

Esame 16/01/2018

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Dati:

- Un file di GO term mapping, che riporta associazioni gene - termine della gene ontology
- Un file di dati di microarray, con livelli di espressione per ciascun gene (riga) in una serie di condizioni
- Un dominio GO (C/F/P)
- Un numero minimo di geni (soglia)

Programma python

Scrivere un programma `expression_by_function` che:

- prenda in ingresso i nomi dei due file, il dominio e la soglia
- per ciascuno dei termini del dominio selezionato con un numero di geni associati pari almeno alla soglia, stampi la media dei profili di espressione dei geni ad esso associati

Esempio Dati di Ingresso

file associazioni GO

Q0085	ATP6	S000007268	C	cytoplasm	GO:0005737	..
Q0085	ATP6	S000007268	C	membrane	GO:0016020	..
Q0085	ATP6	S000007268	C	mitochondrial envelope	GO:000	
Q0085	ATP6	S000007268	C	mitochondrion	GO:0005739	..
Q0085	ATP6	S000007268	F	ATPase activity	GO:0016887	..
...						
YAL001C	TFC3	S000000001	C	chromosome	GO:0005694	..
YAL001C	TFC3	S000000001	C	cytoplasm	GO:0005737	..
YAL001C	TFC3	S000000001	C	mitochondrion	GO:0005739	..
...						

file microarray

YORF	NAME	GWEIGHT	4741-A	4741-B	4741-C	...
EWEIGHT			1	1	1	...
YKL175W	YKL175W	1	3.353900		3.313720	3.329830 ...
YHR069C	YHR069C	1	3.195310		3.155620	3.160450 ...
YDR533C	YDR533C	1	3.353550		3.302160	3.323650 ...
YMR089C	YMR089C	1	3.153010		3.119910	3.142150 ...
YKL176C	YKL176C	1	3.252290		3.218960	3.250450 ...
...						

Esempio esecuzione

```
> python expression_by_function.py
Inserire nome file GO terms: go_slim_mapping_filtered.tab
Inserire nome file espressioni: GSE65666.final.pcl
Inserire dominio (C/F/P): C
Inserire numero minimo geni: 200
```

cytoplasm	3.099	3.070	3.086	3.068	3.041	3.072
membrane	3.058	3.028	3.044	3.025	2.998	3.029
mitochondrion	3.095	3.065	3.081	3.041	3.015	3.045
nucleus 2.976	2.948	2.962	2.961	2.934	2.966	
chromosome	2.790	2.765	2.777	2.765	2.740	2.770
endomembrane	3.153	3.123	3.138	3.118	3.088	3.121
endoplasmic	3.161	3.130	3.147	3.134	3.106	3.137
vacuole 3.089	3.058	3.076	3.057	3.029	3.060	
plasma 3.024	2.995	3.011	3.002	2.977	3.006	

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file GO e restituisca un dizionario che associa ad ogni dominio un dizionario GO term → elenco dei geni annotati con quel termine (term2genes)
- 2 una che legge il file di microarray e restituisca il numero di condizioni testate, ed un dizionario gene → lista dei valori di espressione per ogni condizione (gene2exprs)
- 3 una che data una lista di geni, un dizionario gene2exprs e il numero di condizioni, calcoli la media dei profili di espressione dei geni nella lista
- 4 una che dato un dizionario term2genes, una soglia, un dizionario gene2exprs e il numero di condizioni, stampi per ogni GO term un numero di geni almeno pari alla soglia, la media dei profili di espressione dei suoi geni
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

- Calcolare gli ID delle sequenze all'interno della directory `fastas/` aventi il seguente pattern:
 - un'alanina (A) ad inizio sequenza
 - almeno una metionina (M) in qualsiasi altro punto della sequenza

Risultato atteso

```
1ast_  
1by2_  
1n7uA  
1p5uB
```

Esercizi da linea di comando

- Stampare a schermo i file con ID che cominciano per `li`, in ordine decrescente di lunghezza.

Risultato atteso

```
689 li1iP.f
470 librB.f
408 linp_.f
120 lichA.f
 80 licwA.f
 50 liytA.f
```


Modalita' di esecuzione e consegna

- 1 Avviare la macchina in modalita' `ESAME`
- 2 Autenticarsi con nome utente `sci-esame` e password fornita dal docente
- 3 Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
- 4 Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
- 5 Creare sul Desktop una cartella con *nome_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina