**Εργασία Βιοπληροφορικής ΙΙ Εργαστήριο 3**

**Ονοματεπώνυμο: ΤΑΜΖΙΝΤ ΡΑΚΙ ΣΑΡΙΦ**

**ΑΜ: 02356**

**(Αυτα που ειχαμε για το σπιτι)**

**Εκφώνηση:**

**Α) Ανάλυση γονιδιακής έκφρασης με το GEO2R και ανάλυση εμπλουτισμού με το gProfiler**α) Μεταβείτε στην σελίδα της GEO (https://www.ncbi.nlm.nih.gov/geo/). Βρείτε μια case-control μελέτη στην οποία οι δύο ομάδες αναφέρονται ξεκάθαρα.  
β) Με την μελέτη που βρήκατε στην GEO, πραγματοποιήστε ανάλυση διαφορικής έκφρασης στο GEO2R, ορίζοντας σωστά τα cases και τα controls, και κατεβάστε τα αποτελέσματα της ανάλυσης.  
γ) Επιλέξτε τα top 100 διαφορικά εκφρασμένα γονίδια (p < 0.05 ) από τα αποτελέσματα και πραγματοποιήστε ανάλυση εμπλουτισμού στο gProfiler (<https://biit.cs.ut.ee/gprofiler/gost>)

Η case-control μελέτη που επέλεξα για να πραγματοποιήσω την μελέτη είναι η **GSE41586** και θα παραθέσω με μορφή printscreen τα εκάστοτε αποτελέσματα για κάθε ερώτημα.

Α)

A screenshot of a computer

AI-generated content may be incorrect.

Β)

A screenshot of a computer

AI-generated content may be incorrect.

Γ)

A screen shot of a computer

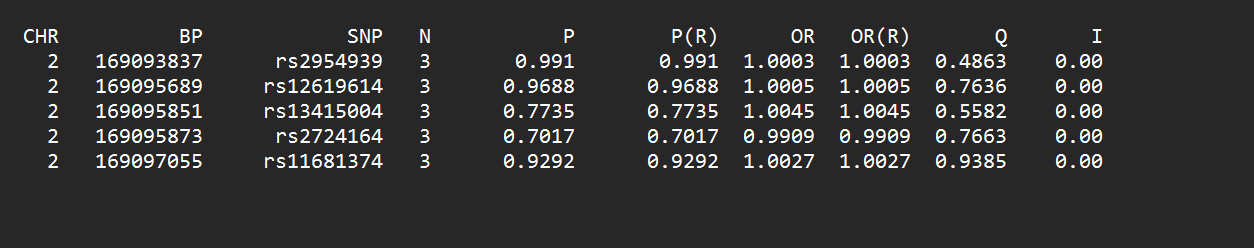
AI-generated content may be incorrect.

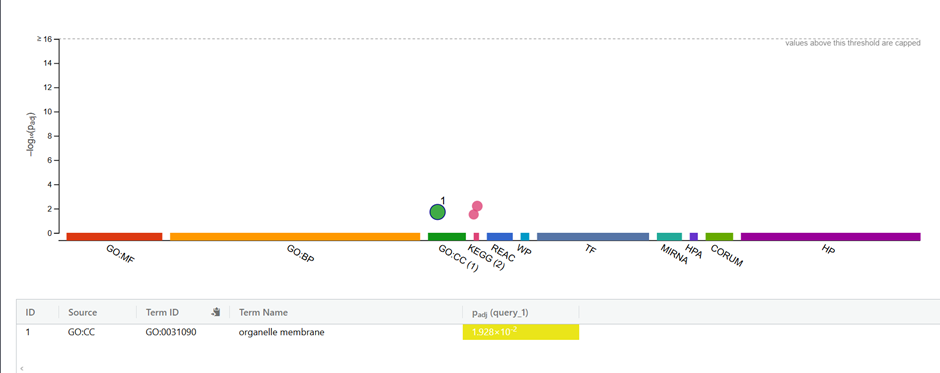
Στο συγκεκριμένο γράφημα γίνεται ανάλυση των αποτελεσμάτων που παρατίθενται στο ερώτημα Β). Συγκεκριμένα, στον άξονα Y (-log10(padj)) αναγράφεται ο αρνητικός λογάριθμος του p-value. Όσο πιο ψηλά είναι οι τιμές, τόσο πιο σημαντικός είναι ο εμπλουτισμός και δηλαδή τόσο πιο χαμηλό p-value έχει το γονίδιο. Στον άξονα Χ αναγράφονται οι κατηγορίες των λειτουργικών όρων. Στην περίπτωση μας, εμφανίζονατι δυο γονίδια που έχουν σημαντικούς εμπλουτισμούς στο Gene Ontology: Cellular Component (**GO:CC**). Το πρώτο είναι το **GO:0071944 “cell periphery”** με **p-value=8.973x10^4<0.05** και ανήκει στα γονίδια που εντοπίζονται στην μεμβράνη του κυττάρου ή κοντά σε αυτή. Το δεύτερο είναι το **GO:0097209 “epidermal lamellar body”** με **p-value=3.006x10^-2<0.05** και ανήκει στα γονίδια με επιδερμικές λειτουργίες.

**Β) GWAS μετα-ανάλυση με το PLINK και ανάλυση εμπλουτισμού με το gProfiler**  
ΠΡΟΣΟΧΗ : Για αυτή την άσκηση, θα πρέπει να εκτελέσετε το PLINK σε περιβάλλον Linux (είτε με WSL, Cygwin, VirtualBox, κλπ.)  
α) Κατεβάστε το PLINK v.1.9 (Stable, Linux 64-bit) από την ακόλουθη διεύθυνση: (https://www.cog-genomics.org/plink/1.9/)  
β) Κατεβάστε τα ακόλουθα σύνολα δεδομένων από την διεύθυνση : https://github.com/gmanios/bioplhroforiki\_II/tree/main/PLINK\_EXAMPLE\_data  
γ) Πραγματοποιήστε μετα-ανάλυση με το PLINKκαι ανάλυση εμπλουτισμού με τα στατιστικά σημαντικά SNPs (p < 1e-8 ) στο gProfiler. Ερμηνεύστε και σχολιάστε τα αποτελέσματα της μετα-ανάλυσης και της ανάλυσης εμπλουτισμού.

Γ)

**Ανάλυση αποτελεσμάτων**





Στο παραπάνω γράφημα αναπαρίστανται τα αποτελέσματα από την ανάλυση των πληροφοριών στην παραπάνω φωτογραφία. Συγκεκριμένα, στον άξονα Y (-log10(padj)) αναγράφεται ο αρνητικός λογάριθμος του p-value. Αντίστοιχα και εδώ όσο πιο ψηλά είναι οι τιμές, τόσο πιο σημαντικός είναι ο εμπλουτισμός και δηλαδή τόσο πιο χαμηλό p-value έχει το γονίδιο. Στον άξονα Χ αναγράφονται οι κατηγορίες των λειτουργικών όρων. Στο γράφημά μας εδώ, έχουμε σημαντικό εμπλουτισμό ξανά για το Gene Ontology: Cellular Component (**GO:CC**) για το γονίδιο **GO:0031090 “organelle membrane”** με **p-value=1.928x10^-2<0.05**. Το συγκεκριμένο γονίδιο εντοπίζεται πιο συχνά σε μεμβράνες οργανιδίων, όπως πχ σε μιτοχόνδρια, πυρήνας, ενδοπλασματικό δίκτυο κ.λ.π.