

Índice general

1	Práctica TMT: Evaluation of NCI-7 Cell Line Panel as a Reference Material for Clinical Proteomics	2
---	---	---

FragPipe es un software que engloba varios programas. El buscador es MSFragger, pero cuenta también con otras herramientas como Philosopher, PTM-Shepherd, etc. Lo descargamos de GitHub y descargamos las licencias de MSFragger, IonQuant y diaTracer.

1. Práctica TMT: Evaluation of NCI-7 Cell Line Panel as a Reference Material for Clinical Proteomics

En este experimento había 7 líneas celulares y se querían ver las fosforilaciones de las líneas cancerígenas en comparación con el Wild-Type. Se hizo una digestión TMT, marcando los péptidos para saber si la muestra es KO o WT. Se hizo un LC-MS y se buscó contra la base de datos RefSeq. De esta forma se tienen en cuenta las modificaciones fijas y variables. Se cargometilan las cisteínas, quedando marcadas. Estas modificaciones se deben poner en la búsqueda.

En FragPipe, en la pestaña de Workflow, seleccionamos el workflow de TMT10. Luego pulsamos el botón de "Add files" y cargamos el fichero "01_CPTAC_TMTS1-NCI7_P_JHUZ_20170509_LUMOS.mzML".

MSFragger detecta que es un DDA, es decir, espectro Data Dependent. Como solo hay un fichero raw, no vamos a indicar réplicas biológicas. En la siguiente pestaña (DIA Pseudo MS2) debemos asegurarnos de que no hay nada marcado. Para Database, se pueden descargar las bases de datos de humanos, ratones, levadura y Covid. Por defecto, las bases de datos son de UniProt, y se puede seleccionar si añadir solo aquellas entradas revisadas manualmente, añadir decoys (que ayuda a calcular la FDR para la validación de los péptidos), añadir proteínas contaminantes (queratina, látex, errores experimentales), isoformas, etc. Se recomienda guardar esta base de datos con los raw.

La siguiente pestaña es de MSFragger. Se divide en varias secciones:

■ Peak Matching

Por un lado, se especifica la tolerancia de la masa del precursor. Se pone el rango de tolerancia en Daltons o PPM para la ventana de masa del precursor o fragmento. En el paper, se especifica que se han incluido 20 ppm tolerancia al precursor y 0.06 Da tolerancia al ion del fragmento. El estándar está en -20 y 20 en PPM, pero debemos cambiar la masa de tolerancia de los fragmentos a 0.06 Da, como se indica en el paper. Para que el programa no tarde mucho, se recomienda quitar la calibración y optimización.

■ Protein Digestion

En el paper se especifica que los péptidos se buscaron con dos missed cleavages. Esto lo recreamos en el programa: ponemos la enzima tripsina que corta en KR y 2 missed cleavage en C.

■ Modifications

Aquí se dividen las modificaciones variables y fijas. La publicación, al ser un TMT 10, ya hay 229.16293 masa adicional en el N-terminal de la lisina. De igual forma, hay una carbometilación en cisteína que aumenta la masa en 57.02146. Esto lo debemos reflejar en modificaciones fijas: 229.16293 en N-Term Peptide y en K (lysine), y 57.02146 en C (cysteine), aunque esto último debería estar por defecto.

Cuando se considera una modificación variable, deben tenerse en cuenta todos los posibles sitios de localización con todas las distribuciones posibles de esa modificación. Esto conduce a un fuerte aumento del tamaño del espacio de búsqueda, que escala exponencialmente con la inclusión de modificaciones adicionales, y con el consiguiente tiempo de computación y recursos. Para esta parte, no vamos a incluir la modificación variable de fosforilación, debido al tiempo de cómputo. No obstante, la oxidación en la metionina sí la mantenemos.

■ Opciones avanzadas

Existen más opciones avanzadas dependiendo del tipo de búsqueda, pero no vamos a profundizar en ellos, solo en **Advances Output Options**. Se permiten especificar los formatos de salida, siendo recomendada la salida más completa: TSV_PEPXML_PIN.

Con esto, hemos terminado la parte del buscador. La siguiente pestaña es de Validación. Son métodos estadísticos o de ML para validar y comprobar que las asignaciones péptido-proteína son las correctas. Vamos a quitar MSBooster, dejando todo lo demás de forma predeterminada.

La pestaña de PTMs debe estar desactivada, al igual que Glyco y Quant (MS1). En la pestaña de Quant (Isobaric) se extraen las intensidades de los espectros, llegando a cuantificarlo comparado con las muestras. TMT-Integrator extrae y combina abundancias de canales de múltiples muestras marcadas con TMT. Activamos la pestaña y pulsamos Edit/Create. Ahí seleccionamos TMT10 y pulsamos Load into Table. Se genera un fichero en el que se van a ir guardando los canales. En Basic Options, seleccionamos Quant level 2, Define Reference Virtual, Group By All y Normalization MD (median centering).

Las pestañas de Spec Lib, Quant DIA y Skyline deben estar desactivadas. En Run, seleccionamos la carpeta donde queremos que se guarden los resultados. Finalmente, pulsamos el botón "RUN". En caso de que dé error por falta de memoria, en MSFragger se pueden ampliar los splits de la base de datos.

Entre los resultados, está psm.tsv. Lo abrimos y vemos que hay distintas columnas:

- Spectrum: identificador del espectro MS/MS, sigue el formato (nombre de archivo).(scan).(scan).(charge)
- Spectrum File: espectro nombre del archivo de identificación de origen
- Peptide: secuencia de aminoácidos del péptido sin incluir ninguna modificación
- Modified Peptide: secuencia peptídica que incluye las modificaciones; los residuos modificados van seguidos de paréntesis que contienen la masa entera (en Dalton) del residuo más la modificación; en blanco si el péptido no está modificado.
- Hyperscore: puntuación de similitud entre los espectros observados y teórico; los valores más altos indican una mayor similitud.
- Nextscore: puntuación de similitud (hyperscore) de la segunda posición más alta para el espectro
- PeptideProphet Probability: puntuación de confianza determinada por PeptideProphet, los valores más altos indican mayor confianza.
- Number of missed cleavages: número de sitios potenciales de escisión
- ...