

# Aplicación de técnicas de aprendizaje supervisado a la selección de biomarcadores para la enfermedad de Parkinson

Jaime Níguez Baeza

Juan Antonio Botía Blaya  
Universidad de Murcia



Laura Ibáñez Lladó  
Washington University in St. Louis



UNIVERSIDAD  
DE MURCIA

Máster Universitario en Bioinformática

# Contenido

- |  |   |                     |
|--|---|---------------------|
| 1. Enfermedad de Parkinson                               | } | Introducción        |
| 2. Biomarcadores y RNA circulares                        |   |                     |
| 3. Selección de circRNA mediante aprendizaje supervisado | } | Métodos, Resultados |
| 4. Análisis de enriquecimiento                           |   |                     |
| 5. Caracterización de biomarcadores candidatos           |   | Discusión           |
| 6. Conclusiones  |   | Conclusiones        |



# 1. Parkinson



## Definición y características

Conjunto de trastornos de origen variado caracterizados por afecciones motoras y no motoras.

Enfermedad **neurodegenerativa** con pérdida de neuronas dopaminérgicas en la *substantia nigra* del mesencéfalo.

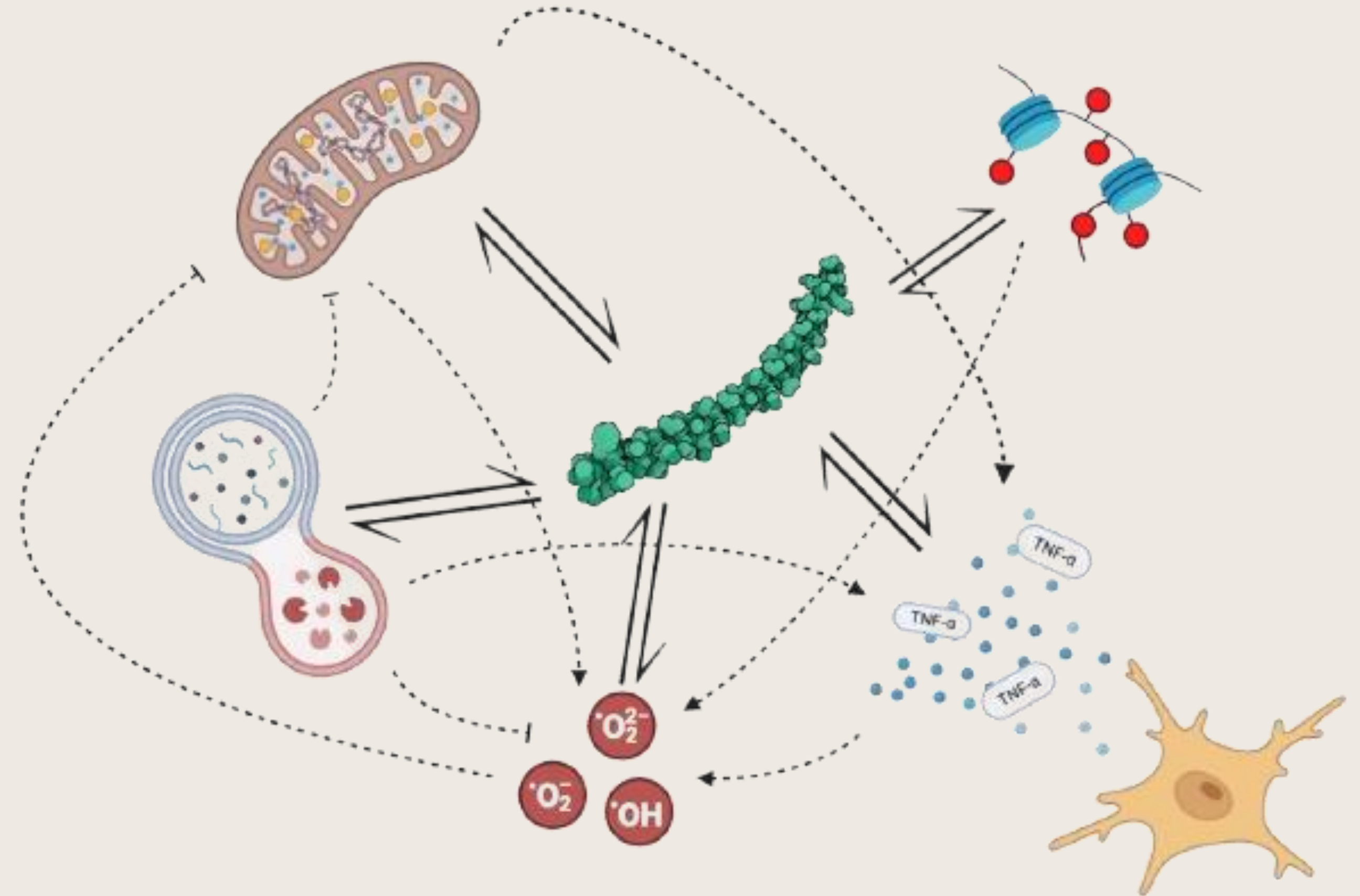


Parkinsonismo

- **Bradicinesia progresiva.**
- **Rigidez.**
- **Temblor en reposo.**
- **Pérdida del control de la postura.**
- **Afecciones del estado de ánimo y el comportamiento.**
- **Depresión.**
- **Trastornos cognitivos.**
- **Pérdida de autonomía.**

## Mecanismos biológicos

- $\alpha$ -Sinucleína
- Estrés oxidativo
- Mitocondria y Lisosoma
- Neuroinflamación
- Epigenética





# 2. RNA circulares como Biomarcadores



## RNA circulares

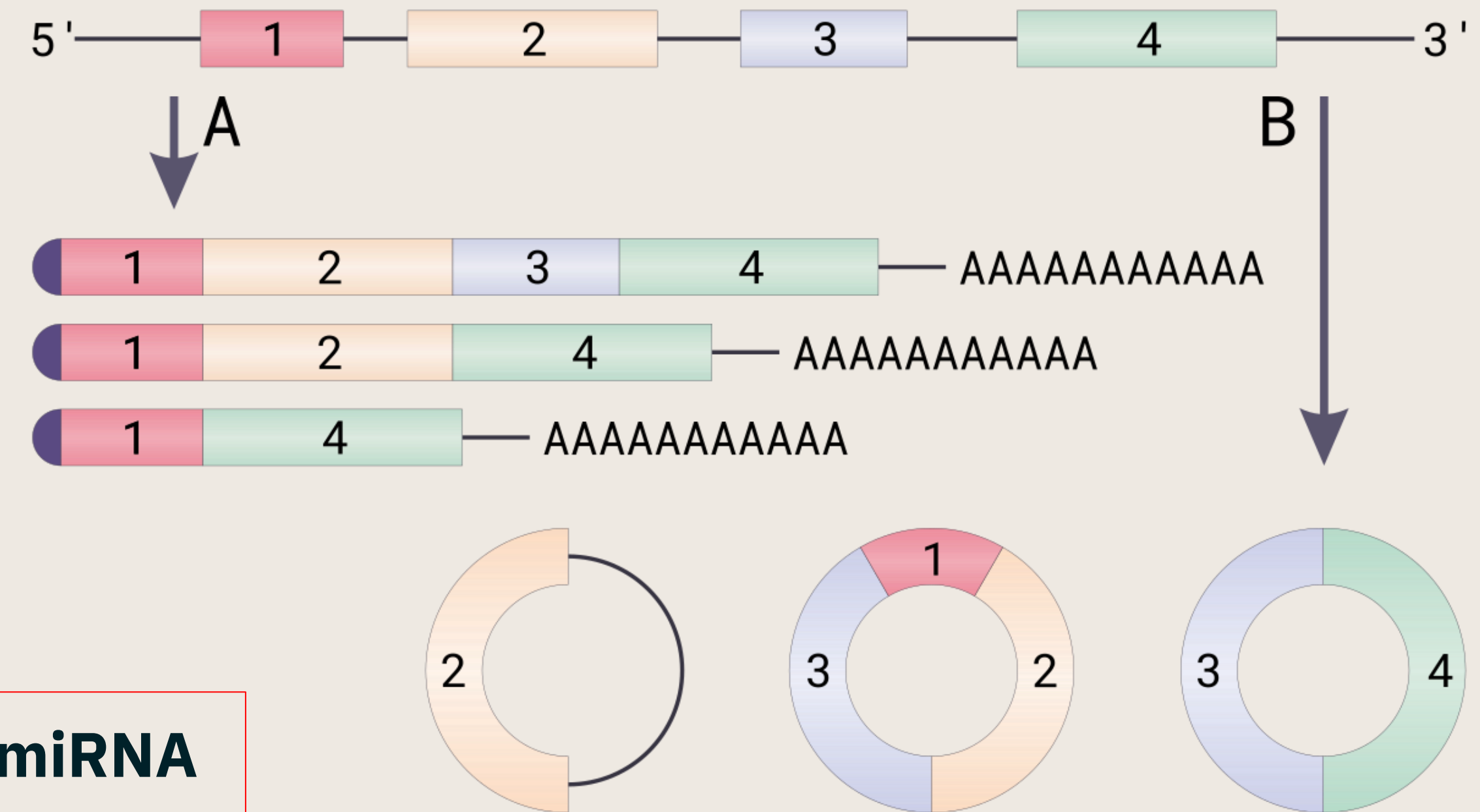
Estructura en forma de anillo completamente cerrada sin extremos 3' o 5' libres.

Rol en procesos patológicos como el cáncer, enfermedades cardiovasculares y neurodegenerativas.

Enriquecidos en el cerebro.

Detectables en la sangre.

- **Esponjas miRNA**
- **Unión a proteínas**
- **Unión a RNA**
- **Codificantes**



## Características biomarcadores

*Un indicador mensurable de algún estado o condición biológica que se mide y evalúa objetivamente para examinar procesos biológicos normales, procesos patógenos o respuestas farmacológicas a una intervención terapéutica*

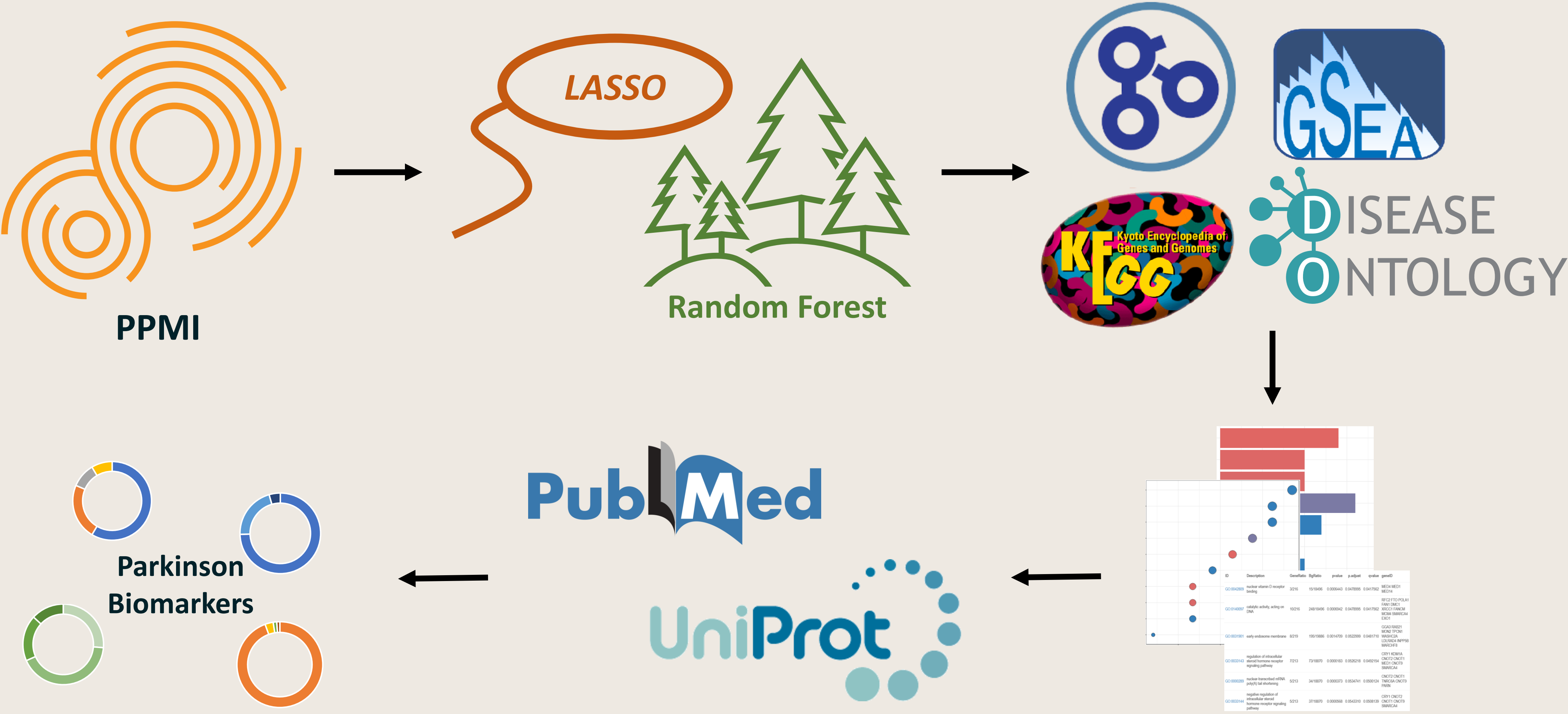
**National Institutes of Health (1998)**

- Diagnóstico temprano
- Selección de terapias
- Seguimiento de la enfermedad

- Estabilidad
- Especificidad
- Precisión
- Reproducibilidad
- Sensibilidad



Workflow



# 3. Selección de circRNA mediante técnicas de Machine Learning





Parkinson's  
Progression  
Markers  
Initiative



Washington  
University in St. Louis



THE MICHAEL J. FOX FOUNDATION  
FOR PARKINSON'S RESEARCH



UNIVERSIDAD  
DE MURCIA

## Contexto clínico

Expresión 2176 circRNA en sangre

*RNA-Seq*

858 pacientes Totales

*PD* & *Healthy Control*  
(538) (320)

Datos preprocesados  
(Normalizados, NA's, NearZeroVar)



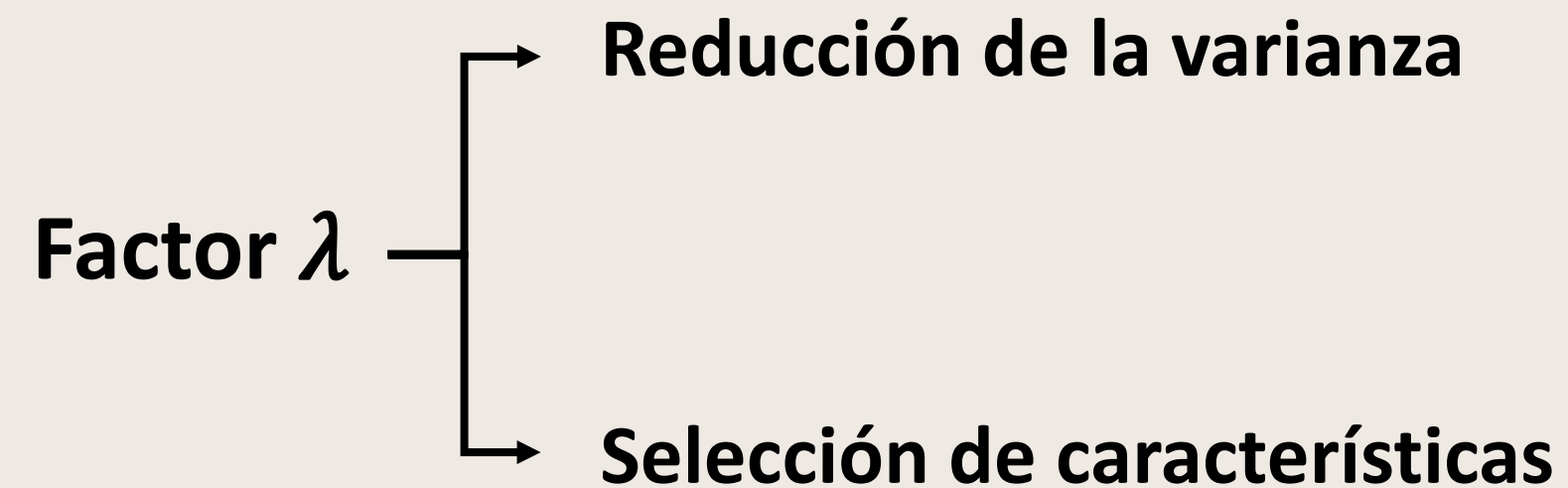
### Retos

- Alta dimensionalidad
- Datos ruidosos
- Clases desbalanceadas

# Machine Learning

## Least Absolute Shrinkage and Selection Operator

**LASSO**



10-fold CV → ↓ MAE / Deviance

Train 80: Test 20

- Clasificación binaria
- Alta interpretabilidad
- Reducir dimensionalidad
- Robustez frente a ruido

Rendimiento  
(Acc > NIR)



Árboles de decisión + *Bootstrap*

Hiperparámetros:

- *Mtry*
- *Splitrule*
- *Min.Node.Size*

10-fold CV → ↑ F- Score

Train 90: Test 10

No selección de características



Rendimiento de los modelos



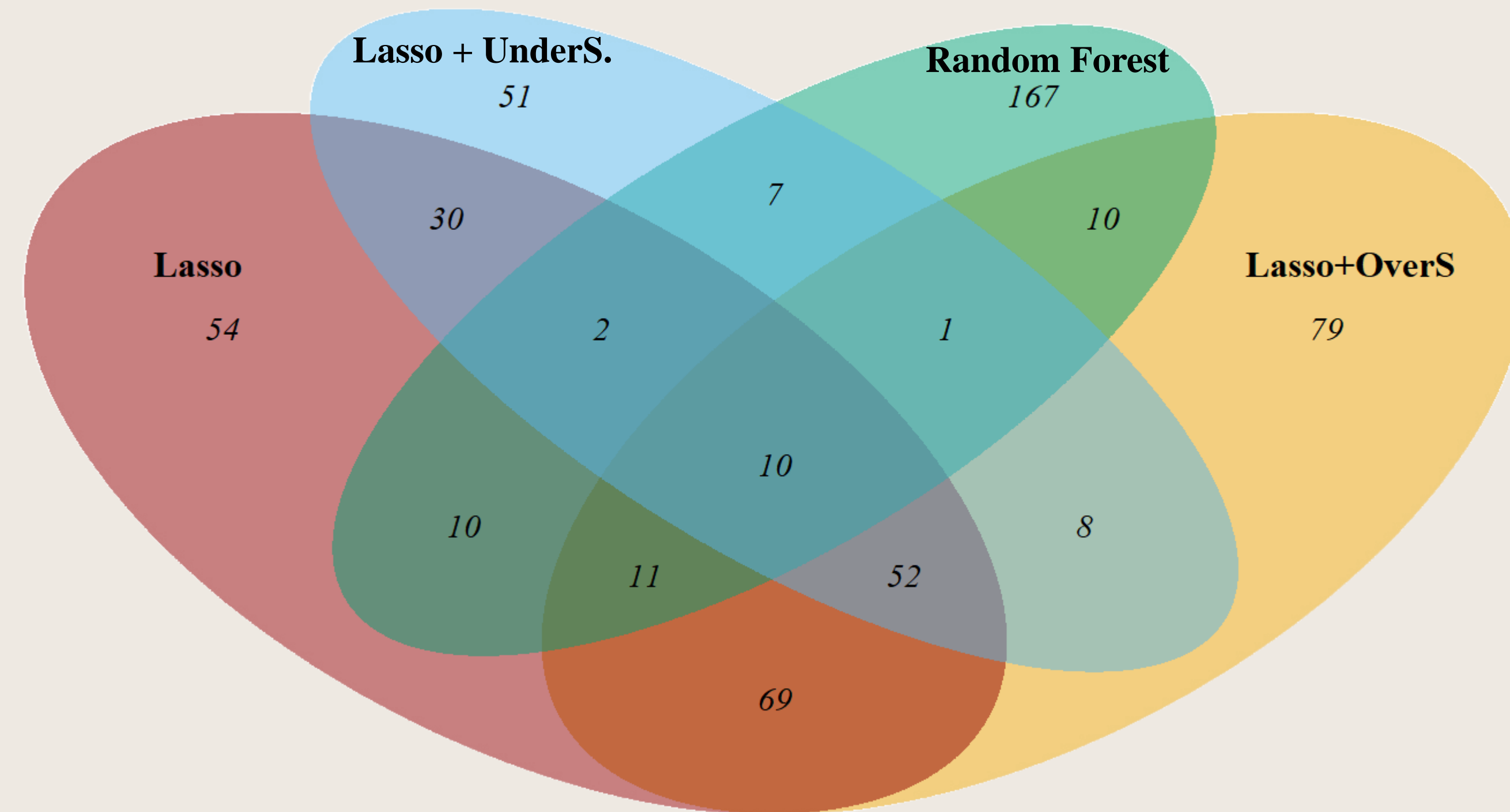
Predictores	238	240	160	2173 (218)
$\lambda$	0.014	0.013	0.019	—
Accuracy	0.672	0.667	0.579	0.729
NIR	0.626	0.626	0.626	0.623
p-value (Acc>NIR)	0.117	0.152	0.909	0.026

Rendimiento de los modelos

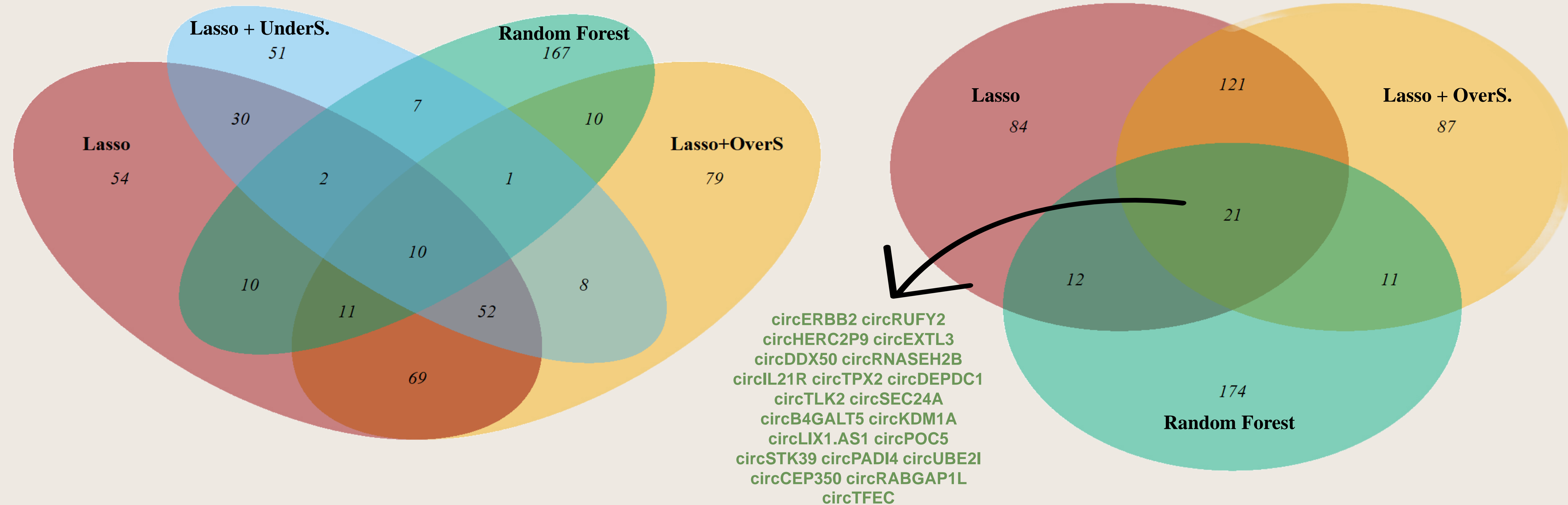
	<div>LASSO</div>	<div>LASSO OverS.</div>	<div>LASSO UnderS.</div>	<div> Random Forest</div>
Predictores	★★★ 238	★★★ 240	☆☆☆ 160	★★★ 2173 (218)
$\lambda$	0.014	0.013	0.019	—
Accuracy	0.672	0.667	0.579	0.729
NIR	0.626	0.626	0.626	0.623
p-value (Acc>NIR)	0.117	0.152	0.909	0.026



## Distribución de los predictores en los distintos modelos



## Distribución de los predictores en los distintos modelos



# 4. Análisis de enriquecimiento y selección de términos relevantes





# Análisis de enriquecimiento



Lista de genes:

- Predictores *Lasso*
- Predictores *Random Forest*  
(218: Importancia:  $q\ 0.9$ )



Ranking de genes:

- Coeficientes *Lasso*
- Importancia *Random Forest*

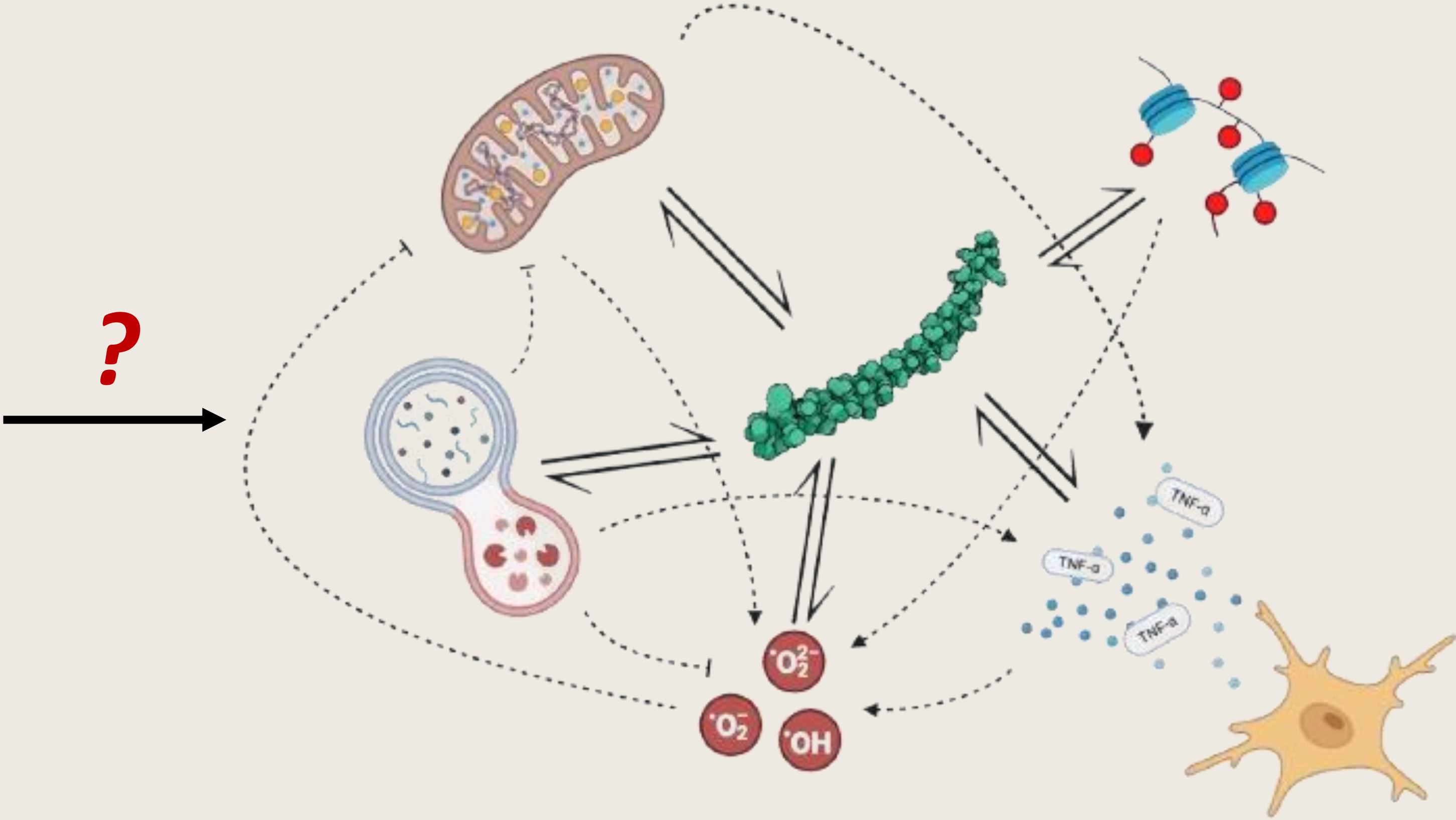
$qvalue \leq 0.05$

635  $\longrightarrow$  36

Términos	Genes enriquecidos
<i>RASopathy</i>	BRAF PTPN11 SHOC2 MAPK1 NF1
<i>negative regulation of cell cycle</i>	WAPL ATRX CHEK2 RBL1 LATS1 UIMC1 BRCA1 BARD1 PTPN11 RNASEH2B CDK5RAP2 DTL WAC TERF2 MDM2 BUB1
<i>transcription corepressor activity</i>	ATF7IP HDAC9 PHF12 MECP2 MIER2 MED1 NSD1 CASP8AP2 SMARCA4
<i>Ubiquitin mediated proteolysis</i>	UBE2K FANCL BRCA1 HERC1 UBE3C UBE2I MDM2

# Identificación de términos relevantes

Términos	Genes enriquecidos
<i>RASopathy</i>	BRAF PTPN11 SHOC2 MAPK1 NF1
<i>negative regulation of cell cycle</i>	WAPL ATRX CHEK2 RBL1 LATS1 UIMC1 BRCA1 BARD1 PTPN11 RNASEH2B CDK5RAP2 DTL WAC TERF2 MDM2 BUB1
<i>transcription corepressor activity</i>	ATF7IP HDAC9 PHF12 MECP2 MIER2 MED1 NSD1 CASP8AP2 SMARCA4
<i>Ubiquitin mediated proteolysis</i>	UBE2K FANCL BRCA1 HERC1 UBE3C UBE2I MDM2



## Identificación de términos relevantes



KEGG, Disease Ontology, GSEA

- Estudio bibliográfico

### *Ubiquitin mediated proteolysis*

The two major protein degradation systems are the ubiquitin-proteasome pathway (UPP) and ALP. Both are affected in synucleinopathies (...) *Reduced levels of proteasome subunits have been observed in PD patients. Several genes that code for proteasome subunits were downregulated in the SN of PD patients (...)*

**Lehtonen et al., 2019**



Gene Ontology

### **PRKN**

- Lewy body [[GO:0097413](#)]
- Transcription corepressor activity [[GO:0003714](#)]
- Regulation of protein targeting to mitochondrion [[GO:1903214](#)]

### **HTRA2**

- Negative regulation of cell cycle [[GO:0045786](#)]
- Negative regulation of neuron apoptotic process [[GO:0043524](#)]
- Unfolded protein binding [[GO:0051082](#)]



# Resultados del enriquecimiento

<i>early endosome</i>	ERBB2 MGRN1 GGA3 RAB21 MON2 TPCN1 WASHC2A STX6 TBC1D2B ALS2 GPR107 LDLRAD4 INPP5B MARCHF8 RABGAP1L ZFYVE26 AP1G1 AP3D1	<i>transcription coactivator activity</i>	JADE1 KAT6B KDM1A KDM4C MED14 MED27 MED4 MMS19 MTDH NCOA6 SMARCA4
<i>negative regulation of cell cycle</i>	WAPL ATRX CHEK2 RBL1 LATS1 UIMC1 BRCA1 BARD1 PTPN11 RNASEH2B CDK5RAP2 DTL WAC TERF2 MDM2 BUB1	<i>N-glycan biosynthesis, complex type</i>	FUT8
<i>pilocytic astrocytoma</i>	BRAF MBP MAPK1 NF1	<i>Glycosaminoglycan biosynthesis, heparan sulfate backbone</i>	PFKP
<i>early endosome</i>	ALS2 ECPAS EEA1 ERBB2 GGA3 INPP5B LDLRAD4 MAPK1 MGRN1 MON2 RAB21 RABGAP1L TBC1D2B TPCN1 WASHC2A WASHC2C	<i>transcription corepressor activity</i>	ATF7IP HDAC9 PHF12 MECP2 MIER2 MED1 NSD1 CASP8AP2 SMARCA4
<i>regulation of chromosome organization</i>	WAPL ATRX XRN1 USP7 CDK5RAP2 MAPK1 TERF2 SETDB2 SLF2 BUB1	<i>EGFR tyrosine kinase inhibitor resistance</i>	PRKCB BRAF FOXO3 ERBB2 MAPK1 NF1
<i>nuclear periphery</i>	POLA1 NUP205 SMARCC1 AKAP8L SMARCD1 ATXN1 HNRNPM SMARCA4	<i>Thyroid hormone signaling pathway</i>	PRKCB MED13 NCOA1 SLC16A10 RHEB MAPK1 MDM2
<i>forebrain development</i>	KIF14 ATRX ASPM WDR37 ZEB1 PCM1 KDM1A CRKL SCYL2 NCOA1 HERC1 NF1 STIL	<i>Ubiquitin mediated proteolysis</i>	UBE2K FANCL BRCA1 HERC1 UBE3C UBE2I MDM2
<i>nuclear androgen receptor binding</i>	KDM1A KDM4C NSD1 SMARCA4	<i>early endosome membrane</i>	GGA3 RAB21 MON2 TPCN1 WASHC2A LDLRAD4 INPP5B MARCHF8



Resultados del  
enriquecimiento

$\alpha$ -Syn  
Mitocondria  
Degradación

Transcripción  
Neuroinflamación  
Neurodegeneración

early endosome	ERBB2 MGRN1 GGA3 RAB21 MON2 TPCN1 WASHC2A STX6 TBC1D2B ALS2 GPR107 LDLRAD4 INPP5B MARCHF8 RABGAP1L ZFYVE26 AP1G1 AP3D1	transcription coactivator activity	JADE1 KAT6B KDM1A KDM4C MED14 MED27 MED4 MMS19 MTDH NCOA6 SMARCA4
negative regulation of cell cycle	WAPL ATRX CHEK2 RBL1 LATS1 UIMC1 BRCA1 BARD1 PTPN11 RNASEH2B CDK5RAP2 DTL WAC TERF2 MDM2 BUB1	N-glycan biosynthesis, complex type	FUT8
pilocytic astrocytoma	BRAF MBP MAPK1 NF1	Glycosaminoglycan biosynthesis, heparan sulfate backbone	PFKP
early endosome	ALS2 ECPAS EEA1 ERBB2 GGA3 INPP5B LDLRAD4 MAPK1 MGRN1 MON2 RAB21 RABGAP1L TBC1D2B TPCN1 WASHC2A WASHC2C	transcription corepressor activity	ATF7IP HDAC9 PHF12 MECP2 MIER2 MED1 NSD1 CASP8AP2 SMARCA4
regulation of chromosome organization	WAPL ATRX XRN1 USP7 CDK5RAP2 MAPK1 TERF2 SETDB2 SLF2 BUB1	EGFR tyrosine kinase inhibitor resistance	PRKCB BRAF FOXO3 ERBB2 MAPK1 NF1
nuclear periphery	POLA1 NUP205 SMARCC1 AKAP8L SMARCD1 ATXN1 HNRNPM SMARCA4	Thyroid hormone signaling pathway	PRKCB MED13 NCOA1 SLC16A10 RHEB MAPK1 MDM2
forebrain development	KIF14 ATRX ASPM WDR37 ZEB1 PCM1 KDM1A CRKL SCYL2 NCOA1 HERC1 NF1 STIL	Ubiquitin mediated proteolysis	UBE2K FANCL BRCA1 HERC1 UBE3C UBE2I MDM2
nuclear androgen receptor binding	KDM1A KDM4C NSD1 SMARCA4	early endosome membrane	GGA3 RAB21 MON2 TPCN1 WASHC2A LDLRAD4 INPP5B MARCHF8





# 5. Caracterización de biomarcadores candidatos



## Genes enriquecidos



- En todos los modelos entrenados
- $\uparrow$  Rendimiento =  $\uparrow$  Nº genes únicos
- $\approx$  Nº Total genes entre modelos
- Diferencias según algoritmo

Biomarcadores favoritos

93 Predictores → 5 Seleccionados

BRAF	MBP	MAPK1	NF1	ST6GAL2	FUT8
PFKP	PRKCB	FOXO3	ERBB2	MED13	
NCOA1	SLC16A10	RHEB	MDM2	UBE2K	
FANCL	BRCA1	HERC1	UBE3C	UBE2I	
MGRN1	GGA3	RAB21	MON2	TPCN1	
WASHC2A	STX6	TBC1D2B	ALS2	GPR107	
LDLRAD4	INPP5B	MARCHF8	RABGAP1L		
ZFYVE26	AP1G1	AP3D1	WAPL	ATRX	CHEK2
RBL1	LATS1	UIMC1	BARD1	PTPN11	
RNASEH2B	CDK5RAP2	DTL	WAC	TERF2	
BUB1	ECPAS	EEA1	WASHC2C	XRN1	USP7
SETDB2	SLF2	POLA1	NUP205	SMARCC1	
AKAP8L	SMARCD1	ATXN1	HNRNPM	SMARCA4	
KIF14	ASPM	WDR37	ZEB1	PCM1	KDM1A
CRKL	SCYL2	STIL	KDM4C	NSD1	JADE1
KAT6B	MED14	MED27	MED4	MMS19	MTDH
NCOA6	ATF7IP	HDAC9	PHF12	MECP2	
MIER2	MED1	CASP8AP2			

- circERBB2:**
- Control de la transcripción en cáncer de vesícula (Huang et al. 2019).
  - *Early endosome* y *EGFR tyrosine kinase inhibitor resistance*.
- circUBE2I:**
- Inducción de ferroptosis en un contexto de estrés oxidativo (Yi et al. 2023).
  - *Ubiquitin mediated proteolysis*.
- circRABGAP1L:**
- RABGAP1L: Enriquecido en *Rab GTPase binding* (Yin et al. 2022), fusión de membranas/transp. Vesículas, relacionados con PD.
  - *Early endosome*.
- circRNASEH2B:**
- Asociado con Alzheimer en estudios de expresión diferencial (Dube et al. 2019).
  - RNASEH2B, relacionado con neuropatías, función anti neuroinflamatoria y en neurogénesis en *Aditi et al. 2020*.
  - *Negative regulation of cell cycle*.
- circKDM1a:**
- KDM1a: Desmetilación de H3K4me, sobrerrepresentado en pacientes de Parkinson (Song et al. 2023).
  - *forebrain development, nuclear androgen receptor binding, transcription coactivator activity*.

# 6. Conclusiones



### Uso de Machine Learning para seleccionar biomarcadores relacionados con el Parkinson

- Capacidad de seleccionar información relevante para la enfermedad.
- Uso de modelos lineales y no lineales para entender mejor las relaciones establecidas con la enfermedad.
- Alternativa y complemento a otros enfoques.

### Papel de los RNA circulares en procesos patológicos

- Función reguladora en procesos biológicos.
- Relación documentada con otras enfermedades.

### RNA circulares candidatos

- Selección mediante criterios predictivos y biológicos.
- Algunos de ellos ya han sido considerados en estudios previos.
- Posibilidad de afinar la selección en futuros estudios *in silico*, *in vitro*/*in vivo*.



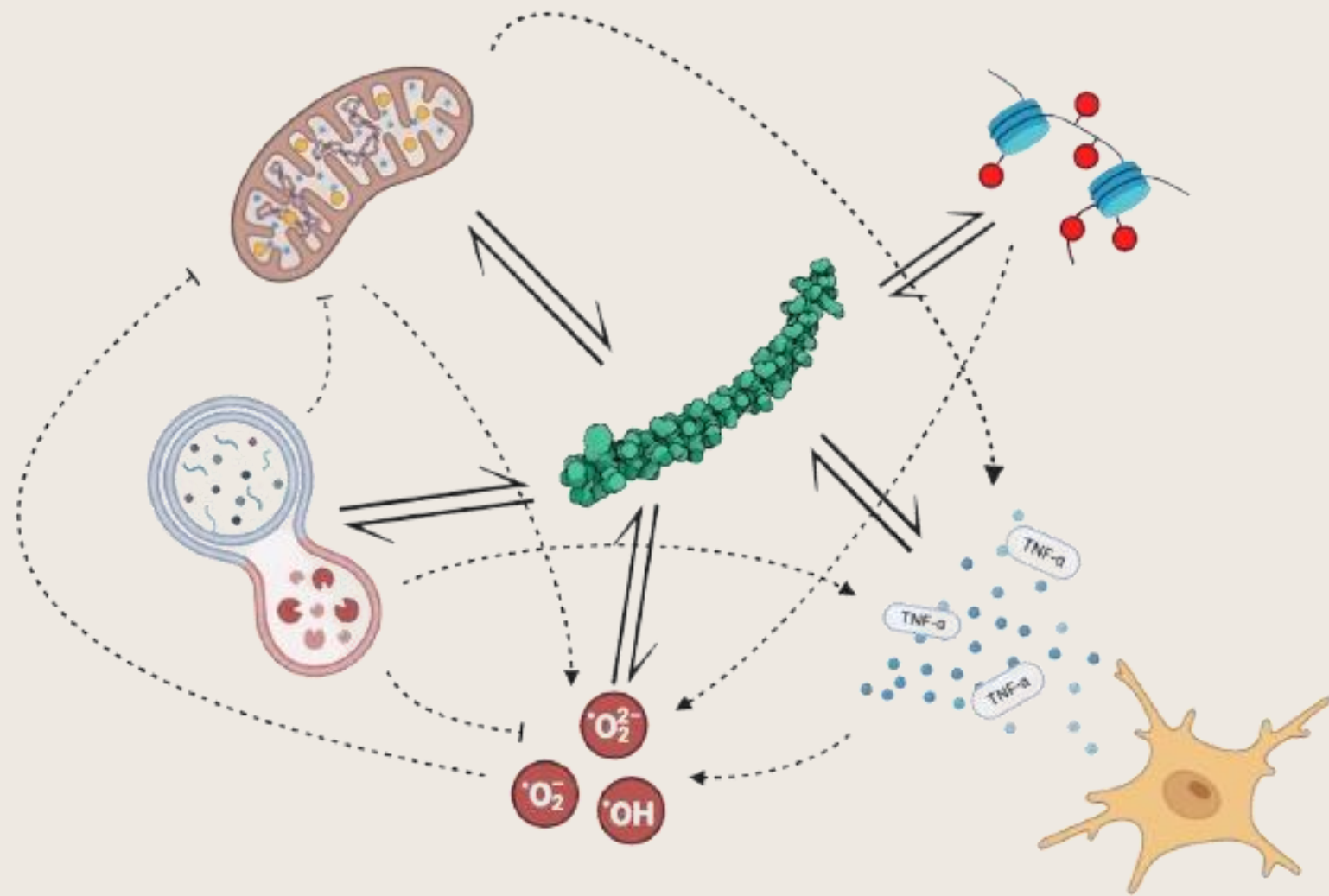
# ¡Muchas Gracias!



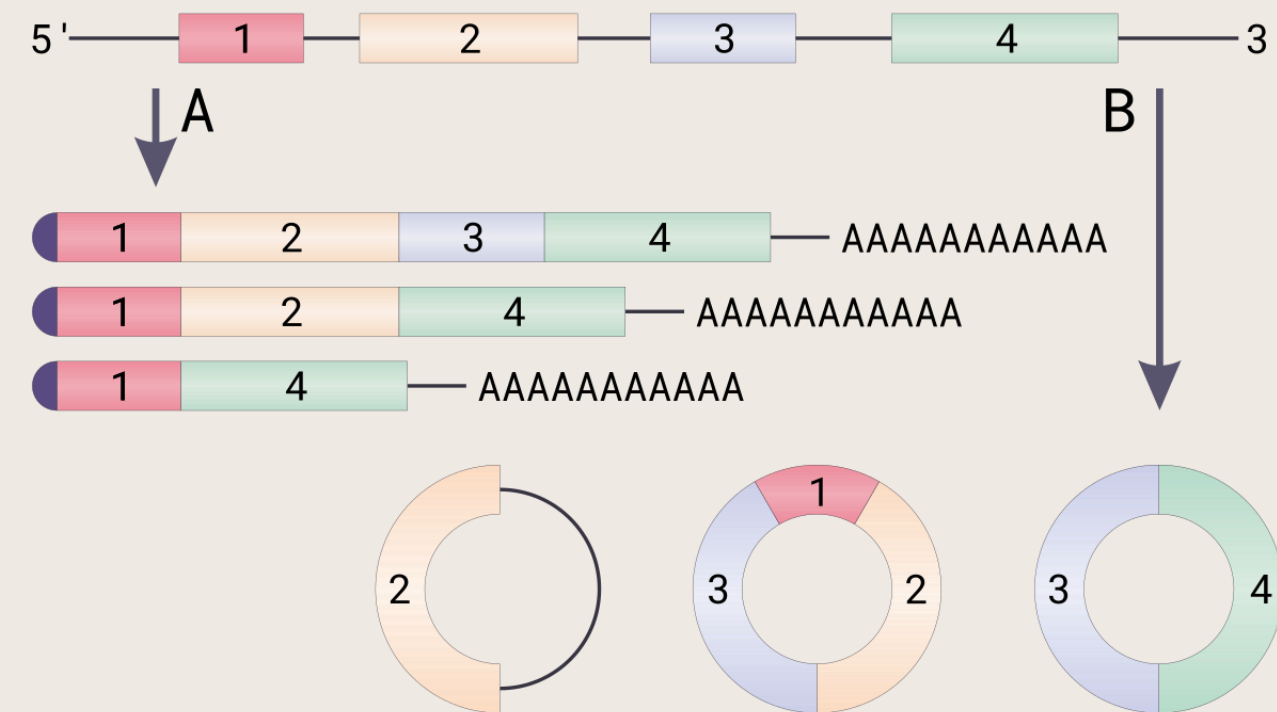
UNIVERSIDAD  
DE MURCIA

Máster Universitario en Bioinformática





Created with BioRender.com



<https://commons.wikimedia.org/wiki/File:CircRNA.svg>



[https://en.wikipedia.org/wiki/File:UniProt\\_\(logo\).png](https://en.wikipedia.org/wiki/File:UniProt_(logo).png)

Parkinson's  
Progression  
Markers  
Initiative

<https://www.ppmi-info.org/>

**LASSO**

<https://www.youtube.com/@statquest>



THE MICHAEL J. FOX FOUNDATION  
FOR PARKINSON'S RESEARCH

[https://en.wikipedia.org/wiki/The\\_Michael\\_J.\\_Fox\\_Foundation#/media/File:The\\_Michael\\_J.\\_Fox\\_Foundation\\_logo.svg](https://en.wikipedia.org/wiki/The_Michael_J._Fox_Foundation#/media/File:The_Michael_J._Fox_Foundation_logo.svg)



[https://en.wikipedia.org/wiki/KEGG#/media/File:KEGG\\_database\\_logo.gif](https://en.wikipedia.org/wiki/KEGG#/media/File:KEGG_database_logo.gif)



<https://github.com/GSEA-MSigDB>



<https://www.hetop.eu/hetop/rep/es/DO/>



<https://es.m.wikipedia.org/wiki/Archivo:US-NLM-PubMed-Logo.svg>



<http://berkeleybop.github.io/project/gene-ontology/>



[https://en.m.wikipedia.org/wiki/File:Olin\\_Business\\_School\\_at\\_Washington\\_University\\_in\\_St.\\_Louis\\_Logo\\_2023.png](https://en.m.wikipedia.org/wiki/File:Olin_Business_School_at_Washington_University_in_St._Louis_Logo_2023.png)



UNIVERSIDAD  
DE MURCIA

Máster Universitario en Bioinformática