

FOSFOPÈPTIDS COM A BIOMARCADOR DE DIFERENCIACIÓ DELS SUBTIPUS TUMORALS MSS I PD: Informe d'anàlisi

L'anàlisi s'ha dut a terme amb el conjunt de dades TIO2+PTYR-Human-MSS+MSI vs PD.xlsx estret del *repositòri de github* [metaboData/Datasets/2018-Phosphoproteomics](https://github.com/metaboData/Datasets/2018-Phosphoproteomics).

1. PREGUNTA BIOLÒGICA:

L'experiment pretén respondre si existeixen diferències a nivell dels fosfopèptids específics en els subtipus tumorals MSS (*Microsatellite Stable*) i PD (*Poorly Differentiated*). I si és el cas saber que indiquen aquests fosfolípids sobre les vies de senyalització diferenciades en cada grup.

D'aquesta manera el principal objectiu seria identificar fosfopèptids específics que permetin diferenciar entre els subgrups tumorals derivats de pacients PDX (*Patient-Derived Xenografts*).

2. DISSENY EXPERIMENTAL

L'estudi es planteja com un estudi comparatiu entre els dos subgrups tumorals utilitzant mostres PDX.

S'han seleccionat dos grups de models PDX que es representen de la següent manera:

- MSS: Mostres M1, M5 i T49
- PD: Mostres M42, M43 i M64

Les mostres es van estudiar per duplicat el que va resultar en un total de 12 mostres (3+3)x2.

Les mostres van ser enriquides amb TIO2 i amb PTYR i van ser sotmeses a un anàlisi de fosfoproteòmica mitjançant cromatografia líquida i espectrofotometria de masses (LC-MS).

3. OBTENCIÓ DE LES DADES,

Les dades crues es van obtenir a partir de les abundàncies extretes de la LC-MS. Aquest conjunt correspon al nombre de mostres anteriors correlacionades amb el nivell d'abundància de 1400 fosfopèptids trobats en alguna de les mostres.

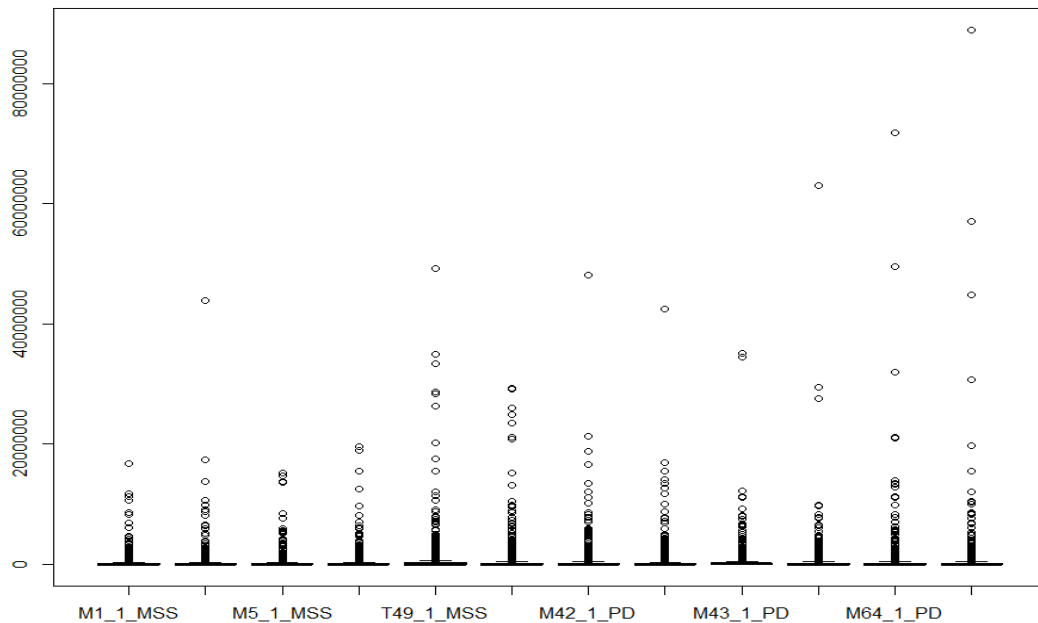
Aquestes van ser llegides a partir del document .XLSX anomenat anteriorment i es va realitzar una primera visualització de les dades. Aquestes comptaven amb un total de 1438 observacions amb 18 variables.

Amb l'objectiu de poder estructurar les dades explorades de manera més òptima i evitar així risc d'errors es va crear un objecte contenidor mitjançant *SummarizedExperiment* de *BiocManager*, el qual contindrà de manera separada les dades de les abundàncies (12 variables) i les metadades de l'experiment (6 variables).

L'objecte contenidor resultant va ser guardat en format binari .Rda i pujat al repositori *github* de l'estudi (punt 6).

4. CONTROL DE QUALITAT I NORMALITZACIÓ:

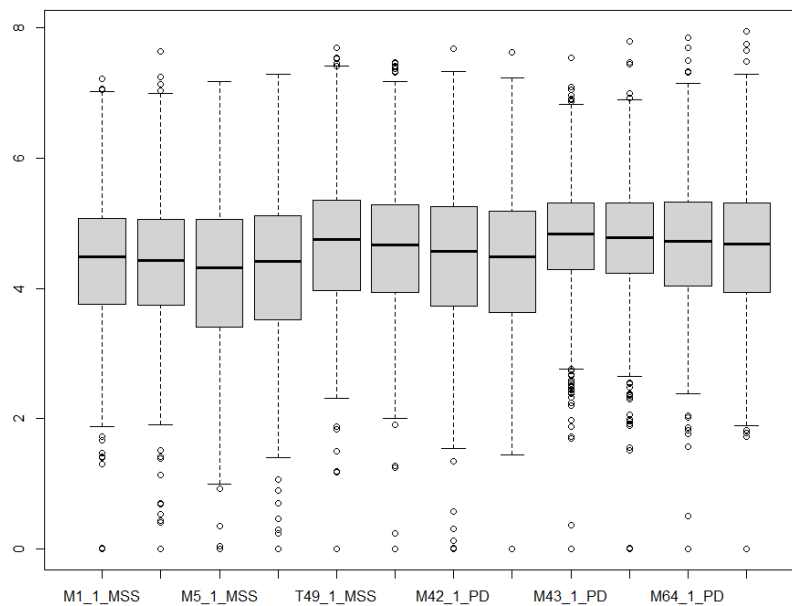
Posteriorment a l'obtenció de les dades i a la creació de l'objecte contenidor es va realitzar un control de qualitat d'aquestes dades a nivell d'anàlisi visual mitjançant un *boxplot* (il·lustració 1).



Il·lustració 1: Gràfic de caixes de les dades de l'estudi en cru

Aquest, juntament amb l'anàlisi d'*outliers* mostrava que existien 15 punts que podien resultar de risc i per això es va aplicar una logarimització en base 10 de les dades ja que aquestes tendeixen a valors propers a zero i ens anàlisi estadístics posteriors poden ser influenciats per aquest fet i estudiar si varien aquests *outliers*.

Per tant un cop realitzar aquest anàlisi s'aplica una transformació i es repeteix el boxplot (il·lustració 2) obtenint el següent resultat:



Il·lustració 2: Gràfic de caixes de les dades d'estudi amb transformació Log10

Amb la transformació logarítmica es pot observar amb més detall les caixes i els bigotis de cadascuna de les distribucions de dades de les mostres. Tot hi això es repeteix l'anàlisi dels *outliers* i segueixen apareixent punts de risc.

5. Anàlisi estadístic

Finalment es va decidir tirar endavant amb els *outliers* trobats en l'apartat anterior i procedir a realitzar l'anàlisi estadístic.

Aquest es va realitzar utilitzant l'objecte creat i el paquet de BiocManager Limma. Es va utilitzar aquest paquet per aplicar un model lineal a les dades i identificar els fosfopèptids més presents en cada mostra. Durant l'ajustament del model s'aplica un ajust bayesià empíric per tal de millorar la precisió dels resultats, ja que en el nostre cas tenim un gran nombre de dades i en aquest casos es recomana utilitzar aquest ajust.

D'altra banda a l'anàlisi s'ajusta el p-valor obtingut per tal d'evitar risc d'obtenció de falsos positius als resultats, degut també al gran nombre de dades.

El resultat de l'anàlisi ha estat el següent:

Taula 1: Resultats de l'anàlisi estadístic realitzat amb limma i aplicant un model lineal.

	mainEff	AveExpr	F	P.Value	adj.P.Val
O00193.1	802048.92	641359.81	34.87586	0.0001157082	0.1663884
P35269.1	-435826.87	412752.57	23.39308	0.0005694049	0.2040053
P16144.3	313748.10	270927.30	22.75343	0.0006322067	0.2040053
Q5T0Z8.1	225849.08	176503.85	21.01902	0.0008488023	0.2040053
Q9H1C7	677600.13	446716.73	20.98220	0.0008542838	0.2040053
P84103	27775.59	18777.87	20.70335	0.0008971889	0.2040053

La taula 1 mostra els 6 fosfopèptids més destacats a les mostres i l'efecte en cadascun d'ells entre les mostres estudiades, el qual pot suggerir una diferència entre els grups.

En aquest cas el P-valor ajustat es > 0.05 el que significa que no es pot afirmar que existeixen diferències entre els grups estudiats pel que fa a l'abundància de cada fosfopèptid.

6. Resposta a la pregunta biològica

Després de realitzar l'anàlisi de les dades de l'experiment, l'estudi ha permès observar que amb les dades recollides no existeix diferència entre els subgrups tumorals MSS i PD i que per tant no es possible diferenciar fosfopèptids que ajudin a diferenciar els dos subtipus tumorals.

7. Creació del repositori Github

Un cop acabat l'anàlisi i la redacció d'aquest informe s'ha creat un *repositori de github* per tal de recopilar tot el material utilitzar per dur a terme l'estudi. Aquest es compon de:

- Informe d'anàlisi
- Objecte contenidor amb les dades i metadades (format .Rda)
- Codi R utilitzat (format .txt)
- Metadades (format .md)

Al repositori s'hi pot accedir mitjançant l'enllaç següent:

https://github.com/Nillopez/L-pep-Gil_Nil-PEC1.git