

AO_PEC1

Amaia Carrascal Minino

2024-11-04

Contents

Abstract	1
Objetivos del estudio	1
Materiales y Métodos	2
Resultados e interpretacion	2
Generación de un contenedor en formato SummarizedExperiment	2
Exploración de los datos	3
Discusión y limitaciones y conclusiones del estudio	6
Direccion del repositorio Git	7

Abstract

En este estudio se analizaron muestras de xenoinjertos mediante espectrometría de masas para investigar diferencias en los perfiles de fosfopéptidos entre dos fenotipos, MSS y PD. El objetivo fue identificar fosfopéptidos que permitan diferenciar ambos fenotipos. Para ello, calculamos la diferencia de medias y representamos los fosfopéptidos con diferencias significativas en abundancia ($\alpha = 0.05$), encontrando un total de 50 fosfopéptidos distintos entre los dos grupos.

El análisis visual mediante el gráfico de barras y el gráfico de volcano reveló distintos grados de variación en las abundancias de fosfopéptidos entre MSS y PD. Al realizar un análisis de componentes principales (PCA) para explorar las relaciones entre los fosfopéptidos, observamos que el componente principal 1 (PC1) captura una gran parte de la variabilidad, aunque algunas muestras están muy cercanas entre sí. Este análisis sugiere que ciertos fosfopéptidos son característicos de cada grupo; sin embargo, podrían ser necesarias otras técnicas de análisis o un mayor número de muestras para corroborar su potencial como biomarcadores.

Objetivos del estudio

Extraer los datos y crear un contenedor en formato SummarizedExperiment para facilitar el análisis comparativo entre los grupos. Nuestro objetivo es identificar fosfopéptidos que permitan diferenciar claramente los dos grupos tumorales mediante análisis estadístico y visualización.

Materiales y Métodos

En este experimento se han analizado muestras procedentes de modelos de xenoinjertos derivados de pacientes (Patient-Derived Xenografts, PDXs) pertenecientes a dos subtipos tumorales distintos. Se utilizaron muestras enriquecidas en fosfopéptidos, y se llevó a cabo un análisis de espectrometría de masas por cromatografía líquida acoplada a espectrometría de masas (LC-MS) con dos réplicas técnicas por muestra. Estas réplicas fueron tratadas como muestras independientes tanto para el análisis como para la visualización de los datos para este estudio.

El conjunto de resultados consiste en conteos normalizados de señales de MS para aproximadamente 1400 fosfopéptidos. Como se indicó anteriormente, el objetivo es identificar fosfopéptidos que permitan diferenciar los dos grupos tumorales.

Los datos se proporcionaron en un archivo Excel: TIO2+PTYR-human-MSS+MSIvsPD.XLSX.

Los grupos se definen de la siguiente manera:

- Grupo MSS: Muestras M1, M5 y T49.
- Grupo PD: Muestras M42, M43 y M64. Cada muestra cuenta con dos réplicas técnicas, lo que da un total de seis muestras por grupo.

La primera columna, “SequenceModification”, contiene las modificaciones para los distintos fosfopéptidos. Existen otras columnas que hemos decidido omitir para el presente análisis.

Para visualizar los datos, hemos creado un gráfico de volcano que muestra las diferencias entre los fosfopéptidos, destacando aquellos con cambios significativos entre los grupos. Posteriormente, generamos un gráfico de barras que compara los fosfopéptidos con diferencias estadísticamente significativas. Ajustamos los valores p utilizando el método de la Tasa de Descubrimiento Falso (False Discovery Rate, FDR) para controlar el riesgo de obtener resultados falsos positivos cuando al realizar múltiples pruebas estadísticas al mismo tiempo.

A continuación, realizamos un análisis de componentes principales (PCA) para explorar si las combinaciones de fosfopéptidos permiten diferenciar claramente las muestras de ambos grupos. Finalmente, creamos un gráfico de las cargas (loadings) de cada componente principal para identificar los fosfopéptidos que más contribuyen a la variabilidad entre las muestras.

Resultados e interpretación

Generación de un contenedor en formato SummarizedExperiment

Extraemos los datos del archivo Excel y creamos un objeto en formato SummarizedExperiment para facilitar el análisis estructurado. El objeto tiene la siguiente estructura:

- **Clase:** SummarizedExperiment
- **Dimensiones:** 1438 filas x 12 columnas
- **Metadatos (2):**
 - **conditions:** Información sobre las condiciones experimentales.
 - **description:** Descripción general del experimento.
- **Ensayos (1):**
 - **counts:** Conjunto de datos que contiene los counts de cada fosfopéptido en cada muestra.

- **Nombres de filas (1438 fosfopéptidos):**
 - Ejemplos:
 - * LYPELSQYMGLSLNEEEIR[2] Phospho | [9] Oxidation
 - * VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho
 - * YQDEVFGGFVTEPQEESEEEVEEPEER[17] Phospho
 - * YSPSQNSPIHHIPSR[1] Phospho | [7] Phospho
- **Nombres de los datos de filas (rowData):**
 - No hay metadatos adicionales asociados a las filas.
- **Nombres de columnas (12 muestras):**
 - Ejemplos:
 - * M1_1_MSS, M1_2_MSS, ..., M64_1_PD, M64_2_PD
- **Nombres de los datos de columnas (colData) (2):**
 - **SampleID:** Identificador único de cada muestra.
 - **Phenotype:** Fenotipo al que pertenece cada muestra (por ejemplo, MSS o PD).

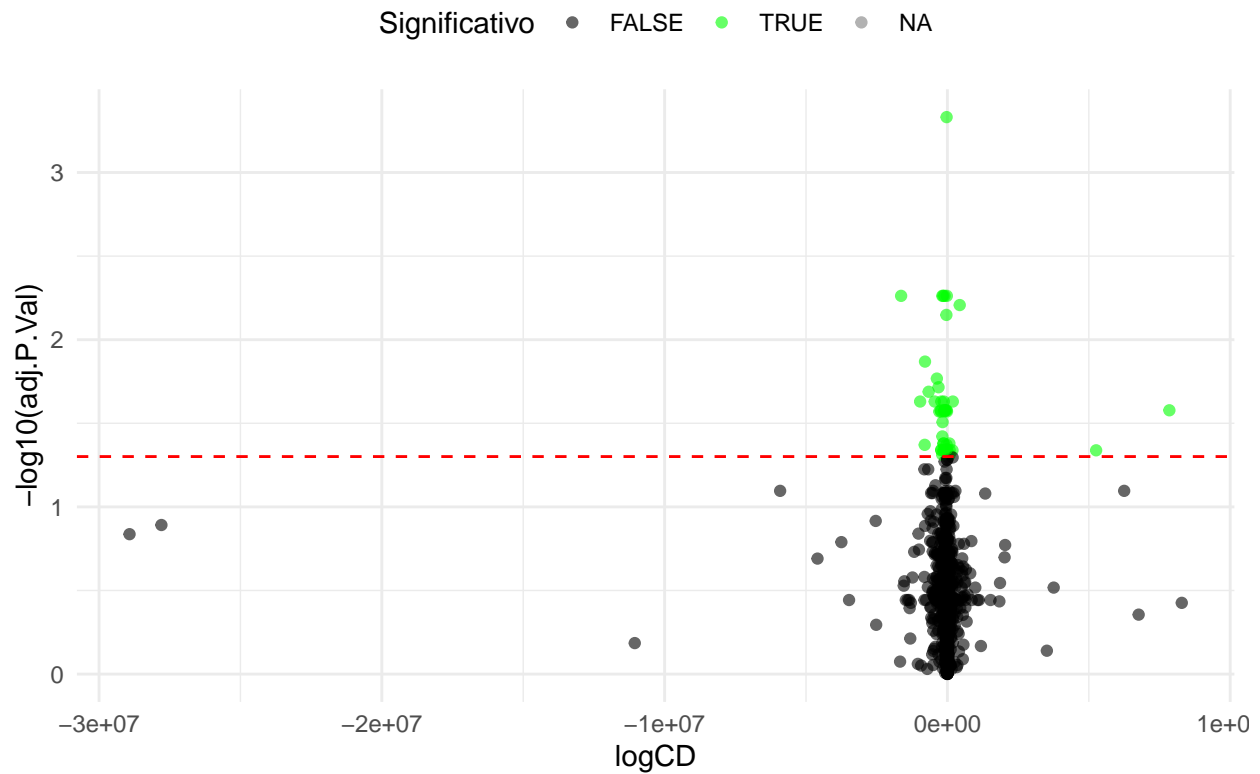
Exploración de los datos

El primer gráfico es un gráfico de volcano que se utiliza para visualizar las diferencias en la abundancia de fosfopéptidos entre dos grupos (MSS y PD). En este gráfico, cada punto representa un fosfopéptido y su posición en los ejes proporciona información sobre la magnitud y la significancia de las diferencias.

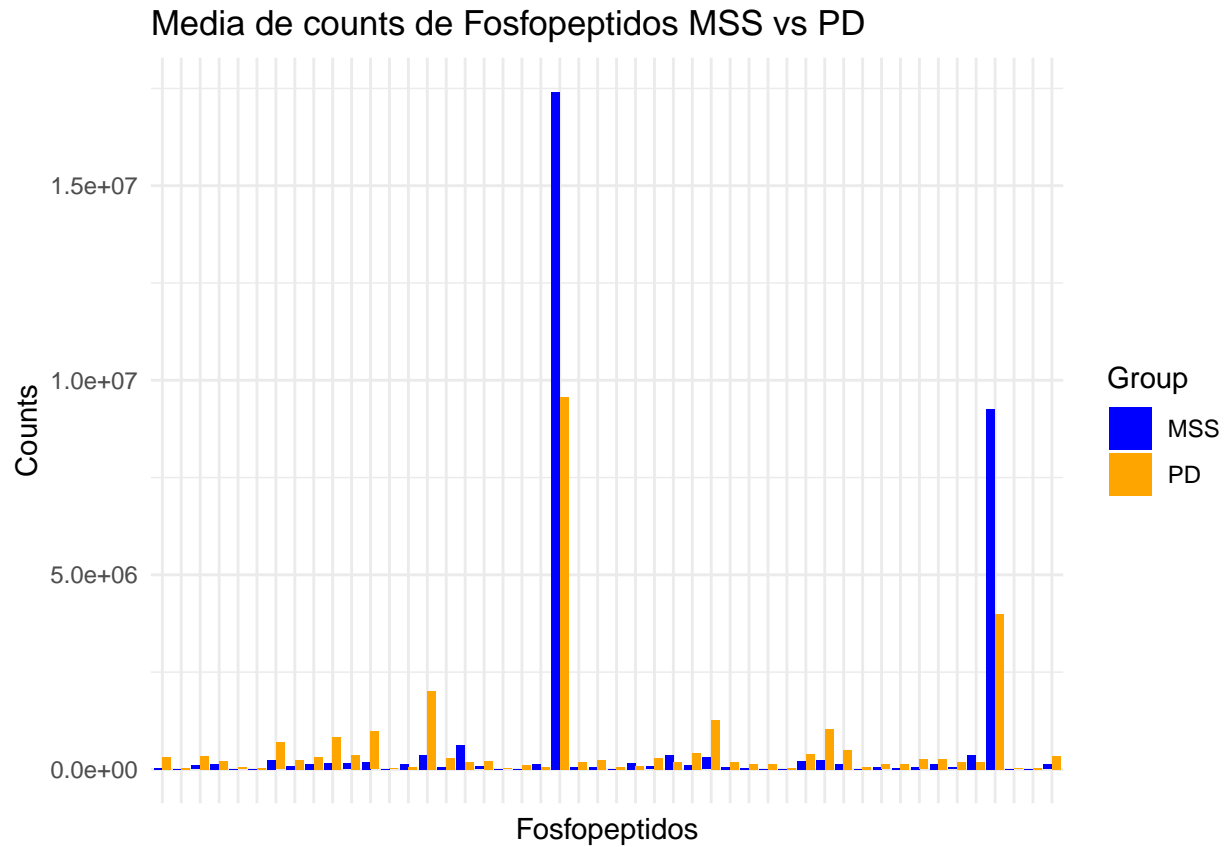
Los puntos en color verde representan los fosfopéptidos que cumplen con el criterio de significancia establecido (ajustado $p < 0.05$). Estos fosfopéptidos muestran diferencias significativas entre los dos grupos. Los puntos en negro o gris representan fosfopéptidos sin diferencias estadísticamente significativas entre los grupos. Como se puede ver bastantes fosfopéptidos muestran diferencias significativas.

Para visualizar estas diferencias mas claramente, filtramos los fosfopéptidos que presentan diferencias y los representamos en un grafico barras. Este gráfico de barras permite una visualización clara de cómo varía la abundancia de cada fosfopéptido entre los dos grupos. Los fosfopéptidos que muestran diferencias marcadas en abundancia entre MSS y PD podrían ser potenciales indicadores de diferencias biológicas entre los grupos y pueden ser relevantes.

Gráfico de Volcano Diferencias de fosfopéptidos



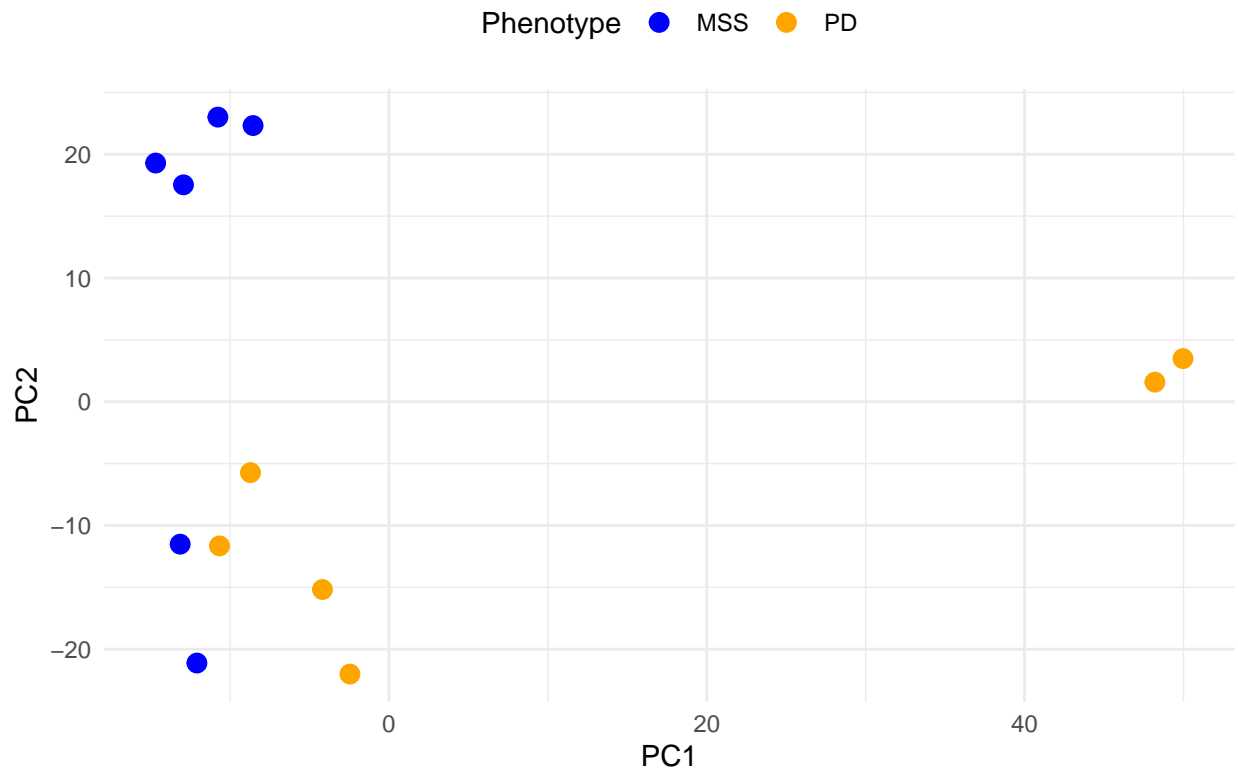
Para visualizar estas diferencias mas claramente, filtramos los fosfopéptidosque presentan diferencias y los representamos en un grafico barras.Este gráfico de barras permite una visualización clara de cómo varía la abundancia de cada fosfopéptido entre los dos grupos. Los fosfopéptidos que muestran diferencias marcadas en abundancia entre MSS y PD podrían ser potenciales indicadores de diferencias biológicas entre los grupos y pueden ser relevantes.



El ultimo gráfico representa un Análisis de Componentes Principales (PCA) sobre los dos fenotipos, MSS (en azul) y PD (en naranja). Cada punto en el gráfico corresponde a una muestra individual, y su posición en el espacio bidimensional está determinada por sus valores en los dos primeros componentes principales, PC1 y PC2.

Las muestras de MSS están agrupadas en el lado izquierdo del gráfico, mientras que las muestras de PD están agrupadas hacia la derecha en el eje de PC1. Pero algunas de las muestras de PD estan hacia la izquierda tambien,

PCA de Recuentos de fosfopéptidos



Imprimimos los factores por orden de importancia.

```
## EEKDKDDQEWESPPKPTVFISGVIA[14] Phospho
##                                0.04296131
##      QKSEEPSVSIPFLQTALLR[9] Phospho
##                                0.04295236
##      IGSDPLAYEPK[3] Phospho
##                                0.04291960
##      VDINTPDVDVHGPDWHLK[5] Phospho
##                                0.04286703
##      DKDDQEWESPPKPTVFISGVIA[11] Phospho
##                                0.04286372
##      TASELLDR[3] Phospho
##                                0.04282435
```

Discusión y limitaciones y conclusiones del estudio

Como se observa en el gráfico de PCA, el estudio presenta ciertas limitaciones. Existe un grupo de muestras que no es fácilmente distinguible en los componentes principales PC1 o PC2, lo cual puede deberse al número reducido de muestras analizadas. Para intentar mejorar la diferenciación, hemos considerado las réplicas como muestras independientes; sin embargo, esto introduce un posible sesgo, ya que las réplicas son muy similares entre sí, como se puede ver en el gráfico.

Este resultado indica que la mayor parte de la variabilidad observada es de carácter biológico, y que algunas muestras podrían clasificarse incorrectamente si solo se emplea esta técnica de análisis. Esto sugiere que

sería necesario complementar el PCA con otras técnicas o aumentar el número de muestras para obtener una clasificación más robusta y precisa.

Dirección del repositorio Git

<https://github.com/acarrascalmi/CARRASCAL-Minino-Amaia-PEC1>