

Metadades del conjunt de dades *dades_fosforilacio*

Introducció

El conjunt de dades *dades_fosforilacio* es va crear mitjançant espectrometria de masses per identificar i quantificar modificacions post-traduccionals (PTMs) en proteïnes. Aquestes modificacions són essencials per a la regulació de processos biològics, i el seu estudi pot aportar informació important sobre malalties com el Parkinson.

Descripció general del conjunt de dades

Aquest conjunt de dades inclou informació sobre les proteïnes i les seves modificacions en dues condicions experimentals:

- *MSS*: Condicions que simulen un entorn sanguini saludable.
- *PD*: Condicions que simulen la malaltia de Parkinson.

La taula conté 1.438 files (cada una representant una observació) i 18 columnes que inclouen metadades de les proteïnes i les intensitats de les modificacions observades en les diferents mostres.

Principals variables del conjunt de dades

Les variables més rellevants inclouen:

1. *SequenceModifications*: Tipus de modificació en la seqüència d'aminoàcids.
2. *Accession*: Identificador únic de la proteïna.
3. *Description*: Descripció de la proteïna i el seu origen.
4. *Score*: Puntuació de qualitat de la identificació.
5. *CLASS i PHOSPHO*: Categoria de modificació i indicador si és fosforilació o no.

Les columnes addicionals representen les mostres, i cada valor numèric reflecteix la intensitat de la modificació de cada proteïna en una mostra concreta.

Passos seguits en l'anàlisi

1. *Importació de dades:* Vam importar les dades en un entorn de treball R, creant una taula on vam poder visualitzar tant les metadades com les mesures de cada mostra.

2. *Organització i classificació de les mostres:* Les mostres es van classificar automàticament en dos grups segons la condició experimental (MSS o PD), la qual cosa permet comparar fàcilment les dades entre els dos estats.

3. *Creació d'un Objecte per a l'Anàlisi:* Per facilitar l'anàlisi, vam compilar les dades numèriques, juntament amb les metadades, en un format estructurat. Això permet un accés ràpid i consistent a les dades, tant de les proteïnes com de les mostres, en un sol objecte.

4. *Anàlisi i visualització de dades:* Vam generar visualitzacions per explorar la distribució de les intensitats de fosforilació en les diferents mostres, utilitzant gràfics com histogrames i boxplots. Això va donar-nos una primera visió general de les dades i de les possibles diferències entre els grups.

5. *Preprocessament:* Per garantir l'eficàcia de l'anàlisi, es van eliminar valors perduts i es van normalitzar les intensitats, fent les dades comparables entre mostres i condicions.

6. *Emmagatzematge per a futures anàlisis:* Finalment, vam guardar l'objecte estructurat amb les dades i les metadades en un fitxer per facilitar les anàlisis futures sense necessitat de repetir el preprocessament.

Resum dels grups de mostres

Els grups es divideixen en:

- MSS: Dades de condicions normals.
- PD: Dades sota condicions associades al Parkinson.

Aquestes comparacions poden ajudar a identificar patrons de modificacions que estan potencialment associats amb la malaltia.

Importància del preprocessament

Els passos de *preprocessament*, com l'eliminació de valors perduts i la normalització de les intensitats, són fonamentals per assegurar que les comparacions entre les mostres siguin fiables i reflecteixin de manera precisa les diferències entre condicions. Això permet obtenir resultats significatius que poden aportar informació valuosa per a la investigació biomèdica.

Apartat de resultats dades_fosforilacio

Estadística general

Hem obtingut un resum estadístic de les intensitats de senyal per a cada mostra en les condicions **MSS** (model normal) i **PD** (model Parkinson):

- **Intensitats mínimes:** En ambdues condicions, algunes observacions mostren un valor mínim de 0, fet que pot reflectir absència de senyal o dades mancants.
- **Mitjanes i medianes:** Les intensitats mostren una mitjana i una mediana més elevades en les mostres del grup PD, amb variabilitat notable.
- **Valors màxims:** Es van registrar valors màxims significativament alts en les mostres PD, amb un pic de més de 88 milions en una mostra específica, suggerint diferències notables en les modificacions proteiques entre els dos grups.

Dades de fosforilació

Les primeres observacions mostren modificacions de fosforilació en proteïnes identificades com **Syntenin-1**, **Mitogen-activated protein kinase**, i **Claudin-3**. La puntuació de qualitat (Score) associada a aquestes identifications és elevada, indicant un grau de confiança en la detecció de la modificació.

Comparació Estadística entre Condicions

Mitjançant un **test de t de Welch**, es va comparar la mitjana de les intensitats de senyal entre les condicions MSS i PD:

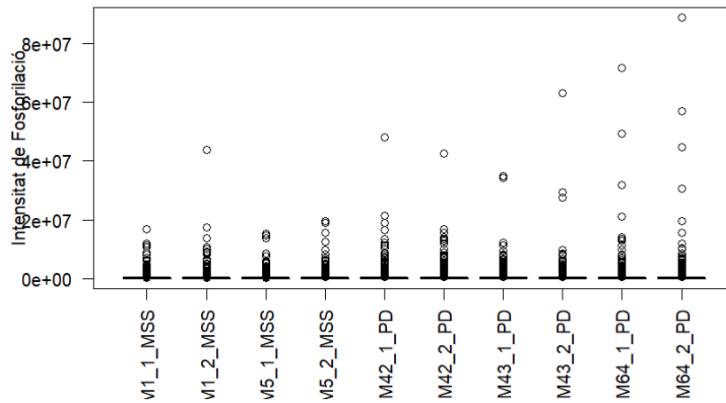
- **Resultats del test:** Es va obtenir un valor de p de **0.009**, indicant una diferència estadísticament significativa entre les intensitats mitjanes de les dues condicions.
- **Interval de confiança del 95%:** Aquesta diferència es va estimar entre -268,319 i -38,241, amb una mitjana de 244,257 en MSS i 397,537 en PD, fet que apunta a una intensitat superior en les mostres de la condició PD.

Conclusions

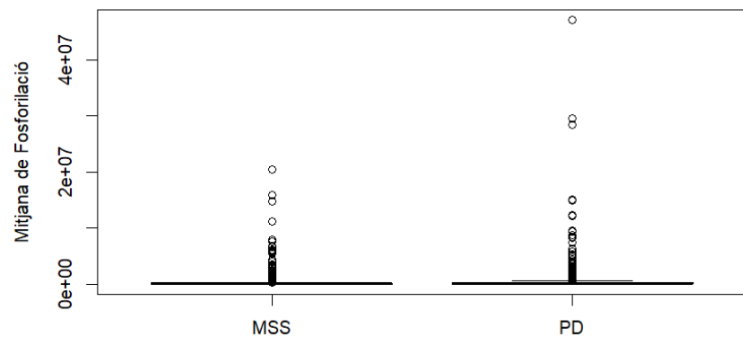
Els resultats suggereixen una variabilitat substancial en les intensitats de modificacions de fosforilació en proteïnes entre els grups MSS i PD. Aquestes diferències podrien estar associades amb les condicions patològiques de Parkinson, justificant futures investigacions per determinar el rol de les modificacions post-traduccionals en la patogènesi de la malaltia.

Taules obtingudes

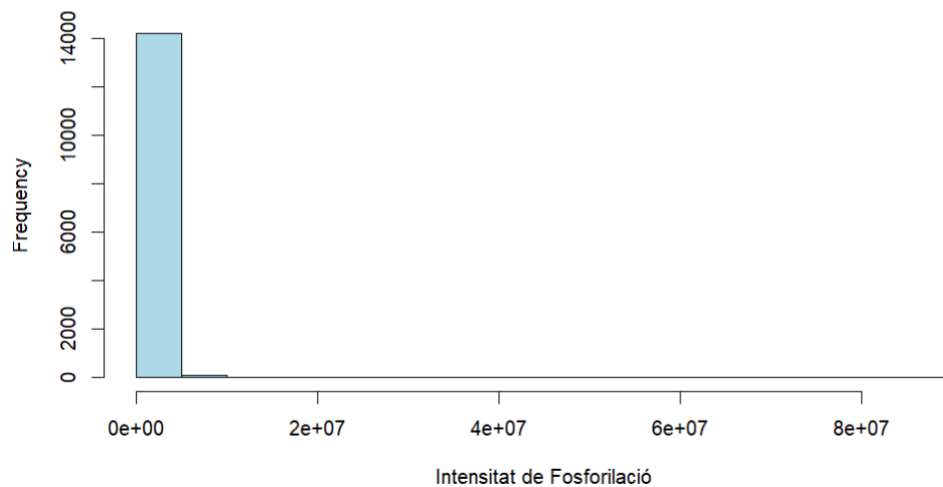
Distribució de Fosforilació per Mostra



Comparació de Mitjanes entre Grups



Distribució de les Mesures de Fosforilació



URL: <https://github.com/Marc110193/Vila-Muntadas-Marc-PAC1/upload>