

UN NUEVO MÉTODO PERMITE GARANTIZAR LA REPRODUCIBILIDAD EN EXPERIMENTOS COMPUTACIONALES

- **Científicos del Centro de Regulación Genómica (CRG) en Barcelona, han desarrollado un sistema de gestión de flujos de trabajo que evita los problemas de reproducibilidad al analizar grandes bases de datos genómicas por ordenador.**
- ***Nextflow* contribuye a la adopción de buenas prácticas científicas y ofrece un importante marco para aquellos proyectos de investigación que requieren el análisis de grandes conjuntos de datos para tomar decisiones, por ejemplo, en la medicina de precisión.**

La reproducibilidad en la investigación es crucial para poder avanzar en ciencia. Lamentablemente, y de acuerdo con estudios y encuestas recientes*, el número de experimentos que no se pueden reproducir va incrementando y la reproducibilidad en investigación ya es uno de los principales retos que científicos, instituciones, entes financiadores y editoriales deben abordar para asegurar la credibilidad y el avance de la ciencia.

Para comprender los datos genómicos, los científicos dependen cada vez más de una combinación de programas informáticos llamados *pipelines*. Estos *pipelines* procesan los datos, los analizan y dan lugar a resultados como, por ejemplo, el riesgo de padecer una enfermedad genética. Desafortunadamente los resultados de estos *pipelines* no son siempre reproducibles y, en la era de la medicina de precisión, esta reproducibilidad limitada puede tener implicaciones importantes para nuestra salud.

Ahora, un equipo de investigadores en el Centro de Regulación Genómica (CRG) en Barcelona, liderados por Cedric Notredame, han desarrollado un sistema de gestión de flujos de trabajo que asegura la reproducibilidad en los experimentos computacionales. El sistema, llamado *Nextflow*, se describe en el último número de la revista *Nature Biotechnology*. “Existen diminutas variaciones entre plataformas computacionales que pueden inducir inestabilidad numérica, lo que da lugar a la irreproducibilidad de los experimentos computacionales. *Nextflow* permite que los científicos eviten estas variaciones y contribuye a establecer buenas prácticas científicas en experimentos por ordenador”, explica Cedric Notredame, autor principal del trabajo.

“Una pequeña variación puede que no parezca un problema cuando se está utilizando una cantidad enorme de datos genómicos para un proyecto de investigación, pero incluso las variaciones más pequeñas pueden ser cruciales si pretendemos utilizar los resultados de nuestro análisis para tomar decisiones, por ejemplo, médicas”, añade Paolo Di Tommaso, primer autor del trabajo. “La irreproducibilidad será una cuestión fundamental para la medicina de precisión” concluye.

Conteniendo la irreproducibilidad

La principal razón por la que existe irreproducibilidad en experimentos computacionales es la complejidad de los ordenadores modernos. Con la gran cantidad de archivos y programas que contienen, los ordenadores son como máquinas hechas de miles de millones de partes en movimiento. Incluso al utilizar exactamente el mismo *pipeline* y los mismos datos, existen pequeñas variaciones entre ordenadores que pueden dar lugar a irreproducibilidad de los resultados.

La solución a este problema pasa por proporcionar no sólo los datos y el software sino también todo el ambiente pre-configurado para su ejecución, gracias los contenedores de software, a una tecnología de virtualización de nueva generación. El equipo del CRG ha implementado *Nextflow* como herramienta que permite gestionar el flujo de trabajo computacional utilizando este tipo de contenedores. “Es como si congeláramos el experimento, así todo aquel que quiera reproducirlo, podría hacerlo exactamente de la misma manera y en las mismas condiciones, sin tener que re-introducir manualmente ninguna configuración compleja. Trabajar de este modo garantiza que un mismo conjunto de datos dará lugar a los mismos resultados independientemente de dónde se analicen”, explican los autores.

En el artículo que acaba de publicar *Nature Biotechnology*, los autores muestran cómo *Nextflow* permite integrar los recursos más sofisticados para asegurar la reproducibilidad como, *Zenodo* para los datos, *Github* y *Docker* para el software y la computación en la nube. Este trabajo supone un punto de inflexión en la reproducibilidad de experimentos computacionales y un impulso para las buenas prácticas en el análisis de grandes conjuntos de datos. El CRG está comprometido con ayudar a promover este aspecto clave de la biología moderna entre la comunidad científica poniendo este nuevo recurso a disposición de la investigación académica pero también de producción clínica y comercial. También ha organizado una serie de talleres y cursos dedicados al uso de *Nextflow*.

Referencia:

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)

(*) Otros artículos citados en esta nota de prensa:

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).

Información sobre la financiación:

Este proyecto ha contado con el apoyo del Centro de Regulación Genómica, el Plan Nacional del Ministerio de Economía, Industria y Competitividad, 'Centro de Excelencia Severo Ochoa 2013-2017', el proyecto Elixir EXCELERATE y el proyecto OpenRiskNet de la Comisión Europea en el programa marco H2020. Maria Chatzou y Evan W Floden han contado con el apoyo de la Fundación Bancaria La Caixa Foundation. También se agradece el apoyo del programa CERCA de la Generalitat de Catalunya.

Imágenes disponibles en:

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

Pie de foto y créditos:

Nextflow CRG team.jpg => Investigadores del CRG desarrollan Nextflow, un nuevo método para que garantiza la reproducibilidad de los experimentos computacionales. De izquierda a derecha: Emilio Palumbo, Paolo Di Tommaso, Evan W Floden, Cedric Notredame y Pablo Prieto © CRG.

Contacto para medios:

Laia Cendros, Oficina de prensa, Centro de Regulación Genómica (CRG)

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