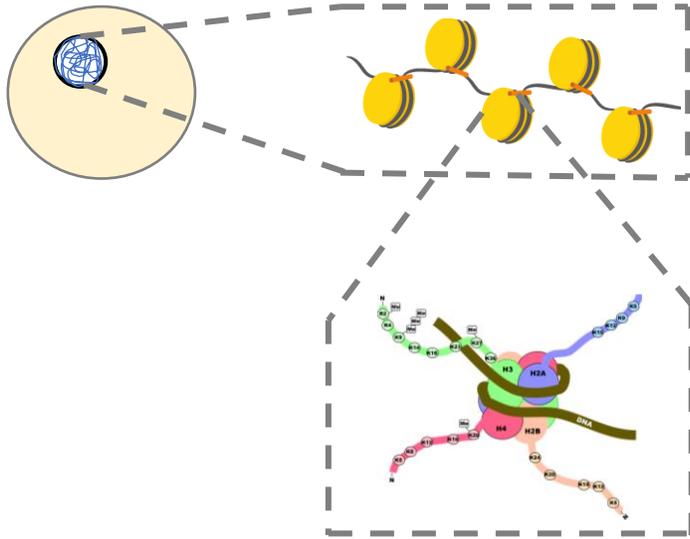


Reducción de dimensionalidad para la construcción de modelos predictivos con datos epigenéticos

Juan José Alba Linares, Cancer Epigenetics Lab, Oviedo

Contexto biológico



Células del linaje hematopoyético

Datos

- H3K9me3
- H3K27me3
- H3K36me3
- H3K4me3
- H3K4me1
- H3K27ac

N > 5000

N > 1500

	S1_H3K9me3	S2_H3K27me3	S2_H3K27me3
A	100	15	200
B	5000	0	600
C	1500	800	0
D	6000	0	10000

Objetivo

Identificar la combinación mínima de elementos (filas de la matriz) que es capaz de lograr una mejor clasificación de cada tipo celular.

Problema

Las columnas de la matriz representan los niveles de los elementos para una muestra (tipo celular) y una marca de histonas concreta.

Reducción de dimensionalidad para la construcción de modelos predictivos con datos epigenéticos

Juan José Alba Linares, Cancer Epigenetics Lab, Oviedo

Solución Parcial

- H3K9me3
- H3K27me3
- H3K36me3
- H3K4me3
- H3K4me1
- H3K27ac

Integración multi-omica

Generación de factores abstractos
(combinación lineal de marcas de histonas)

Selección de elementos con mayor
contribución por marca de histona

Support Vector Machine

Problema

Este proceso se repetiría por cada tipo celular por separado, de forma que no nos aseguramos la selección de los mejores elementos clasificadores.