

**FULL DE RESUM DEL TREBALL DE FI DE GRAU DEL BDBI (ESCI-UPF)**

<b>TÍTOL DEL PROJECTE:</b> Imputing eQTL from tissue specific and across tissues data	
<b>AUTOR/A:</b> Laura Aviñó Esteban	<b>NIA:</b> 102777
<b>CURS ACADÈMIC:</b> 2018-2019	
<b>DATA:</b> 18 Juny de 2019	
<b>TUTOR/S:</b> Daniel Zerbino	
<b>PARAULES CLAU (mínim 3)</b>	
<ul style="list-style-type: none"> <li>• Català: Predicció, imputació, eQTL, Hi-C, epigenòmica, xarxes reguladores, aprenentatge automàtic, expressió de RNA, GTEX.</li> <li>• Castellà: Predicción, imputación, eQTL, Hi-C, epigenómica, red reguladora, aprendizaje automático, expresión de RNA, GTEX.</li> <li>• Anglès: Prediction, imputation, eQTL, Hi-C, epigenomics, regulatory networks, machine learning, RNA expression, GTEX.</li> </ul>	
<b>RESUM DEL PROJECTE (extensió màxima: 100 paraules per llengua)</b>	
<ul style="list-style-type: none"> <li>• Català: Tot i tenir coneixement d'on es troben els gens i les regions del genoma que els regulen, anomenades enhancers; encara no sabem quines interactua amb quin gen. En aquest treball he generat models basats en aprenentatge automàtic utilitzant diferents subconjunts dades, incloent-hi Hi-C, epigenòmica i eQTLs, per predir aquestes connexions. També he creat un mètode per imputar les interaccions utilitzant tècniques de reducció de dimensionalitat en l'expressió de RNA. En el conjunt, he demostrat que integrant diferents fonts de dades en el mateix mètode augmenta el rendiment d'aquest.</li> <li>• Castellà: A pesar de tener conocimiento de dónde se encuentran los genes y las regiones del genoma que los regulan, llamadas enhancers; todavía no sabemos qué region interactúa con qué gen. En este trabajo he generado modelos basados en aprendizaje automático utilizando diferentes subconjuntos datos, incluyendo Hi-C, epigenómica y eQTLs para predecir estas conexiones. También he creado un método para imputar las interacciones utilizando técnicas de reducción de dimensionalidad en la expresión de RNA. En conjunto, he demostrado que integrando diferentes fuentes de datos en el mismo método aumenta el rendimiento del mismo.</li> <li>• Anglès: Despite knowing where genes and the genomic regions regulate them, called enhancers; we still do not know which one interacts with which gene. In this work, I have generated models based on machine learning techniques using different subsets data, including Hi-C, epigenomics and eQTLs, to predict such connections. I have also created a method to impute interactions using dimensionality reduction techniques in the expression of RNA. All in all, I have shown that integrating different data sources in the same method increases its performance.</li> </ul>	