

En busca de proteínas que definen la cilia primaria

Una estrategia basada en datos independientes de secuencia y aprendizaje automático

Autores: Emilia Torriglia¹, Florencia Irigoien^{2, 3}, Laura Romanelli-Cedrez⁴, Flavio Pazos Obregón¹

1. Unidad de Bioquímica Analítica y Proteómica, IIBCE - Institut Pasteur de Montevideo

2. Unidad Académica de Histología y Embriología - Facultad de Medicina, UDELAR

3. Laboratorio de Genética Molecular Humana - Institut Pasteur de Montevideo

4. Laboratorio de Biología de Gusanos - Institut Pasteur de Montevideo

XV Jornadas de la SBBM, 17 de septiembre de 2026

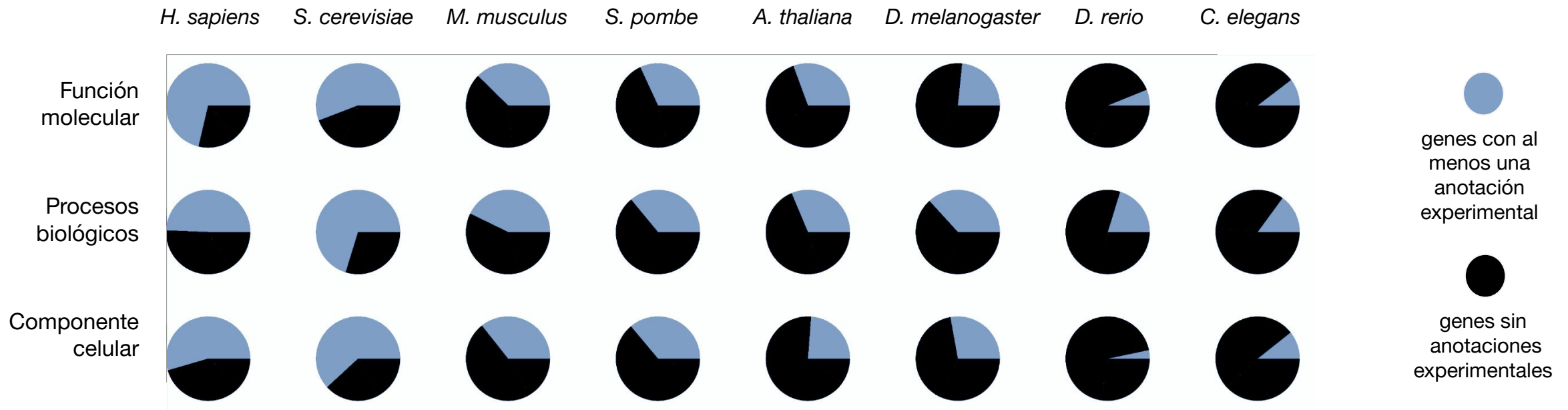


- Anotación funcional de genes
- La cilia primaria
- Dos enfoques para la predicción de función
- Conclusiones y perspectivas



Brecha en la anotación funcional

- La generación de datos genómicos supera la capacidad de caracterización funcional experimental. **La caracterización experimental de la función génica requiere de tiempo y recursos**
- Muchos genes **carecen de una anotación funcional completa.**
- **En organismos modelo:**

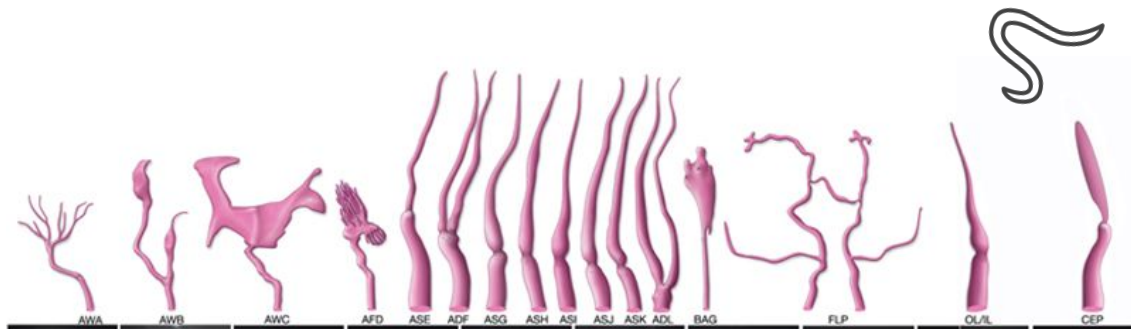


adaptado de Wue y Rhee, 2023

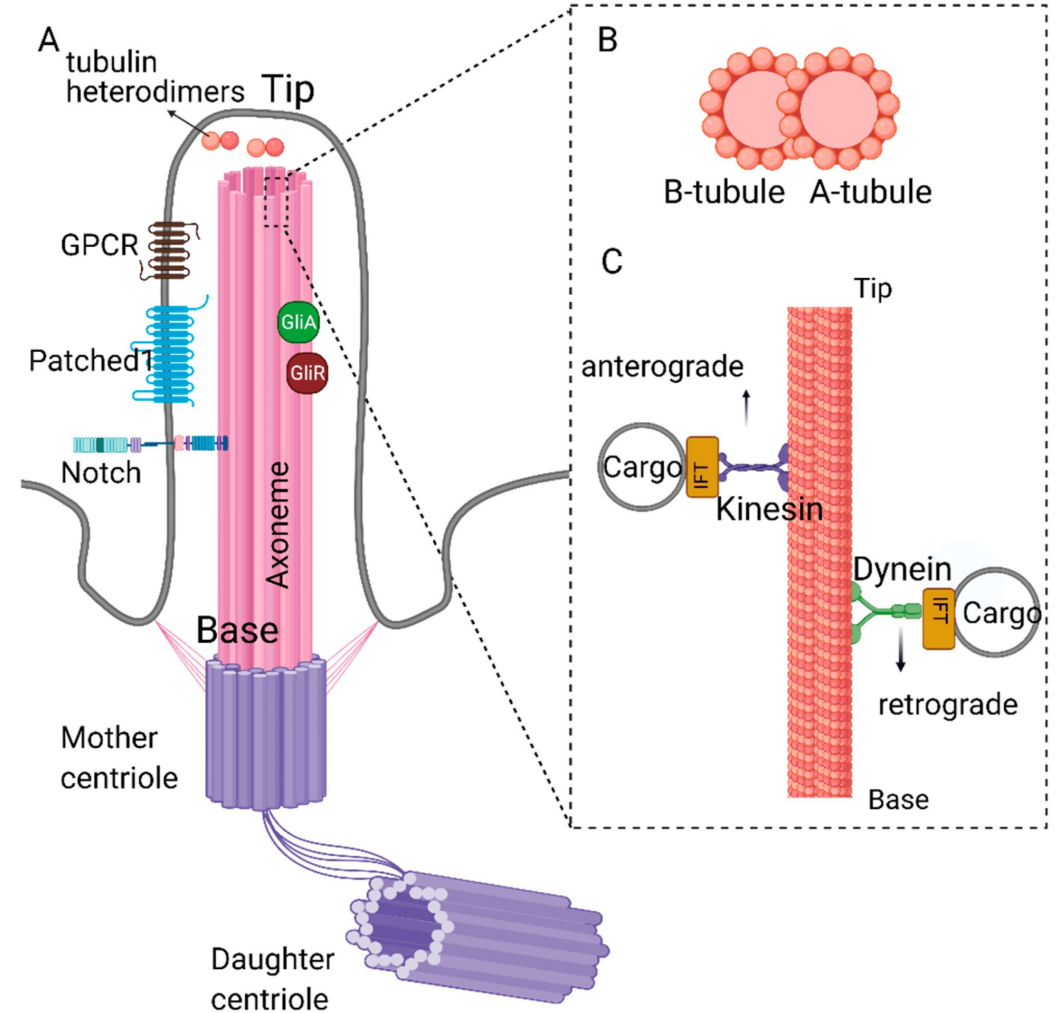
- La homología **no siempre es suficiente para inferir función.** Por ese motivo, se busca incorporar datos biológicos **independientes de la homología** para la predicción funcional

Caso de estudio: la cilia primaria

- Organelo **sensorial** conservado evolutivamente que actúa como antena celular para vías de señalización
- Su disfunción genera patologías sistémicas graves denominadas **ciliopatías**
- Composición proteica característica – pero no descrita exhaustivamente
- **Ausencia de secuencias consenso o dominios estructurales universales** para identificar genes ciliares.



WormAtlas

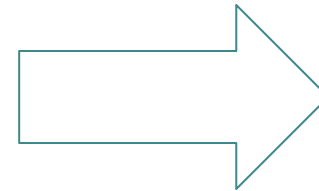


Yanardag S, Pugacheva EN (2021)

Dos enfoques para la predicción de función

Primer paso: elección de “genes ciliares conocidos” en *Caenorhabditis elegans*

Genes anotados con términos GO
relacionados a cilia, con evidencia
experimental



88 genes

Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Perfiles de expresión similares pueden indicar funciones biológicas compartidas

A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single cell resolution

~400-500 min

Jonathan S. Packer^{1,*}, Qin Zhu^{2,*}, Chau Huynh¹, Priya Sivaramakrishnan³, Elicia Preston³, Hannah Dueck^{3,8}, Derek Stefanik⁴, Kai Tan^{3,5,6,7}, Cole Trapnell¹, Junhyong Kim^{4,#}, Robert H. Waterston^{1,#}, John I. Murray^{3,#}

10^x GENOMICS

Cell

CellPress

Resource

Molecular topography of an entire nervous system

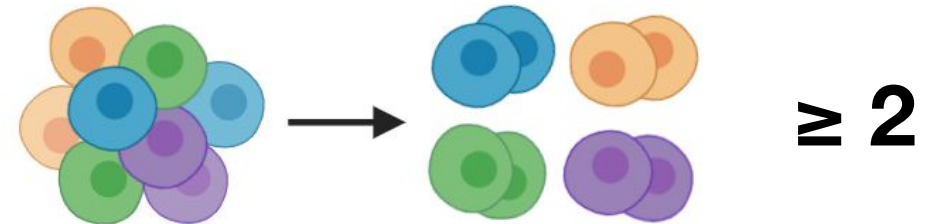
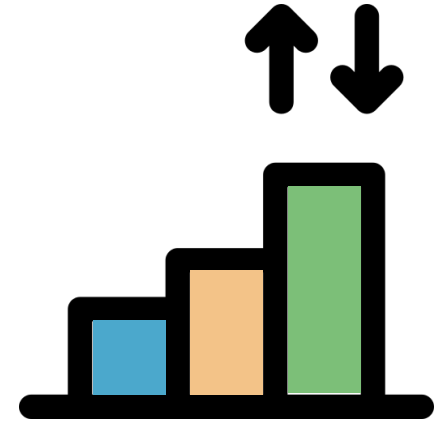
Seth R. Taylor,¹ Gabriel Santpere,^{2,3,10} Alexis Weinreb,^{2,4,10} Alec Barrett,^{2,4,10} Molly B. Reilly,^{5,6,10} Chuan Xu,^{2,10} Erdem Varol,^{7,10} Panos Oikonomou,^{5,8,10} Lori Glenwinkel,^{5,6} Rebecca McWhirter,¹ Abigail Poff,¹ Manasa Basavaraju,^{2,4} Ibnul Rafi,^{5,6} Eviatar Yemini,^{5,6} Steven J. Cook,^{5,6} Alexander Abrams,^{2,4} Berta Vidal,^{5,6} Cyril Cros,^{5,6} Saeed Tavazoie,^{5,8} Nenad Sestan,² Marc Hammarlund,^{2,4,*} Oliver Hobert,^{5,6,*} and David M. Miller III^{1,9,11,*}

Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión **y poca variabilidad** entre tipos celulares.



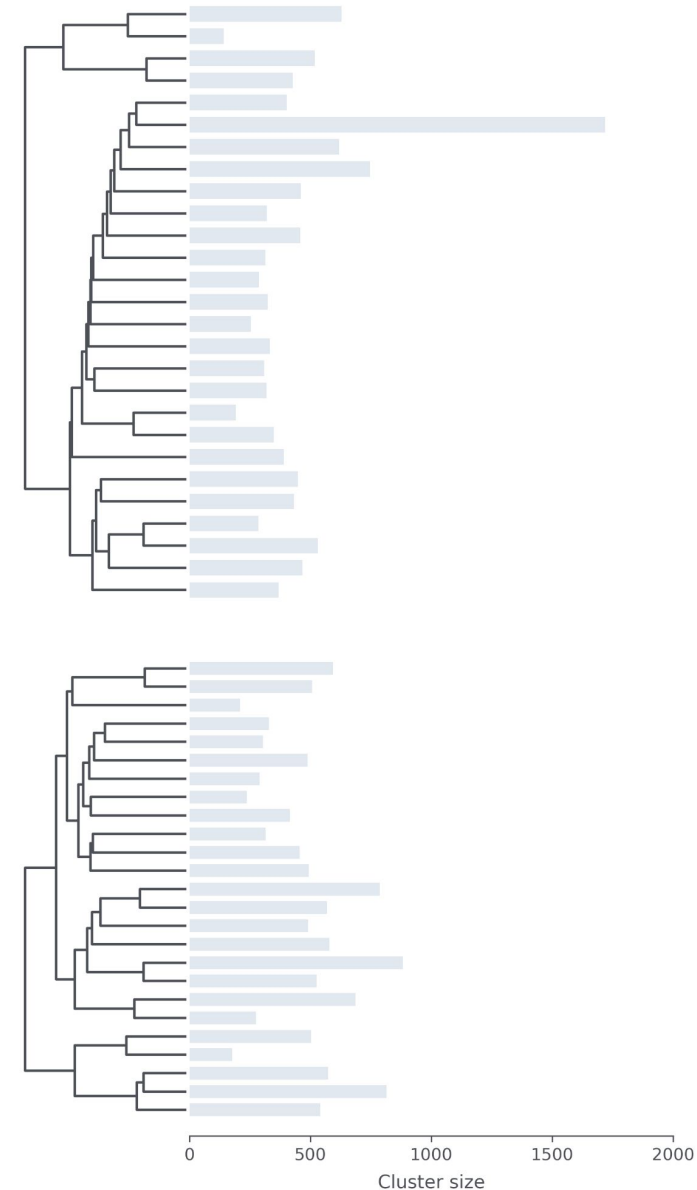
Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión **y poca variabilidad** entre tipos celulares.

Clustering jerárquico (12040 genes, neuronas ciliadas) y **análisis de enriquecimiento** en genes ciliares



Embrión

Adulto

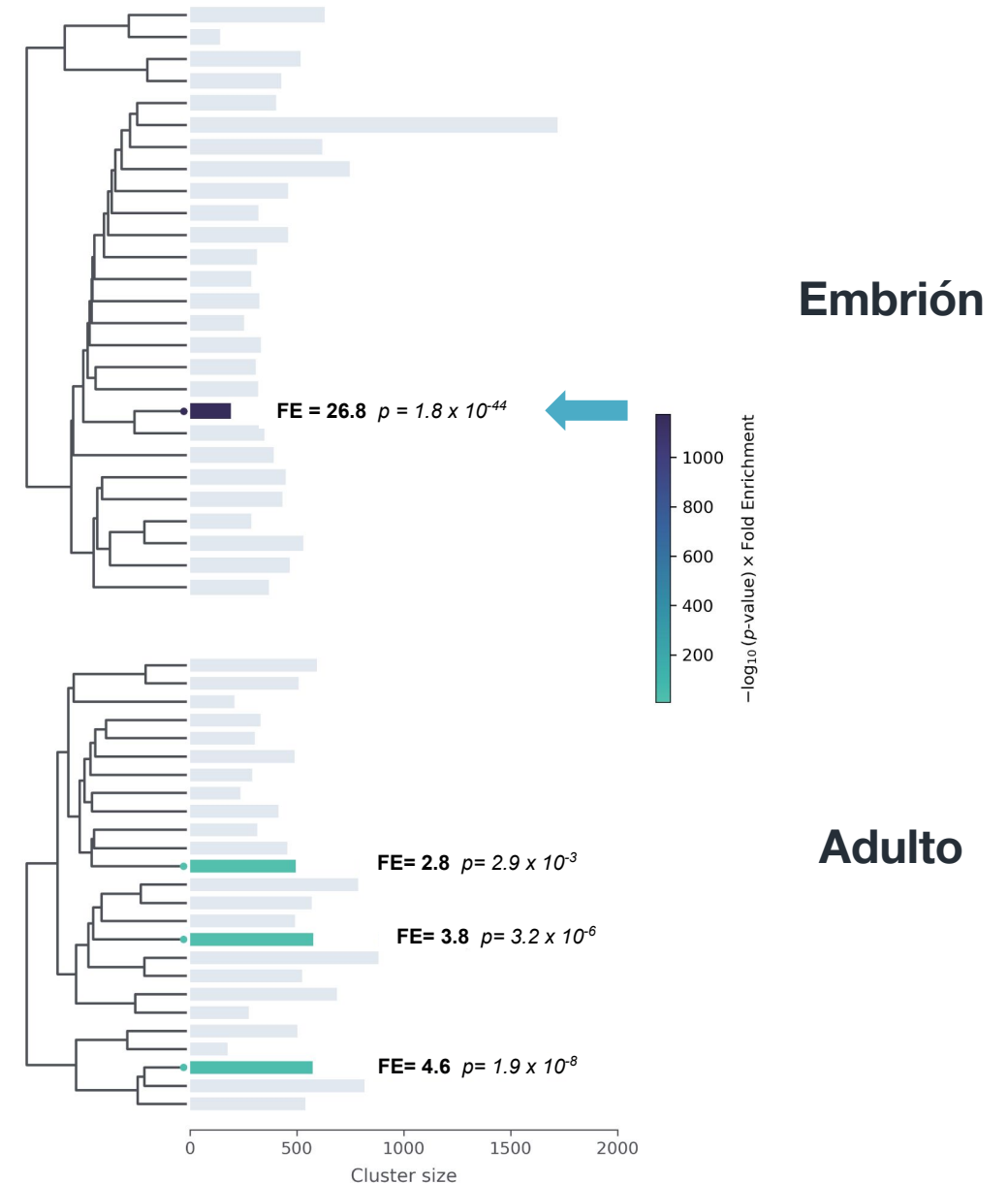
Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión **y poca variabilidad** entre tipos celulares.

Clustering jerárquico (12040 genes, neuronas ciliadas) y **análisis de enriquecimiento** en genes ciliares



Dos enfoques para la predicción de función

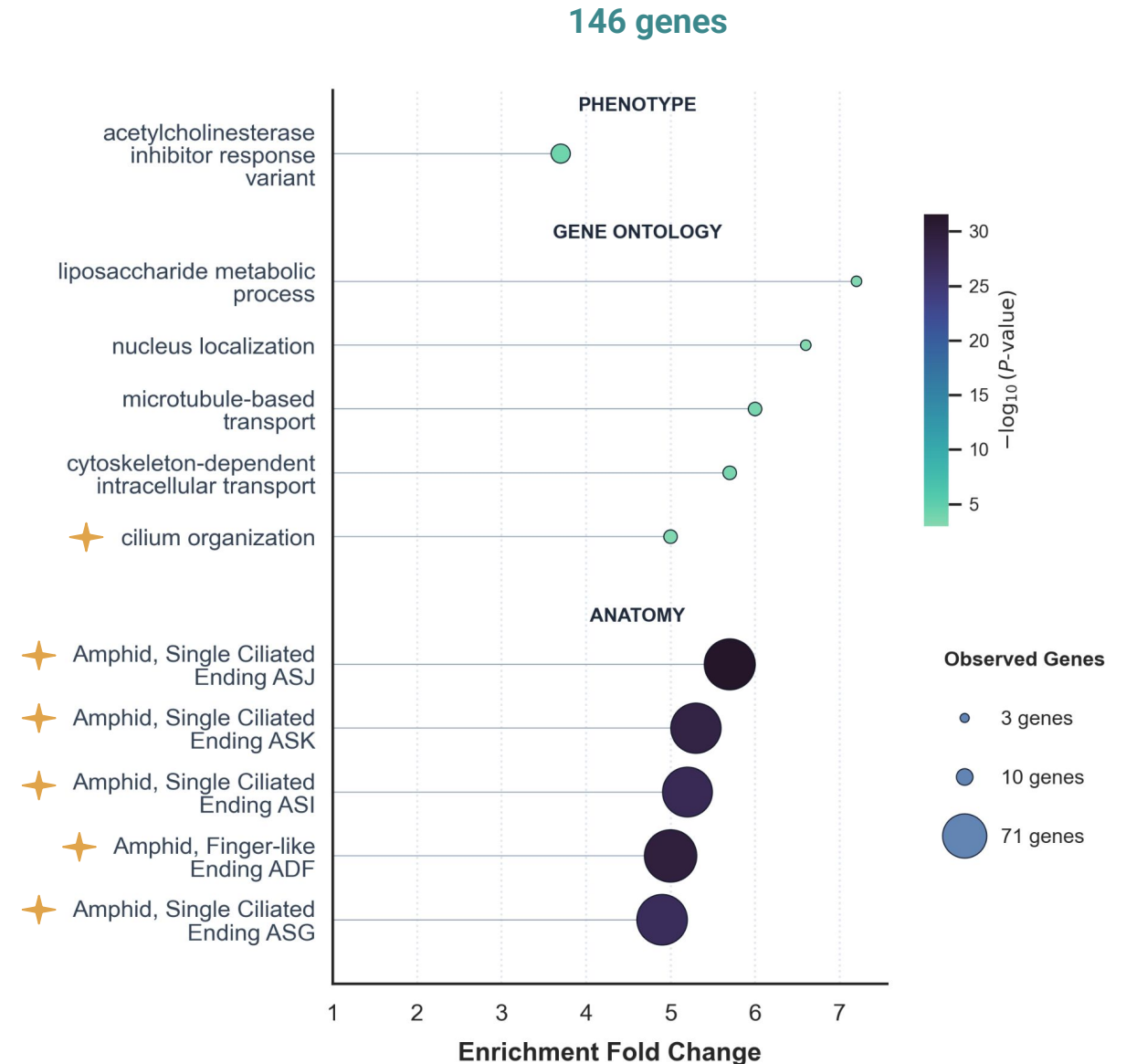
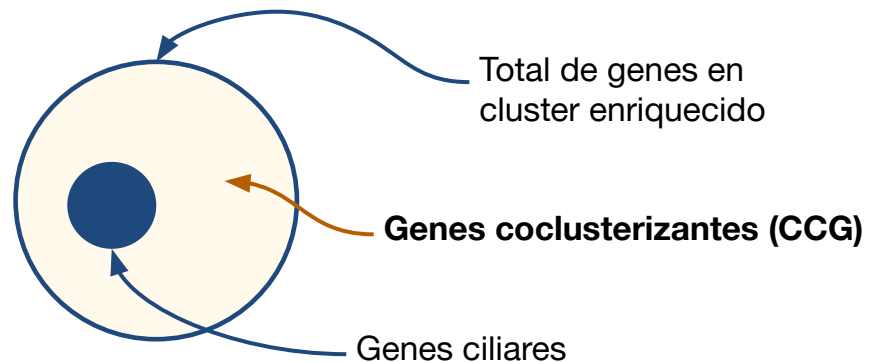
Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión y poca variabilidad entre tipos celulares.

Clustering jerárquico (12040 genes, neuronas ciliadas) y **análisis de enriquecimiento** en genes ciliares

Identificación de **genes coclusterizantes** y caracterización de **clusters enriquecidos** en genes ciliares



Dos enfoques para la predicción de función

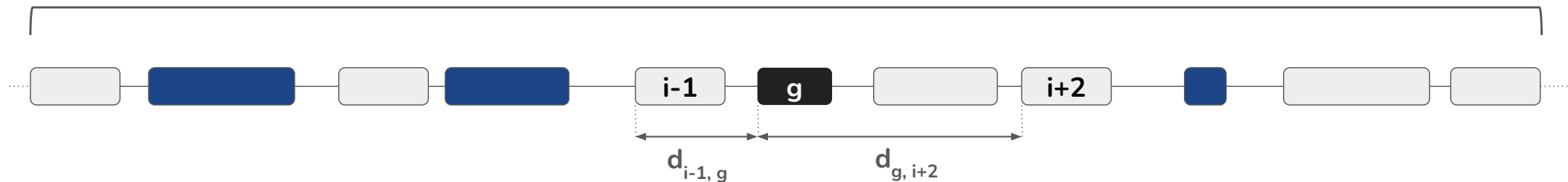
Paisaje funcional

Cálculo de Weighted Functional Landscape Arrays

Los genes **no están distribuidos al azar** en el genoma, existen patrones de ubicación relacionados a la función que se pueden representar de forma **matricial**

■ Genes anotados con término GO t

$w = 5$



GO terms

Windows

■			
■			
■			
■			

$$E(g, w, t) = \frac{f_{\text{obs}}(g, w, t)}{f_{\text{expected}}(t)}$$

$$f_{\text{expected}}(t) = \frac{\text{number of genes with GO term } t}{\text{number of genes on chromosome}}$$

$$f_{\text{obs}}(g, w, t) = \frac{\sum_{i \in W} p_i \text{GO}_{i,t}}{\sum_{i \in W} p_i}$$

$$p(d) = e^{-d/L}$$

Dos enfoques para la predicción de función

Paisaje funcional

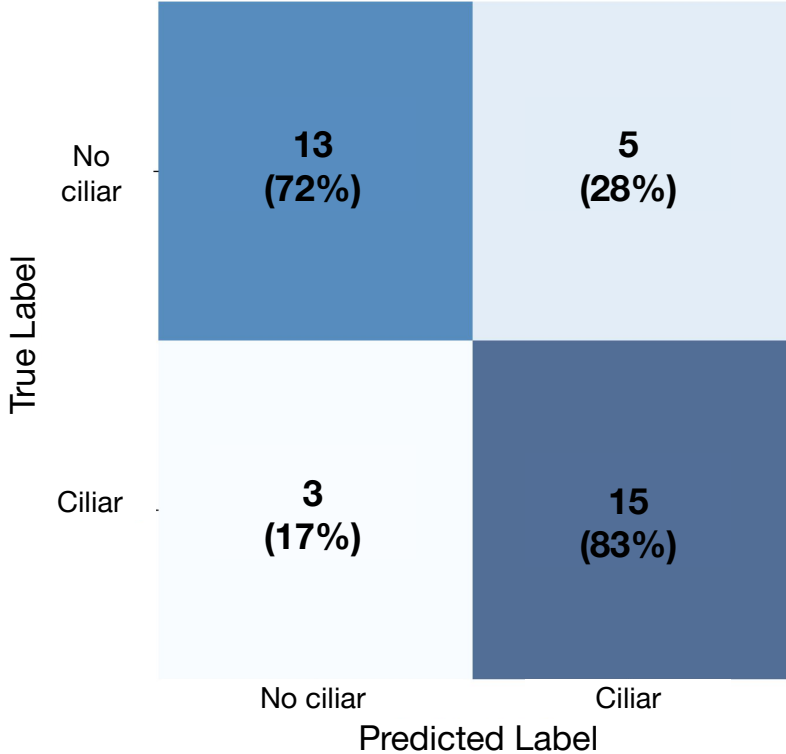
Cálculo de Weighted Functional Landscape Arrays

Selección de genes **positivos** y **negativos**, separados en **entrenamiento** y **evaluación** (80:20)

RandomForest: **entrenamiento**, optimización de parámetros y umbral de decisión

Evaluación del modelo

	Optimized threshold (0.57)
Precision	0.87
Recall	0.72
F1 score	0.79
ROC-AUC	0.88



Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión **y poca variabilidad entre tipos celulares.**

Clustering jerárquico (12040 genes, neuronas ciliadas) y **análisis de enriquecimiento** en genes ciliares

Identificación de **genes coclusterizantes** y caracterización de **clusters enriquecidos** en genes ciliares

Paisaje funcional

Cálculo de Weighted Functional Landscape Arrays

Selección de genes **positivos** y **negativos**, separados en **entrenamiento** y **evaluación** (80:20)

Random Forest: **entrenamiento**, optimización de parámetros y umbral de decisión

Evaluación del modelo

Dos enfoques para la predicción de función

Transcriptómica *single cell*

Recopilación de datos de scRNA-seq (Packer *et al.*, 2019, Taylor *et al.* 2021)

Filtrado de genes con bajos niveles de expresión **y poca variabilidad entre tipos celulares.**

Clustering jerárquico (12040 genes, neuronas ciliadas) y **análisis de enriquecimiento** en genes ciliares

Identificación de **genes coclusterizantes** y caracterización de **clusters enriquecidos** en genes ciliares

Paisaje funcional

Cálculo de Weighted Functional Landscape Arrays

Selección de genes **positivos** y **negativos**, separados en **entrenamiento** y **evaluación** (80:20)

Random Forest: **entrenamiento**, optimización de parámetros y umbral de decisión

Evaluación del modelo

Genes coclusterizantes



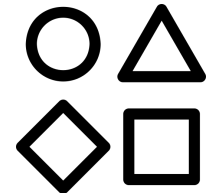
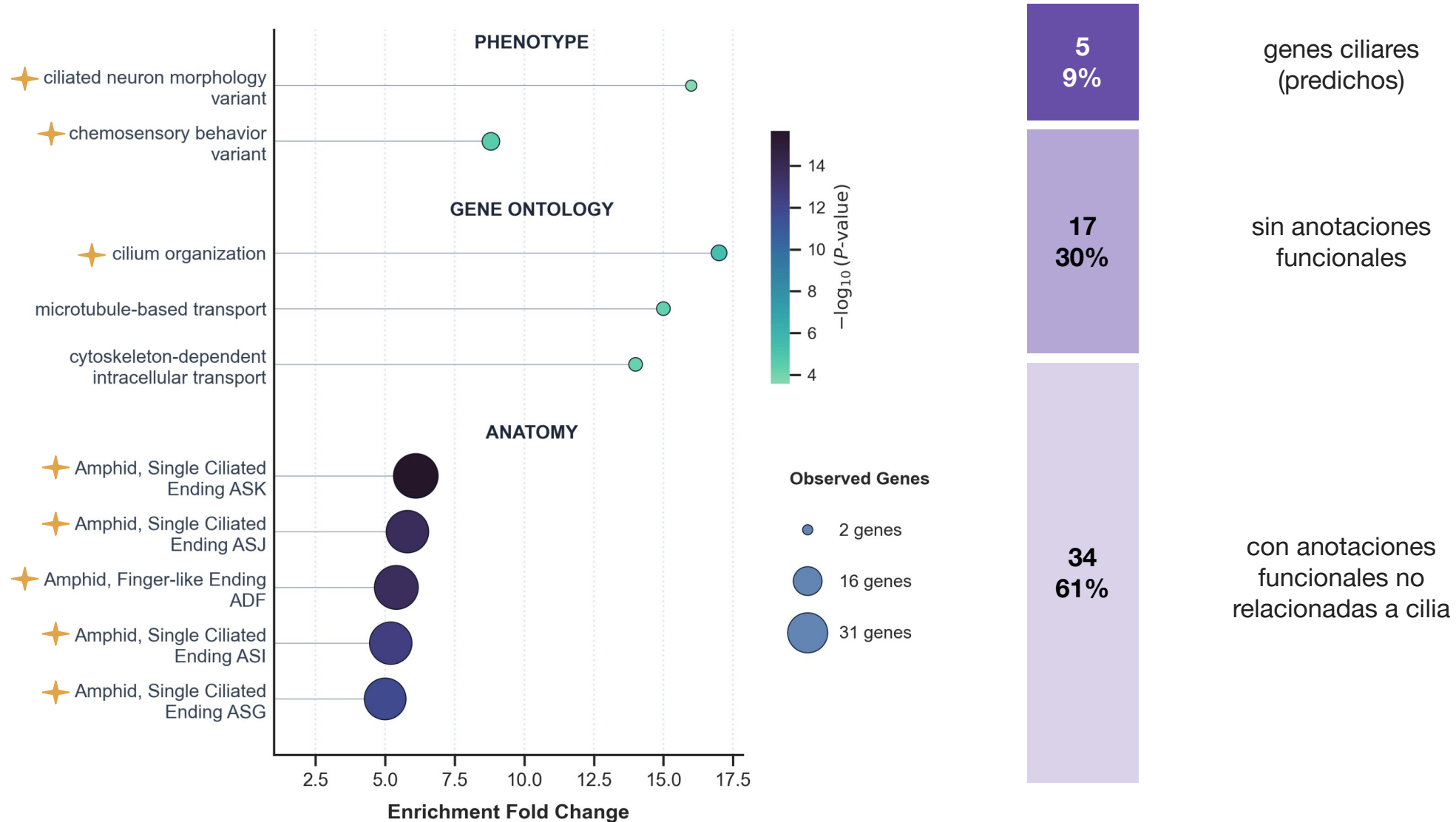
Clasificador entrenado con matrices de paisaje funcional



Genes candidatos seleccionados a partir de aproximaciones independientes

Dos enfoques para la predicción de función

Candidatos preliminares



Conclusiones y perspectivas

- Se identificaron genes con **perfiles de expresión** similares a los genes ciliares conocidos
- Se logró entrenar un clasificador con buen poder predictivo únicamente a partir del **paisaje funcional** de un gen
- Se combinaron ambos acercamientos para llegar a una lista preliminar de **56 genes candidatos** a ser genes ciliares
- A continuación, se plantea el **refinamiento** del modelo de paisaje funcional, así como el uso de **otros tipos de datos** para la priorización de candidatos

¡Muchas gracias!