

# Eina Shiny per a l'anàlisi de rutes

Vasyl Druchkiv

13 de juny, 2019

- ① Que és l'anàlisi de rutes?
- ② Què necessitem per començar?
- ③ Quins mètodes hi ha?
  - ORA
  - GSEA
  - Topologia de rutes
- ④ Instal·lació de l'aplicació
- ⑤ Exemple d'ús

# Qué és l'anàlisi de rutes?

Un Pathway és un conjunt de gens que actuen junts per dur a terme un procés biològic.

# Què és l'anàlisi de rutes?

- Reducció de la informació/complexitat resultant de l'experiment d'expressió genètica
- Millor comprensió del funcionament mecànic de la biologia de la condició estudiada
- Agrupació de gens relacionats en un conjunt (*set*)
  - Funció biològica
  - Funció molecular
  - Component cel·lular
- Permet valorar:
  - Relació entre els gens
  - Relativa importància de cada gen etc.
- Més potència estadística via reducció del nombre d'hipòtesis múltiples

# Què necessitem per començar?

- Dades d'un experiment de *microarrays*

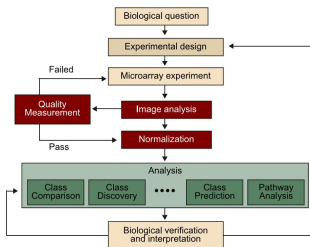


Figure 1: El procés d'anàlisi de microarrays.

- Llista ordenada de gens anotats amb Entrez ID
- *log Fold Change* per a cada gen
- Un *subset* de gens diferencialment expressats

# Què necessitem per començar?

- Una base de dades d'anotació de les rutes/conceptes
  - Gene Ontology (GO)
    - ▶ Biological proces
    - ▶ Molecular function
    - ▶ Cellular component
  - KEGG
    - ▶ Metabolism
    - ▶ Genetic Information Processing
    - ▶ Environmental Information Processing
    - ▶ Cellular Processes
    - ▶ Organismal Systems
    - ▶ Human Deseases
    - ▶ Drug Development
  - Reactome

# Quins mètodes hi ha?

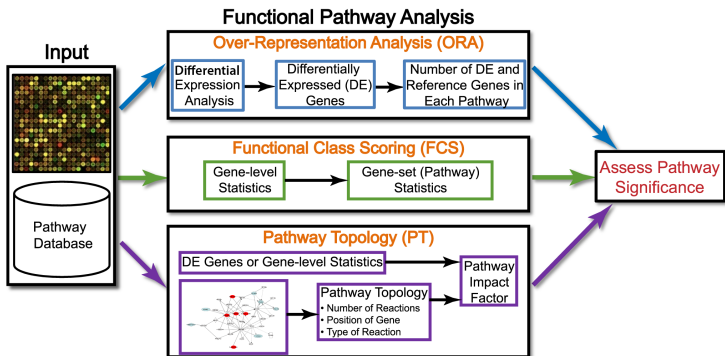


Figure 2: Mètodes per a anàlisi de les rutes.

# Quins mètodes hi ha?

## ORA

### BIOINFORMATICS APPLICATIONS NOTE

Vol. 20 no. 18 2004, pages 3710–3715  
doi:10.1093/bioinformatics/bth456



**GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes**

Elizabeth I. Boyle<sup>1</sup>, Shuai Wang<sup>1</sup>, Jeremy Gollub<sup>2</sup>, Hong Jin<sup>2</sup>, David Botstein<sup>1</sup>, J. Michael Cherry<sup>1</sup> and Gavin Sherlock<sup>1,\*</sup>

<sup>1</sup>Department of Genetics and <sup>2</sup>Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA

Received on August 5, 2003; revised on May 10, 2004; accepted on June 29, 2004  
Advance Access publication August 5, 2004

- 1 Selecció dels gens diferencialment expressats.
- 2 Determinar si algunes rutes anoten la llista especificada de gens amb la freqüència més alta que la que s'esperaria per casualitat.

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

- 3 Ajustament de valors de P per a *Multiple hypothesis testing*



# Quins mètodes hi ha

ORA. Exemple.

```
d <- matrix(c(25, 237, 177, 11323),  
            nrow = 2,  
            dimnames = list(c("DE", "Not DE"),  
                             c("In GS", "Not in GS")))
```

d

	In GS	Not in GS
DE	25	177
Not DE	237	11323

```
fisher.test(d, alternative = "greater")$p.value
```

```
[1] 2.815289e-12
```

# Quins mètodes hi ha?

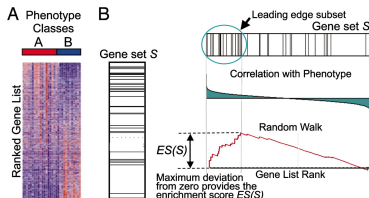
## GSEA

### Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

Aravind Subramanian<sup>1,2</sup>, Pablo Tamayo<sup>1,3</sup>, Vamsi K. Mootha<sup>4,5</sup>, Sayan Mukherjee<sup>1</sup>, Benjamin L. Ebert<sup>1,6</sup>, Michael A. Gillette<sup>1,7</sup>, Amanda Paulovich<sup>1</sup>, Scott L. Pomeroy<sup>1</sup>, Todd R. Golub<sup>8,9</sup>, Eric S. Lander<sup>10,11,12</sup>, and Jill P. Mesirov<sup>1,3</sup>

<sup>1</sup>Broad Institute of Massachusetts Institute of Technology and Harvard, 300 Charles Street, Cambridge, MA 02141; <sup>2</sup>Department of Systems Biology, Albert 136, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02446; <sup>3</sup>Institute for Genome Sciences and Policy, Center for Interdisciplinary Engineering, Medicine, and Applied Sciences, Duke University, 105 Science Drive, Durham, NC 27708; <sup>4</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, 44 River Street, Boston, MA 02115; <sup>5</sup>Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; <sup>6</sup>Roswell Park Cancer Research Center, 1900 Fairview Avenue North, G-250, P.O. Box 19024, Seattle, WA 98119-0224; <sup>7</sup>Department of Neurology, Boston 200, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115; <sup>8</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and <sup>9</sup>Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142

Contributed by Eric S. Lander, August 2, 2005



- 1 Càlcul de la puntuació d'enriquement ( $ES$ : *Enrichment Score*).
- 2 Estimació del nivell de significació per a la puntuació  $ES$ .
- 3 Càlcul del valor de  $P$  ajustat.