

MEMÒRIA DEL TREBALL DE FI DE GRAU DEL GRAU (ESCI-UPF)

**Upstream open reading frames identification and classification
in the human genome**

AUTOR/A: Ferriol Calvet Riera	NIA: 103633
GRAU: Bachelor's Degree in Bioinformatics	
CURS ACADÈMIC: 2019-2020	
DATA: 22 de juny del 2020	
TUTOR/S: Dr. Jonathan M. Mudge i Dr. Adam Frankish	

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

TÍTOL DEL PROJECTE: Upstream open reading frames identification and classification in the human genome	
AUTOR/A: Ferriol Calvet Riera	NIA: 103633
CURS ACADÈMIC: 2019-2020	
DATA: 22 de juny del 2020	
TUTOR/S: Dr. Jonathan M. Mudge i Dr. Adam Frankish	
PARAULES CLAU (mínim 3)	
<ul style="list-style-type: none"> • Català: marcs oberts de lectura <i>upstream</i> (<i>uORFs</i>), anotació genòmica, classificació • Castellà: marcos abiertos de lectura <i>upstream</i> (<i>uORFs</i>), anotación genómica, clasificación • Anglès: upstream open reading frames, genomic annotation, classification 	
RESUM DEL PROJECTE (extensió màxima: 100 paraules per llengua)	
<ul style="list-style-type: none"> • Català: A les regions no traduïdes dels gens que codifiquen per proteïnes, podem trobar-hi marcs oberts de lectura (<i>uORFs</i>, de les sigles en anglès) que poden tenir un paper important en la regulació del procés de traducció. Hem generat un recull dels 104497 possibles <i>uORFs</i>, que conté informació addicional sobre cadascun d'ells i també una puntuació individual associada que ens informa de com de probable es que siguin funcionals. A part d'això també donem eines per poder actualitzar els resultats a noves versions de l'anotació, i poder visualitzar els <i>uORFs</i> a UCSC genome browser. • Castellà: En las regiones no traducidas de los genes que codifican para proteínas, encontramos marcos abiertos de lectura (<i>uORFs</i>, de las siglas en inglés) que pueden ser elementos reguladores importantes del proceso de traducción. Hemos generado un conjunto de 104497 posibles <i>uORFs</i>, que contiene información adicional sobre cada uno de ellos y también una puntuación individual indicando cómo de probable es que sea funcional. A parte de esto, también damos las herramientas necesarias para poder actualizar los resultados a nuevas versiones de la anotación y para poder visualizar los 	

uORFs en UCSC genome browser.

- **Anglès:**

Upstream open reading frames are genomic features present in the 5' UTRs of protein-coding transcripts that are known to be important for translation regulation. We have provided a dataset with 104497 unique possible uORFs in the GENCODEv34 transcriptome. Each of them has additional features and a score showing how likely it is to be real that was computed based on them. The tools and pipelines we have implemented are easily replicable and can allow fast updates to the current annotation or data version. Apart from all these, we provide a custom track to visualize uORFs in the UCSC genome browser.