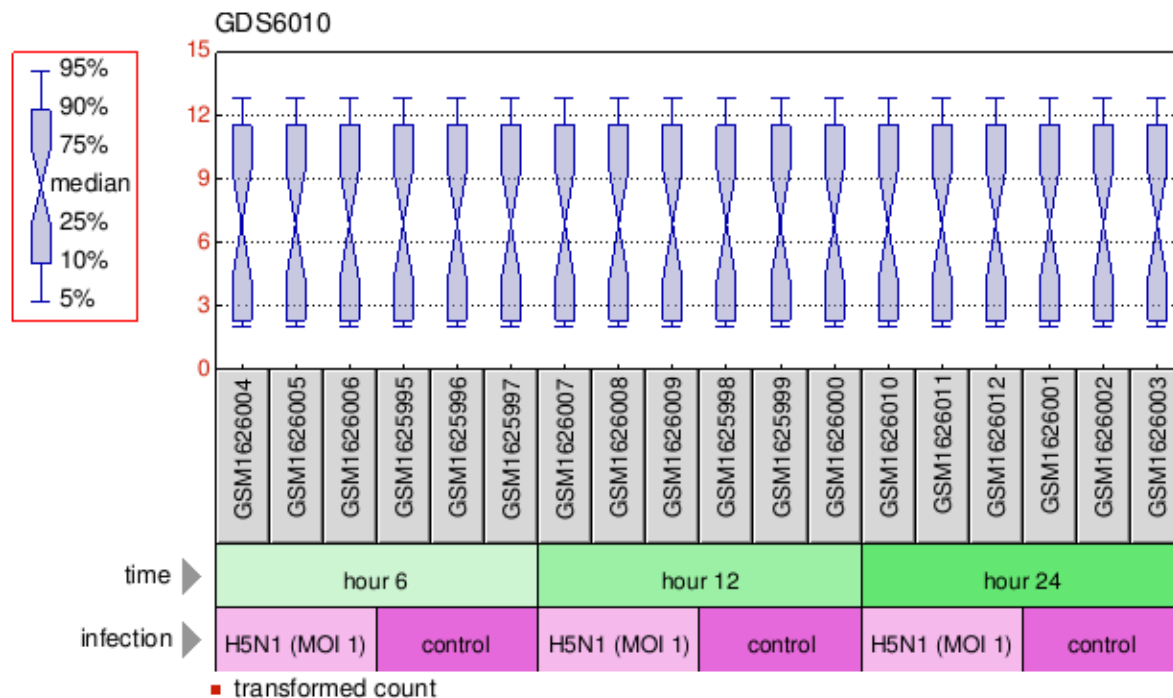


To dataset GDS6010 μετράει τις εκφράσεις γονιδίων σε άτομα προσβλημένα από influenza virus H5N1 και υγιή άτομα σε astrocytes U251 (cell lines). Η μέτρηση έγινε σε τρεις χρονικές στιγμές: 6, 12 και 24 ώρες μετά τη μόλυνση. Για τις αντίστοιχες μετρήσεις έχουμε και τα controls. Το παρακάτω διάγραμμα δείχνει τον πειραματικό σχεδιασμό.



Το link για το dataset είναι εδώ

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS6010>

1. Κατεβάστε το dataset και φορτώστε το στην R
2. Χρησιμοποιήστε μόνο τα δείγματα που αντιστοιχούν στις 6 και στις 24 ώρες και χρησιμοποιήστε αυτές τις κατηγορίες ως παράγοντα (factor) με 2 levels: 6 και 24.
3. Βρείτε ποια γονίδια διαφοροποιούνται μεταξύ των 6 και 12 ωρών παίρνοντας υπόψιν και τον παράγοντα infection (γραμμικά μοντέλα με 2 παράγοντες με αλληλεπίδραση).
4. Θεωρείτε ότι η μόλυνση με H5N1 προκαλεί συστηματική αντίδραση στον οργανισμό (δηλαδή επηρεάζονται πολλά γονίδια) ή επηρεάζονται μόνο λίγα γονίδια;
5. Χρησιμοποιήστε την lm συνάρτηση ώστε να δείτε σε ένα μοντέλο με αλληλεπίδραση ποιοι παράγοντες είναι σημαντικοί.
6. Βρείτε ποια γονίδια επηρεάζονται μεταξύ control και H5N1 παίρνοντας υπόψιν τις ώρες μετά τη μόλυνση (πάρτε όλα τα δείγματα και χρησιμοποιήστε τον χρόνο όχι σαν ποσοτική μεταβλητή αλλά σαν ποιοτική. πχ όπως το φύλο, αλλά με 3 επίπεδα).
7. Για τα 100 γονίδια με το μικρότερο p-value, κατεβάστε τους υποκινητές τους με την biomaRt και βρείτε αν υπάρχει κάποιο TFBS (transcription factor binding site) που

υπεραντιπροσωπεύεται έναντι ενός τυχαίου συνόλου 1000 γονιδίων (γι αυτό πάαρτε 1000 τυχαία γονίδια από το microarray πείραμα και βρείτε τους υποκινητές τους).

8. *** bonus *** για το γονίδιο με το μικρότερο p-value βρείτε πόσα γονίδια συνεκρφάζονται με αυτό (χρησιμοποιώντας τη συνάρτηση cor, μπορείτε να υπολογίσετε το συντελεστη συσχέτισης μεταξύ 2 vector). Επαναλάβετε την ανάλυση 7 γι αυτά τα γονίδια ώστε να βρείτε τους TFBS.