# 440\_211\_PubMed

Primera proba:

* Datos crudo no PRIDE, carpeta TXT que é saída de MaxQuant.
* Seleccionamos ProteinGroup.txt, onde hai 4 grupos chamados Group1, Group 2, Group 3, Group 4. Non sabemos a correspondencia destes grupos cos grupos de estudo.
* Analizamos os datos con Perseus como di no artículo:
  + Seleccionamos LFQ intensities
  + Eliminamos as proteínas identificadas coa base de datos de proteínas reversas (reverse).
  + Conservamos as proteínas que aparecen mínimo en tres mostras de polo menos un dos 3 grupos de estudo.
* Imputamos dunha distribución normal con width 0.3 e shift 1.8.
* Facemos test t entre todos os grupos a pares xa que non sabemos cal é cal, corrixindo por BH.
* Non coinciden os resultados cos indicados no artículo, o que pode deberse á imputación que non se pode replicar cos datos do artículo, se non se indica a semente usada ao mostrear aleatoriamente da distribución tampoco se vai a lograr a réplica perfecta.

Resultados obtidos das comparacions:

* Grupo 1 vs Grupo 2: 0 DEPs
* Grupo 1 vs Grupo 3: 0 DEPs
* Grupo 1 vs Grupo 4: 2 DEPs
* Grupo 2 vs Grupo 4: 1 DEPs
* Grupo 3 vs Grupo 4: 39 DEPs

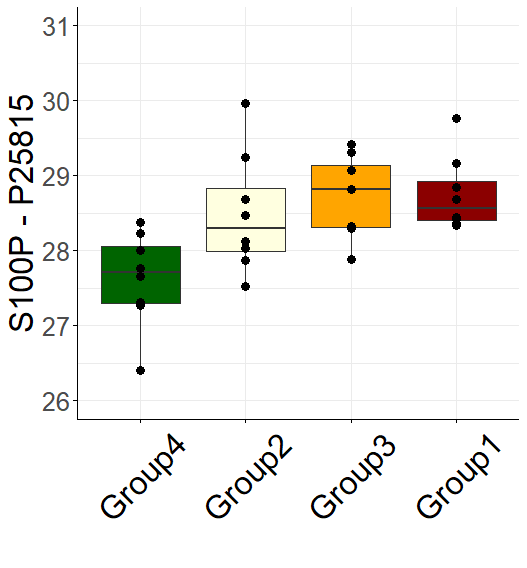
Polos resultados, ten pinta de que o control é o Grupo 4. Ainda así, moi diferentes aos do artículo.

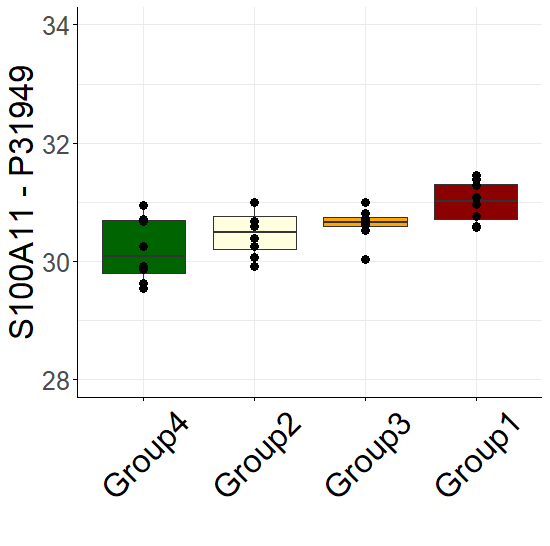
Segundo intento

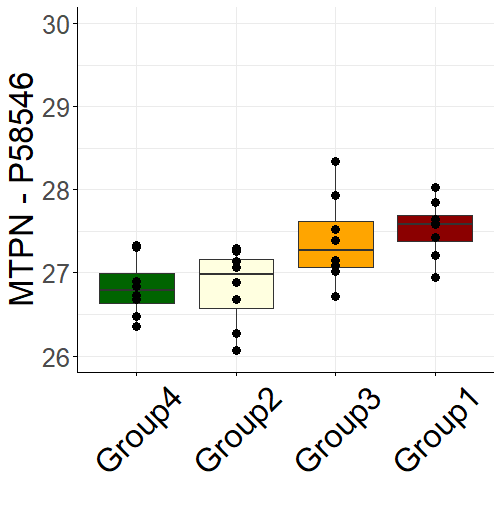
* Quedámonos soamente coas proteínas con datos de LFQ intensity completos (140).
* Tanto en Perseus como R, repetimos os diferentes test t e vemos que concorda por completo os resultados.
* Despois, seleccionamos dentro das 140 proteínas con datos completos, aquelas que aparezan representadas na figura suplementaria 4.
* Facendo os boxplots da LFQ intensity (con transformación logarítmica) nas proteínas común e comparando estos boxplots cos da figura S4 lógrase identificar a correspondencia entre os grupos:
  + Control = Grupo 4
  + T2DM = Grupo 2
  + NPDR = Grupo 3
  + PDR = Grupo 1

A continuación, temos as representación:

GRAPH FROM FIGURE S4 vs GRAPH CREATED WITH R USING DATA IN proteinGroups.txt

O resto das proteínas (faltan 4) atópanse na seguinte imaxe (datos do artículo) ou ben en R>3.Extraction>1\_Estimations.R (datos do proteinGroups.txt



Nova estratexia de análise:

* Seleccionar LFQ intensities, eliminar as reverse, quedarnos soamente con proteínas que teñan datos completos en grupo control e grupo T2DM (máximo 3 missing values permitidos por grupo).
* Transformación logarítmica e non imputar
* Estimar test-t, BH e logFC.