

Descàrrega de les dades

Mitjançant la eina *Bash* de l'ordinador he creat una carpeta connectada amb git anomenada **dadesomiques**. En aquesta carpeta inclouré els projectes en els que vagi treballant al llarg del curs.

Posteriorment he clonat les dades d'interès i he creat un projecte a R studio connectat amb git. Per a fer-ho, s'ha d'obrir un nou projecte que es connectarà amb git mitjançant la url obtinguda del projecte a git.

Creació del contenidor

Per a la creació del contenidor del tipus *SummarizedExperiment* s'han de transposar les files per les columnes de les dades a analitzar. D'aquesta manera romandran les mostres en columnes i les files contindran la informació de cadascuna d'elles (Accession number, origen de la mostra...). No es pot transposar la matriu de dades obtinguda directament ja que hi ha columnes no numèriques.

Primerament, he afegit una columna més que indica el número de pèptid (en total 1438).A continuació, he obtingut cada columna en format llista. Un cop obtingudes les llistes, les he unit per formar un dataframe on les columnes són les diferents mostres i les files les característiques d'aquestes mostres. D'aquestes dades la part més rellevant és el valor de intensitat de cada pic de pèptid fosforil.lat, que ja es troba corregit convenientment.

El contenidor l'he guardat en la extensió rda al repositori del projecte.

Exploració de les dades

Cada tumor s'ha intectat en tres ratolins diferents i s'han extret duplicats de les mostres obtingudes en cadascun dels 3. És important veure si hi ha diferència en cadascun dels duplicats, per tant es realitza un test estadístic t-student de dades aparellades per detectar si hi ha diferències en les mitgens dels duplicats.

En els duplicats 1-2,9-10 i 11-12 no hi ha diferències significatives entre ells, per tant serien uns duplicats vàlids, tot i que faltaria calcular la desviació típica d'aquests duplicats. En els altres duplicats si hi ha diferències significatives respecte als duplicats.

Respecte a la normalitat de les dades, cap de les mostres segueix un patró de normalitat, com s'infereix del test Lillie aplicat a totes les mostres.

La pregunta biològica seria si existeix diferències en la expressió de fosfopèptids en els dos tumors que s'han estudiat.

Les dades que han proporcionat els investigadors han estat normalitzades amb pèptoids no fosforil.lats control.

Per manca de temps, no puc aprofundir en les diferències entre els dos subtipus de tumors, però a priori observant el boxplot es podria inferir que si hi ha un patró de pèptids que es produiria amb més quantitat en el subtipus dos respecte al subtipus 1. Per tant valdria la pena repetir les condicions experimentals millorant la obtenció de les dades i obtenint dades més homogènies que permetin el seu estudi amb més fiabilitat.

Reposició de dades en github

https://github.com/mgloriasf/phospho_project.git

Bibliografia

Morgan M, Obenchain V, Hester J, Pagès H (2024). *SummarizedExperiment: A container (S4 class) for matrix-like assays*. R package version 1.36.0, <https://bioconductor.org/packages/SummarizedExperiment>.