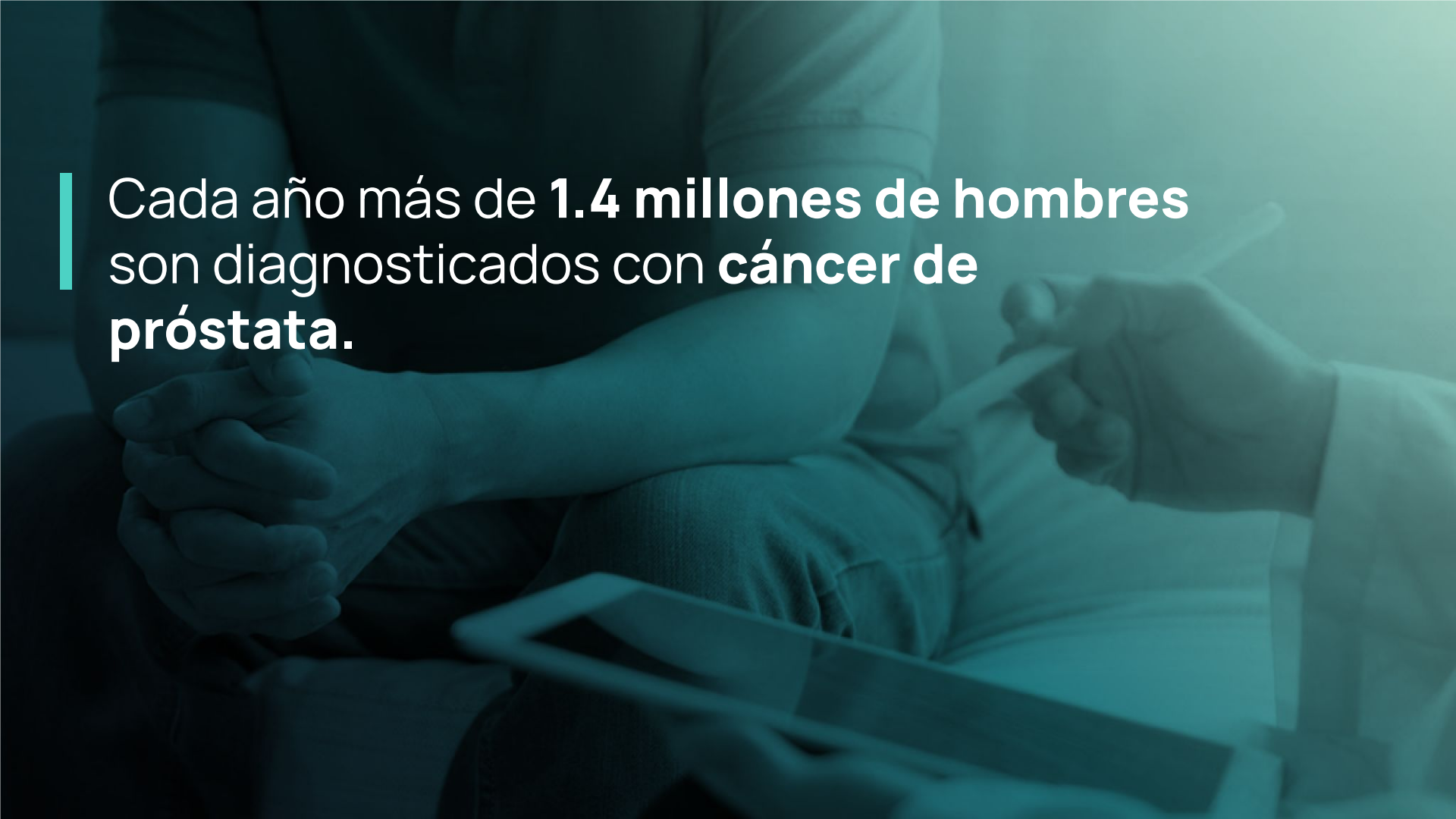




PreGen
PREDICTIVE GENOMIC INTELLIGENCE





Cada año más de **1.4 millones de hombres** son diagnosticados con **cáncer de próstata**.

| El problema



Biopsia



Diagnóstico

El diagnóstico es simple

El problema



Biopsia



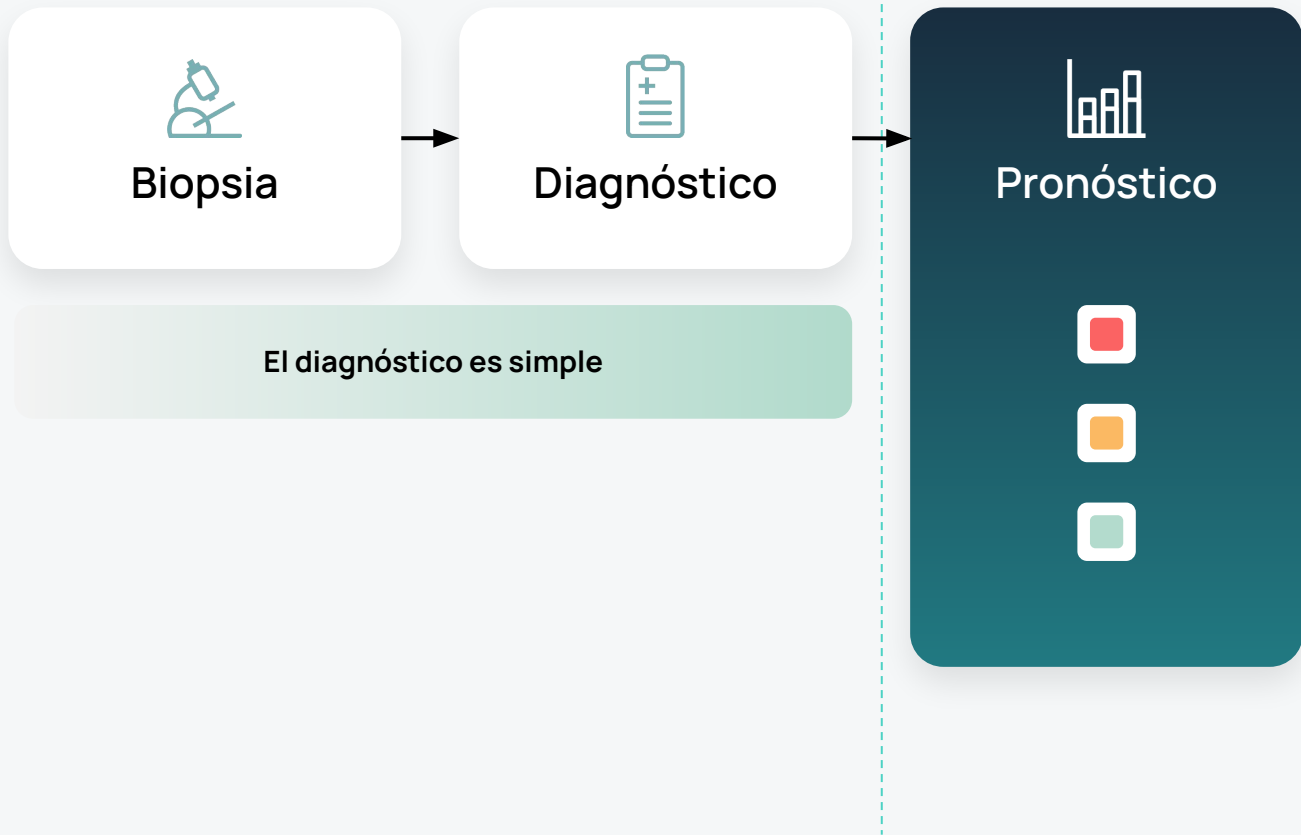
Diagnóstico



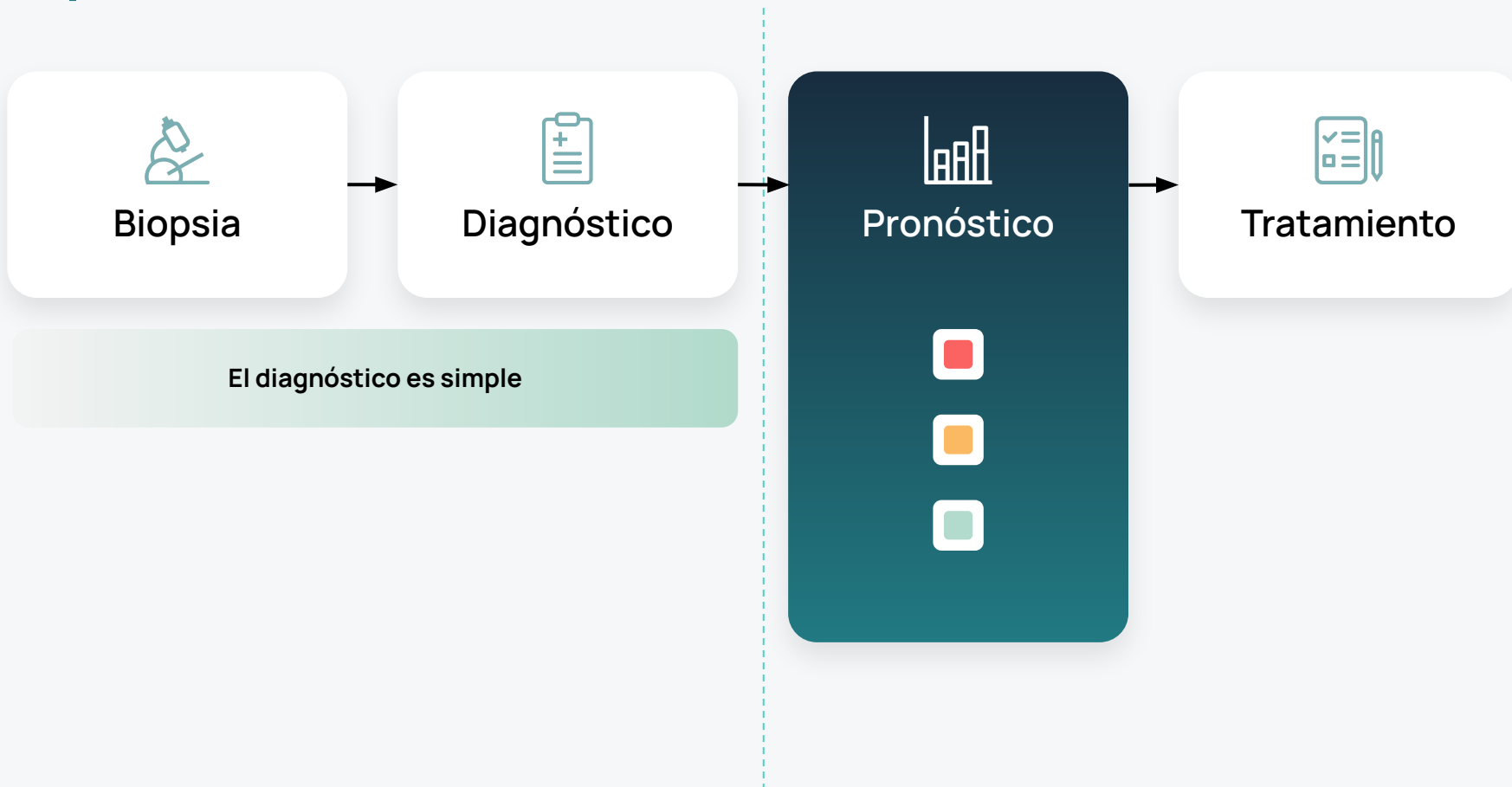
El diagnóstico es simple



El problema



El problema



| El problema



Reciben tratamientos
tóxicos innecesarios

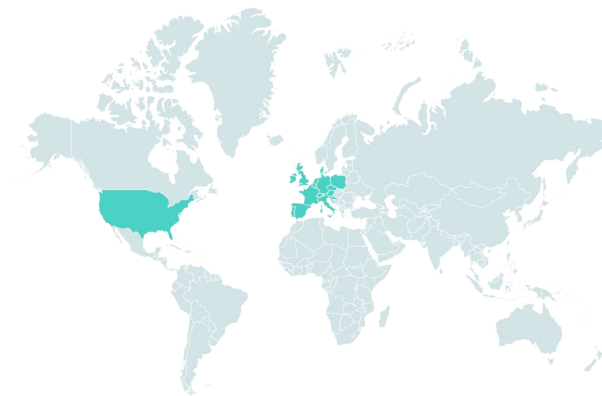


No reciben
el tratamiento adecuado

Tests de pronóstico disponibles

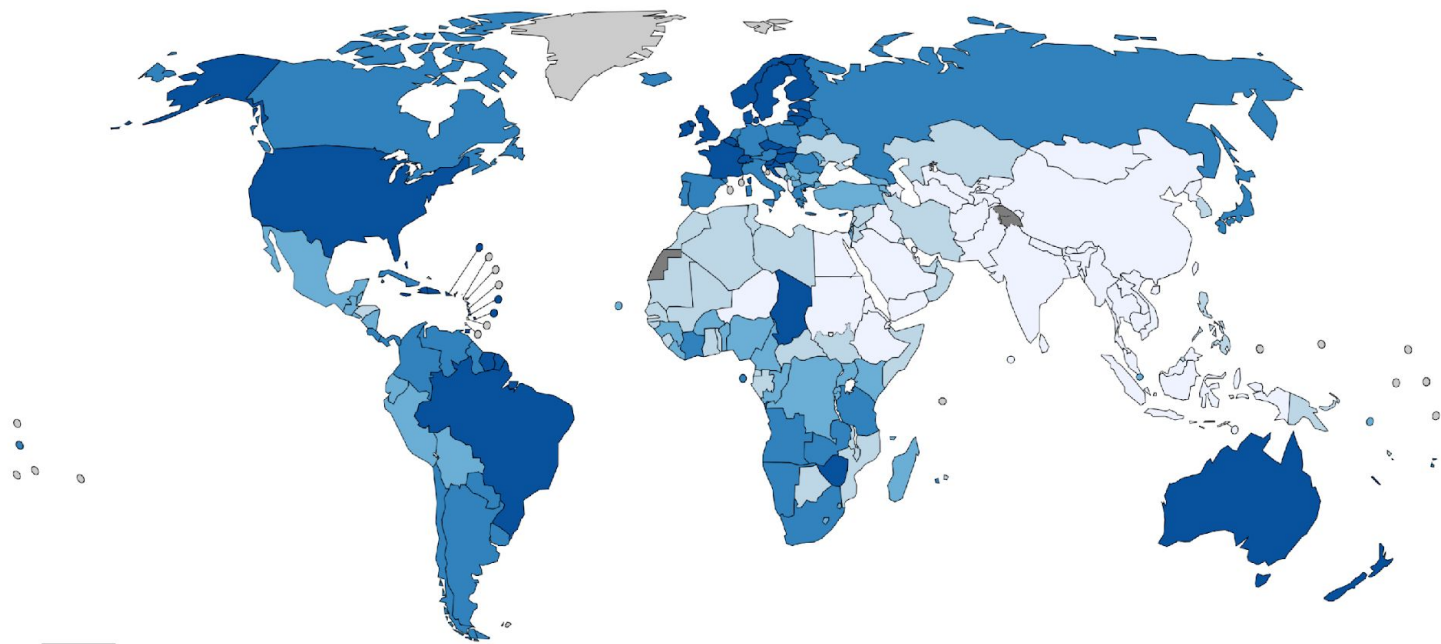


Caros
+US\$ 4000

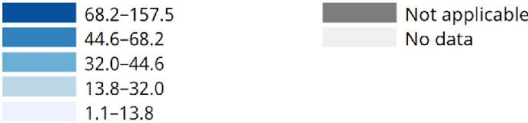


Tecnología centralizada

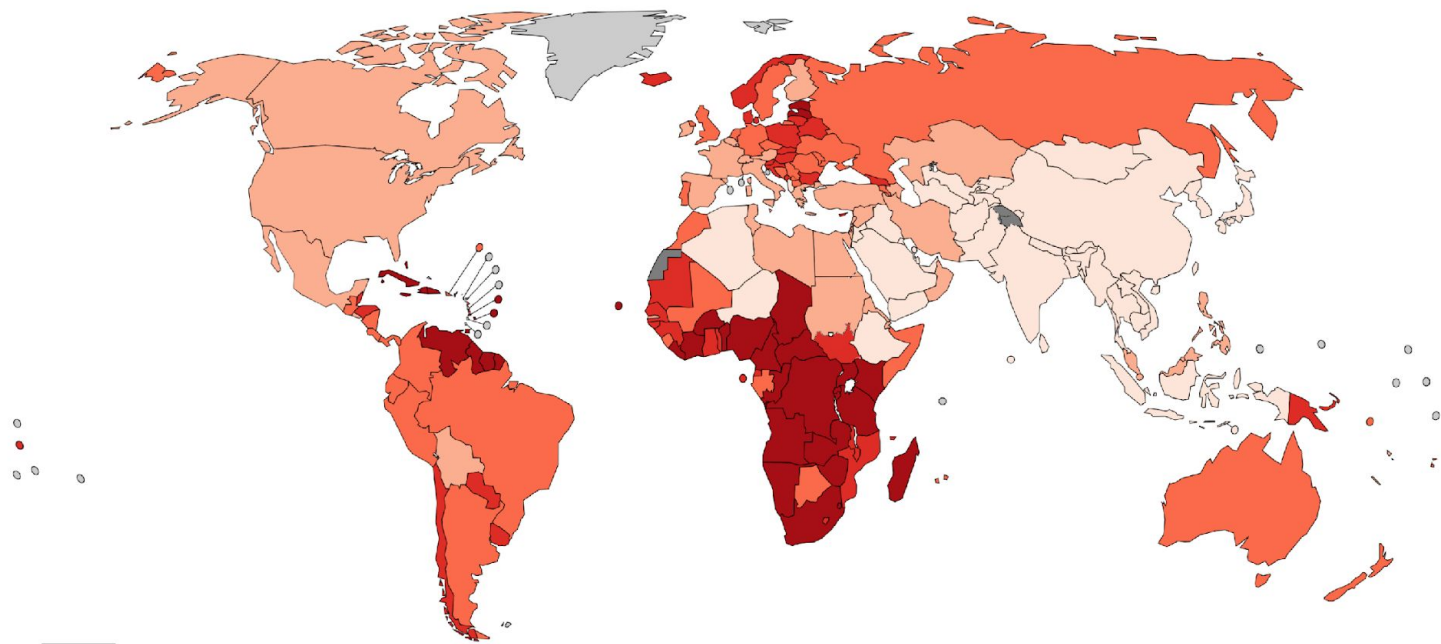
Prostate Cancer Incidence



ASR (World) per 100 000



Prostate Cancer Mortality



ASR (World) per 100 000



Este problema es prevalente
en muchos tipos de tumores.

Tecnología

Pipeline

Literature review



Selection of genes

```
candidate.genes<-c("gene_1",  
"gene_2",  
...,  
"gene_n")
```



Database review



Selection of datasets



Data
curation



Integrative
bioinformatic
analysis



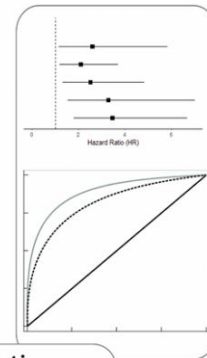
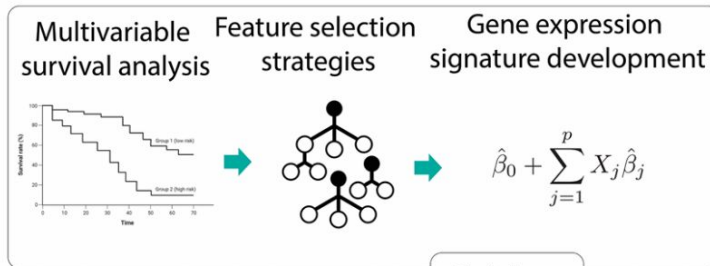
Training
datasets



Validation
datasets



Validation on
independent
datasets



Data review, selection and curation

Literature review



144 stemness-associated genes in Prostate Cancer



- Gene **symbols** and **IDs** curation
- Mapping to HGNC

Database review



Inclusion criteria



9 PCa patients datasets (n=1.730) with molecular and clinical data.
Data **cleansing**, **harmonization** and **normalization**.



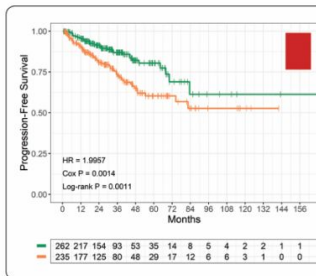
5 training datasets
4 validation datasets

- Remove patients with incomplete data
- Exclude genes with low counts
- Match probes to gene IDs
- Standardize clinical variable formats
- Normalize gene expression data
- Batch effect correction
- Impute missing values
- Check for duplicated samples or patients
- Stratify by relevant clinical variables
- ...

Dataset	Samples	Survival Endpoint	Covariates	Cohort
TCGA-PRAD	497 PCa (RNAseq)	Disease Progression Disease-Free Time (n = 337)	Gleason Group, PSA levels, Clinical T Stage, Targeted	Training
GSE70768	111 PCa (Microarray)	Biochemical Relapse	Age, Gleason Group, PSA levels, T Stage	Training
GSE70769	92 PCa (Microarray)	Biochemical Relapse	Gleason Group, PSA levels, T Stage	Training
GSE116918	248 PCa (Microarray)	Metastasis Development Relapse	Age, Gleason Score, PSA levels, T Stage	Training
GSE16560	281 PCa (Microarray)	Death	Age, Gleason Group	Training
GSE54460	106 PCa (RNA-seq)	Biochemical Relapse	Age, Gleason Score, PSA levels, T Stage	Validation
GSE94767	233 PCa (Microarray)	Biochemical Relapse	Gleason Group, PSA levels, T Stage	Validation
DKFZ	81 PCa (RNA-seq)	Biochemical Relapse	Age, Gleason Score, PSA levels, T Stage	Validation
SU2C-PCF	81 metastatic CRPC (RNA-seq)	Death	Age, Gleason Score, PSA levels	Validation

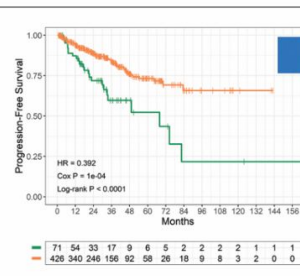
Survival analysis

B



TCGA (PFS) *DBNL*

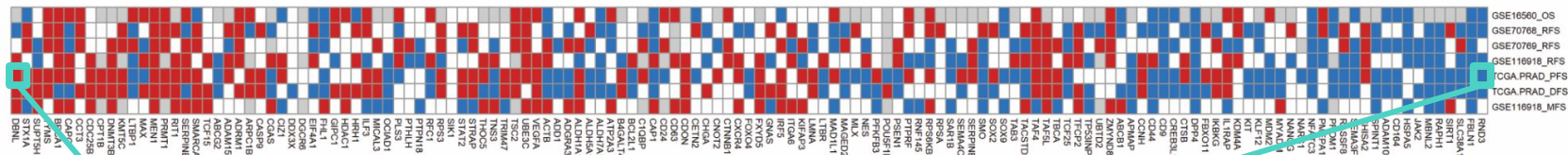
High expression Low expression



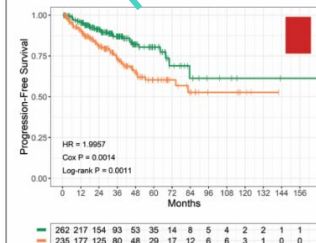
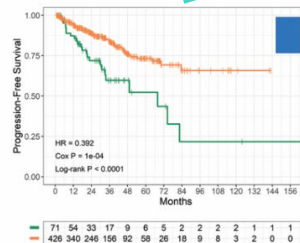
TCGA (PFS) *MBNL2*

Survival analysis

A Univariable analysis



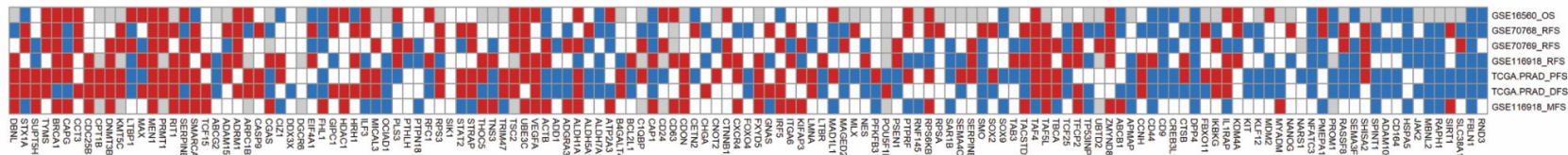
B

TCGA (PFS) **DBNL**TCGA (PFS) **MBNL2**

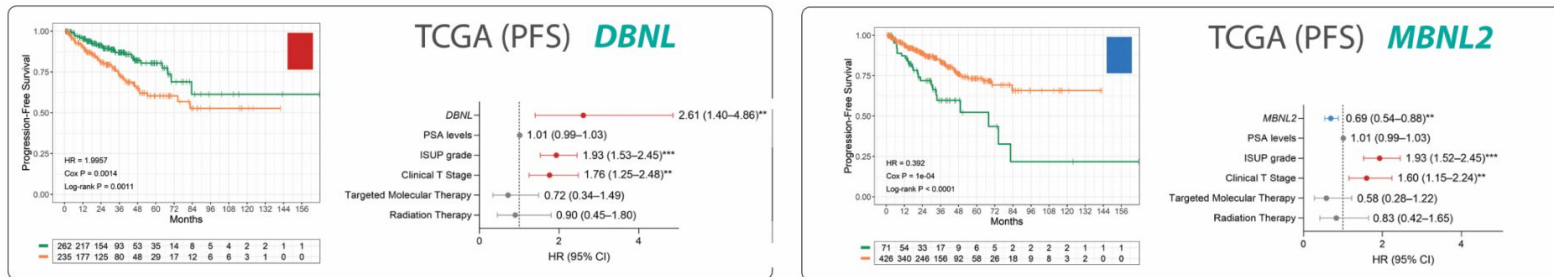
■ High expression ■ Low expression

Survival analysis

A Univariable analysis

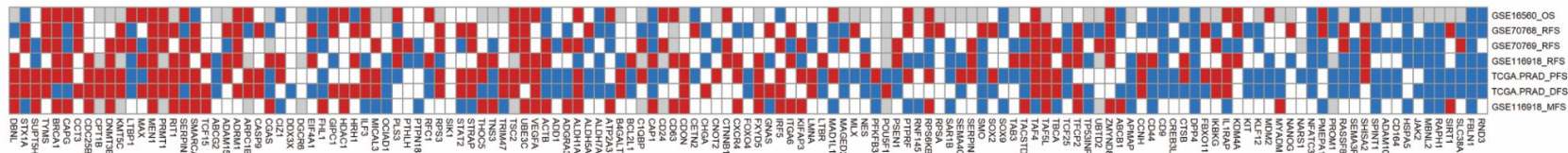


B

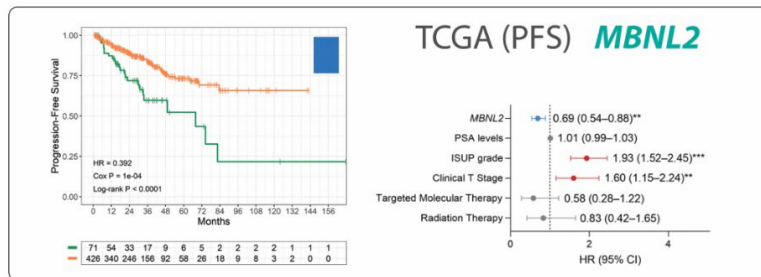
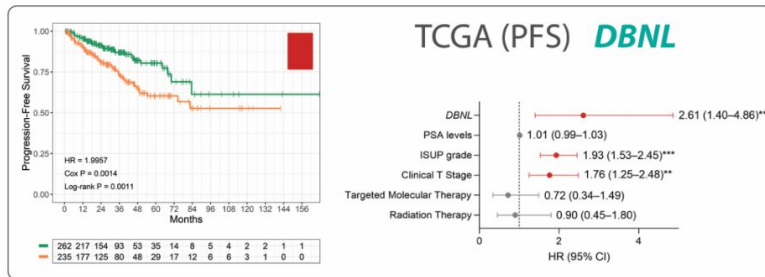


Survival analysis

A Univariable analysis

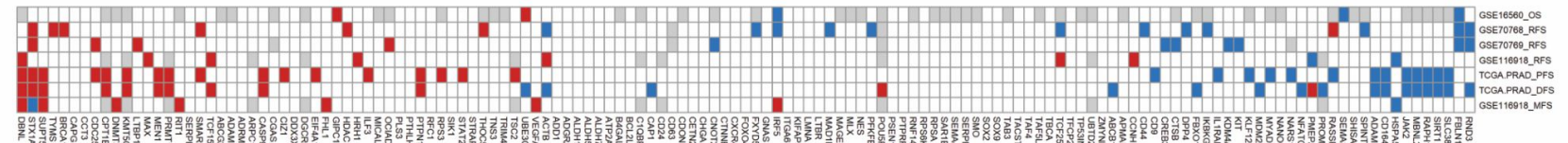


B



Orange High expression Green Low expression

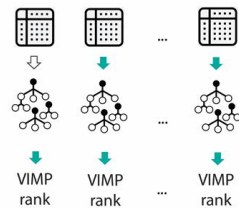
C Multivariable analysis



Red High risk Blue Low risk Grey Not available White N.S.

Feature Selection

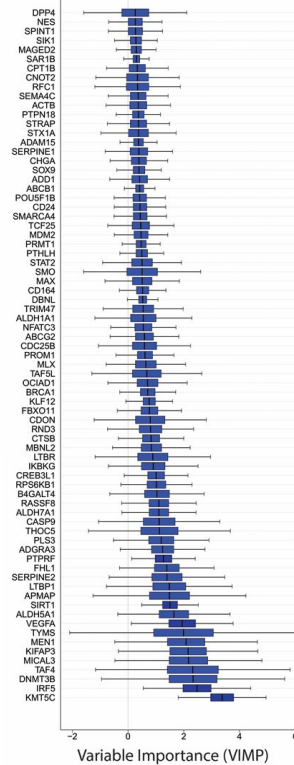
A Random Forest for feature selection



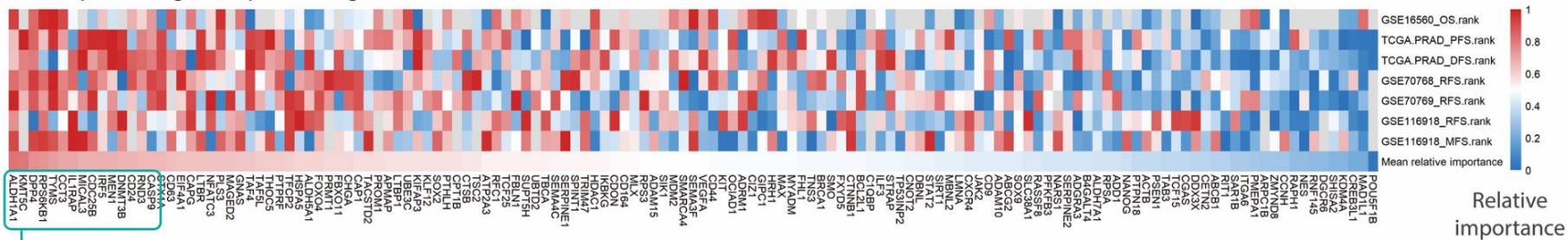
Variable importance
across datasets

Consensus
gene rank

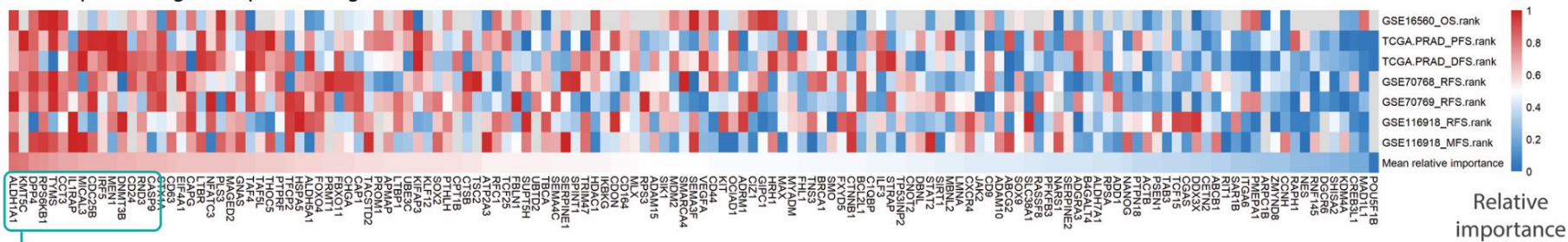
TCGA-PRAD (PFS)



B



B Development of gene expression signature



C

Survival models

Cox

Lasso

Regularized Cox

Ridge

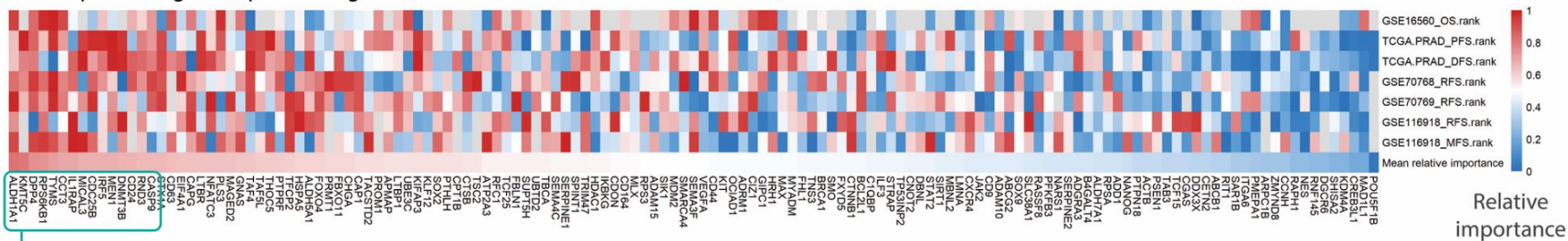
Regularized Cox

$$\hat{\beta}_0 + \sum_{j=1}^p X_j \hat{\beta}_j$$

Coefficient value



B Development of gene expression signature



C

Survival models

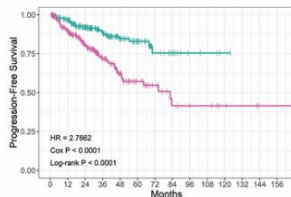
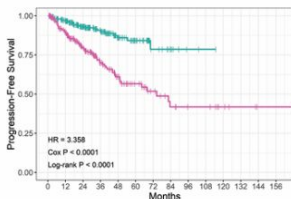
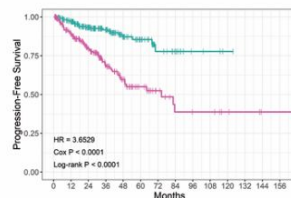
Cox

Lasso
Regularized Cox

Ridge
Regularized Cox

$$\hat{\beta}_0 + \sum_{j=1}^p X_j \hat{\beta}_j$$

Coefficient value



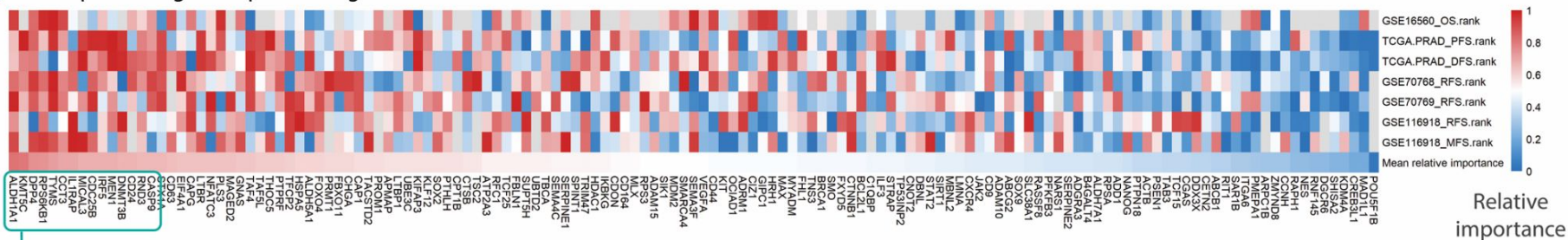
p < 0.001

p < 0.001

High risk score

Low risk score

B Development of gene expression signature



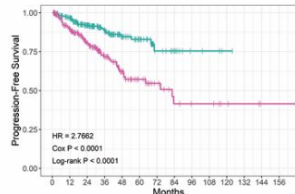
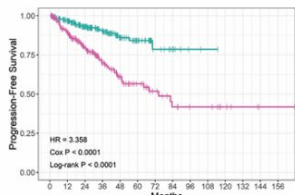
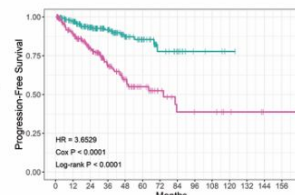
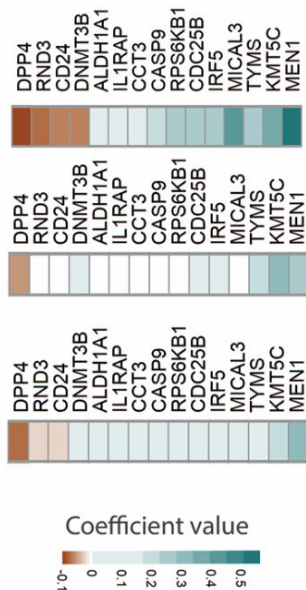
C Survival models

Cox

Lasso
Regularized Cox

Ridge
Regularized Cox

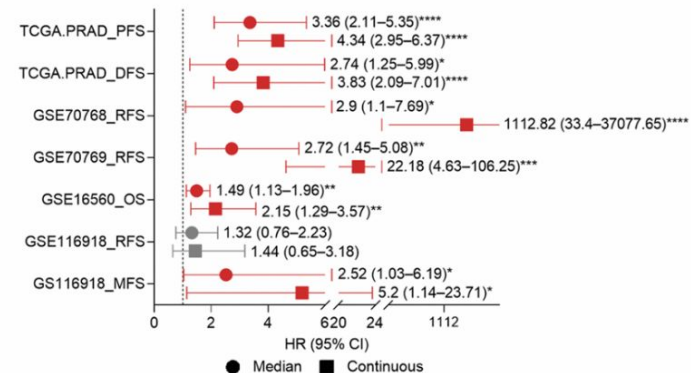
$$\hat{\beta}_0 + \sum_{j=1}^p X_j \hat{\beta}_j$$



Lasso model Risk Score

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

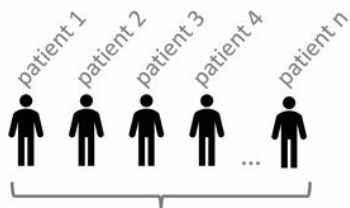
D Evaluation in training datasets



Validation

A

$$\begin{aligned} &0.284 \times \text{KMT5C} + 0.272 \times \\ &\text{MEN1} + 0.218 \times \text{TYMS} + \\ &0.090 \times \text{IRF5} + 0.083 \times \\ &\text{DNMT3B} + 0.048 \times \\ &\text{CDC25B} - 0.060 \times \text{DPP4} \end{aligned}$$



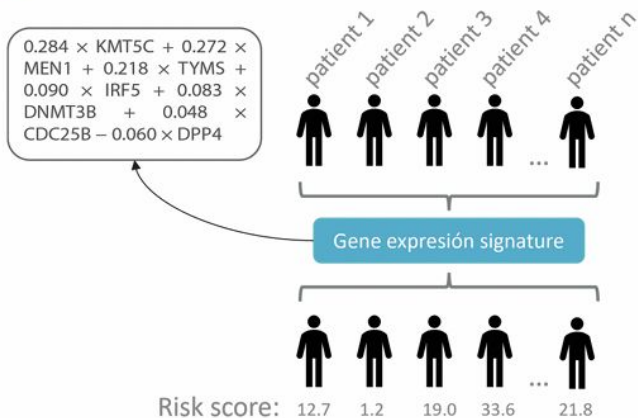
Gene expresión signature



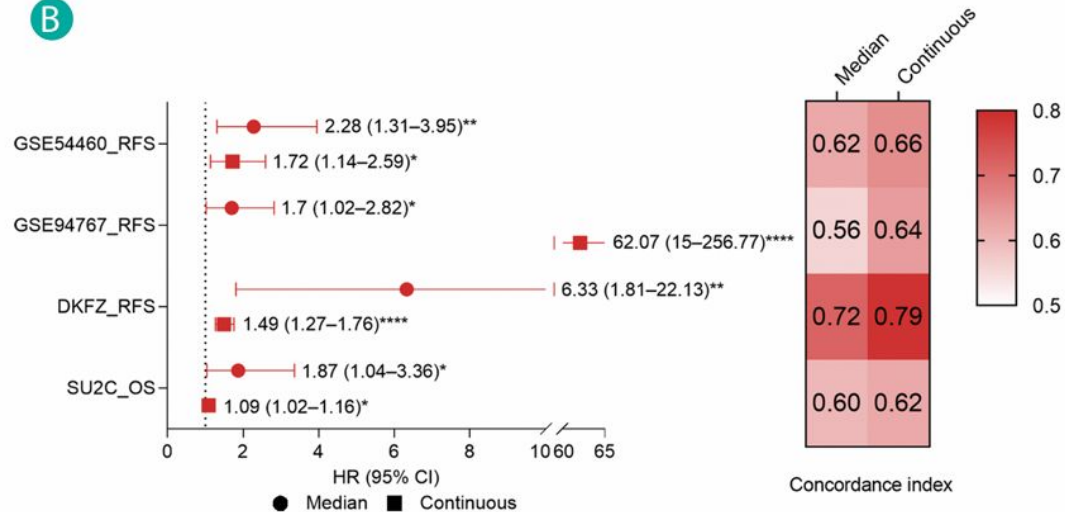
Risk score: 12.7 1.2 19.0 33.6 21.8

Validation

A



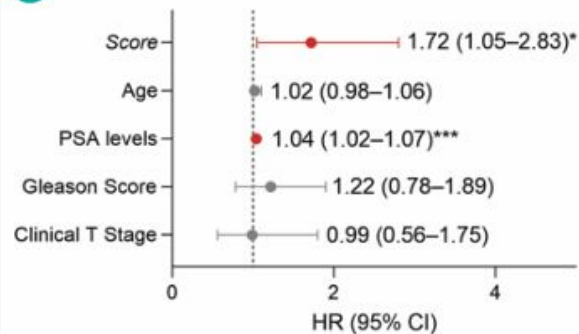
B



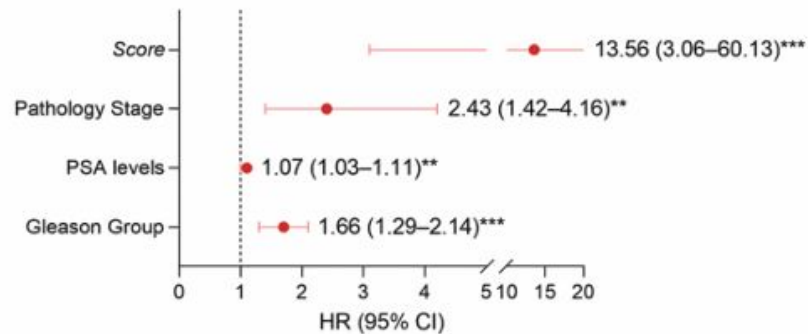
Validation

C

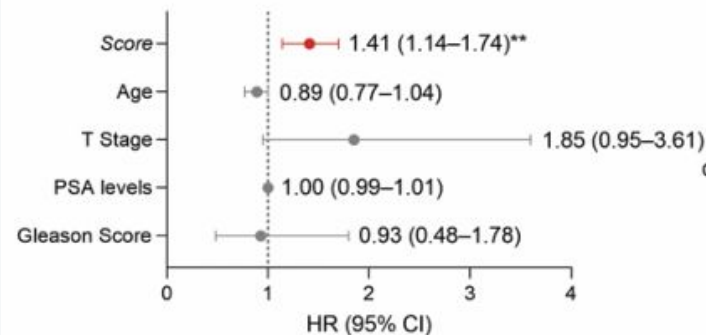
GSE54460 - RFS



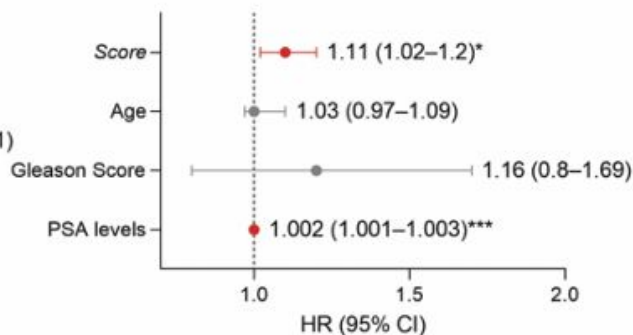
GSE94767 - RFS



DKFZ - RFS

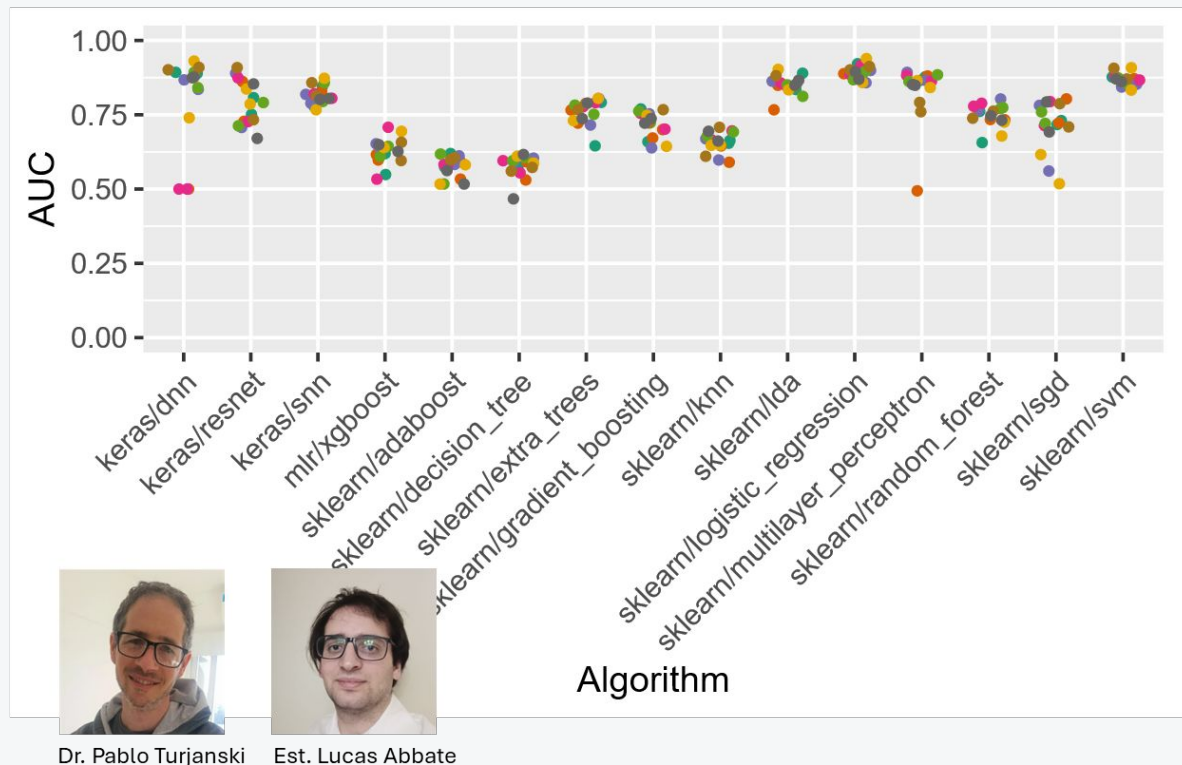


SU2C - OS

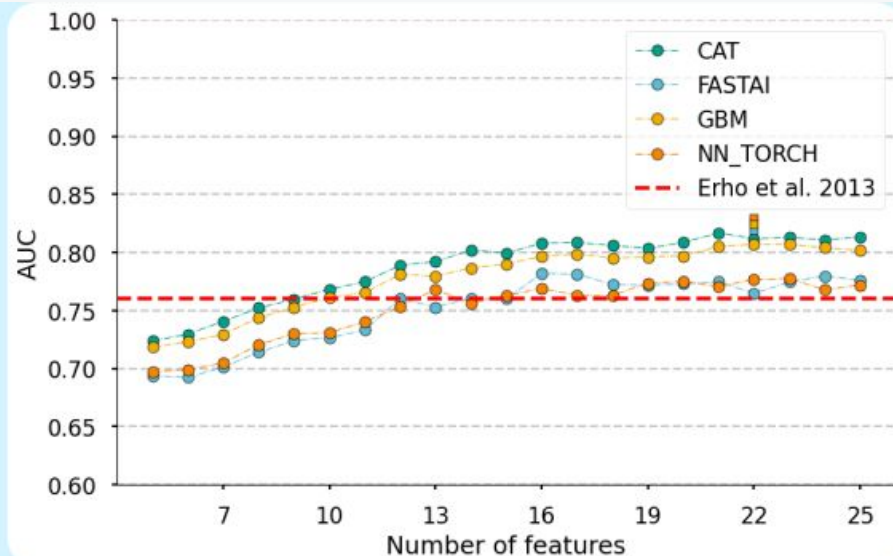
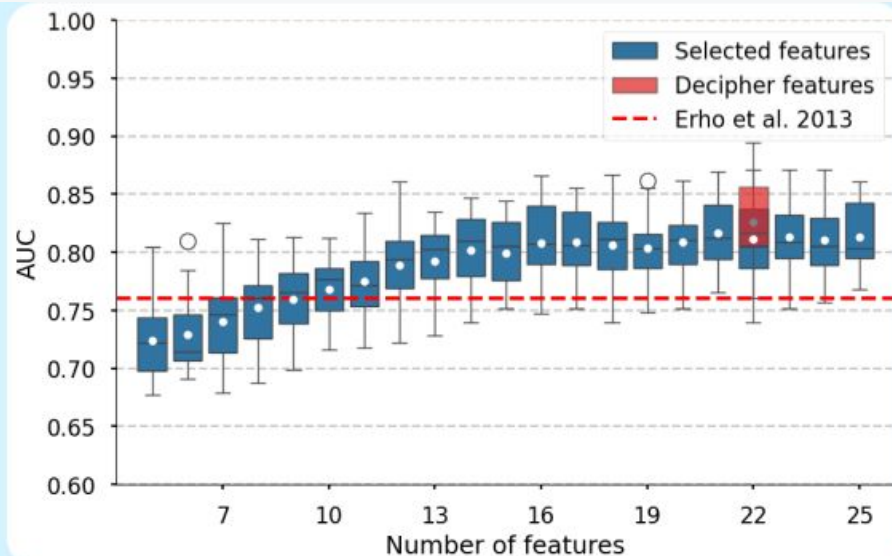


oportunidad para investigación **interdisciplinaria**

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



Evaluación de modelos



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

International Journal of Molecular
Sciences. 2024 Jan;25(21):11356.

Sabater A. et al.



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

International Journal of Molecular
Sciences. 2024 Jan;25(21):11356.

Sabater A. et al.



¿Cómo convertimos esto en algo con impacto?

| Cómo lo hacemos



Biopsia original

| Cómo lo hacemos



Biopsia original



PreGen KIT



Equipos de PCR

| Cómo lo hacemos



Biopsia original



PreGen KIT



Plataforma en la
nube

LOW

INTERMEDIATE

HIGH



Respuesta inmediata

| Solución



**Igual capacidad
pronóstica**



10x más baratos



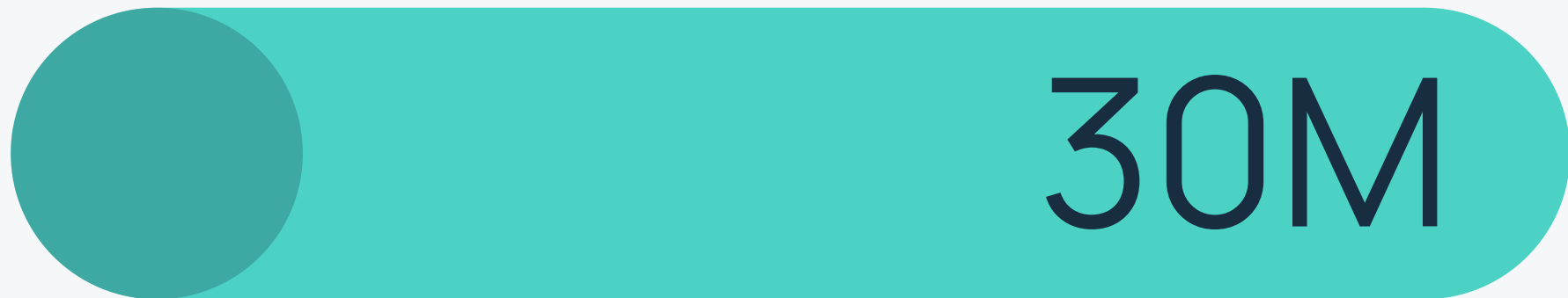
**Completamente
descentralizado**

| El mercado: cáncer de próstata



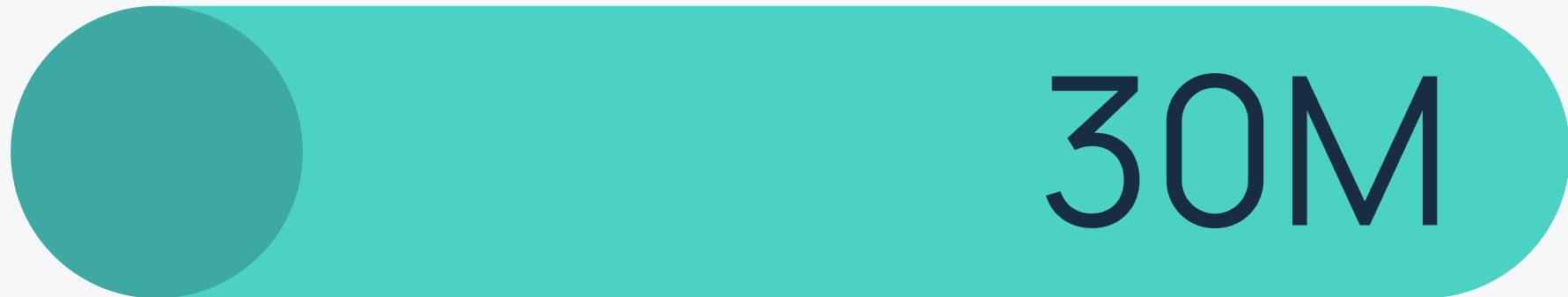
biopsias por año.

| El mercado total



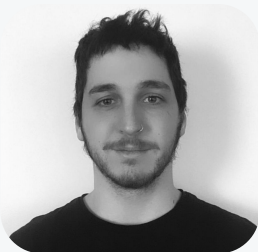
biopsias por año.

| El mercado total



→ U\$\$ 9.0000M

Equipo



Oliver Gibson
MSc, MBA cand.
CEO
Business &
Bioinformática



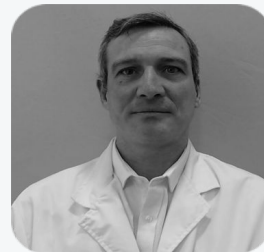
Juan Bizzotto,
PhD
CTO
Biotecnología &
Bioinformática



Geraldine Gueron,
MSc, PhD
CSO
Oncóloga
molecular



Daniel Alonso, MD,
PhD
Oncología
traslacional.
Validación clínica.



Federico Cayol,
MD
Oncólogo
molecular.
Validación clínica.

Advisors



Daniel Nofal
Enzyme VC



Marcelo Martí
Bitgenia

Muchas gracias



PreGen
PREDICTIVE GENOMIC INTELLIGENCE