

Esame 20/01/2020

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che prenda in ingresso un file con un elenco di regioni conservate in UTR e un numero n e stampi:

- I primi n geni ordinati per lunghezza media delle loro regioni conservate
- I primi n cromosomi ordinati per lunghezza media delle regioni conservate nei loro geni

Esempio file di input

```
> cat conserved_regions
```

HGNC	ucsc	chr	start	end	strand	region.length	...
SDF4	uc001adj.1	chr1	1159309	1159325	-	17	...
SDF4	uc001adj.1	chr1	1159341	1159347	-	7	...
DVL1	uc002quu.2	chr1	1275478	1275556	-	79	...
CCNL2	uc001afg.1	chr1	1326239	1330902	-	324	...
CCNL2	uc001afg.1	chr1	1333613	1333621	-	9	...
CCNL2	uc001afg.1	chr1	1333633	1333660	-	28	...
...							

Esempio esecuzione

```
> python regions_stats.py
Inserire nome file: conserved_regions
Inserire numero elementi: 5
Genes with longest conserved on average
RYR1      2619
DDX3X     2549
NOVA1     2074
BCL11A    1718
TSHZ3     1599
Chromosomes with longest conserved on average
chr15     181
chr2       123
chrX       122
chr10      122
chr8       105
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file dati e restituisca una lista con nome gene, cromosoma e lunghezza della regione conservata
- 2 una che data la lista e un'indicazione se usare gene o cromosoma, restituisca un dizionario da gene (o da cromosoma) a elenco di lunghezze di regioni conservate per quel gene (o cromosoma)
- 3 una che dato un dizionario chiave \rightarrow lista di valori, restituisca una lista ordinata di coppie chiave, media dei suoi valori (si può usare la funzione `mean` del modulo `statistics` per calcolare la media di una lista)
- 4 una che data una lista ordinata e un numero n , stampi le prime n coppie come da esempio
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

Calcolare quante sequenze nella directory `fastas` contengono almeno cinquanta acidi aspartici o glutammici (D o E)

Risultato atteso

14

Esercizi da linea di comando

Estrarre dal file `conserved_regions` il gene del cromosoma X che ha il numero maggiore di sequenze conservate (ignorare NA che sta per “Not Available”)

Risultato atteso

11 KDM6A

Modalita' di esecuzione e consegna

- 1 Accendere la macchina ed aspettare (la modalità esame si avvia automaticamente)
- 2 Aprire un browser e collegarsi al sito
`examina.icts.unitn.it`
- 3 Autenticarsi con le credenziali di ateneo
- 4 Scaricare il testo del compito
- 5 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`
- 6 Creare una cartella con *nome_cognome* e metterci i due file realizzati, comprimerla e caricarla su `examina.icts.unitn.it`
- 7 Spegnere la macchina