ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΕΡΓΑΣΤΗΡΙΟ 3  
ΤΑΜΖΙΝΤ ΡΑΚΙ ΣΑΡΙΦ

ΑΜ:02356  
(1ο αρχειο word , με αυτα που καναμε στο εργαστηριο)  
  
  
**Α) Ανάλυση γονιδιακής έκφρασης με το GEO2R και ανάλυση εμπλουτισμού με το gProfiler**

α) Μεταβείτε στην ιστοσελίδα του GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/> )

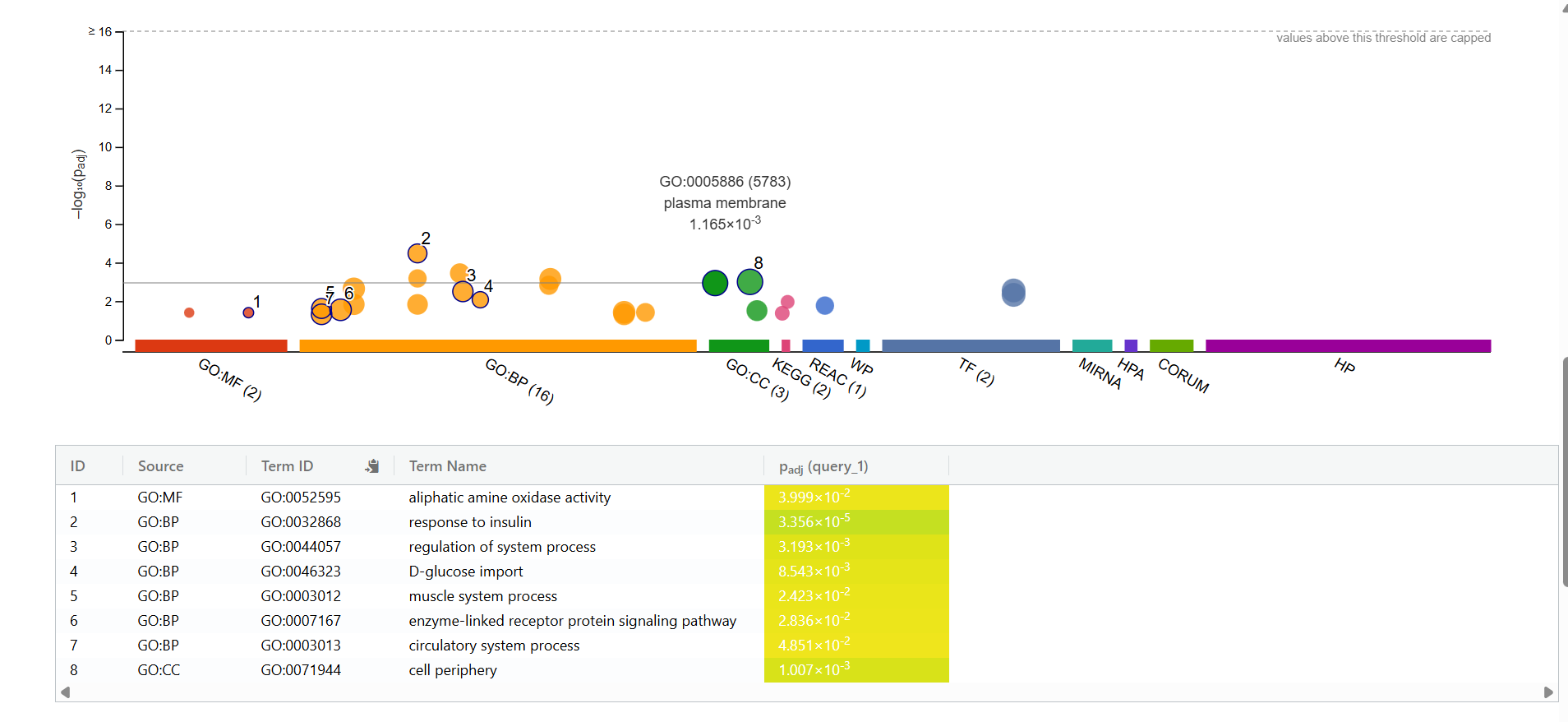
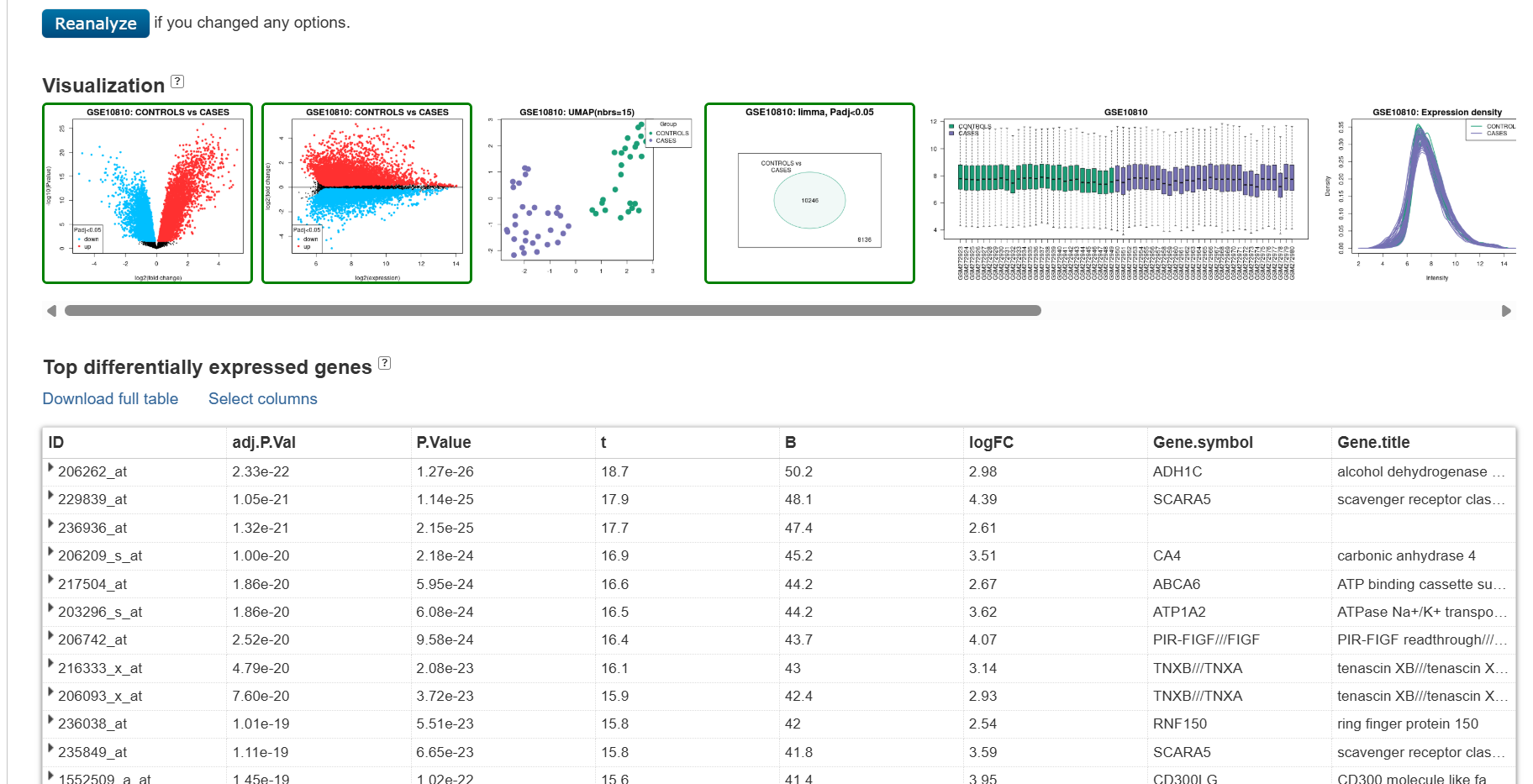
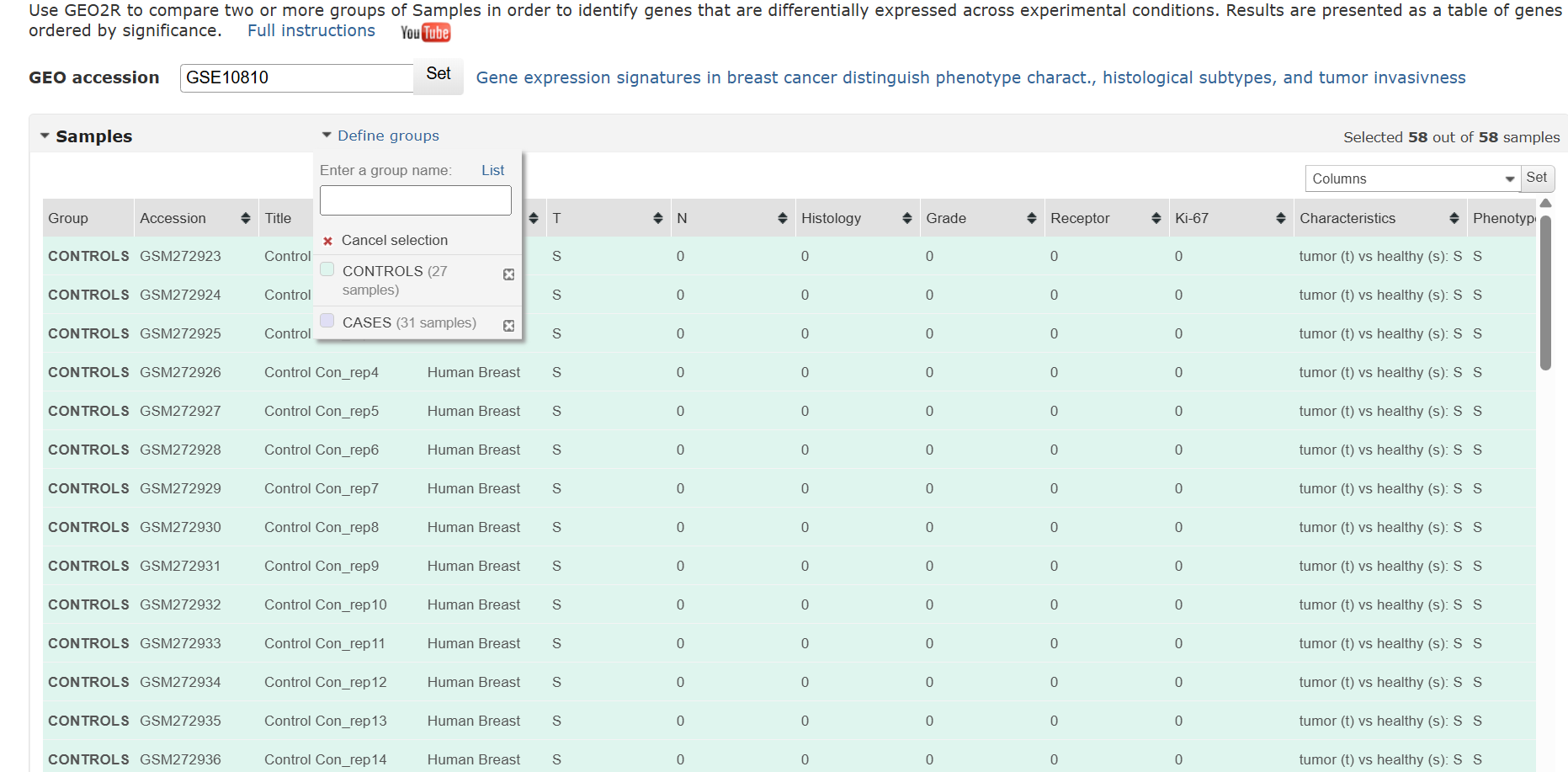
β) Πραγματοποιήστε ανάλυση διαφορικής έκφρασης (Differential Expression Analysis) για την μελέτη **GSE10810**, ορίζοντας σωστά τις ομάδες cases (ασθενείς) και controls (υγιείς) και επιλέγοντας την μέθοδο διόρθωσης **Benjamini & Hochberg (False discovery rate)**.Στην συνέχεια, κατεβάστε τα αποτελέσματα 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>)

**Α) Ανάλυση γονιδιακής έκφρασης με GEO2R και ανάλυση εμπλουτισμού με το gProfiler**

Στην παρούσα άσκηση πραγματοποιήσαμε ανάλυση διαφορικής γονιδιακής έκφρασης και λειτουργική ερμηνεία των αποτελεσμάτων με τα εργαλεία **GEO2R** και **gProfiler**.

1. **Ανάλυση διαφορικής έκφρασης με GEO2R**:
   1. Μεταβήκαμε στην επίσημη ιστοσελίδα του [GEO2R](https://www.ncbi.nlm.nih.gov/geo/geo2r/).
   2. Εισάγαμε τον κωδικό μελέτης **GSE10810**, η οποία αφορά δείγματα καρκίνου του μαστού.
   3. Ορίσαμε τις ομάδες:
      1. **Group 1 (Control)**: υγιή δείγματα.
      2. **Group 2 (Case)**: δείγματα ασθενών.
   4. Επιλέξαμε ως μέθοδο διόρθωσης για τα πολλαπλά tests την **Benjamini & Hochberg (FDR)**, ώστε να ελέγξουμε το ποσοστό ψευδώς θετικών αποτελεσμάτων.
   5. Εκτελέσαμε την ανάλυση και κατεβάσαμε το πλήρες αρχείο αποτελεσμάτων σε μορφή .txt ή .csv.
2. **Επιλογή Top διαφορικά εκφρασμένων γονιδίων**:
   1. Από τα αποτελέσματα της GEO2R, φιλτράραμε τα γονίδια με **p-value < 0.05**.
   2. Επιλέξαμε τα **100 κορυφαία γονίδια** με βάση την στατιστική σημαντικότητα και τα ταξινομήσαμε κατά αύξουσα p-value.
3. **Ανάλυση εμπλουτισμού με το gProfiler**:
   1. Μεταβήκαμε στο εργαλείο [gProfiler (g:GOSt)](https://biit.cs.ut.ee/gprofiler/gost).
   2. Εισάγαμε τη λίστα με τα 100 διαφορικά εκφρασμένα γονίδια στο πεδίο εισόδου.
   3. Επιλέξαμε ως οργανισμό το **Homo sapiens** και εκτελέσαμε την **ανάλυση εμπλουτισμού**.
   4. Το εργαλείο παρείχε λειτουργικές κατηγορίες εμπλουτισμού (π.χ. GO terms, KEGG pathways, Reactome), επισημαίνοντας βιολογικές διεργασίες και μονοπάτια στα οποία συμμετέχουν τα γονίδια αυτά.



**Β) Ανάλυση γονιδιακής έκφρασης με το Cyber-T και ανάλυση εμπλουτισμού με το gProfiler**

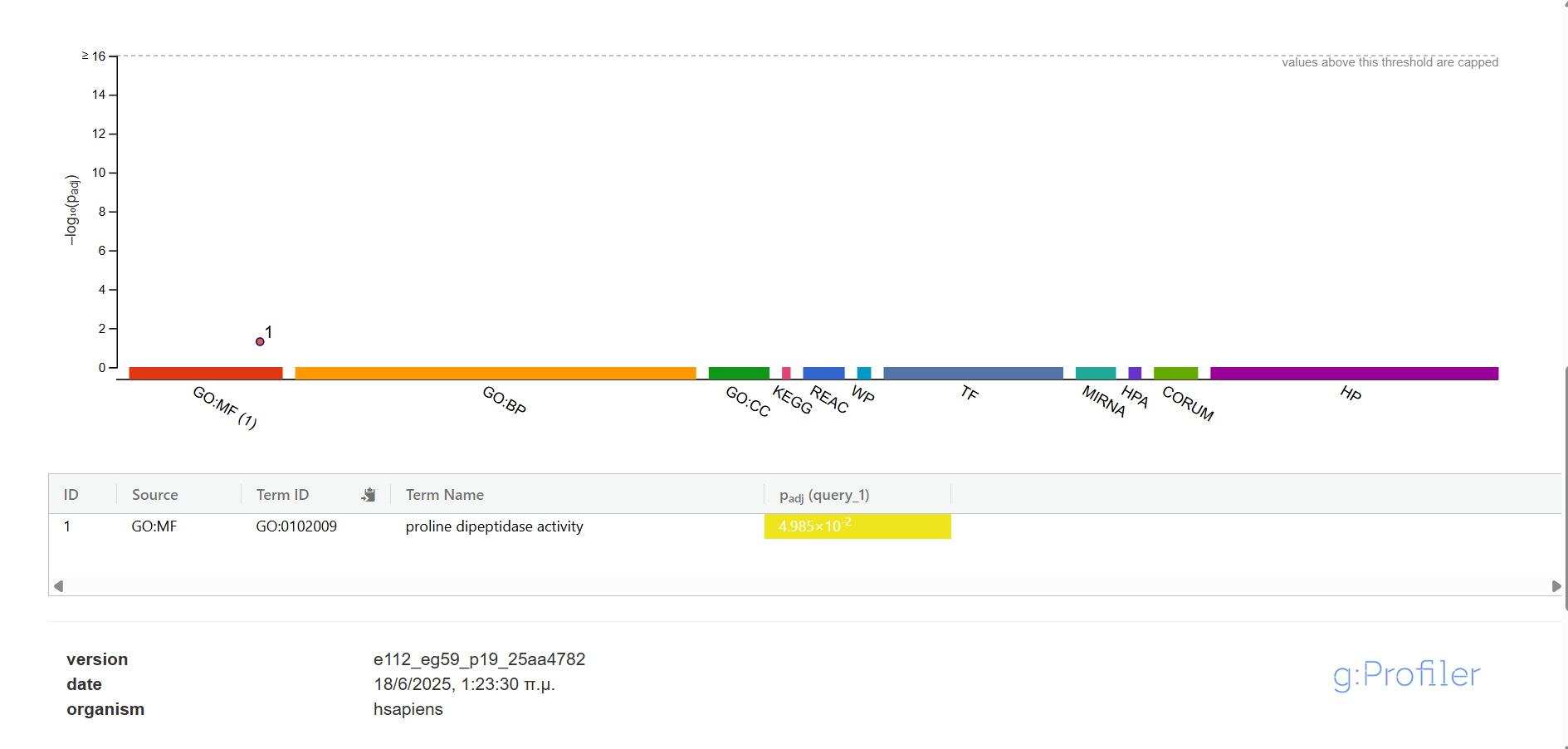
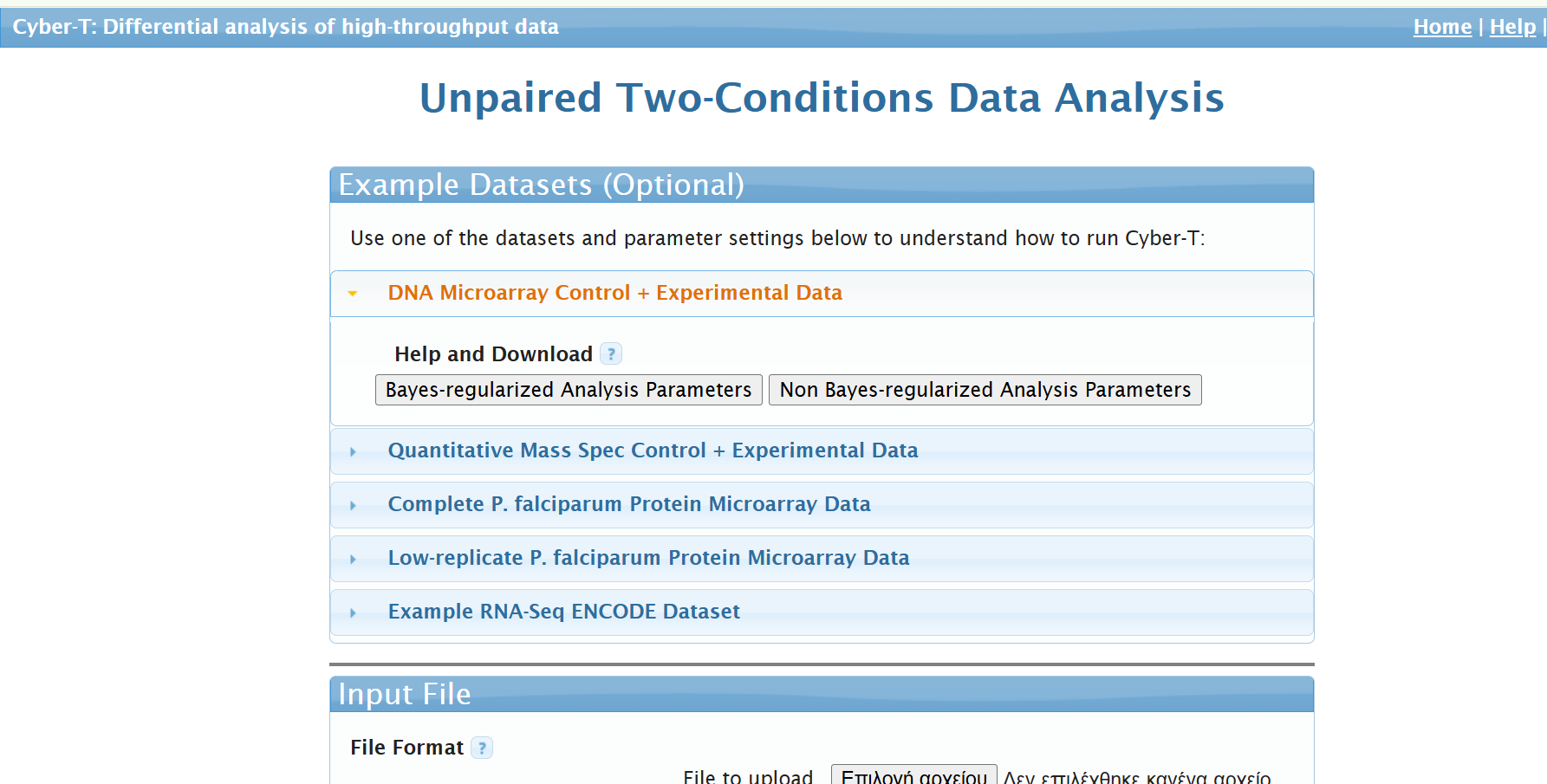
α) Μεταβείτε στην ιστοσελίδα του Cyber-T (<http://cybert.ics.uci.edu/>)

β) Χρησιμοποιήστε το example datasets όπως φαίνονται στην εικόνα και πραγματοποιήστε την ανάλυση με το Cyber-T

### **Β) Ανάλυση γονιδιακής έκφρασης με το Cyber-T και ανάλυση εμπλουτισμού με το gProfiler**

Σε αυτή την άσκηση χρησιμοποιήσαμε το διαδικτυακό εργαλείο **Cyber-T** για να πραγματοποιήσουμε ανάλυση διαφορικής έκφρασης γονιδίων και στη συνέχεια ερμηνεύσαμε τα αποτελέσματα μέσω **ανάλυσης εμπλουτισμού** με το **gProfiler**.

1. **Ανάλυση διαφορικής έκφρασης με το Cyber-T**:
   1. Μεταβήκαμε στην ιστοσελίδα του [Cyber-T](http://cybert.ics.uci.edu/), ένα εργαλείο μικροστοιχειοσειρών που εφαρμόζει bayesian στατιστική για την εύρεση διαφορετικά εκφρασμένων γονιδίων.
   2. Επιλέξαμε το διαθέσιμο **παραδειγματικό dataset** (example dataset) που παρέχει η πλατφόρμα.
   3. Ορίσαμε τις ομάδες σύγκρισης σύμφωνα με τις προκαθορισμένες συνθήκες του παραδείγματος (π.χ. treated vs control).
   4. Διατηρήσαμε τις προεπιλεγμένες παραμέτρους ανάλυσης (όπως Bayesian estimation, unequal variance κ.λπ.) και εκτελέσαμε την **Cyber-T ανάλυση**.
   5. Το εργαλείο παρήγαγε πίνακα αποτελεσμάτων, περιλαμβάνοντας μεταξύ άλλων: **gene names, mean expression values, fold change, p-values και false discovery rates (FDR)**.
   6. Κατεβάσαμε τα αποτελέσματα σε μορφή .txt ή .csv.
2. **Επιλογή σημαντικών γονιδίων και ανάλυση εμπλουτισμού**:
   1. Από τα αποτελέσματα, επιλέξαμε τα **γονίδια με p-value < 0.05**.
   2. Δημιουργήσαμε λίστα με τα **top διαφορετικά εκφρασμένα γονίδια** (ανάλογα με την κατάταξη ως προς το p-value ή το fold change).
   3. Μεταβήκαμε στην ιστοσελίδα του [gProfiler (g:GOSt)](https://biit.cs.ut.ee/gprofiler/gost) και εισάγαμε τη λίστα γονιδίων στο πεδίο εισαγωγής.
   4. Εκτελέσαμε **ανάλυση εμπλουτισμού** επιλέγοντας **Homo sapiens** ως οργανισμό.
   5. Το εργαλείο επέστρεψε εμπλουτισμένες λειτουργικές κατηγορίες (GO terms), μοριακές λειτουργίες, βιολογικές διεργασίες και συμμετοχή σε μονοπάτια (pathways), παρέχοντας πλήρη βιολογική ερμηνεία των αποτελεσμάτων.

**Γ) GWAS ανάλυση και μετα-ανάλυση με το PLINK και ανάλυση εμπλουτισμού με το gProfiler**

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_DEMO_data>

γ) Εκτελέστε την ακόλουθη εντολή στο PLINK για να πραγματοποιήσετε μετα-ανάλυση με τα τις μελέτες που κατεβάσατε στο προηγούμενο βήμα  
  
  
**Γ) GWAS Ανάλυση και Μετα-ανάλυση με το PLINK και Ανάλυση Εμπλουτισμού με το gProfiler**

Σε αυτή την άσκηση χρησιμοποιήσαμε το εργαλείο **PLINK** για να πραγματοποιήσουμε μετα-ανάλυση δεδομένων GWAS από τρεις διαφορετικές μελέτες. Ο σκοπός ήταν να εντοπίσουμε SNPs που σχετίζονται στατιστικά με το φαινότυπο που εξετάζεται.

Αρχικά, κατεβάσαμε την έκδοση 1.9 του PLINK για Linux από τον επίσημο ιστότοπο και του δώσαμε δικαιώματα εκτέλεσης μέσω τερματικού. Έπειτα, κατεβάσαμε τα αρχεία demo\_PLINK\_1.txt, demo\_PLINK\_2.txt και demo\_PLINK\_3.txt από το GitHub, και τα τοποθετήσαμε στον φάκελο εργασίας μας.

Για τη μετα-ανάλυση, χρησιμοποιήσαμε την εντολή:

bash

ΑντιγραφήΕπεξεργασία

./plink --meta-analysis demo\_PLINK\_1.txt demo\_PLINK\_2.txt demo\_PLINK\_3.txt logscale no-allele report-all

Η εντολή αυτή συνδυάζει τα δεδομένα από τις τρεις μελέτες, εντοπίζοντας SNPs που εμφανίζουν στατιστικά σημαντική συσχέτιση. Ελέγξαμε επίσης ότι τα αρχεία ήταν σε σωστή μορφή και encoding (UTF-8).

Στη συνέχεια, επιλέξαμε τα SNPs με p-value < 0.05 και τα εισάγαμε στο **gProfiler**, για να πραγματοποιήσουμε **ανάλυση εμπλουτισμού**. Εκεί πήραμε λειτουργικές πληροφορίες για τα σημαντικά SNPs, όπως συμμετοχή σε βιολογικές διεργασίες (GO terms), μονοπάτια, κ.λπ.

Με αυτή τη διαδικασία, μπορέσαμε να συνδυάσουμε αποτελέσματα από διαφορετικά datasets και να τα ερμηνεύσουμε βιολογικά.