

DExMA: meta-análisis de expresión génica e imputación de genes faltantes

Juan Antonio Villatoro-García^{1,2}, Jordi Martorell-Marugán^{1,2,3}, Daniel Toro-Domínguez¹, Yolanda Román-Montoya², Pedro Femia² and Pedro Carmona-Sáez^{1,2}

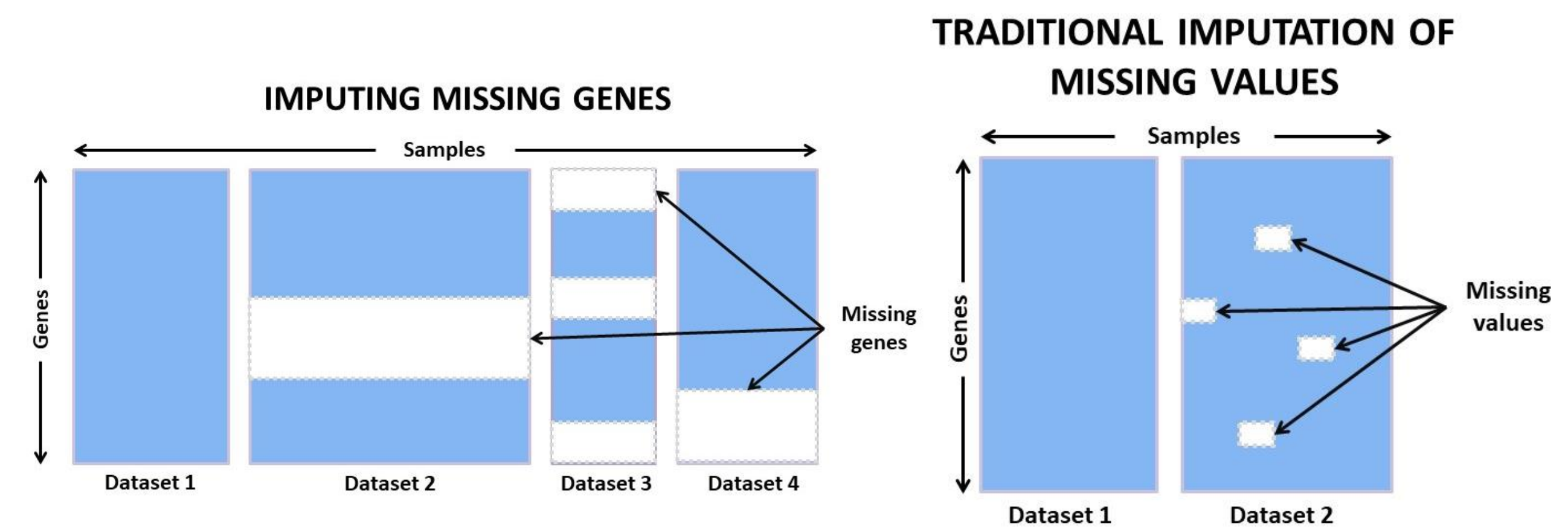
¹ GENYO. Centro de Genómica e Investigación Oncológica: Pfizer / University of Granada / Andalusian Regional Government. PTS Granada - Avenida de la Ilustración, 114 - 18016, Granada, Spain

² Departamento de Estadística e Investigación Operativa, Universidad de Granada, España.

³ Fundación para la investigación Biosanitaria de Andalucía Oriental-Alejandro Otero (FIBAO), Granada, 18012, España

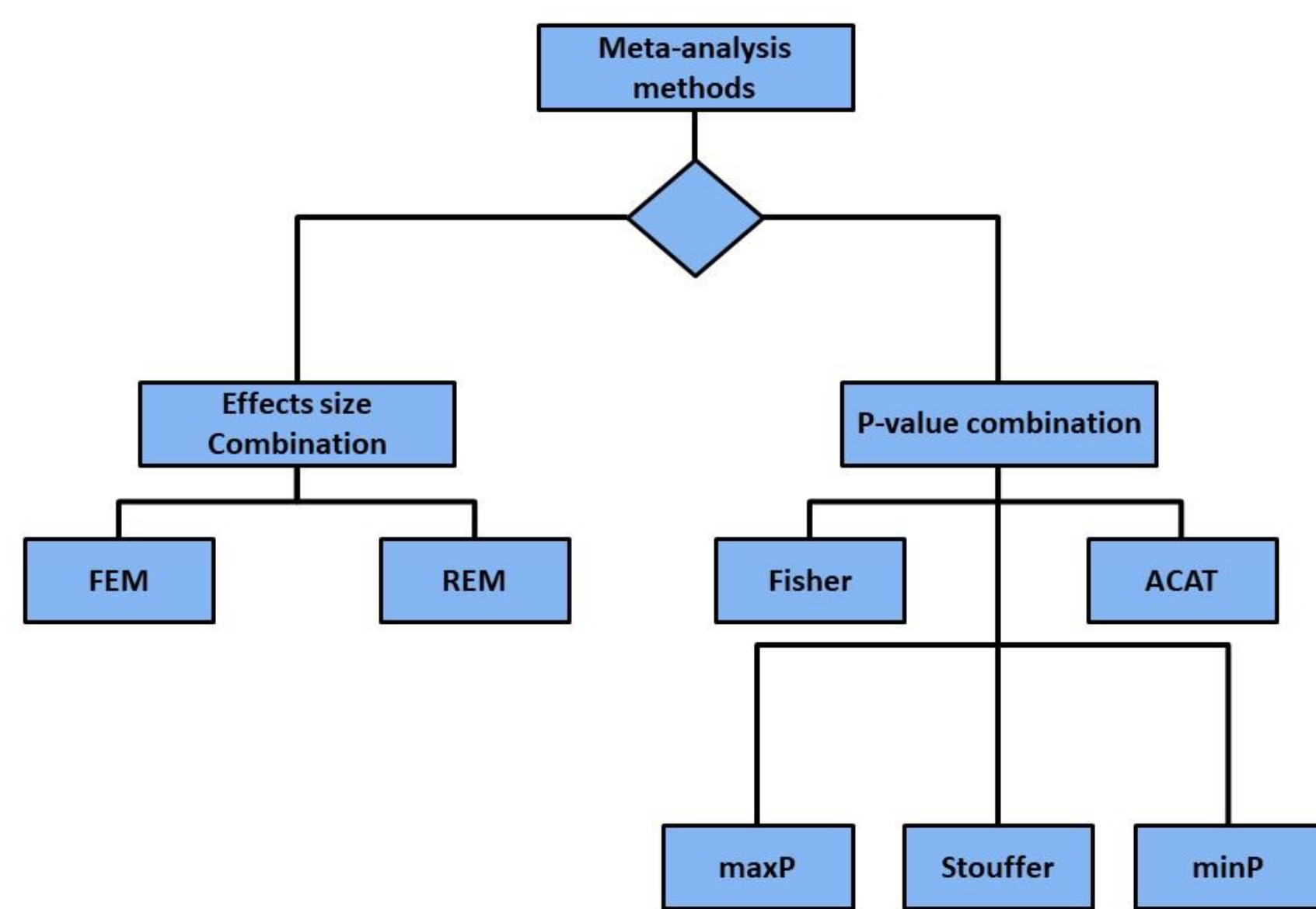
Introducción

En los últimos años, las técnicas de meta-análisis se han vuelto muy populares porque permiten a los investigadores combinar los resultados de diferentes estudios para obtener un efecto común. En el caso específico del campo de las transcriptómicas, el meta-análisis de expresión génica ha ganado gran relevancia debido al incremento del número de conjunto de datos que están disponibles en repositorios públicos. Sin embargo, el amplio uso de estas técnicas han revelado algunos problemas que normalmente ocurren debido al uso inadecuado de estos métodos. Uno de estos problemas es la posible existencia de genes faltantes, ya que trabajar solo con genes comunes puede producir pérdida de información. Para controlar este problema, hemos desarrollado DExMA, un paquete de R que realiza todos los pasos del meta-análisis de expresión génica. Además, DExMA permite controlar la existencia de valores faltantes y su imputación.

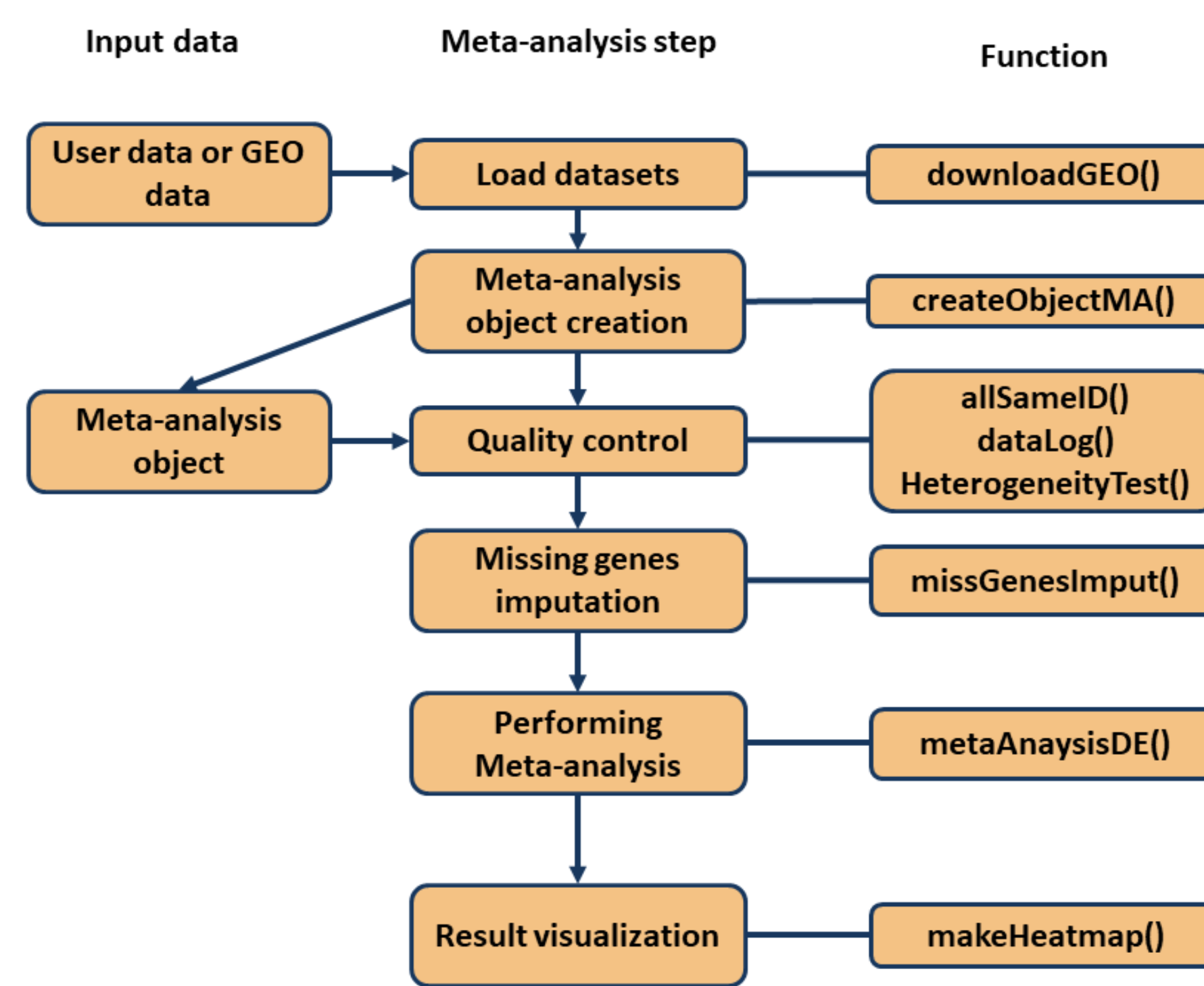


Métodos implementados

Los principales métodos del meta-análisis de expresión génica han sido implementados:

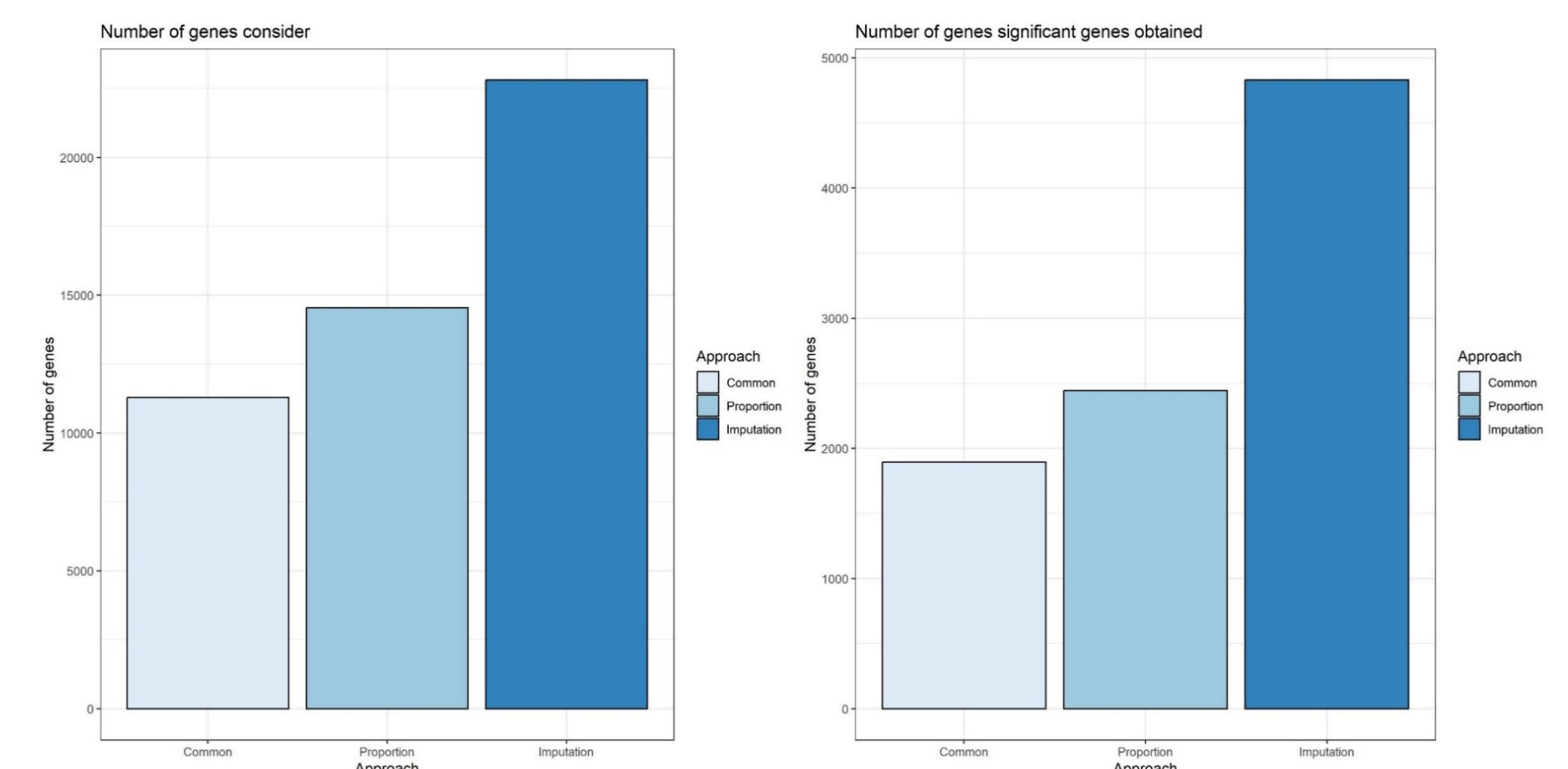


Flujo de Trabajo



Resultados

- En cada uno de los enfoques, se compararon el número de genes bajo estudio y el número de genes significativos obtenidos.
- Posteriormente, para verificar la significación biológica de los resultados, se llevó a cabo un análisis funcional



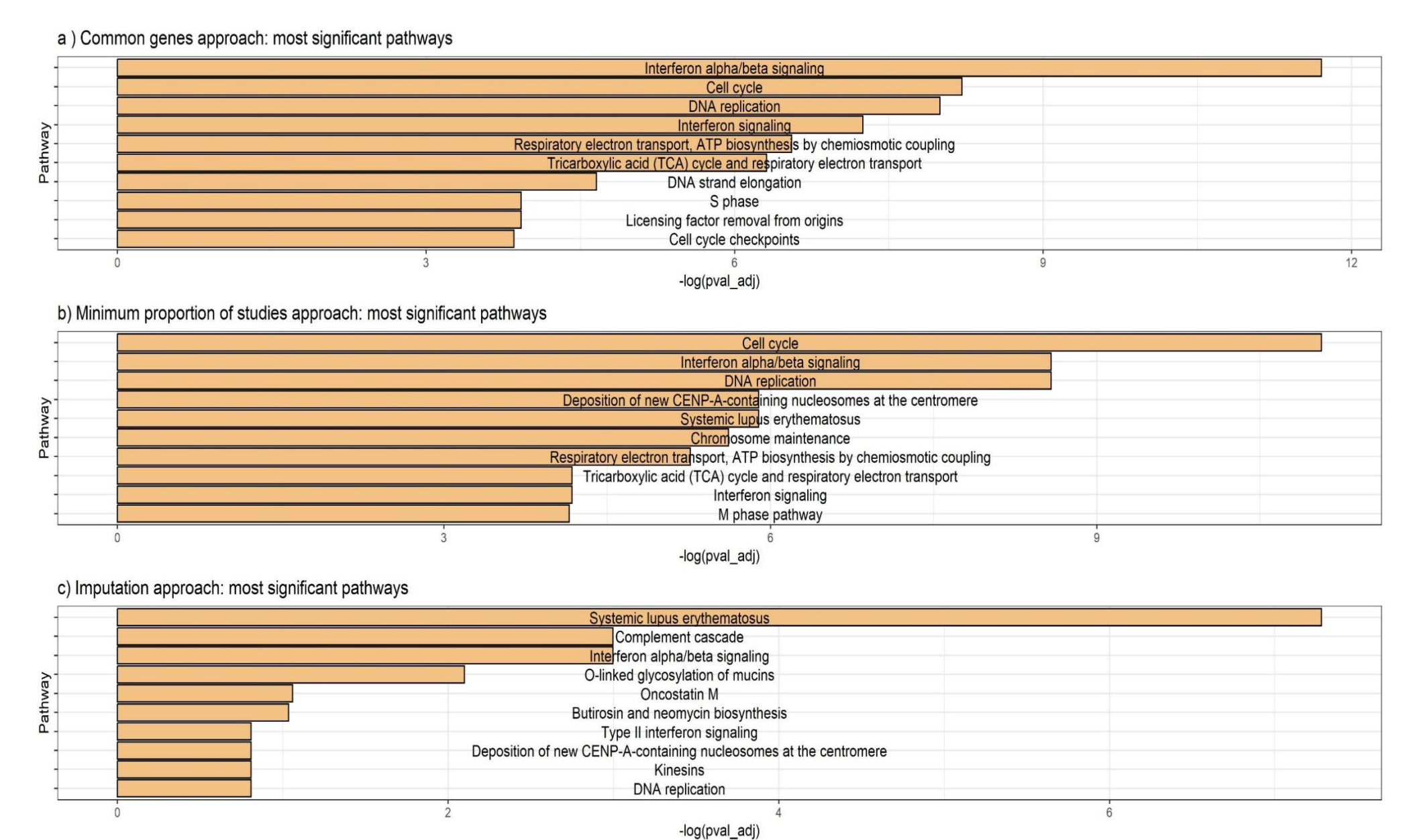
Ejemplo con datos reales

DExMA contiene dos métodos diferentes para controlar los genes faltantes:

- El primer enfoque consiste en llevar a cabo un meta-análisis con aquellos genes presentes en un mínimo número (o proporción) de estudios.
- La segunda técnica consiste en aplicar el método KNN en el espacio de las muestras para imputar la expresión de los genes faltantes.

DExMA ha sido utilizado para analizar tres conjuntos de datos públicos de Lupus Eritematoso Sistémico (LES) extraídos de la base de datos GEO. Para demostrar la utilidad del paquete, el meta-análisis se aplicó desde tres enfoques diferentes:

- Usando solo los genes comunes
- Considerando aquellos genes que estaban presentes en al menos dos conjuntos de datos.
- Realizando una imputación previa de los genes faltantes



Conclusiones

- El amplio uso del meta-análisis de expresión génica ha revelado algunos inconvenientes, como la pérdida de información debido a la existencia de conjuntos de datos con genes faltantes.
- El paquete DExMA permite realizar los diferentes pasos del meta-análisis de expresión génica controlando la posible existencia de genes faltantes.
- El uso de DExMA en datos reales demuestra la importancia de mantener tantos genes como sea posible para evitar la obtención de resultados poco realistas.

Publicación:

Villatoro-García, J.A.; Martorell-Marugán, J.; Toro-Domínguez, D.; Román-Montoya, Y.; Femia, P.; Carmona-Sáez, P. DExMA: An R Package for Performing Gene Expression Meta-Analysis with Missing Genes. *Mathematics* 2022, 10, 3376. <https://doi.org/10.3390/math10183376>

Financiación:

Ayudas del plan propio UGR 2022 PPIJA2022-14



SCAN ME