

Regulación de la transcripción y de la reparación del DNA mediante remodelación de cromatina

Objetivo del proyecto

Analizar distintos aspectos de la cromatina, especialmente en relación con el posicionamiento nucleosomal, y cómo esto puede afectar a la transcripción y a la reparación del DNA.

*El objetivo del proyecto es desarrollar un conjunto de datos bien curado para cada secuencia nucleosomal y para secuencias de control para el organismo *S pombe*, en el que cada entrada contenga información estadística sobre todas las variables que sospechamos que pueden ser relevantes para el posicionamiento nucleosomal, por ejemplo: frecuencia de nucleótidos y dinucleótidos, ocurrencias de pares de dinucleótidos a cierta distancia, asimetría en las frecuencias, índices de ciclabilidad y flexibilidad, coordenadas genómicas o distancia media a la diáda.*

Este conjunto de datos servirá como base para el entrenamiento de una red neuronal residual, que cuenta con una arquitectura donde la entrada inicial se realimenta en distintas capas intermedias para evitar perder la relación con los datos de entrada. Los resultados se validarán buscando una capacidad de predicción mínima de un 80-90%. Otras redes entrenadas previamente tanto por nuestro laboratorio como por terceros solo con datos de la firma nucleosomal no superaban una fiabilidad del 70%. Una vez entrenada, analizaremos los pesos las conexiones de entrada de esos datos para identificar cuáles de todas las variables candidatas han sido favorecidas en el entrenamiento.

Todo este procedimiento computacional se realizará en SCAYLE cuando sea necesario por el volumen de cálculo. Una vez identificadas las variables candidatas, comprobaremos experimentalmente en el laboratorio las predicciones informáticas mediante análisis bioquímicos y genómicos.

Participantes del proyecto

Instituto de Biología Funcional y Genómica, <https://ibfg.usal-CSIC.es/francisco-antequera.html>

Universidad de Salamanca, www.usal.es

SCAYLE, Supercomputación Castilla y León (España), www.scayle.es

Ejecución: 2024 al 2027.

Financiación del proyecto

Ministerio de Ciencia, Innovación y Universidades 2023.

Funciones de SCAYLE

En particular, respecto a SCAYLE, están intentando caracterizar las secuencias nucleosomales de las levaduras *Schizosaccharomyces pombe* y *Saccharomyces cerevisiae* (unas 50.000) en función

de distintas propiedades como flexibilidad, frecuencia nucleotídica, ciclabilidad, etc. mediante redes neuronales residuales. Esto les permitirá entender mejor qué propiedades genómicas afectan al posicionamiento nucleosomal. Este tipo de análisis es bastante prospectivo, están probando distintos tipos de redes y distintas configuraciones, además de distintos conjuntos de entrenamiento. Por tanto, es relativamente costoso en tiempo de cálculo y la potencia de SCAYLE les permite hacer grandes baterías de pruebas.



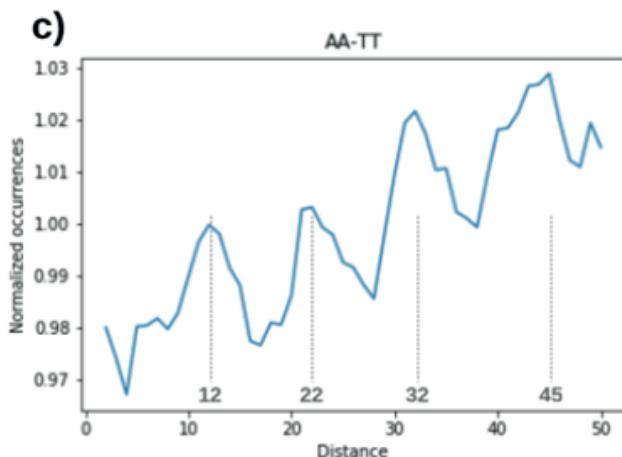
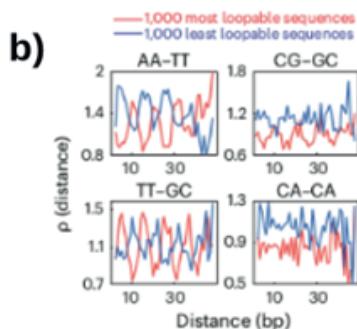
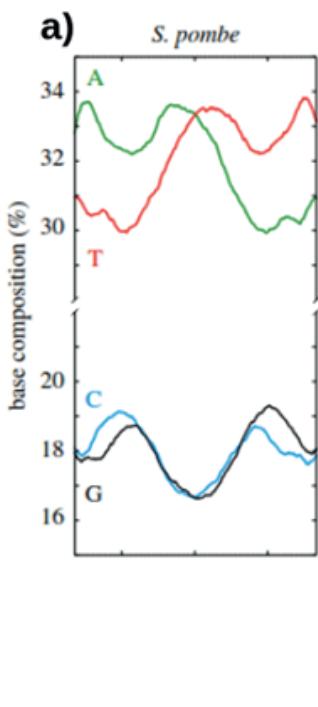
Referencia: PID2023-149512NB-I00

Justificación del proyecto

El laboratorio ha encontrado y publicado con anterioridad distintos patrones que se observan en la secuencia de DNA característicos de las posiciones nucleosomales. Esta información ha podido ser corroborada y utilizada para modificar la posición de nucleosomas, favoreciendo o dificultando la transcripción, y permitiendo profundizar más en el conocimiento sobre los patrones de secuencia involucrados. Dichas modificaciones surgen de modificaciones ('remasterizaciones') de la secuencia genómica silvestre. Llegados a este punto, consideran que dichos patrones pueden refinarse para conseguir los mismos objetivos con una modificación menos invasiva del DNA. Sin embargo, estos patrones son emergentes, pero difíciles de detectar en secuencias individuales, por lo que las técnicas de análisis actuales no son suficientes y están investigando si la IA puede ayudar en este sentido.

Líder del proyecto

Instituto de Biología Funcional y Genómica. El Instituto de Biología Funcional y Genómica (IBFG) es un Centro Mixto de investigación de titularidad compartida entre el Consejo Superior de Investigaciones Científicas (CSIC) y la Universidad de Salamanca (USAL). Su línea general de investigación refleja el interés actual del Centro en estudiar los mecanismos reguladores de las funciones celulares y su integración en el contexto del genoma a través de aproximaciones metodológicas avanzadas de biología celular, molecular y genómica, estructurándose en tres unidades básicas: I. Biología Funcional y de Sistemas, II. Biología del Genoma, y III. Biotecnología y Biomedicina.



a) Firma nucleotídica caracterizada por nuestro grupo en la levadura *S. pombe* (Gonzalez et al. 2016) b) Evidencia de ocurrencia de distintos dinucleótidos en secuencias muy flexibles y poco flexibles (Basu et al 2022). c) Evidencias iniciales encontradas en *S. pombe* de co-ocurrencia de dinucleótidos a distancias similares a las de b). Todas estas evidencias son propiedades emergentes pero las diferencias son pequeñas para su caracterización en secuencias individuales. Nuestro trabajo actual trata de encontrar características como las de c) en secuencias individuales mediante inteligencia artificial y un curetaje previo de datos estadísticos de la secuencia de alta calidad.