

Εργασία 2

Μέρος 1ο

Ανάλυση Μικροσυστοιχιών

Κατεβάστε τα δεδομένα από το dataset

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS3709>

Το αρχείο περιέχει θηλυκά, αρσενικά άτομα που χωρίζονται σε καπνιστές και μη καπνιστές.

Ερώτηση 1:

Βρείτε τα γονίδια που η έκφραση τους επηρεάζεται

1. από το φύλο
2. από το κάπνισμα

Χρησιμοποιείστε μοντέλο αλληλεπίδρασης (FactorA*FactorB)

Ερώτηση 2:

Κάντε το ίδιο αλλά χρησιμοποιήστε αθροιστικό μοντέλο (FactorA+FactorB)

Ερώτηση 3:

Κάντε το ίδιο με t-test ενώνοντας αντίστοιχα τους καπνιστές/μη καπνιστές (ώστε να απαντήσετε το 1) και τα φύλα ώστε να απαντήσετε το 2.

Ερώτηση 4:

Βρείτε με 2 τρόπους, όποιους θέλετε (δικαιολογήστε την επιλογή σας) τα γονίδια που επηρεάζονται από το κάπνισμα στα θηλυκά άτομα. Συγκρίνετε τα αποτελέσματα.

Μέρος Β

Ανάλυση αλληλουχιών

Κατεβάστε τα γονιδιώματα με τις upstream αλληλουχίες από το UCSC.

Οι δύο πρώτες είναι από άνθρωπο και χιμπατζή
οι επόμενες δύο από ποντίκι και αρουραίο.

Βρείτε όλα τα μοτίβα από κάθε είδος και το πλήθος τους (μήκος 6 βάσεις)

α) το συνολικό τους πλήθος (μετρώντας το πολλές φορές αν υπάρχει πάνω από μία φορά σε κάποια περιοχή)

β) μετρώντας το μόνο μία φορά αν υπάρχει σε κάποια περιοχή

Βρείτε αν κάποια μοτίβα υπάρχουν περισσότερες φορές στα πρωτεύοντα από ότι στα τρωκτικά

<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/upstream1000.fa.gz>

<https://hgdownload.soe.ucsc.edu/goldenPath/panTro6/bigZips/upstream1000.fa.gz>

<https://hgdownload.soe.ucsc.edu/goldenPath/mm39/bigZips/upstream1000.fa.gz>

<https://hgdownload.soe.ucsc.edu/goldenPath/rn7/bigZips/upstream1000.fa.gz>