

Εργαστηριακή Άσκηση 3

A) Ανάλυση γονιδιακής έκφρασης με το GEO2R και ανάλυση εμπλουτισμού με το gProfiler

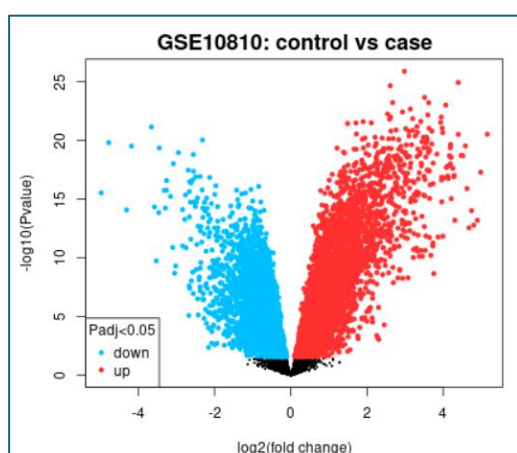
Μπαίνοντας στο GEO, κάνω αναζήτηση για την μελέτη GSE10810 και αφού τα δείγματα είναι σαφώς διαχωρισμένα (control-case):

GSM272948 Control Con_rep26
GSM272949 Control Con_rep27
GSM272950 Tumor Tum_Ph1_rep1
GSM272951 Tumor Tum_Ph1_rep2

Κλπ,

πατάω Analyze with GEO2R, το οποίο βρίσκεται κάτω από τα citations και ανακατευθύνομαι στο GEO2R. Ορίζω τις ομάδες control και case και βάζω το κάθε γονίδιο στην κατάλληλη:

Αφού κάνω Analyze και βγουν τα αποτελέσματα της ανάλυσης της διαφορικής έκφρασης, τα κατεβάζω πατώντας Download full table, που βρίσκεται ακριβώς πάνω από τον πίνακα.

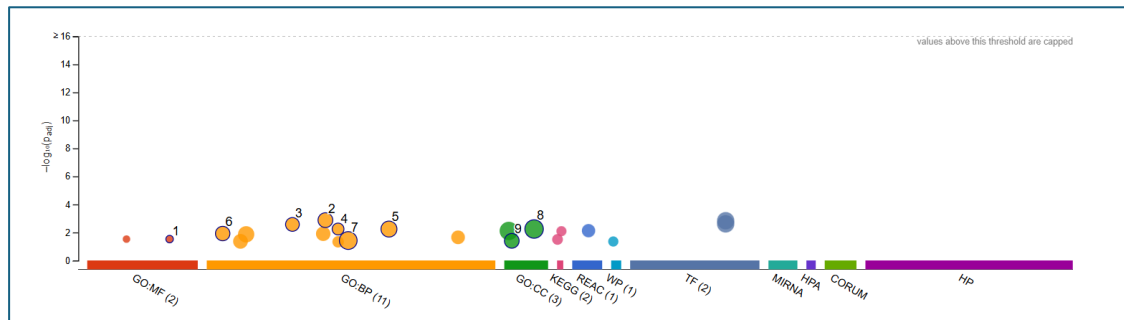


- Με μαύρο χρώμα φαίνονται τα ασήμαντα γονίδια, όπου $p\text{-value} > 0.05$.
- Με κόκκινο είναι οι κουκίδες των γονιδίων που υπερεκφράζονται, είναι δηλαδή σημαντικά αυξημένα.
- Με μπλε είναι οι κουκίδες των γονιδίων που υποεκφράζονται, είναι δηλαδή σημαντικά μειωμένα.

Επομένως, η ύπαρξη μεγάλου αριθμού διαφοροποιημένων γονιδίων υποδηλώνει ότι η υπό μελέτη βιολογική κατάσταση επηρεάζει σημαντικά την έκφραση των γονιδίων.

Περνάω στη συνέχεια το αρχείο στο Excel, αφαιρώ τις ασαφείς ή άδειες καταχωρήσεις και αντιγράφω τις 100 πρώτες καταχωρήσεις, αφού τα δεδομένα είναι ήδη ταξινομημένα από το πιο σημαντικό γονίδιο προς το λιγότερο σημαντικό, με βάση το

p-value και adj.p.val. Πηγαίνω στο gProfiler, κάνω επικόλληση τη λίστα με τα γονίδια αυτά στο λευκό πλαίσιο, πατάω Run query και βγαίνουν τα αποτελέσματα της ανάλυσης εμπλουτισμού.



Κάθε κύκλος στο παραπάνω διάγραμμα είναι ένας εμπλουτισμένος όρος, στους σημαντικότερους των οποίων είναι αναγεγραμμένος ο αριθμός των γονιδίων που περιέχουν και αναλύονται στον πίνακα παρακάτω. Το ύψος του κύκλου δηλώνει το πόσο στατιστικά σημαντικός είναι ένας όρος και το μέγεθός του το πόσα γονίδια περιέχει.

ID	Source	Term ID	Term Name	$p_{\text{adj}}(\text{query}_1)$
1	GO:MF	GO:0052595	aliphatic amine oxidase activity	2.987×10^{-1}
2	GO:BP	GO:0044057	regulation of system process	1.328×10^{-1}
3	GO:BP	GO:0032868	response to insulin	2.688×10^{-1}
4	GO:BP	GO:0046323	D-glucose import	5.699×10^{-1}
5	GO:BP	GO:0071495	cellular response to endogenous stimulus	5.734×10^{-1}
6	GO:BP	GO:0003012	muscle system process	1.199×10^{-2}
7	GO:BP	GO:0048731	system development	3.773×10^{-2}
8	GO:CC	GO:0071944	cell periphery	5.694×10^{-1}
9	GO:CC	GO:0015629	actin cytoskeleton	3.981×10^{-2}

Ο πίνακας αυτός λοιπόν, εμφανίζει τα σημαντικότερα αποτελέσματα εμπλουτισμού. Μας δίνει την κατηγορία του γονιδίου, τον κωδικό της λειτουργίας του, τη λειτουργία και το p-value του.

Επομένως, η ανάλυση εμπλουτισμού αποκαλύπτει στατιστικά σημαντικό εμπλουτισμό σε διάφορες βιολογικές διεργασίες, μοριακές λειτουργίες και κυτταρικά συστατικά. Αυτό σημαίνει ότι η μελετώμενη βιολογική κατάσταση επηρεάζει κρίσιμες κυτταρικές και συστημικές λειτουργίες, με έμφαση στον μεταβολισμό και την κυτταρική δομή.