# RFLOMICS: Interactive web application for multi-omics data analysis

Nadia Bessoltane [*] Audrey Hulot [*] Gwendal Cueff [*] Christine Paysant-Le-Roux[†] Delphine Charif [*]

## Résumé (max 300 mots)

Afin de mieux comprendre les processus biologiques complexes, des données moléculaires à différentes échelles de la cellule, appelées « données multi-omiques », sont produites en masse. Ces données omiques décrivent par exemple l'expression des gènes (données transcriptomique), l'abondance des protéines (données protéomique) ou l'abondance des métabolites (données métabolomique). Ces données sont hétérogènes, de grande dimension et souvent bruitées. Elles sont acquises selon un plan d'expérience commun construit autour de l'étude d'un ou plusieurs processus biologiques. On distingue les analyses single-omics (une table) des analyses multi-omics (plusieurs tables).

L'étude de chaque couche omique constitue une première étape intéressante pour explorer et extraire la variabilité biologique pertinente au regard du processus étudié et réduire la dimension des tables.

Une fois cette étape effectuée, l'analyse conjointe des tables est envisageable, pour lier entre elles les différentes couches d'omiques. L'analyse multi-omics constitue un domaine de recherche actif, les caractéristiques de ces données demandant des traitements particuliers.

Une telle analyse de données multi-tableaux hétérogènes, reste un défi technique qui nécessite des méthodes pertinentes et des paramètres adaptés aux données, ainsi que des méthodes de visualisations adéquates et une gestion rigoureuse de l'environnement d'analyse.

C'est dans ce contexte que RFLOMICS a été développé : pour permettre d'harmoniser les pratiques, de gagner du temps sur le code et de garantir la reproductibilité des analyses en s'appuyant sur un pipeline utilisant des méthodes et des paramètres expertisés. RFLOMICS est un package R avec une interface shiny, qui permet d'analyser de manière guidée trois types d'omiques (transcriptomique, protéomique et metabolomique), acceptant plusieurs tables par type d'omique. Il permet de prendre en compte jusqu'à trois facteurs biologiques et deux facteurs techniques. L'interface guide l'utilisateur dans toutes les étapes de l'analyse, de la création de son modèle jusqu'à la génération d'un rapport html contenant toutes les étapes qu'il aura effectuées.

**Mots-clefs** : Biostatistiques – omiques – Rshiny – Reproductibilité - package

## Développement

RFLOMICS prend comme données d'entrée des tableaux d'omiques issus d'un même plan expérimental : les variables descriptives des individus sont communes à tous les tableaux d' omiques. Le package offre la possibilité de réaliser toutes les étapes classiques de l'analyse single-omics : contrôle qualité des données, analyse multivariée, normalisation et transformation, régression linéaire généralisée, clustering, aide à l'interprétation biologique.

* Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin, 78000, Versailles, France, {prenom.nom}@inrae.fr

† Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France, christine.paysant-le-roux@inrae.fr

Les méthodes utilisées sont adaptées en fonction des types d'omiques indiqués par l'utilisateur pour chaque table. A l'issue des analyses single-omics, l'interface offre la possibilité de l'analyse multi-omics, où une méthode supervisée et une méthode non supervisée sont proposées à l'utilisateur. Pour toutes ces méthodes, les fonctions de RFLOMICS s'appuient sur des packages du CRAN et Bioconductor.

Le package est composé de deux couches : une première couche constituée des méthodes du package, utilisables en ligne de commande, et une deuxième couche composant l'interface shiny.

Les méthodes ont été codée en programmation orientée objet. L'utilisation de la classe d'objets MultiAssayExperiment (MAE) [1], de type S4, spécialement conçu pour des analyses en biologie, permet de gérer plusieurs tableaux de natures différentes. Chaque table et ses informations sont consignées sous la forme d'un SummarizedExperiment (SE) [2]. Les analyses spécifiques de chacun des tableaux sont ajoutées aux slots *metadatas* des SE correspondants. Les analyses multi-omiques sont, quant à elles, stockées dans les *metadatas* du MAE global. A l'exception de fonctions utilitaires, la majorité des méthodes de RFLOMICS sont implémentées dans des fonctions prenant en entrée un SE ou un MAE et retournant ces mêmes classes, permettant une flexibilité dans les analyses effectuées. L'étape de clustering pouvant être demandeuse en ressources, un lien avec un cluster de calcul peut être fait via le package clustermq [3] pour déporter les calculs.

En ce qui concerne la couche shiny, chaque étape d'analyse est programmée sous forme d'un module shiny faisant appel aux méthodes appropriées côté serveur. Cette programmation en modules permet à l'utilisateur d'analyser les différentes tables de données en parallèle. Cette architecture est flexible et facilement adaptable, permettant de rajouter de nouveau types d'omiques ou de nouvelles étapes d'analyse selon les demandes et l'évolution des technologies. Le public cible du package comportant des personnes peu habituées à l'analyse en ligne de commande, la plupart des paramètres des méthodes ont été expertisés et entrés par défaut et n'apparaissent pas au niveau de l'interface, rendant celle-ci plus conviviale pour l'utilisateur.

Enfin, un rapport Rmarkdown peut être généré par l'utilisateur. A chaque module d'analyse correspond un rapport 'child'. La compilation du rapport complet est donc possible quelle que soit l'étape d'analyse et ne comportera que les étapes validées par l'utilisateur.

Le package est actuellement disponible sur un dépôt gitlab. Pour pallier les changements de version des packages et les problèmes de compatibilités, une solution de déploiement d'une image docker est envisagée, notamment via le projet sk8 [4].

## Références

[1] Ramos M, *et al*. Software for the Integration of Multiomics Experiments in Bioconductor. Cancer Res. 2017 Nov 1;77(21):e39-e42. doi: 10.1158/0008-5472.CAN-17-0344.
[2] Morgan M, Obenchain V, Hester J, Pagès H (2022). *SummarizedExperiment: SummarizedExperiment container*. R package version 1.28.0, https://bioconductor.org/packages/SummarizedExperiment.
[3] M Schubert. clustermq enables efficient parallelisation of genomic analyses. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz284
[4] https://sk8.inrae.fr/index.html