

UN NOU MÈTODE PERMET GARANTIR LA REPRODUCTIBILITAT EN EXPERIMENTS COMPUTACIONALS

- **Científics del Centre de Regulació Genòmica (CRG), han desenvolupat un sistema de gestió de fluxos de treball que evita els problemes de reproductibilitat en analitzar grans bases de dades genòmiques per ordinador.**
- ***Nextflow* contribueix a la instauració de bones pràctiques científiques i ofereix un important marc per aquells projectes d'investigació que requereixen l'anàlisi de grans conjunts de dades per prendre decisions, per exemple, en la medicina de precisió.**

La reproductibilitat en la investigació és crucial per poder avançar en ciència. Lamentablement, i d'acord amb estudis i enquestes recents*, hi ha un increment en el nombre d'experiments que no es poden reproduir i la reproductibilitat en investigació ja és un dels principals reptes que científics, institucions, ens finançadors i editorials científiques han d'abordar per assegurar la credibilitat i el progrés de la ciència.

Per comprendre les dades genòmiques, els científics depenen cada vegada més d'una combinació de programes informàtics anomenats *pipelines*. Aquests *pipelines* processen les dades, les analitzen i donen lloc a resultats com, per exemple, el risc de patir una malaltia genètica. Malauradament els resultats d'aquests *pipelines* no són sempre reproduïbles i, en l'era de la medicina de precisió, aquesta reproductibilitat limitada pot tenir implicacions importants per a la nostra salut.

Ara, un equip d'investigadors al Centre de Regulació Genòmica (CRG) a Barcelona, liderats per Cedric Notredame, han desenvolupat un sistema de gestió de fluxos de treball que assegura la reproductibilitat en els experiments computacionals. El sistema, anomenat *Nextflow*, es descriu en l'últim número de la revista *Nature Biotechnology*. "Hi ha diminutes variacions entre plataformes computacionals que poden induir inestabilitat numèrica, el què dóna lloc a la irreproductibilitat dels experiments computacionals. *Nextflow* permet que els científics evitin aquestes variacions i contribueix a establir bones pràctiques científiques en experiments per ordinador", explica Cedric Notredame, autor principal del treball.

"Una petita variació pot no semblar un problema quan s'estan utilitzant una quantitat enorme de dades genòmiques per a un projecte d'investigació, però fins i tot les variacions més petites poden ser crucials si pretenem utilitzar els resultats de la nostra anàlisi per prendre decisions, per exemple, mèdiques ", afegeix Paolo Di Tommaso, primer autor del treball. "La irreproductibilitat serà una qüestió fonamental per a la medicina de precisió", conclou.

Contenir la irreproductibilitat

La principal raó per la qual existeix irreproductibilitat en experiments computacionals és la complexitat dels ordinadors moderns. Amb la gran quantitat d'arxius i programes que contenen, els ordinadors són com màquines fetes de milers de milions de parts en moviment. Fins i tot en utilitzar exactament el mateix *pipeline* i les mateixes dades, hi ha petites variacions entre ordinadors que poden donar lloc a irreproductibilitat.

La solució a aquest problema passa per proporcionar no només les dades i el programari sinó també tot l'ambient pre-configurat per a la seva execució, gràcies als contenidors de programari, una tecnologia de virtualització de nova generació. L'equip del CRG ha implementat *Nextflow* com a eina que permet gestionar els fluxos de treball computacional

utilitzant aquest tipus de contenidors. "És com si congeléssim l'experiment, així tot aquell que vulgui reproduir-lo, podria fer-ho exactament de la mateixa manera i en les mateixes condicions, sense haver de re-introduir manualment cap configuració complexa. Treballar d'aquesta manera garanteix que un mateix conjunt de dades donarà lloc als mateixos resultats independentment d'on s'analitzin", expliquen els autors.

Nextflow permet integrar els recursos més sofisticats per assegurar la reproductibilitat com ara: Zenodo per a les dades, Github i Docker per al programari i la computació en el núvol. Aquest treball suposa un punt d'inflexió en la reproductibilitat d'experiments computacionals i un impuls per a les bones pràctiques en l'anàlisi de grans conjunts de dades. El CRG està compromès amb ajudar a promoure aquest aspecte clau de la biologia moderna posant aquest nou recurs a disposició de la recerca i també de la producció clínica i comercial. També ha organitzat un seguit de tallers i cursos dedicats a l'ús de *Nextflow* adreçats a la comunitat científica.

Referència:

Di Tommaso et al. *Nextflow enables reproducible computational workflows*. *Nature Biotechnology* 35, 316-319 (2017). doi: [10.1038/nbt.3820](https://doi.org/10.1038/nbt.3820)

(*) Altres articles citats en aquesta nota de premsa:

Baker, M. *1,500 scientists lift the lid on reproducibility*. *Nature* **533**, 452–454 (2016).

Munafò MR et al. *A manifesto for reproducible science*. *Nature Human Behaviour* 1, 0021 (2017).

Informació sobre el finançament:

Aquest projecte ha comptat amb el suport del Centre de Regulació Genòmica, el Plan Nacional del Ministeri de Economia, Indústria i Competitivitat, 'Centro de Excelencia Severo Ochoa 2013-2017', el projecte Elixir EXCELERATE i el projecte OpenRiskNet de la Comissió Europea en el programa marc H2020. Maria Chatzou i Evan W Floden han comptat amb el suport de la Fundació Bancària La Caixa. També s'agraeix el suport del programa CERCA de la Generalitat de Catalunya.

Imatges disponibles a:

<https://www.dropbox.com/sh/d4ow6i3ttgdg8fte/AAAIAY7gRI9lp2vAK7cnMjWJa?dl=0>

Peu de foto i crèdits:

Nextflow CRG team.jpg => Investigadors del CRG desenvolupen Nextflow, un nou mètode que garanteix la reproductibilitat dels experiments computacionals. D'esquerra a dreta: Emilio Palumbo, Paolo Di Tommaso, Evan W Floden, Cedric Notredame i Pablo Prieto © CRG.

Contacte per a mitjans:

Laia Cendrós, Oficina de premsa, Centre de Regulació Genòmica (CRG)

Tel.+34 93 316 02 37 – Mòbil +34 607 611 798 – E-correu: laia.cendros@crg.eu