

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*

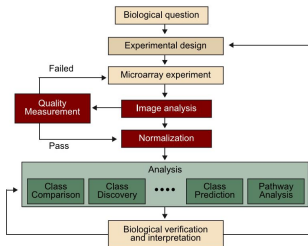


Figure 1: El procés d'anàlisi de microarrays.

- 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?

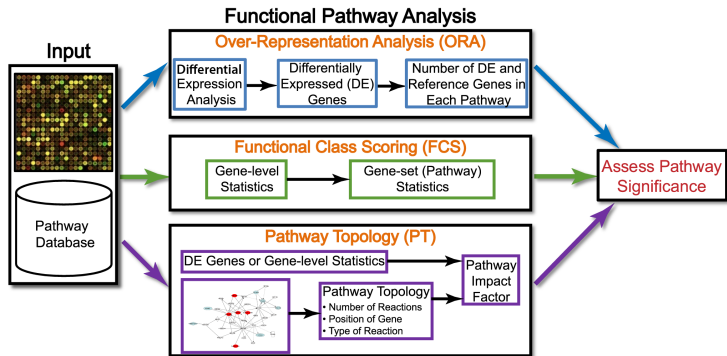


Figure 2: Mètodes per a anàlisi de les rutes.

Quins mètodes hi ha?

ORA

BIOINFORMATICS APPLICATIONS NOTE

Vol. 20 no. 18 2004, pages 3710–3715
doi:10.1093/bioinformatics/bth455



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. Boyle¹, Shuai Weng¹, Jeremy Gollub², Heng Jin², David Botstein¹, J. Michael Cherry¹ and Gavin Sherlock^{1,*}

¹Department of Genetics and ²Department 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?

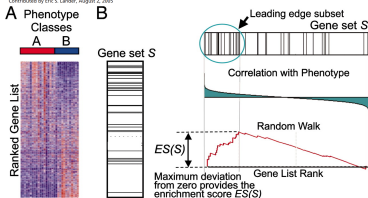
GSEA

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

Aravind Subramanian^{1,2}, Pablo Tamayo^{3,4}, Vamsi K. Mootha^{5,6}, Sayan Mukherjee¹, Benjamin L. Ebert^{1,2}, Michael A. Gillette^{1,2}, Amanda Paulovich^{1,2}, Scott L. Pomeroy¹, Todd R. Golub^{1,2}, Eric S. Lander^{1,2,3,4,5,6}, and Jill P. Mesirov^{1,2}

¹Royal Institute of Massachusetts Institute of Technology and Harvard, 330 Charles Street, Cambridge, MA 02141; ²Department of Systems Biology, Alpert 536, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02145; ³Institute for Genome Sciences and Policy, Center for Transdisciplinary Engineering, Medicine, and Applied Sciences, Duke University, 151 Science Drive, Durham, NC 27708; ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Ervin Street, Boston, MA 02115; ⁵Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02115; ⁶Rod Hutchinson Cancer Research Center, 1180 Fairview Avenue North, CO-623, P.O. Box 19834, Seattle, WA 98196-1624; ⁷Department of Neurology, Enders 200, Children's Hospital, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115; ⁸Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and ⁹Whitehead 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.

Quins mètodes hi ha?

Visualització ORA i GSEA

- Bar Plot
- Dot Plot
- GSEA Plot

Topologia de les rutes

- Gràfic de les rutes KEGG
- Gràfic de les rutes Reactome.

Altres gràfics interessants

- GO Plots
 - Relacions *is a*, *part of*, *regulate*
- Enrichment Map
 - Reduir/simplificar la informació obtinguda mitjançant els mètodes ORA o GSEA
- Gene-Concept Network
 - Els gens compartits entre les categories/rutes

Instal·lació de l'aplicació

L'aplicació està empaquetada i guardada en GitHub

```
if(!require(devtools))install.packages("devtools")
if(!require(BiocManager))install.packages("BiocManager")
devtools::install_github("vdruchkiv/TFM/5_Packages/PathwayApp/PathwayApp")
PathwayApp::runPathwayApp()
```

Instal·lació de l'aplicació

Els paquets que s'instal·laran

Paquet	Font
clusterProfiler	Bioconductor
ReactomePA	Bioconductor
pathview	Bioconductor
pathviewPatched	GitHub vdruchkiv/TFM
dplyr	CRAN
ggplot2	CRAN
knitr	CRAN
kableExtra	CRAN
formattable	CRAN
shiny	CRAN
shinydashboard	CRAN
shinyhelper	CRAN
shinycssloaders	CRAN

L'usuari ha de baixar les bases d'anotacions d'interès!!!

L'exemple d'ús