

Análisis Funcional de Genes Mitocondriales usando g:Profiler

Aissa Omar El Hammouti Chachoui

Herramientas y Algoritmos en Bioinformática

4 de Noviembre 2025

Abstract

Este trabajo presenta un análisis de enriquecimiento funcional de tres genes mitocondriales (COX4I2, ND1 y ATP6) utilizando la herramienta g:Profiler. El análisis identificó 136 términos significativamente enriquecidos, confirmando la participación de estos genes en procesos fundamentales de la respiración celular y la producción de energía. Los resultados demuestran una alta coherencia biológica y permiten comprender las funciones compartidas de estos genes en la fosforilación oxidativa mitocondrial.

Palabras clave: Análisis funcional, Enriquecimiento, Genes mitocondriales, g:Profiler, Bioinformática

1 Introducción

El análisis funcional de genes es una metodología fundamental en bioinformática que permite identificar procesos biológicos, funciones moleculares y componentes celulares asociados a un conjunto de genes de interés. Esta aproximación es especialmente útil para comprender las relaciones funcionales entre genes y su papel en sistemas biológicos complejos.

En este trabajo se analizaron tres genes mitocondriales: **COX4I2** (Citocromo c oxidasa subunidad 4 isoforma 2), **ND1** (NADH deshidrogenasa subunidad 1) y **ATP6** (ATP sintasa F0 subunidad 6). Estos genes son componentes clave de la cadena de transporte de electrones mitocondrial y desempeñan roles esenciales en la producción de ATP celular.

El objetivo de este análisis es identificar y caracterizar los términos funcionales significativamente enriquecidos asociados a estos genes, utilizando múltiples bases de datos de anotación funcional.

2 Metodología

2.1 Herramienta de Análisis

Se utilizó **g:Profiler**, una herramienta web de análisis de enriquecimiento funcional que integra múltiples bases de datos de anotación:

- **Gene Ontology (GO):** Ontología estructurada que describe funciones génicas en tres categorías: procesos biológicos (BP), funciones moleculares (MF) y componentes celulares (CC).
- **KEGG:** Base de datos de rutas metabólicas y de señalización.
- **Reactome:** Base de datos curada de rutas y procesos biológicos.
- **WikiPathways:** Colección de rutas biológicas mantenida por la comunidad científica.

2.2 Análisis Estadístico

g:Profiler realiza un análisis de sobre-representación (ORA, *Over-Representation Analysis*) utilizando el test exacto de Fisher o el test hipergeométrico. Este método compara la proporción de genes en la lista de entrada que están anotados con un término funcional específico contra la proporción esperada por azar en el genoma completo.

Parámetros del análisis:

- Bases de datos: GO:BP, GO:MF, GO:CC, KEGG, REAC, WP

2.3 Implementación

El análisis se implementó mediante un script en Python que utiliza la librería `gprofiler-official` para interactuar con la API de g:Profiler. El script acepta como entrada un archivo con genes y genera como salida un archivo TSV con los términos enriquecidos.

Ejecución:

```
python analisis-funcional.py -i genes_input.txt -o resultados.tsv
```

3 Resultados

3.1 Resumen Estadístico

El análisis identificó un total de **136 términos significativamente enriquecidos** ($p < 0.05$) distribuidos en las diferentes bases de datos consultadas:

- GO:BP (Procesos Biológicos): 65 términos (47.8%)
- GO:MF (Funciones Moleculares): 23 términos (16.9%)
- GO:CC (Componentes Celulares): 21 términos (15.4%)
- KEGG (Rutas Metabólicas): 12 términos (8.8%)
- REAC (Reactome): 9 términos (6.6%)
- WP (WikiPathways): 6 términos (4.4%)

3.2 Términos Principales

La Tabla 1 presenta los 10 términos más significativamente enriquecidos ordenados por p-valor. Estos términos reflejan las funciones principales de los genes analizados.

Table 1: Top 10 términos enriquecidos significativamente

BD	ID	Término	P-valor	Genes
GO:CC	GO:0098803	Respiratory chain complex	4.17×10^{-6}	3/102
GO:MF	GO:0015078	Proton transmembrane transporter activity	2.29×10^{-5}	3/186
GO:BP	GO:0009060	Aerobic respiration	2.45×10^{-5}	3/199
GO:BP	GO:0006119	Oxidative phosphorylation	2.45×10^{-5}	3/146
GO:BP	GO:1902600	Proton transmembrane transporter	2.45×10^{-5}	3/190
GO:BP	GO:0045333	Cellular respiration	3.31×10^{-5}	3/242
KEGG	KEGG:00190	Oxidative phosphorylation	5.78×10^{-5}	3/134
KEGG	KEGG:04714	Thermogenesis	7.57×10^{-5}	3/232
KEGG	KEGG:05208	Chemical carcinogenesis - ROS	7.57×10^{-5}	3/221
KEGG	KEGG:05415	Diabetic cardiomyopathy	7.57×10^{-5}	3/202

4 Interpretación de Resultados

4.1 Coherencia Biológica

Los resultados demuestran una alta coherencia biológica. Los términos más significativos están directamente relacionados con la función conocida de los tres genes:

1. Localización celular: El término más significativo (“Respiratory chain complex”, $p = 4.17 \times 10^{-6}$) confirma que los tres genes forman parte de complejos proteicos de la cadena respiratoria mitocondrial. Específicamente:

- **COX4I2:** Complejo IV (citocromo c oxidasa)
- **ND1:** Complejo I (NADH deshidrogenasa)
- **ATP6:** Complejo V (ATP sintasa)

2. Función molecular: Los genes comparten la función de “transportador transmembrana de protones” ($p = 2.29 \times 10^{-5}$), esencial para generar el gradiente electroquímico que impulsa la síntesis de ATP.

3. Proceso biológico: Los términos “respiración aeróbica”, “fosforilación oxidativa” y “respiración celular” (todos con $p \approx 2 - 3 \times 10^{-5}$) confirman el papel central de estos genes en la producción de energía celular.

4.2 Implicaciones Clínicas

Resulta interesante observar que varios términos KEGG están relacionados con enfermedades neurodegenerativas (Parkinson, Alzheimer, Huntington) y cardiomiopatía diabética. Esto refleja la importancia de la función mitocondrial en estas patologías, donde disfunciones en la cadena respiratoria pueden contribuir al desarrollo de la enfermedad.

4.3 Significancia Estadística

Todos los términos presentados tienen p-valores altamente significativos ($p < 10^{-4}$), lo que indica que la presencia de los tres genes en estos términos funcionales no es casual. La corrección FDR aplicada garantiza que estos resultados son robustos frente al problema de comparaciones múltiples.

5 Conclusiones

El análisis funcional mediante g:Profiler permitió identificar exitosamente 136 términos enriquecidos asociados a los genes COX4I2, ND1 y ATP6. Los resultados confirman que estos genes:

1. Están localizados en la membrana interna mitocondrial
2. Forman parte de complejos de la cadena respiratoria
3. Participan en la fosforilación oxidativa y producción de ATP
4. Tienen relevancia en patologías relacionadas con disfunción mitocondrial

Este análisis demuestra la utilidad de las herramientas de enriquecimiento funcional para comprender las relaciones funcionales entre genes y su papel en procesos biológicos complejos. La metodología implementada es reproducible, automatizada y puede aplicarse a otros conjuntos de genes de interés.