
Machine Learning para corrección de errores en datos de secuenciación de ADN

Trabajo de Fin de Grado

Amin Kasrou Aouam



**UNIVERSIDAD
DE GRANADA**

26-06-2021

Índice

Resumen	3
Abstract	4
Introducción	4
Motivación	4
Estado del arte	4
NGS	4
Bioinformática (Deep Learning)	4
Metodología	4
Tecnologías	4
Pipeline	4
Reproducibilidad	4
Resultados	4
Conclusiones	4
Futuras mejoras	4

Resumen

Las nuevas técnicas de secuenciación de ADN (NGS) han revolucionado la investigación en genómica. Estas tecnologías se basan en la secuenciación de millones de fragmentos de ADN en paralelo, cuya reconstrucción se basa en técnicas de bioinformática. Aunque estas técnicas se apliquen de forma habitual, presentan tasas de error significantes que son perjudiciales para el análisis de regiones con alto grado de polimorfismo. En este estudio se implementa un nuevo método computacional, *locimend*, basado en *Deep Learning* para la corrección de errores de secuenciación. Se aplica al análisis de la región determinante de complementariedad 3 (CDR3) del receptor de linfocitos T (TCR), generada *in silico* y posteriormente sometida a un simulador de secuenciación con el fin de producir errores de secuenciación. Empleando estos datos, entrenamos una red neuronal convolucional (CNN) con el objetivo de generar un modelo computacional que permita la detección y corrección de los errores de secuenciación.

Abstract

Introducción

Motivación

Estado del arte

NGS

Bioinformática (Deep Learning)

Metodología

Tecnologías

Pipeline

Reproducibilidad

Resultados

Conclusiones

Futuras mejoras