Endeavour
Gene prioritization

Sandra Alandes Esteve
Unidad de Bioinformática y Bioestadística

17 junio 2019

Unidad de Bioinformática y Bioestadística

PRINCIPE FELIPE CENTRO DE INVESTIGACION

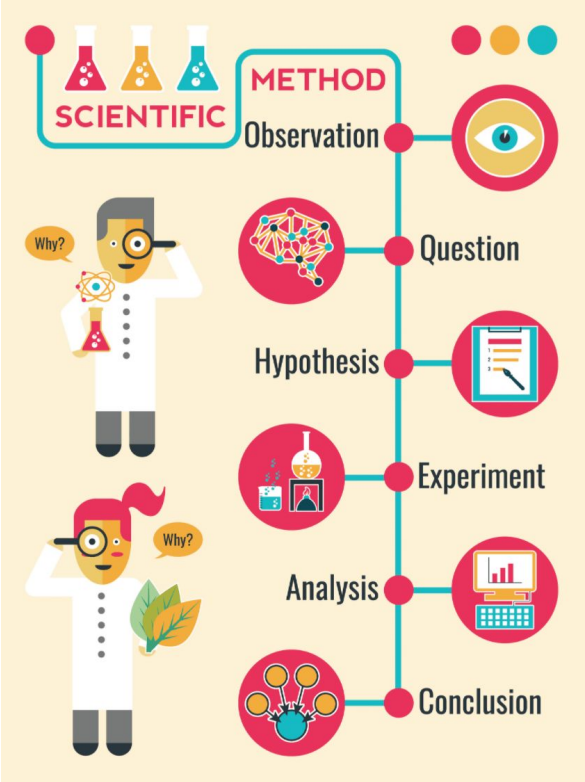
CSIC Consejo Superior de Investigaciones Científicas

WODA
WEB-BASED OMICS DATA ANALYSIS

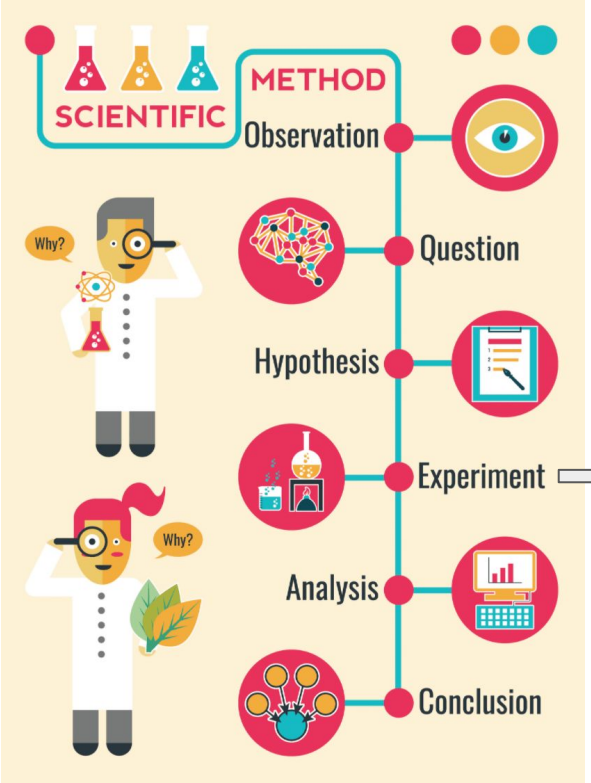
Índice

ENDEAVOUR

1. Introducción



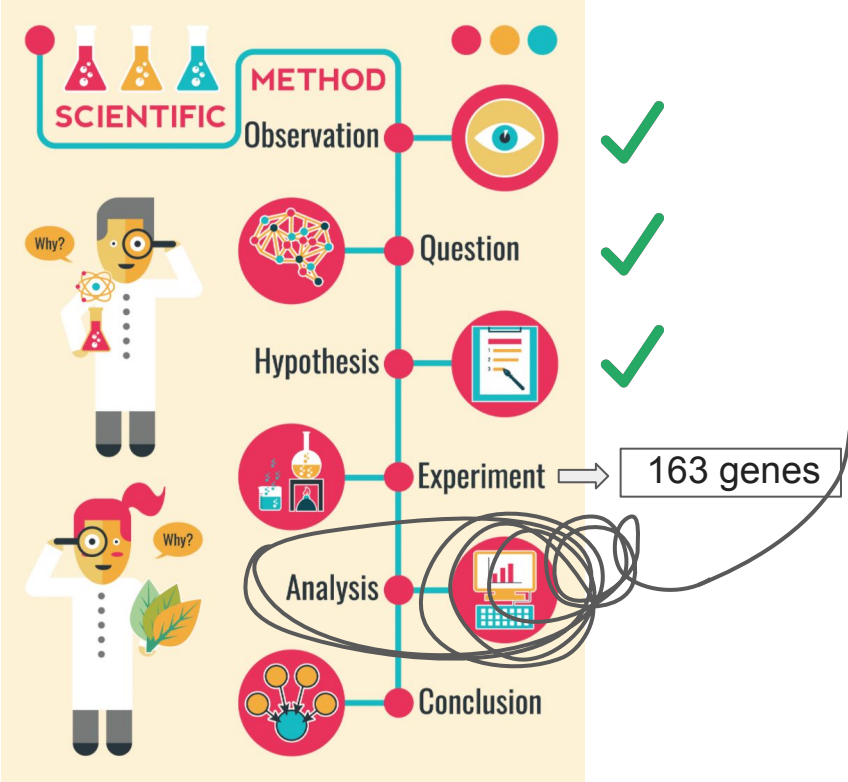
1. Introducción



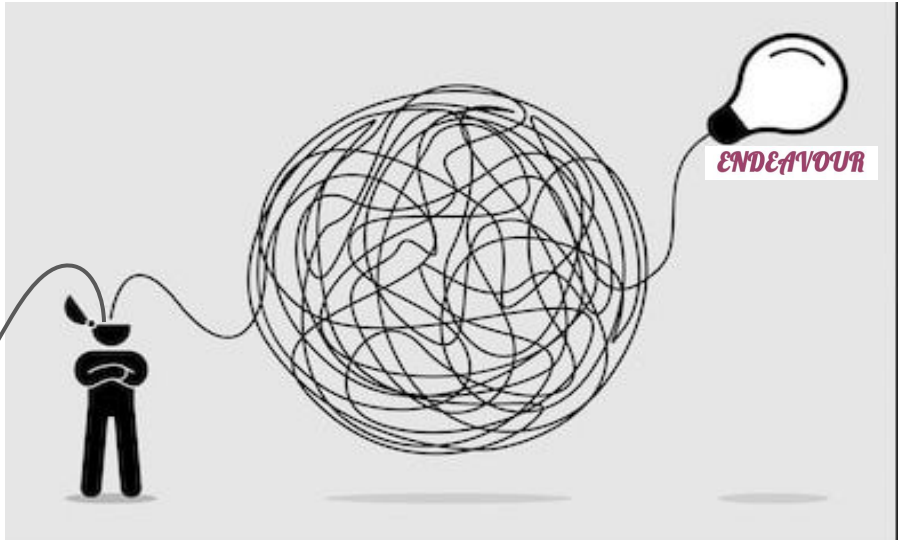
163 genes

ENDEAVOUR

1. Introducción



163 genes



PRIORIZACIÓN

ENDEAVOUR

1. Introducción

1. Introducción

Larga lista de genes candidatos

ENDEAVOUR

1. Introducción

Larga lista de genes candidatos

Falta de dinero

ENDEAVOUR

1. Introducción

Larga lista de genes candidatos

Falta de dinero

Falta de tiempo

ENDEAVOUR

1. Introducción

Larga lista de genes candidatos

Falta de dinero

Falta de tiempo

Hay que empezar

ENDEAVOUR

1. Introducción

Larga lista de genes candidatos

Falta de dinero

Falta de tiempo

Hay que empezar

PRIORIZACIÓN

ENDEAVOUR

ENDEAVOUR: A web resource for gene prioritization in multiple species

The identification of key genes involved in health and disease remains a formidable challenge. Experimental approaches often produce lists of candidate genes, among which disease causing genes are hidden (i.e., their disease associations are still unknown). These lists of candidate genes can be rather large, and thus experimentally validating each candidate gene would be too expensive and time-consuming. There is therefore the need to predict the most promising candidate genes as to be able to maximise the yield of the experimental validation, which has been defined as 'gene prioritization'.

We have developed a bioinformatics approach to prioritize candidate genes underlying biological processes or diseases, and implemented it into a software application termed 'Endeavour'. Our strategy is based on how similar a candidate gene is to a profile derived from genes already known to be involved in the process of interest. Our approach relies on the integration of multiple heterogeneous sources (e.g., coding sequence, gene expression, functional annotation, literature, regulatory information) that cover what we currently know about these genes.

More precisely, Endeavour consists of three stages: training, scoring and fusion. In the first stage, information about the training genes (genes already known to play a role in the process under study) are retrieved from the genomic data sources in order to build models (one per data source). In the second stage, the models are then used to score the candidate genes and to rank them according to their scores. In the last stage, the rankings (one per data source) are fused into a global ranking using Order Statistics. Endeavour is currently available for human, mouse, rat, fruit fly, zebra fish and worm.

We have successfully used Endeavour to prioritize a DiGeorge syndrome associated region, a congenital heart defects associated region, and to optimize a genetic screen in Drosophila melanogaster. Researchers have also used Endeavour to look for genes involved in cleft lip / cleft palate from aCGH data and to analyze the proteome of adipocytes. Please browse our reference section to find a list of Endeavour related publications.

Our Approach

DATA

Data from multiple heterogeneous sources are collected and integrated in our local database, for better performances.

Read more

ALGORITHMS

Our algorithm uses basic machine learning techniques to model the biological process under study and then to prioritize the candidate genes.

Read more

ALTERNATIVES

There exists an alternative in case the already known disease genes can not easily be identified.

Read more

1. Introducción

- Herramienta de priorización de genes

ENDEAVOUR

1. Introducción

- Herramienta de priorización de genes

Lista de genes

TOMM40
IDE
APP
APOE
CLU
UBB
KLC1
PSEN1
MAPT
BCHE

ENDEAVOUR

1. Introducción

- Herramienta de priorización de genes

Lista de genes
TOMM40
IDE
APP
APOE
CLU
UBB
KLC1
PSEN1
MAPT
BCHE

Ranking de genes	
1º	APOE
2º	APP
3º	BCHE
4º	CLU
5º	MAPT
6º	PSEN1
7º	IDE
8º	KLC1
9º	UBB
10º	TOMM40

ENDEAVOUR

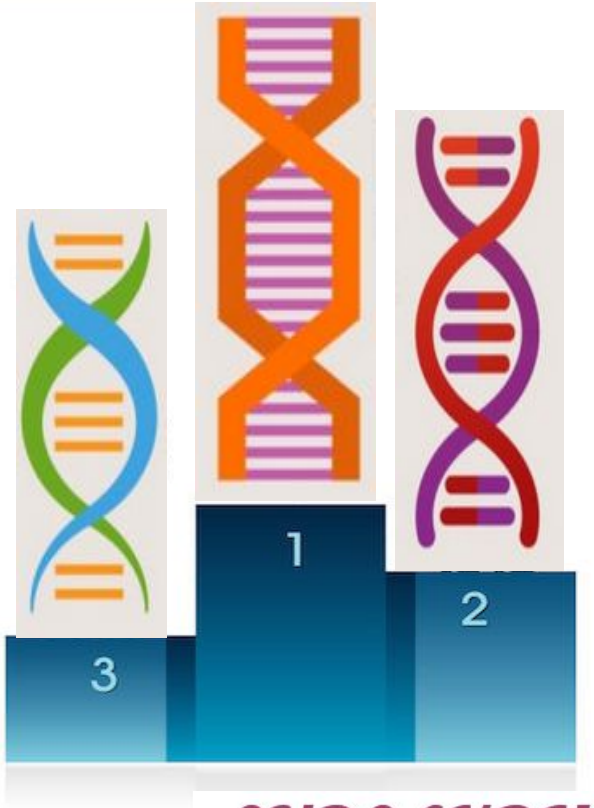
1. Introducción

- Herramienta de priorización de genes

Lista de genes
TOMM40
IDE
APP
APOE
CLU
UBB
KLC1
PSEN1
MAPT
BCHE

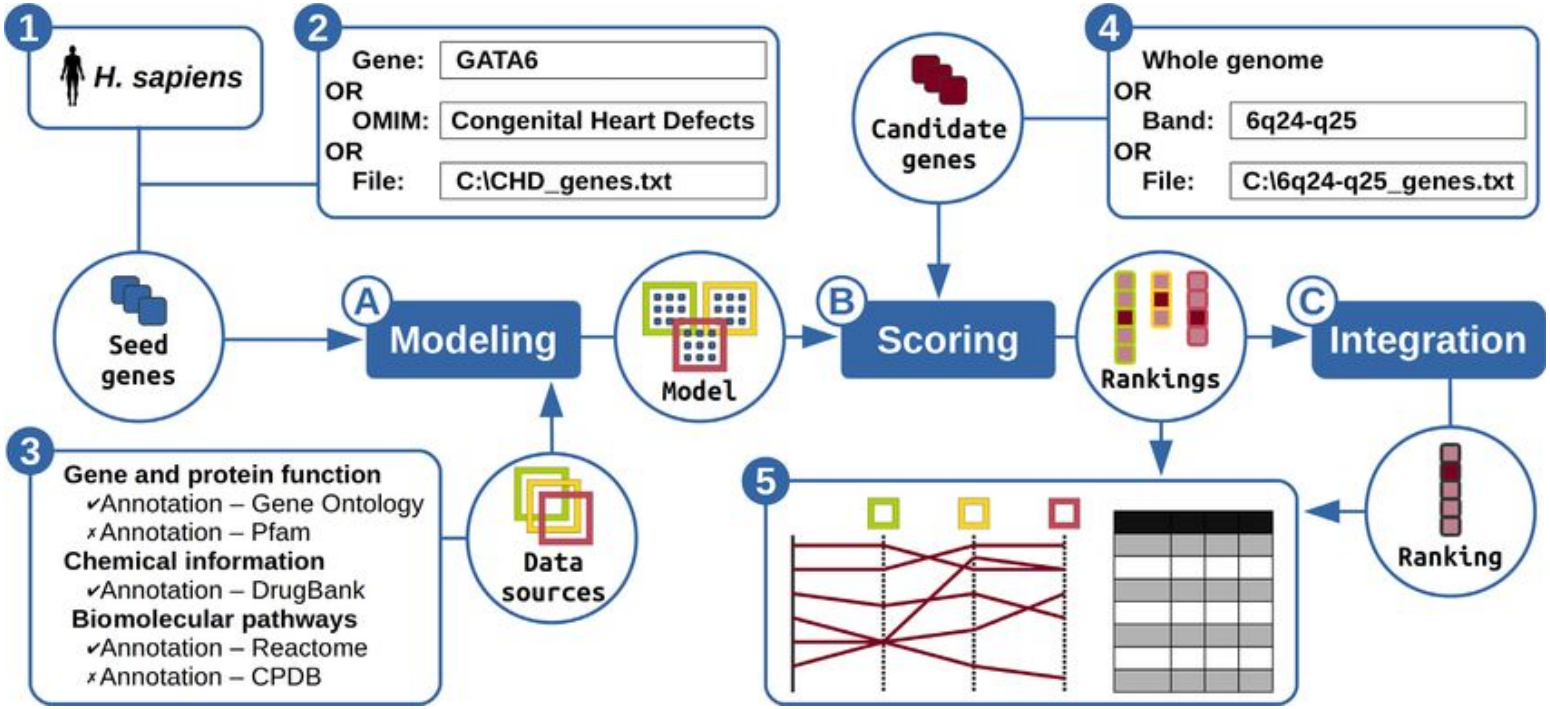
→

Ranking de genes	
1º	APOE
2º	APP
3º	BCHE
4º	CLU
5º	MAPT
6º	PSEN1
7º	IDE
8º	KLC1
9º	UBB
10º	TOMM40



ENDEAVOUR

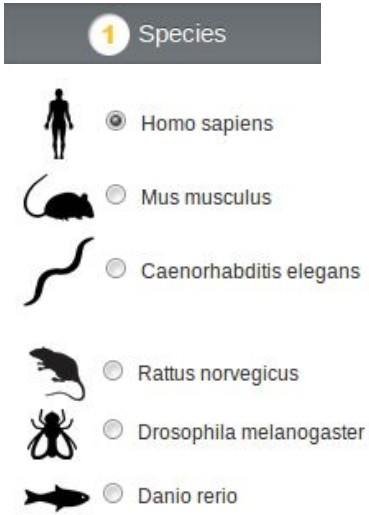
2. Algoritmos



ENDEAVOUR

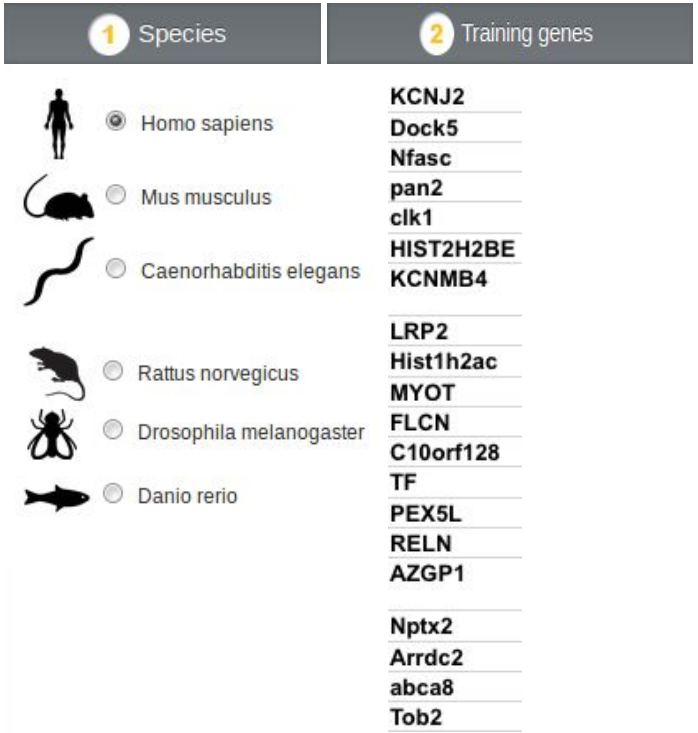
3. Selección de la información

3. Selección de la información



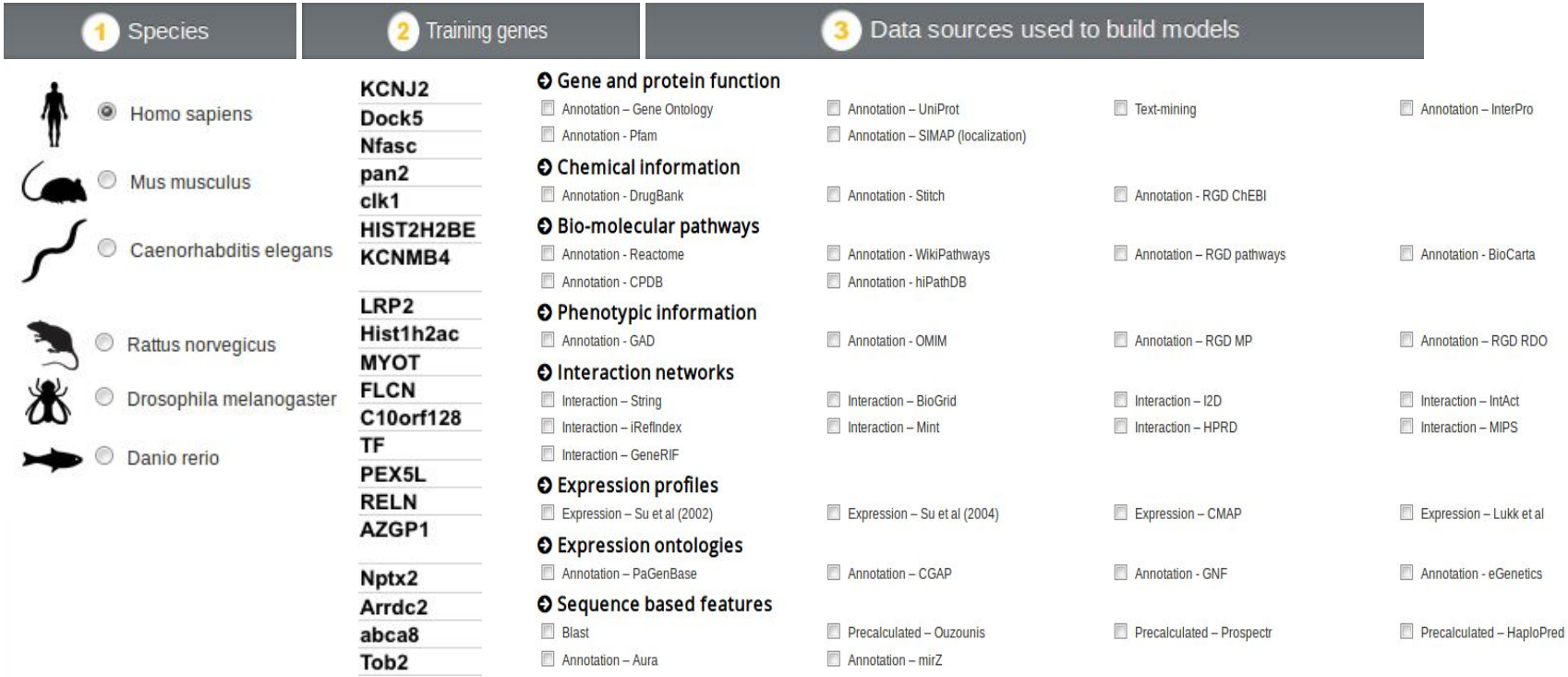
ENDEAVOUR

3. Selección de la información



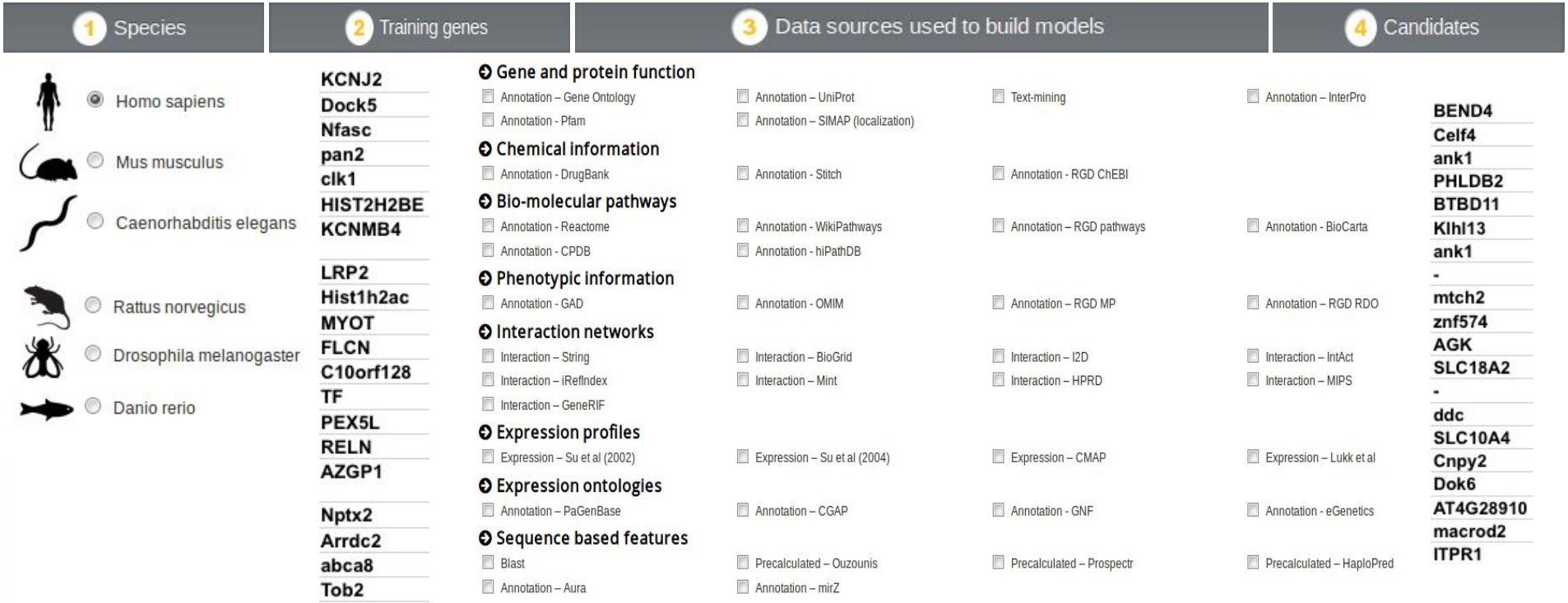
ENDEAVOUR

3. Selección de la información



ENDEAVOUR

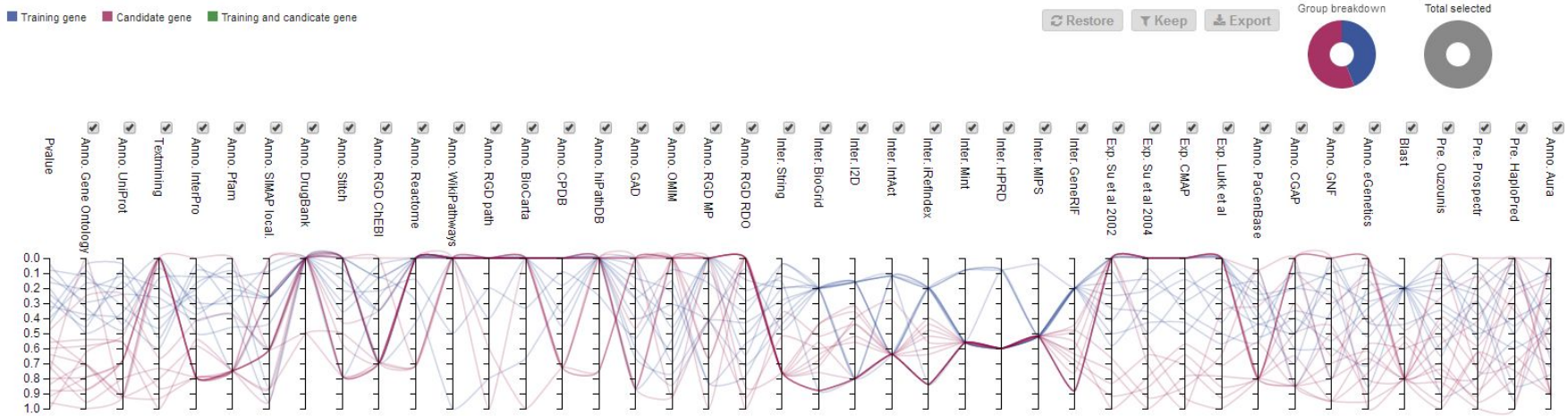
3. Selección de la información



ENDEAVOUR

4. Interpretación de resultados

4. Interpretación de resultados



Gene name	Group	P-value	Annotati...	Annotati...	Text-min...	Annotati...	Annotati...	Annotati...	Annotati...	Annotati...	Annotati...	Annotati...	Annotati...	Annotati...
CYP7B1	Training gene	0.24	0.4167	0.16	0.4	0.25	0.75	0.2609	0	0.2143	0.2174	0.1429	0.5	0.2
HSPD1	Training gene	0.28	0.4583	0.08	0.5333	0.0417	0.3333	0.6087	0	0.3571	0.1304	0.2857	1	0.8
PNPLA6	Training gene	0.32	0.375	0.4	0.6	0.0833	0.1667	0.2609	0	0.0714	0.3478	0	0	0.4
KIAA0196	Training gene	0.36	0.0417	0.04	0.3333	0.2083	0.0417	0.8696	0	0	0.6957	0	0	0
ZFYVE27	Training gene	0.4	0.1667	0.2	0	0.5	0.4583	0.4348	0	0	0.6957	0	0	0
NIPA1	Training gene	0.44	0.0833	0.44	0	0.3333	0.25	0.2609	0	0	0.1304	0	0	0
KIF1A	Candidate gene	0.48	0.25	0.24	0.6667	0.4167	0.375	0.6087	0.5	0.7857	0.6957	0	0	0
SNED1	Candidate gene	0.52	0.7083	0.92	0	0.7917	0.75	0.6087	0	0.7857				
PASK	Candidate gene	0.56	0.5417	0.56	0.8667	0.7917	0.75	0	0	0.4286				

Showing all 25 rows

ENDEAVOUR

5. Consejos

Genes semilla:

- Un conjunto mayor de genes semilla proporciona resultados más fiables. Se recomienda un mínimo de 5 genes y máximo 40 genes.

Fuentes de datos:

- La lista de fuentes de datos disponible depende de la especie seleccionada
- Seleccionar los recursos que mejor se adapten a su problema específico.
- Evitar usar conjuntamente recursos redundantes.
- No hay límite en el número de fuentes de datos, salvo el tiempo de ejecución.

Genes candidatos:

- No hay límite en el número de genes candidatos.

6. Buscador de genes

Sandra Alandes Esteve

salandes@cipf.es

ENDEAVOUR