

València, 10 de juny de 2024

## **El CSIC colidera un estudi pioner sobre la seqüenciació llarga del transcriptoma, clau en la malaltia i l'envelliment**

- **L'Institut de Biologia Integrativa de Sistemes (CSIC-UV) impulsa un estudi internacional sobre els mètodes per a analitzar dades obtingudes per seqüenciació de lectura llarga del transcriptoma**
- **Aquest és el sistema més utilitzat actualment per a estudiar la complexitat de la vida. Els resultats, publicats en 'Nature Methods', revelen una diversitat de molècules d'ARN molt major de l'esperada**



Foto de grup del personal investigador que ha desenvolupat l'estudi. Crèdits: I2SysBio (CSIC – Universitat de València).

Un grup internacional d'investigació coliderat per l'Institut de Biologia Integrativa de Sistemes (I2SysBio), centre mixt del Consell Superior d'Investigacions Científiques (CSIC) i la Universitat de València (UV), publica en *Nature Methods* el major estudi comparatiu realitzat fins a la data sobre els mètodes que analitzen dades obtingudes per seqüenciació de lectura llarga del transcriptoma humà. Es van analitzar diferents tecnologies i diverses eines informàtiques disponibles per a la seqüenciació de lectura llarga de les molècules de l'ARN, molècules fonamentals perquè els gens complisquen la

seua funció. L'estudi va trobar una major diversitat d'ARN de l'esperada, la qual cosa podria tindre grans repercussions en l'estudi de malalties, envelliment i la mateixa complexitat de la vida en la Terra.

Durant anys, un consorci internacional conegut com a Projecte d'Avaluació d'Anotació del Genoma de Lectura Llarga d'ARN-Seq (LRGASP, per les seues sigles en anglés), va avaluar els mètodes i tecnologies en experiments de seqüenciació de lectura llarga d'ARN. Ara, aquest consorci global on CSIC té un paper fonamental ha publicat els resultats d'aquest esforç, oferint orientació per al futur de l'experimentació i l'anàlisi de seqüenciació d'ARN. El treball, publicat a la revista *Nature Methods*, avalua les fortaleses i debilitats de les dues principals plataformes de seqüenciació de lectura llarga d'ARN, Oxford Nanopore Technologies i Pacific Biosciences, així com els mètodes computacionals utilitzats per a avaluar les dades.

L'ARN és el compost molecular de les cèl·lules que transmet la informació de l'ADN a les proteïnes mitjançant els processos de transcripció i traducció, que són universals per a tots els éssers vius. La seqüenciació de lectura llarga de l'ARN permet observar molècules completes d'ARN i identificar xicotets canvis en la forma en la qual els gens donen lloc a proteïnes. Eixos xicotets canvis són crítics per a la constitució d'organismes complexos com els éssers humans, i fallades en la seua síntesi s'associen a diverses malalties. La seqüenciació de lectura llarga de l'ARN s'usa per a identificar aquests canvis i associar-los amb processos biològics diversos.

“Encara que el genoma humà ha estat seqüenciat de punta a punta, encara ens enfrontem a grans desafiaments per a definir amb exactitud com els gens donen lloc a l'enorme diversitat de molècules d'ARN i proteïnes que formen un ésser viu. Aquest coneixement és molt important, perquè xicotets canvis en el pas d'ADN a ARN poden donar lloc a patologies”, explica **Ana Conesa**, professora d'investigació del CSIC a l'I2SysBio i una de les investigadores que han liderat aquest consorci. El seu equip avaluat les prediccions d'ARN proposades per 14 laboratoris bioinformàtics de tot el món, utilitzant el programari SQANTI3 desenvolupat per aquest grup a l'I2SysBio, una de les eines bioinformàtiques de referència en el camp.

### Diversitat d'ARN major de l'esperat

Es van generar més de 427 milions de seqüències de lectura llarga que van ser analitzats en l'estudi. Les dades procedien d'humans, ratolins i manatins. L'ús de dades de manatins va permetre provar els mètodes en una espècie sense un genoma de referència. “Era important provar les tècniques en una espècie no model, ja que cada vegada és més comú veure estudis amb seqüenciació de lectura llarga d'ARN en aquests organismes no tan ben estudiats. Aquesta falta d'informació prèvia ha de ser tinguda en compte durant les anàlisis perquè pot afectar directament els nostres resultats”, assegura **Francisco J. Pardo Palacios**, investigador predoctoral de l'I2SysBio i primer signant d'aquest treball.

Després d'una extensa recopilació i anàlisi de dades, el consorci va produir un conjunt de recomanacions per a la seqüenciació d'ARN. En general, els enfocaments de

seqüenciació de lectura llarga funcionen molt millor que la seqüenciació de lectura curta, sent la qualitat de les lectures, més que la seua abundància, el factor clau de precisió. Ademés, van trobar una quantitat sorprenent de trànscripats no documentats en els genomes humans i de ratolins. “Hem vist que existeix una diversitat d'ARNs molt més gran d'ho pensàvem. Estem veient que cada individu, fins i tot cada cèl·lula, té un transcriptoma propi i personal. El següent pas és esbrinar la rellevància que té això en la malaltia, l'envelliment i la diversitat d'espècies”, resumeix Ana Conesa.

El treball conclou que no existeix un millor enfocament únic per a la seqüenciació de lectura llarga d'ARN. L'article descriu les millors pràctiques en funció dels diferents objectius que puguen tindre els estudis individuals. Les diferents tecnologies existents presenten diferències en les taxes d'error, rendiment de seqüenciació i la longitud de lectura, per la qual cosa els investigadors han de prioritzar quin és més important per a la seua àrea d'estudi. “Crec que això ajudarà a moltes persones que volen desenvolupar encara més la tecnologia, encara hi ha marge de millora en molts d'aquests mètodes”, finalitza **Angela Brooks**, investigadora de la Universitat de Califòrnia Santa Cruz (EUA) i coautora de l'estudi.

#### Referència:

Pardo-Palacios, F.J., Wang, D., Reese, F. *et al.* ***Systematic assessment of long-read RNA-seq methods for transcript identification and quantification.*** *Nat Methods* (2024). [DOI: 10.1038/s41592-024-02298-3](https://doi.org/10.1038/s41592-024-02298-3)