, en hombres, en 2020

Cáncer de próstata: grupos de riesgo

Grupo de Riesgo	Características clínico/patológicas
Muy bajo	▪ T1c y ▪ Gleason ≤ 6/grupo grado 1 y ▪ PSA < 10 ng/mL y ▪ < 3 cilindros de la biopsia positivos, $\leq 50\%$ afección de cada cilindro y ▪ densidad del PSA < 0.15 ng/mL/g
bajo	▪ T1 a T2a y ▪ Gleason ≤ 6/grupo grado 1 y ▪ PSA < 10 ng/mL
Intermedio favorable	▪ T2b a T2c o ▪ Gleason 3+4 = 7/grupo grado 2 o ▪ PSA 10 a 20 ng/mL y ▪ $< 50\%$ cilindros positivos
Intermedio desfavorable	▪ T2b a T2c o ▪ Gleason 3+4 = 7/grupo grado 2 o Gleason 4+3 = 7/grupo grado 3 o ▪ PSA 10 a 20 ng/mL
Alto	▪ T3a o ▪ Gleason 8/grupo grado 4 o Gleason 4+5 = 9 /grupo grado 5 o ▪ PSA > 20 ng/mL
Muy alto	▪ T3b a T4 o ▪ Patrón primario Gleason 5 o ▪ > 4 cilindros con Gleason 8 a 10/grupo grado 4 o 5

- Grupo de pacientes heterogéneo. Pronóstico variable.
- Desafío para decidir un tratamiento óptimo.

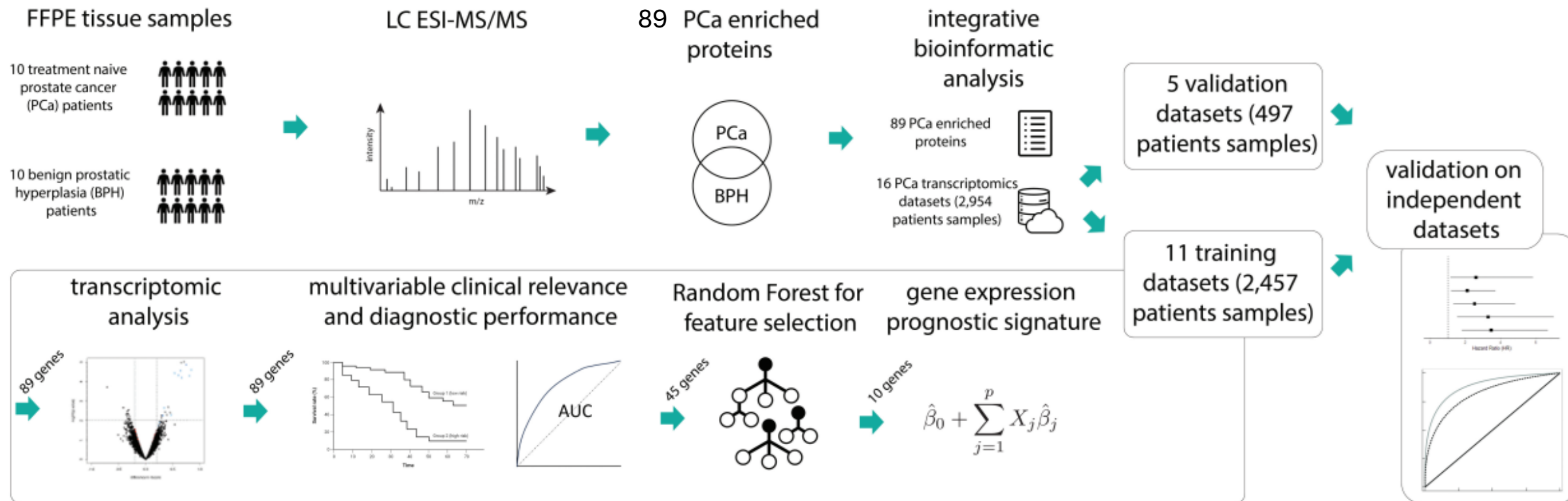
Hipótesis

Una firma molecular, podría ayudar a definir el riesgo de enfermedad agresiva dentro lo un cáncer de próstata (PCa) de grado intermedio y mejorar la predicción de la progresión de la enfermedad.

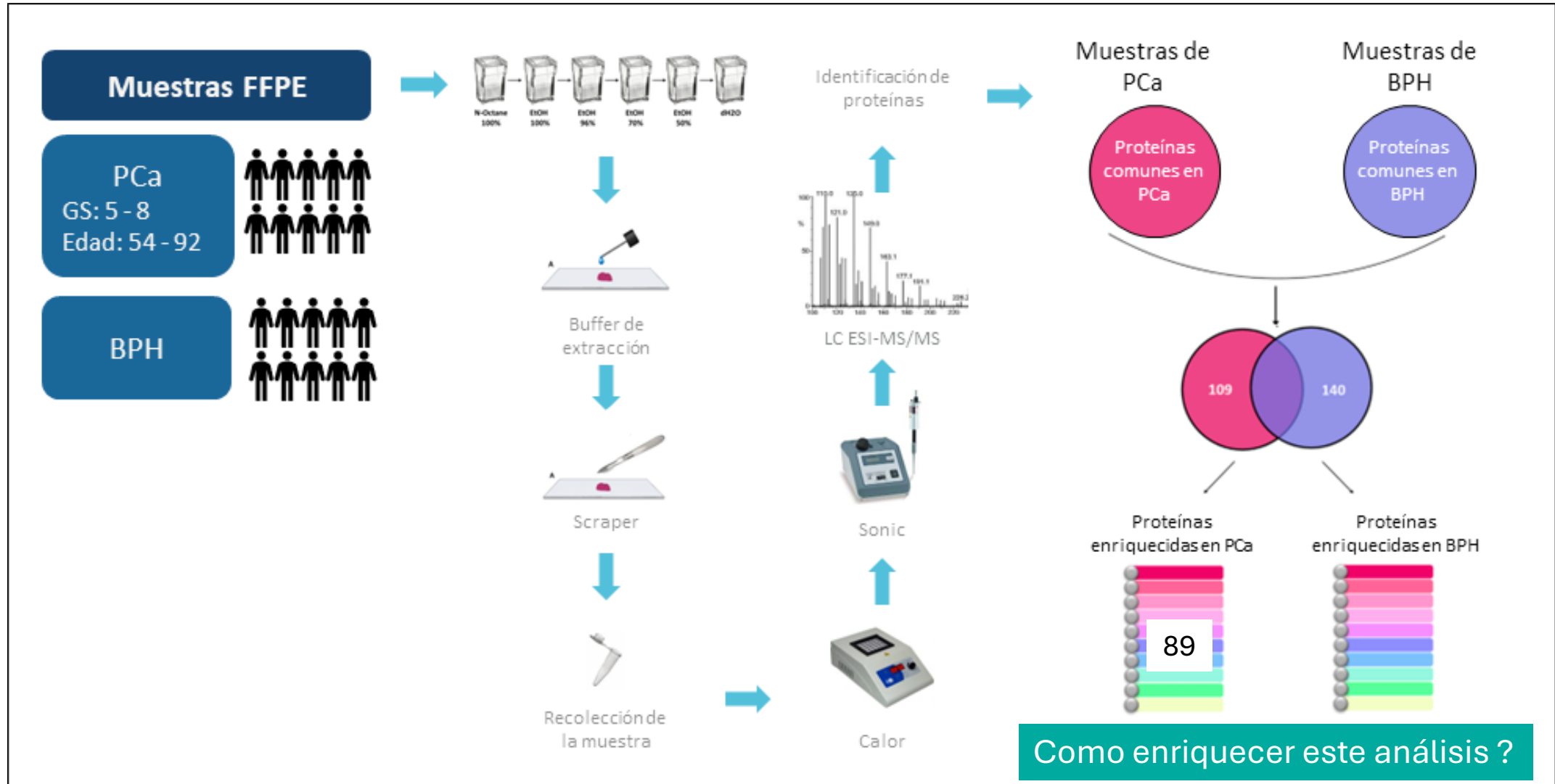
Objetivo

Desarrollar una firma molecular para la estratificación del riesgo del PCa agresivo, independiente de otros parámetros clínicos y patológicos y reconocer, de esta manera, grados intermedios que pueden tener más riesgo de progresar.

Esquema general del trabajo



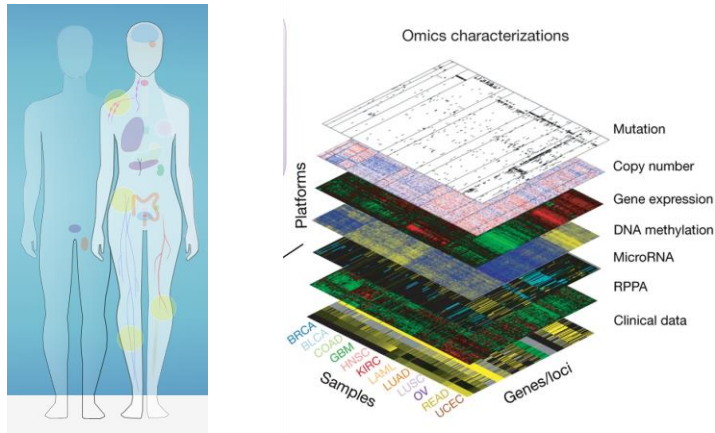
Análisis por **espectrometría de masas** de muestras de tejido de Cáncer de Próstata (PCa) e Hiperplasia Benigna Prostática (BPH)



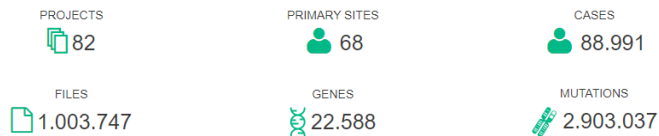
Análisis de datos ómicos en repositorios públicos

Large-scale projects generating cancer genomic datasets

TCGA, ICGC, PCAWG, LINCS, CCLE, CPTAC, Human Protein Atlas, GENIE, CAMELYON, TARGET



Data Portal Summary [Data Release 38.0 - August 31, 2023](#)



Data repositories hosting cancer genomics data

GDC, IDC, TCIA, GEO, Array Express, FDC



Total holdings

	Public	Unreleased	Total
Series	212,015	24,129	236,144
Platforms	25,518	188	25,706
Samples	6,783,632	672,777	7,456,409

Análisis de supervivencia univariable

Set de datos	n	Evento
GSE16560	281	OS
SU2C	81	OS
GSE10645	596	DSS, RFS
TCGA - PRAD	497	PFS, RFS
GSE70768	111	RFS
GSE70769	92	RFS

7 sets de datos de PCa con información de tiempo a evento
1,906 muestras de pacientes
10 análisis de supervivencia

OS: Sobrevida Global

DSS: Sobrevida Específica de Enfermedad

RFS: Sobrevida Libre de Recaída

MFS: Sobrevida Libre de Metastasis

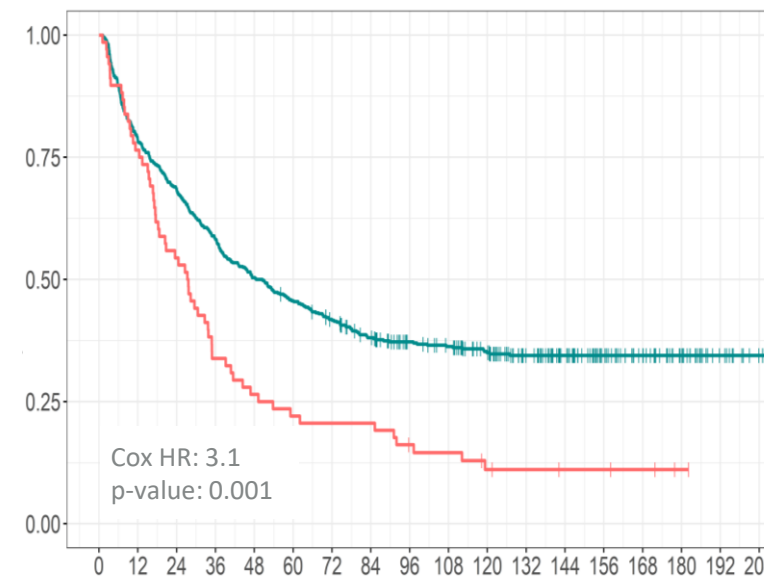
PFS: Sobrevida Libre de Progresión

Expresión génica



■ Baja expresión

■ Alta expresión



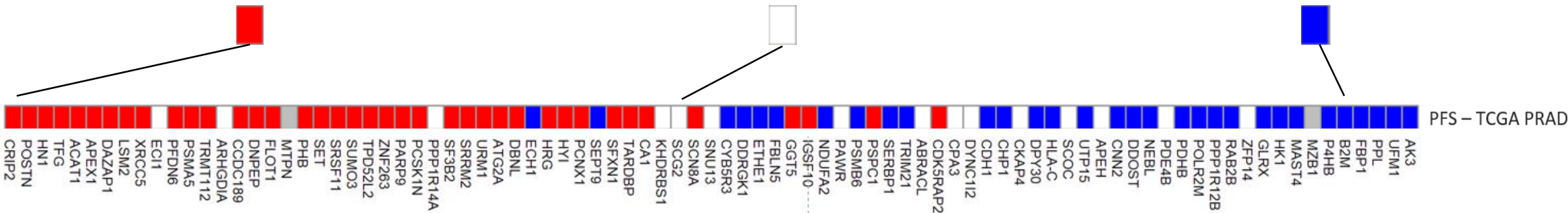
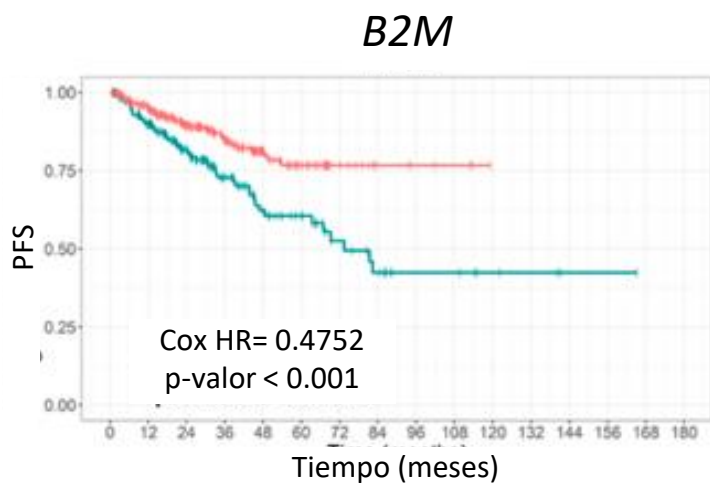
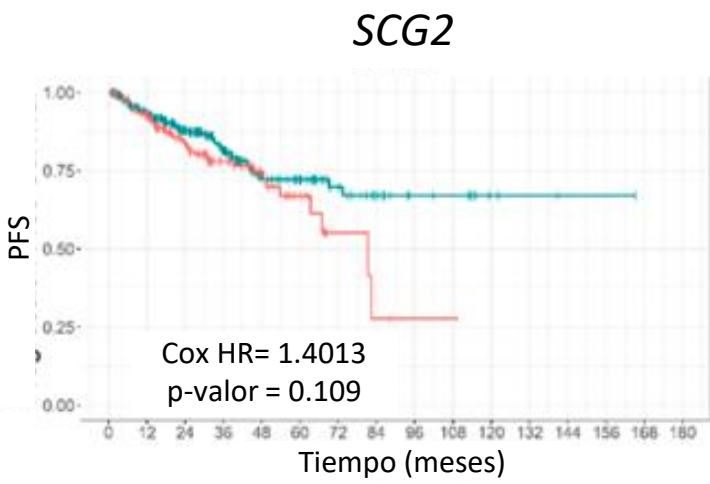
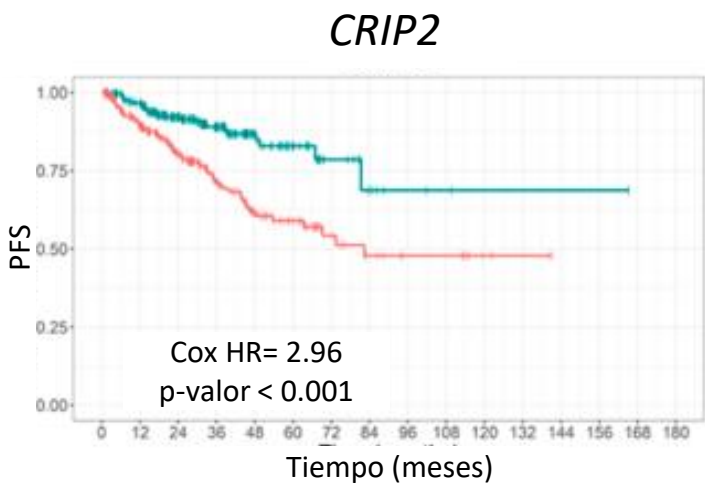
HR>1: alta expresión asociada a alto riesgo de evento

HR<1: alta expresión asociada a bajo riesgo de evento

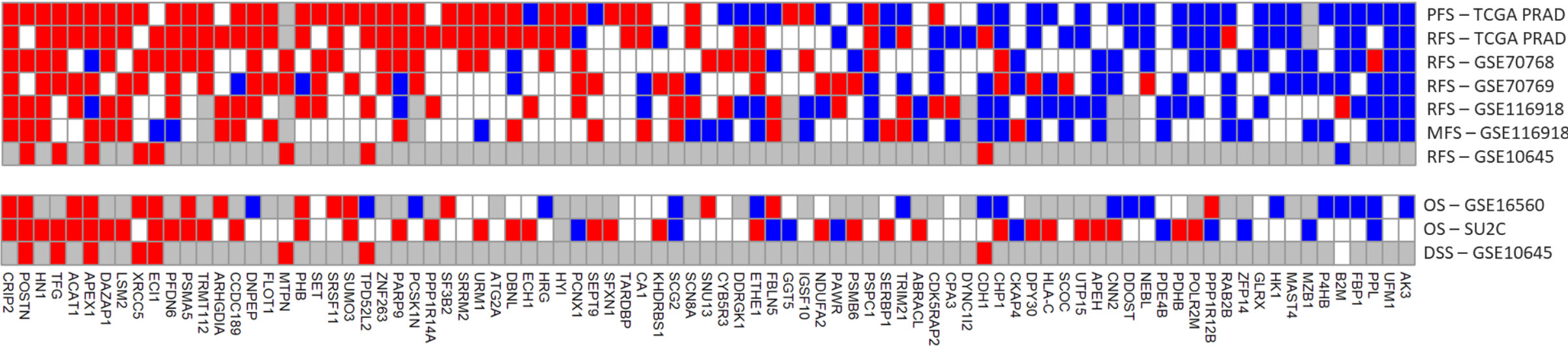
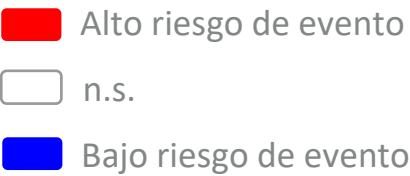
Análisis de supervivencia univariable

TCGA-PRAD (n=497): Hasta 15 años de seguimiento

Baja expresión
Alta expresión



Análisis de supervivencia univariable



Se observa un subgrupo de genes, encabezado por *CRIP2* y *POSTN*, que está consistentemente asociado al riesgo de evento, a través de diversos *sets* de datos

Análisis de supervivencia multivariable

Covariables clínico-patológicas disponibles en TCGA-PRAD:

- Edad
- GS
- PSA
- Estadio T
- Terapia molecular
- Radiación

Supervivencia $\sim$ exp. gen + edad + GS + PSA + estadio T + terapia molecular + radiación

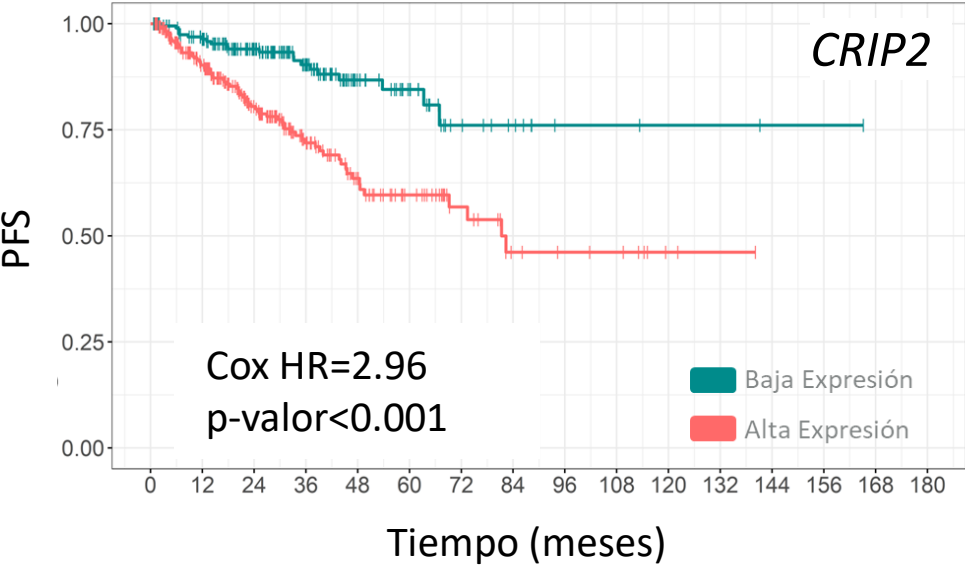
$$h(t) = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \dots + b_px_p)$$

Riesgo de evento = b_1 exp. gen + b_2 edad + b_3 GS + b_4 PSA + b_5 estadio T + b_6 terapia molecular + b_7 radiación

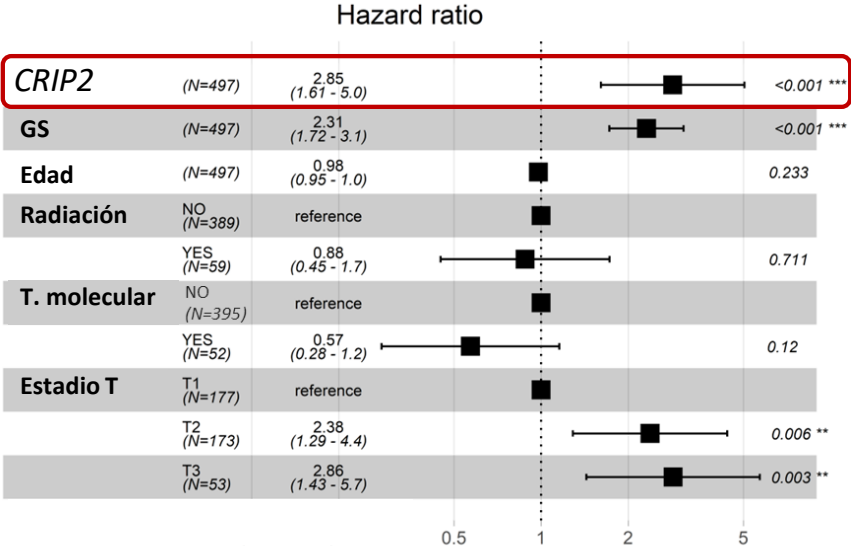
Análisis de supervivencia multivariable

TCGA-PRAD

univariable

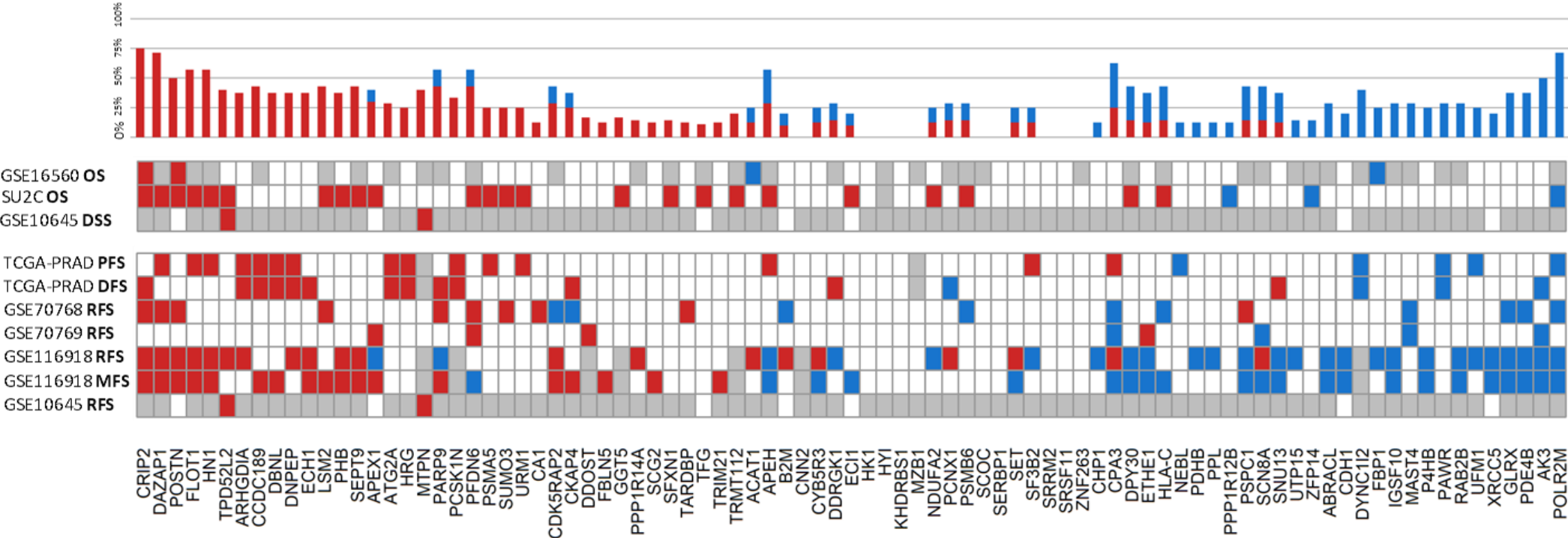


multivariable



Niveles de expresión de *CRIP2* más altos están significativamente asociados con el riesgo de progresión, independientemente de las covariables

Análisis de supervivencia multivariable



	Edad	GS	PSA	Estadio T	Terapia molecular	Terapia radiación	Fusión TMPRSS:ERG2
TCGA-PRAD	x	x		x	x	x	
GSE70768	x	x	x	x			
GSE70769		x	x	x			
GSE116918	x	x	x	x			
GSE10645	x	x		x			
GSE16560	x	x					x
SU2C-PCF	x	x		x			

- Sin cambios en riesgo de evento
- Aumenta el riesgo de evento
- Disminuye el riesgo de evento
- No disponible

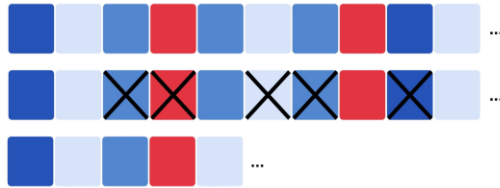
Nuestro análisis multivariable permitió delinear un subgrupo de genes consistentemente asociados a la progresión de la enfermedad independientes de las covariables disponibles

¿Como hago para reducir el número de genes y desarrollar una firma de pocos genes que sea más fácil de implementar?

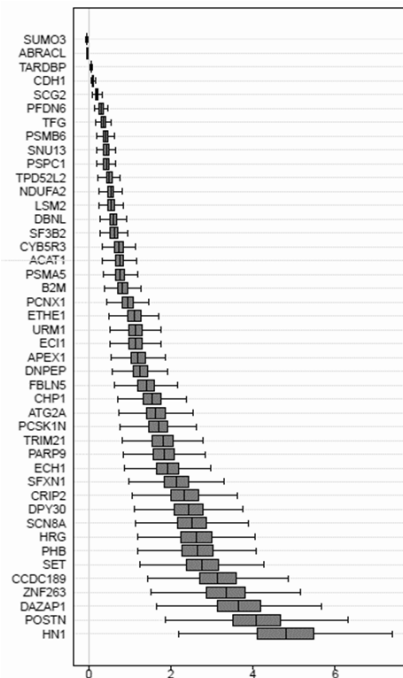
Priorización de genes por estrategias de aprendizaje automático (*Machine Learning*)

Estrategias de machine learning

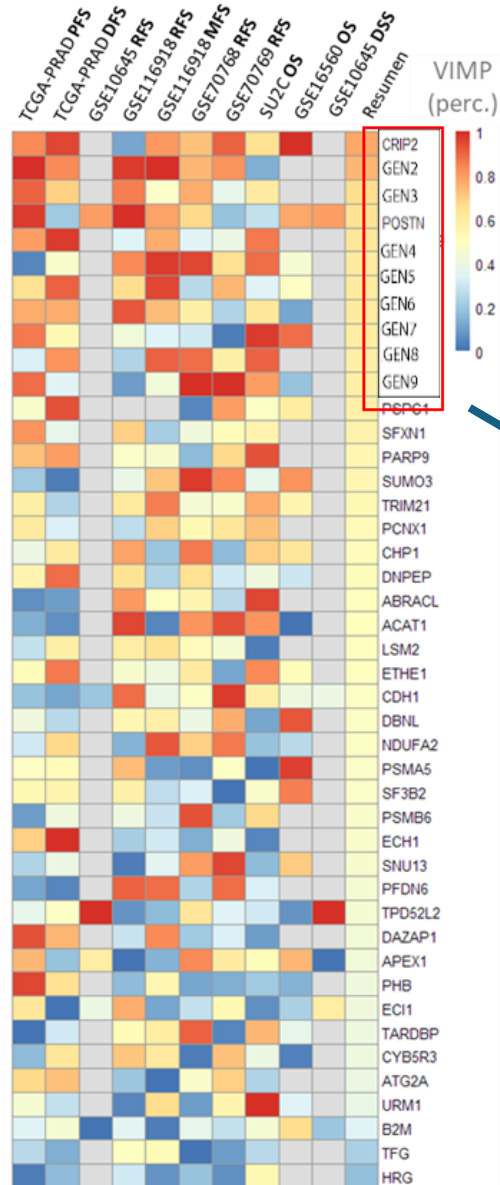
Random forest para la selección de variables



TCGA-PRAD (PFS): cálculo de importancia de variables



(VIMP) Importancia relativa de la variable

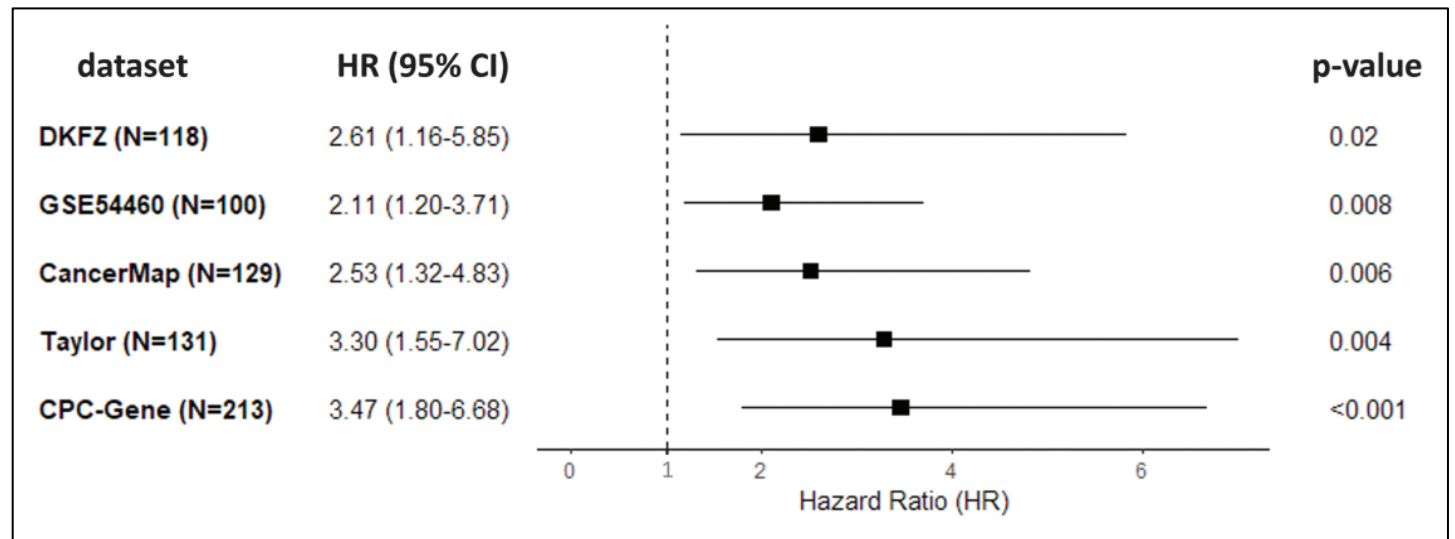
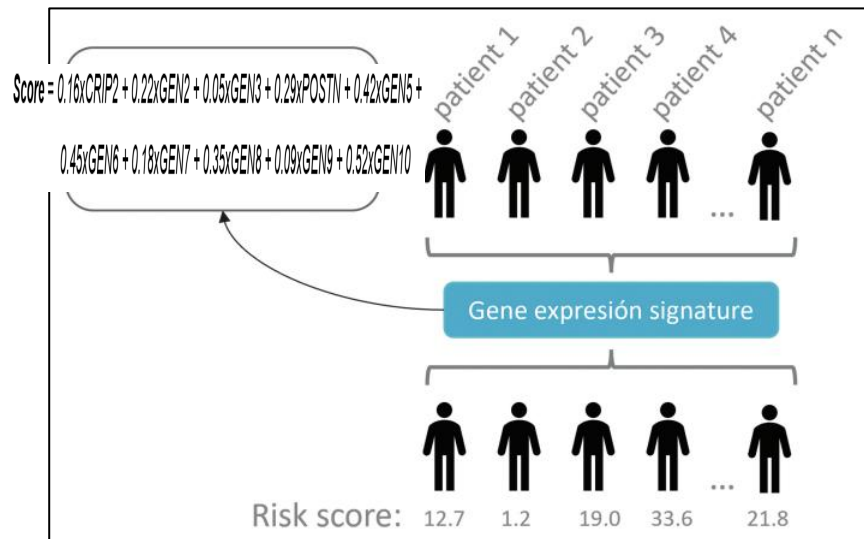


Lasso para el cálculo de los coeficientes

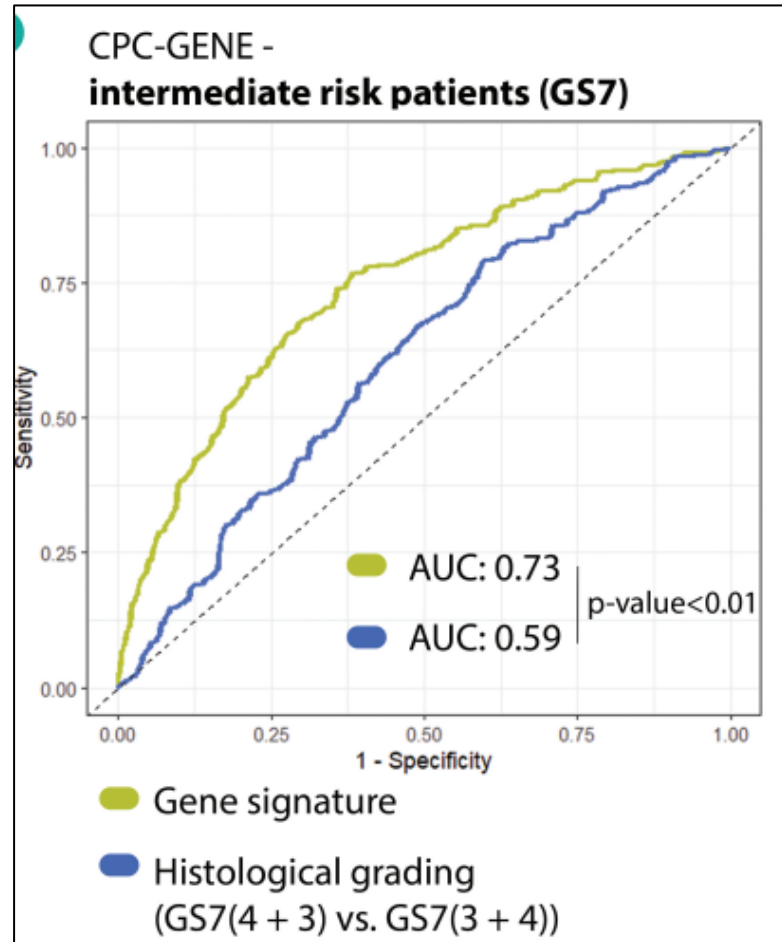
$$\text{Score de riesgo} = b_1 \text{ exp. gen.1} + b_2 \text{ exp. gen.2} + \dots + b_n \text{ exp. gen.n}$$

$$\text{Score} = 0.16 \times \text{CRIP2} + 0.22 \times \text{GEN2} + 0.05 \times \text{GEN3} + 0.29 \times \text{POSTN} + 0.42 \times \text{GEN5} + 0.45 \times \text{GEN6} + 0.18 \times \text{GEN7} + 0.35 \times \text{GEN8} + 0.09 \times \text{GEN9} + 0.52 \times \text{GEN10}$$

Validación en sets de datos independientes



Evaluación del desempeño



	Firma_LIC_10_genes	Decipher	OncoPrint	Prolaris
TCGA.PRAD-PFS (n=497)	0.69	0.71	0.71	0.71
TCGA.PRAD-DFS (n=337)	0.71	0.68	0.72	0.72
SU2C.POL-OS (n=81)	0.62	0.6	0.53	0.67
SU2C.CAP-OS (n=71)	0.53	0.56	0.55	0.61
GSE94767-RFS (n=233)	0.59	0.59	0.63	0.6
GSE70769-RFS (n=92)	0.7	0.69	0.64	0.68
GSE70768-RFS (n=111)	0.66	0.74	0.66	0.6
GSE54460-RFS (n=106)	0.62	0.56	0.62	0.46
GSE21034-RFS (n=140)	0.66	0.76	0.79	0.69
GSE21034-OS (n=140)	0.82	0.69	0.72	0.57
GSE16560-OS (n=281)	0.58	0.58	0.62	0.58
GSE116918-RFS (n=248)	0.56	0.62	0.72	0.52
GSE116918-MFS (n=248)	0.6	0.62	0.76	0.58
GSE107299-RFS (n=65)	0.61	0.66	0.64	0.58

C.index

0.8
0.7
0.6
0.5

Próximos pasos

Evaluar la firma en nuevos sets de datos

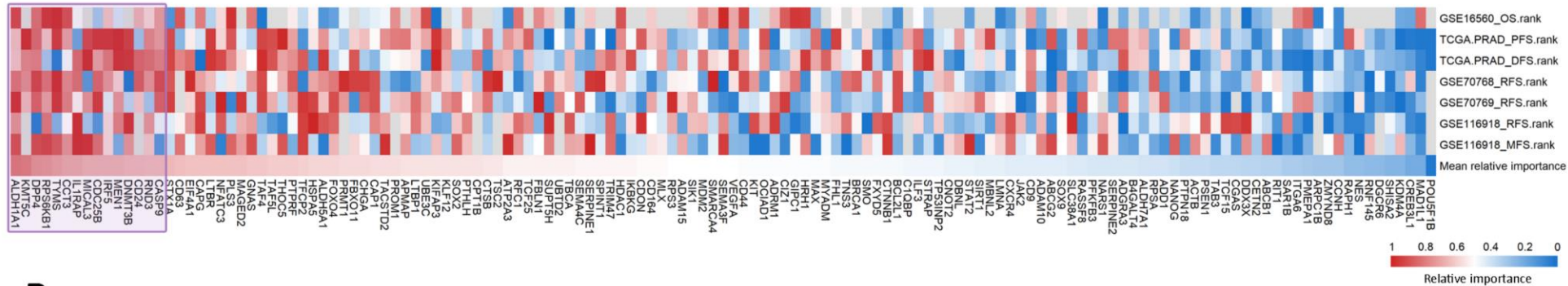
Evaluación de estrategias para su implementación:

- Desarrollo de kit
- Estudios retrospectivos
- Población local

Firma de 7 genes para PCa neuro-endócrino agresivo

A

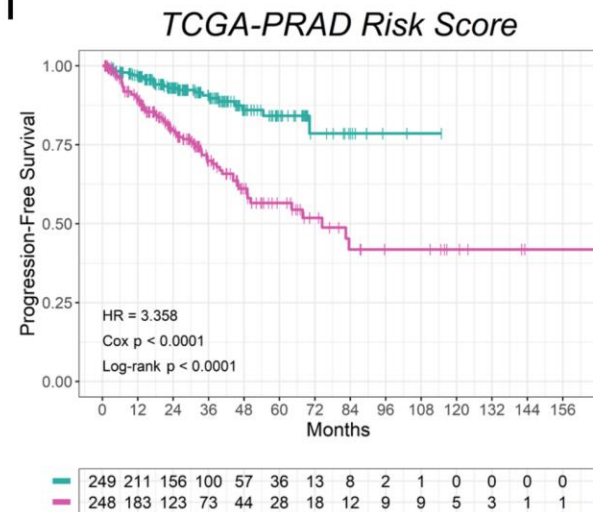
Gene ranking using random forest algorithm



B

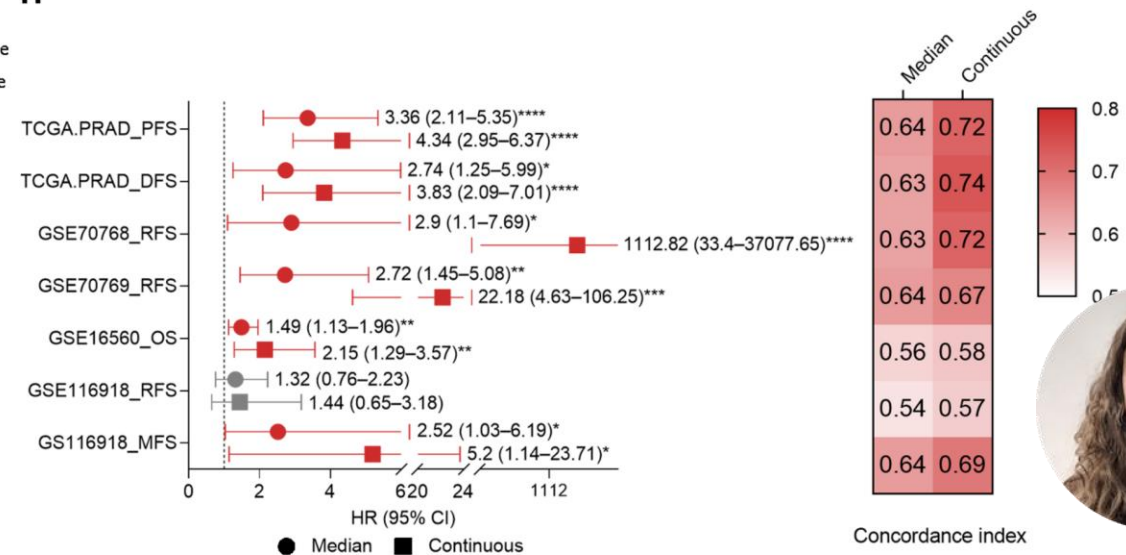
Risk score predictive performance in training datasets

i



$$\text{Risk Score} = 0.284 \times KMT5C + 0.272 \times MEN1 + 0.218 \times TYMS + 0.090 \times IRF5 + 0.083 \times DNMT3B + 0.048 \times CDC25B - 0.060 \times DPP4$$

ii



Lic. Agustina A. Sabater

Unmasking Neuroendocrine Prostate Cancer with a Machine Learning-Driven Seven-Gene Stemness Signature That Predicts Progression

by Agustina Sabater ^{1,2,3}  , Pablo Sanchis ^{1,2,3} , Rocio Seniuk ^{1,2}, Gaston Pascual ^{1,2} , Nicolas Anselmino ⁴ , Daniel F. Alonso ⁵ , Federico Cayol ⁶ , Elba Vazquez ^{1,2} , Marcelo Marti ^{1,2}, Javier Cotignola ^{1,2} , Ayelen Toro ^{1,2} , Estefania Labanca ⁴, Juan Bizzotto ^{1,2,3,*}   and Geraldine Gueron ^{1,2,*}  

¹ Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EGA, Argentina

² Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), CONICET-Universidad de Buenos Aires, Buenos Aires C1428EGA, Argentina

³ Instituto de Tecnología (INTEC), Universidad Argentina de la Empresa (UADE), Buenos Aires C1073AAO, Argentina

⁴ Department of Genitourinary Medical Oncology and The David H. Koch Center for Applied Research of Genitourinary Cancers, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵ Centro de Oncología Molecular y Traslacional y Plataforma de Servicios Biotecnológicos, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal B1876BXD, Argentina

⁶ Sector de Oncología Clínica, Hospital Italiano de Buenos Aires, Buenos Aires C1199ABB, Argentina

* Authors to whom correspondence should be addressed.

Transcriptómica con impacto clínico:
más allá de las firmas de expresión génica

Transcriptómica *discovery-driven*



Dra. Estefanía Labanca



Dr. Pablo Sanchis

Identificación de **variantes y genes de fusión** en leucemia linfoblástica aguda pediátrica

- Identificar y caracterizar alteraciones moleculares con **valor pronóstico y/o terapéutico** en LLA pediátrica
- Desarrollar herramientas y material de referencia que permitan **mejorar la clasificación molecular actual** de los pacientes con LLA en Argentina
- Evaluar la utilidad de **integrar** información de expresión génica a la interpretación de variantes

Protocolo clínico: **ALLIC GATLA 2010**

Multicéntrico, nacional, integrado al consorcio internacional ALLIC
N=1711, 2010-2020

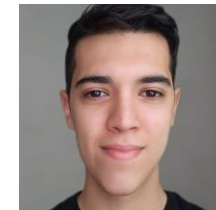
Hospital de niños R. Gutierrez
Hospital de niños Ludovica
Hospital posadas



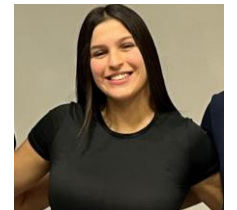
Dr. Javier Cotignola



Dra. Sol Ruiz

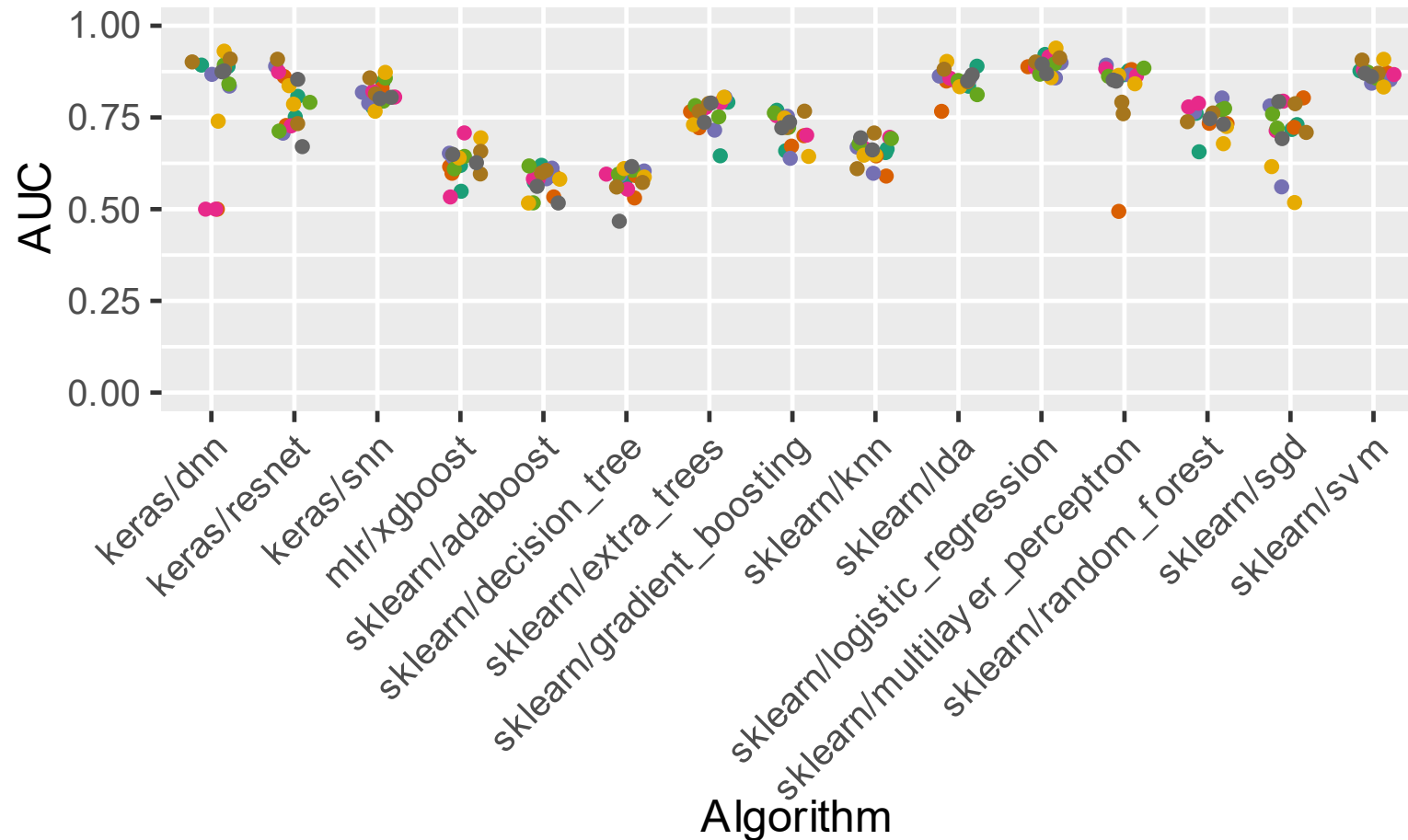


Est. Ignacio Gómez



Est. Lucila Viappiani

ómicas como oportunidad para investigación interdisciplinaria



- Evaluación de **diferentes algoritmos** para **predicción de metástasis** en cáncer de próstata



Dr. Pablo Turjanski



Est. Lucas Abbate

Cerrando...

La **transcriptómica** permite ver la **complejidad** de la biología a nivel célula o tejido

La **informática** permite destilar **información** de esa complejidad

La **colaboración** permite trabajar entre **disciplinas**

Espacios de **encuentro** fomentan la colaboración

Fortalecer nuestra **ciencia**

¡GRACIAS!

III Jornadas de Genómica Clínica

Desarrollo de firmas transcriptómicas para estratificación de riesgo en cáncer

Disertante: **Dr. Juan Bizzotto** (juanantoniobizzotto@gmail.com)



Dra. Geraldine Gueron



Dr. Javier Cotignola



Dra. Elba Vazquez



Dra. Ayelén Toro



Dra. Sol Ruiz



Dr. Juan Bizzotto



Dr. Pablo Sanchis



Lic. Agustina Sabater



Bioq. Gastón Pascual



Lic. Rocío Seniuk



Lic. Sabrina Ledesma



Est. Ignacio Gómez



Est. Julia Lechuga



Est. Mora Gatti



Est. Lucila Viappiani



Est. Nicolas Crespi



Est. Eden Korniansky

Instituciones



Financiamiento



Colaboraciones

