Entrevista a Jose Javier Lucas Lozano, doctor en biología e investigador en el CBMSO sobre enfermedades del sistema nervioso.

**1. ¿Cuál es su formación académica?**

Licenciado en Biología (1989) y Doctor en Biología (1993)

**2. ¿A lo largo de sus años de formación académica tuvo usted la oportunidad de aprender a programar en R o, por el contrario, ha sido más bien un aprendizaje independiente y complementario?**

En los años de mi formación académica, la bioinformática no era una herramienta generalizada y no se estudiaba R en la carrera. En la actualidad, dirijo un equipo de investigación de 8-10 personas ( <http://www.cbm.uam.es/LucasLab> ) y durante los últimos 6 años nos hemos embarcado en proyectos que se apoyan en la bioinformática. Para mí, R y similares eran totalmente desconocidos hasta hace poco, ahora los utilizo de manera rutinaria en el laboratorio, pero las tareas concretas las realizan mis colaboradores cuyas edades oscilan entre los 22 y los 40 años. De ellos, los mayores tampoco estudiaron R en la facultad, mientras que los más jóvenes si lo han estudiado en cierta medida en el grado y/o en el máster. En cualquier caso, cuando vemos que un miembro del equipo va a tener que hacer un uso importante de la bioinformática, le inscribimos en cursos de especialización para profesionales que ofrece nuestro propio organismo (el CSIC o la UAM).

**3. ¿Ha empleado R en algún otro trabajo aparte del actual?**

No

**4. ¿Qué ventajas ofrece R frente a otros lenguajes de programación como Python o Matlab?**

Aunque Python es cada vez más completo en cuanto a librerías de análisis estadístico, R tiene mucho más recorrido y por ello implementa un mayor número de funciones útiles en el tratamiento de datos numéricos. Consecuentemente, es también más sencillo encontrar soluciones a problemas poco frecuentes. R es también más ágil que Python en el procesamiento de estructuras de datos de grandes dimensiones.

**5. Si pudiera cambiar cualquier cosa que no le agrade especialmente de R, ¿qué cambiaría?**

El lenguaje R, si bien no es opaco, dista de la sencillez que ofrece Python para el no iniciado en programación, lo que dificulta alcanzar elevados niveles de manejo con esta herramienta. En nuestro caso, creemos que aproximar R más al propio inglés como hace Python sería un acierto.

**6. ¿Cree usted que R es un lenguaje de programación en auge? ¿Cuál cree que será su importancia dentro de 5 años?**

Creemos que hay diversas evidencias de que R no está en auge en este momento. Incluso aunque haya picos de uso está muy por debajo de la popularidad de Python en términos generales.

**7. ¿Qué importancia cree usted que tiene el saber manejar R para alguien que trabaja en la investigación?**

Actualmente, el saber manejar R tiene mucha importancia por los motivos expuestos en el punto 4. Existen muchas librerías muy potentes exclusivas de R que se usan a diario en publicaciones científicas. Uno de los paquetes de R más extensamente empleados es ggplot2 por poner un ejemplo. En nuestro caso, el análisis de datos de RNA-seq es clave para nuestra investigación y para ello empleamos varios paquetes de R como tximport, DEseq2, edgeR, etc.

**8. ¿En la investigación en la que está trabajando actualmente, para qué emplea usted R?**

Empleamos R para hacer cálculos rápidos sobre tablas extensas de datos, obteniendo desde estadísticos descriptivos, hasta correlaciones y por supuesto una amplia variedad de pruebas estadísticas. También es frecuente llevar a cabo estudios de agrupamientos muestrales jerárquicos a partir de datos genómicos, así como generar gráficos de diferentes tipos.

**9. Finalmente, ¿podría hablarnos brevemente sobre su investigación en el CBMSO acerca de enfermedades del sistema nervioso?**

En mi laboratorio estudiamos los mecanismos moleculares que subyacen a las enfermedades del sistema nervioso, tanto las neurodegenerativas como las psiquiátricas. De la mano de las recientes técnicas de secuenciación masiva de ADN y ARN, tanto de pacientes e individuos control, como de modelos animales y celulares, ha surgido un gran filón de información. Los datos crudos de esos experimentos de secuenciación se van depositando en repositorios de acceso público y, a menudo la información que contienen es útil para proyectos de investigación distintos a los que originaron la secuenciación de las muestras y podemos encontrar la respuesta a nuestras preguntas en experimentos realizados por otros. En otras ocasiones necesitamos generar la muestra biológica y secuenciarla nosotros mismos y también depositamos nuestros datos crudos que pueden ser de utilidad en otros proyectos. Tanto para analizar nuestros experimentos de RNA-seq, como para localizar, descargar y analizar los realizados por otros grupos, es necesaria la bioinformática. Gracias a ella, en los últimos años, en nuestro grupo hemos encontrado importantes claves moleculares sobre el autismo ( <https://www.csic.es/sites/default/files/16agosto2018_autismo_0.pdf> ) y la enfermedad de Huntington ( <https://www.biorxiv.org/content/10.1101/2020.05.11.086017v1> ).