Τα διαγράμματα δημιουργήθηκαν με τη συνάρτηση enrichrVisual, η οποία παίρνει ως είσοδο τον πίνακα δεδομένων [data.visual](#data_visual), τον φάκελο αποθήκευσης των διαγραμμάτων και τον όρο με τον οποίο έγινε εμπλουτισμός (παράδειγμα: “GO MF Terms”) για την προσθήκη των κατάλληλων τίτλων στα διαγράμματα.

1) Density plot-P values of GO MF Terms

* κατανομές των τιμών των στηλών P.value, P.adjust.

2) Top GO MF Terms

* μέτρηση το πόσες φορές (n) εμφανίζεται κάθε όρος (GO MF Term) στον πίνακα data.visual
* συνένωση της πληροφορίας με τον πίνακα data.visual
* επιλογή των στηλών: Term, n, P.adjust
* ομαδοποίηση με βάση τις στήλες Term, n και summarise το P.adjust ώστε να είναι ο μέσος όρος των διαφορετικών τιμών που εμφανίζονται στα διαφορετικά TADs
* ταξινόμηση του πίνακα με φθίνουσα σειρά ως προς τη στήλη n
* αν οι γραμμές του πίνακα ξεπερνούν τις 30, τότε επιλογή των 30 πρώτων γραμμών και αναφορά αυτής της επιλογής ως υπότιτλο στο γράφημα
* δημιουργία ιστογράμματος, στο οποίο ο άξονας x αντιπροσωπεύει τη μέση τιμή P.adjust, ο άξονας y τους όρους GO MF Terms και το χρώμα των ράβδων τη μεταβλητή n.

3) GO MF Terms in different TADs

* χρήση των στηλών Term, n του πίνακα που δημιουργήθηκε για το προηγούμενο γράφημα
* δημιουργία ιστογράμματος, στο οποίο ο άξονας x αντιπροσωπεύει τον αριθμό των TADs, στα οποία βρέθηκε ο κάθε όρος και ο άξονας y τους όρους GO MF Terms.

4) Top 10 GO MF Terms network graph

* μέτρηση το πόσες φορές (n) εμφανίζεται κάθε όρος (GO MF Term) στον πίνακα data.visual
* επιλογή των δέκα GO MF Terms με τις μεγαλύτερες τιμές n
* συνένωση με τον πίνακα data.visual
* επιλογή των στηλών Term, TAD
* δημιουργία γράφου, στον οποίο οι κόμβοι αντιστοιχούν στα TADs (κάθε TAD αντιστοιχεί σε έναν κόμβο), και οι ακμές συνδέουν τους κόμβους που ανήκουν στην ίδια ομάδα, δηλαδή GO MF Term. Σε κάθε ομάδα αντιστοιχεί διαφορετικό χρώμα ακμών.

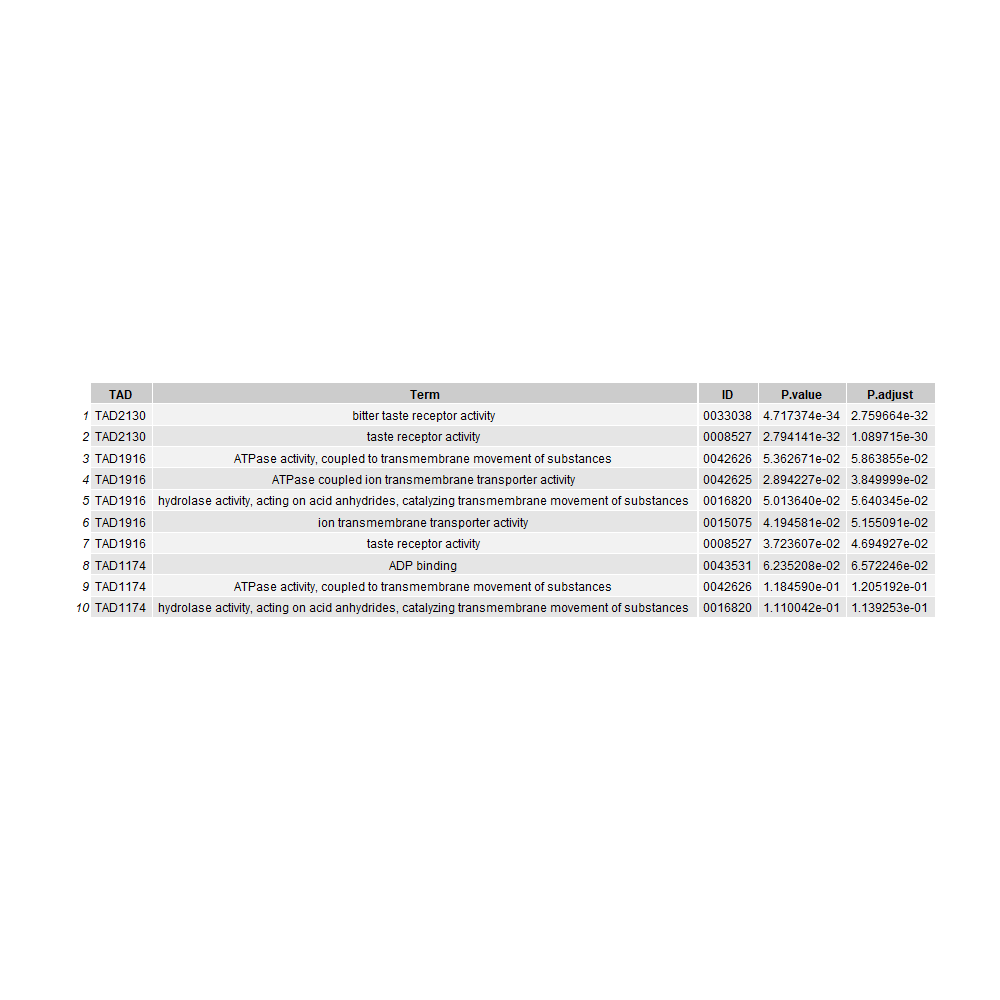
5) Adjusted P values per TAD

* ομαδοποίηση του πίνακα data.visual ανά TAD
* για κάθε TAD:
  + ταξινόμηση των γραμμών κατά φθίνουσα σειρά σύμφωνα με τη στήλη P.adjust
  + αν οι γραμμές του πίνακα ξεπερνούν τις 30, τότε επιλογή των 30 πρώτων γραμμών και αναφορά αυτής της επιλογής ως υπότιτλο στο γράφημα
  + δημιουργία ιστογράμματος για κάθε TAD στο οποίο ο άξονας y αναπαριστά τους GO MF Terms που βρέθηκαν στο TAD και ο άξονας x την πιθανότητα P.adjust που αντιστοιχεί σε κάθε όρο.

6) Adjusted P values per GO MF Terms

* ομαδοποίηση του πίνακα data.visual ανά GO MF Term
* για κάθε Term:
  + ταξινόμηση των γραμμών κατά φθίνουσα σειρά σύμφωνα με τη στήλη P.adjust
  + αν οι γραμμές του πίνακα ξεπερνούν τις 30, τότε επιλογή των 30 πρώτων γραμμών και αναφορά αυτής της επιλογής ως υπότιτλο στο γράφημα
  + δημιουργία ιστογράμματος για κάθε GO MF Term στο οποίο ο άξονας y αναπαριστά τα TADs στα οποία βρέθηκε ο όρος και ο άξονας x αναπαριστά την πιθανότητα εύρεσης P.adjust του όρο σε κάθε TAD.

**Βήματα για την παραγωγή του πίνακα εισόδου data.visual της συνάρτησης enrichrVisual**

****

Στιγμιότυπο του πίνακα data.visual.

**-enrich all**

Μεθοδολογία εμπλουτισμού – enrichAll

* χρήση της συνάρτησης enrichR με εισόδους:
  + τη στήλη Gene\_id των δεδομένων εισόδου (integrated\_table\_with\_sign\_tads.csv / έξοδος της TADiff.R)
  + τις βάσεις δεδομένων σύμφωνα με τις οποίες θα γίνει ο εμπλουτισμός GO\_Molecular\_Function\_2018 , GO\_Biological\_Process\_2018 , KEGG\_2019\_Human
* η συνάρτηση επιστρέφει μία λίστα με τρεις πίνακες, ο καθένας με 9 στήλες (Term, Overlap, P.value, Adjusted.P.value, Old.P.value, Old.Adjusted.P.value, Odds.Patio, Combined.Score, Genes) και γίνεται επιλογή των αποτελεσμάτων με P.value < 0.05

Ανάλυση δεδομένων και υπολογισμός p value – analysisAll, calculatePvalue, produceOutputs

* η ανάλυση γίνεται ξεχωριστά για κάθε πίνακα με τα αποτελέσματα του εμπλουτισμού από τη συνάρτηση enrichAll
* επιλογή των στηλών Term, Overlap,Genes
* η στήλη Overlap είναι τύπου character και έχει τη μορφή: αριθμητής/παρονομαστής, όπου ο αριθμητής αντιπροσωπεύει τον αριθμό γονιδίων από τα δεδομένα εισόδου που αντιστοιχούν σε αυτό τον όρο (αυτά αναγράφονται αναλυτικά στη στήλη Genes) και ο παρονομαστής το συνολικό αριθμό γονιδίων του ανθρώπινου γονιδιώματος που αντιστοιχούν σε αυτό τον όρο.
* γίνεται χωρισμός της στήλης σε δύο επιμέρους στήλες numerator και denominator, ώστε να υπάρχει σε ξεχωριστή στήλη διαθέσιμη η πληροφορία για τον αριθμό των γονιδίων που αντιστοιχούν σε κάθε όρο στη βάση δεδομένων
* χωρισμός της στήλης Genes, ώστε σε κάθε γραμμή να βρίσκεται ένα γονίδιο και συνένωση στις στήλες των γονιδίων (Gene\_id == Genes) με τον πίνακα data.with.genes, ο οποίος έχει παραχθεί από τα δεδομένα εισόδου πριν τον εμπλουτισμό και περιέχει δύο στήλες: Gene\_id, tad\_name. Ως αποτέλεσμα, γίνεται γνωστό σε ποιό/ποιά TADs συναντάται κάθε όρος και σε ποια γονίδια οφείλεται αυτό
* έπειτα εκτελείται hypergeometric test for over-represented terms και υπολογισμός p value για κάθε όρο και σε κάθε TAD με τη χρήση της συνάρτησης phyper(hitInSample-1,hitInPop, failInPop,sampleSize, lower.tail=FALSE)
* hitInSample = αριθμός των γονιδίων του TAD που αντιστοιχούν στον όρο (με χρήση της συνάρτησης count()
* hitInPop = ο αριθμός των γονιδίων που αντιστοιχούν στον όρο στη βάση δεδομένων (η τιμή που αντιστοιχεί στη στήλη denominator και υπολογίστηκε πριν από το Overlap)
* failInPop = Ο συνολικός αριθμός των γονιδίων που καλύπτει η βάση – hitInPop
* sampleSize = Ο αριθμός των γονιδίων από τα δεδομένα εισόδου που αντιστοιχούν σε αυτό το TAD
* και τέλος προσαρμογή p values με τη μέθοδο “fdr”