Este trabajo parte de los resultados de ChromHMM, que usa los picos de experimentos de ChIP-seq para las modificaciones de histonas, usando como referencia el genoma humano en su versión hg19. Con ello, se entrena un modelo para 11 estados, para así predecir las combinaciones de modificaciones de histonas presentes en los resultados. En nuestro caso, se estudia el estado 1, que incluye las marcas de promotor activo H3K4me3 y H3K27ac. Como resultado, se obtienen las probabilidades de cada estado para cada ventana de 200 pbs de cada cromosoma. Se ha diseñado una rutina de trabajo para crear los archivos bed con los intervalos de 200 pbs en los cuales el estado 1 es el más probable. Después, se ha hecho la intersección de los archivos bed resultantes para cada monocito de estudio.

El archivo resultante se ha importado en R y, usando la librería annotatr, se ha procedido a la anotación de cada intervalo, para saber qué elemento genómico está asociado a dicho intervalo en el genoma (usando la versión hg19). Esta librería permite crear un conjunto de datos aleatorios a partir de los de estudio y así comparar el dataset de estudio con el aleatorio, así como las anotaciones correspondientes a cada uno. Se ha podido observar un enriquecimiento de anotaciones de promotores, que no se observaba en el caso aleatorio. En mi caso, he procedido a guardar un archivo txt con todos los genes anotados como promotores.

Por otro lado, se ha pensado que la existencia de modificaciones de histonas típicas de promotores activos significa que dichos genes están expresándose en el tipo celular en cuestión. Para observar si los resultados obtenidos tienen sentido y coinciden con esta observación, se ha procedido a hacer una búsqueda de los genes de expresión típica en monocitos CD14+, como es el caso. Entre ellos, se ha centrado la atención en los genes: CD14, LYN, HLA-A, HLA-C, TP53...

Así, se ha procedido a la visualización de estos genes en el navegador genómico del UCSC, habiendo subido en él el fichero bed con la intersección de segmentos con el estado 1 como más probable entre ambos monocitos, el fichero de marcas hiper e hipometilación y el fichero con los picos de la DNasa de células sanguíneas de ENCODE. Además, se eligió mostrar las modificaciones de histonas de interés, pero en los tipos celulares disponibles asociados al sistema inmune, al cual pertenecen las células de estudio (GM12878 y K562).

De las observaciones realizadas, llama la atención que los intervalos correspondientes a estos genes se ubican en el extremo 5’ de los mismos, en los cuales también se observa hipometilación (asociada a expresión) y solapamiento con intervalos de DNasa. Esto significa que estos promotores están activos y accesibles, para permitir la entrada de la maquinaria de transcripción.

Líneas celulares UCSC:

- GM12878: B-lymphocyte, lymphoblastoid, International HapMap Project - CEPH/Utah - European Caucasion, Epstein-Barr Virus.

- K562: leukemia, "The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises." - ATCC.

/usr/lib/jvm/java-1.11.0-openjdk-amd64