Plant & Food™
Research
Rangahau Ahumāra Kai

Reproducible multi-omics integration with the {moiraine} R package

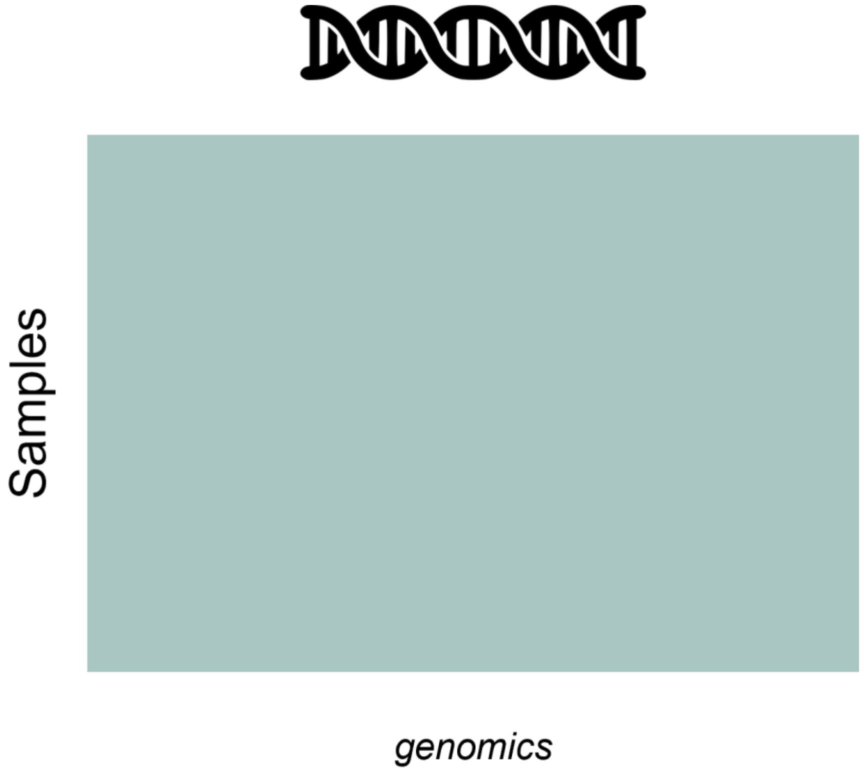
Olivia Angelin-Bonnet

R Exchange 2025 | 15 May 2025

The New Zealand Institute for Plant and Food Research Limited

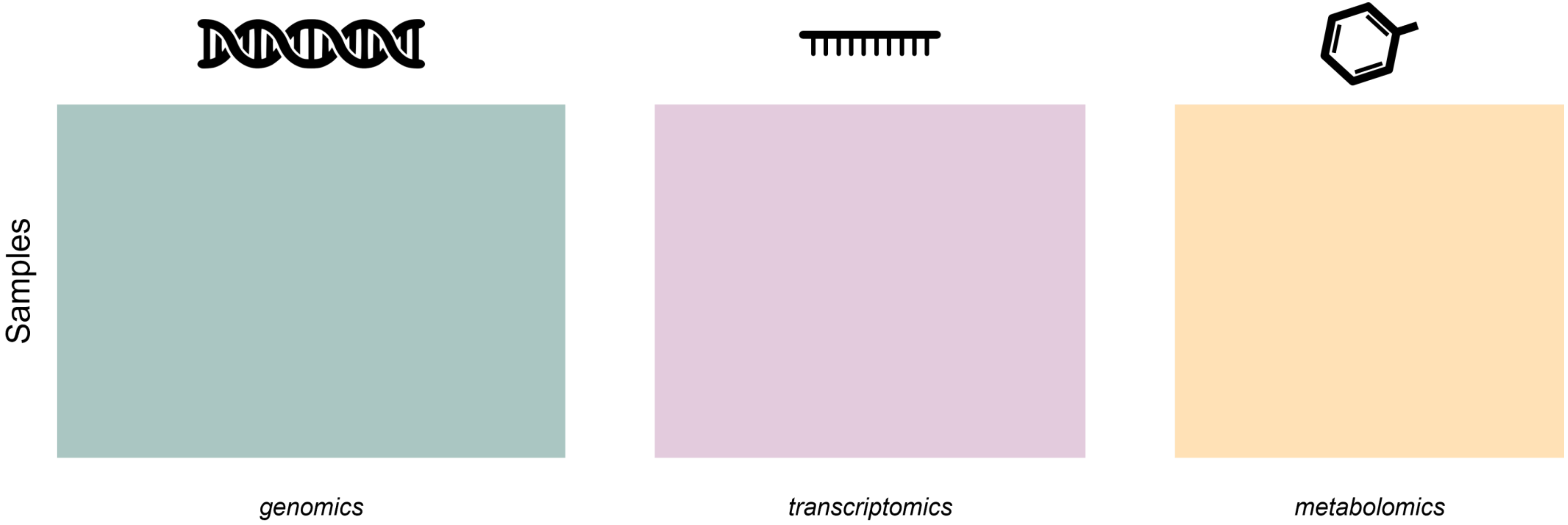
Multi-omics data

Omics data: high-throughput measurement of molecular layer across biological samples

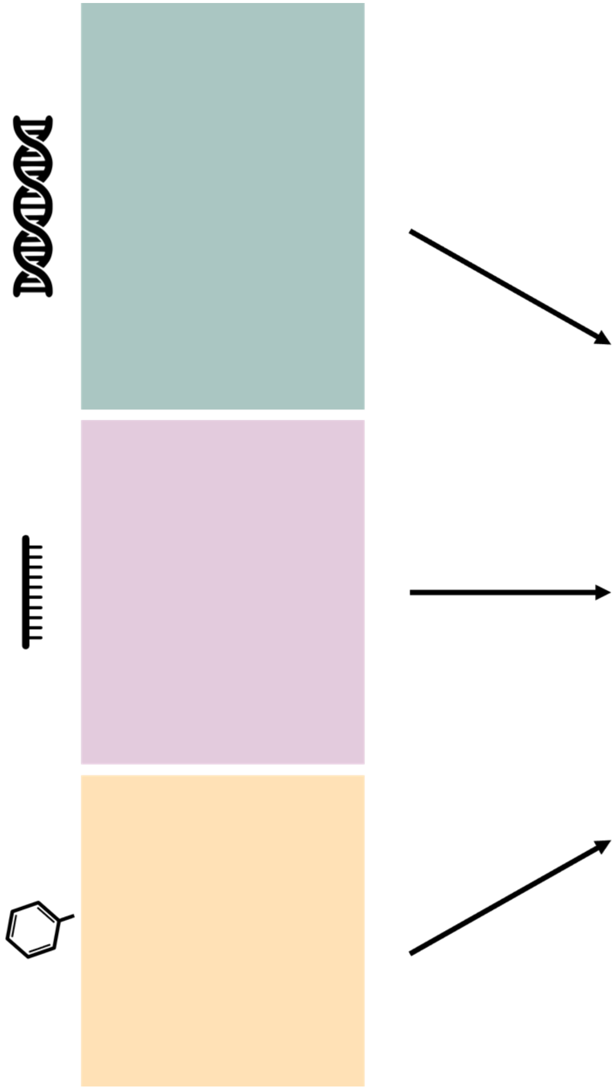


Multi-omics data

Multi-omics data: different omics layers measured on the same biological samples



Multi-omics integration



Multi-omics integration methods:

- DIABLO – {mixOmics}
- MOFA – {MOFA2}
- MCIA – {omicade4}
- omeSOM
- MFA – {FactoMineR}
- …

Technical challenges

Independent implementation of methods leads to:

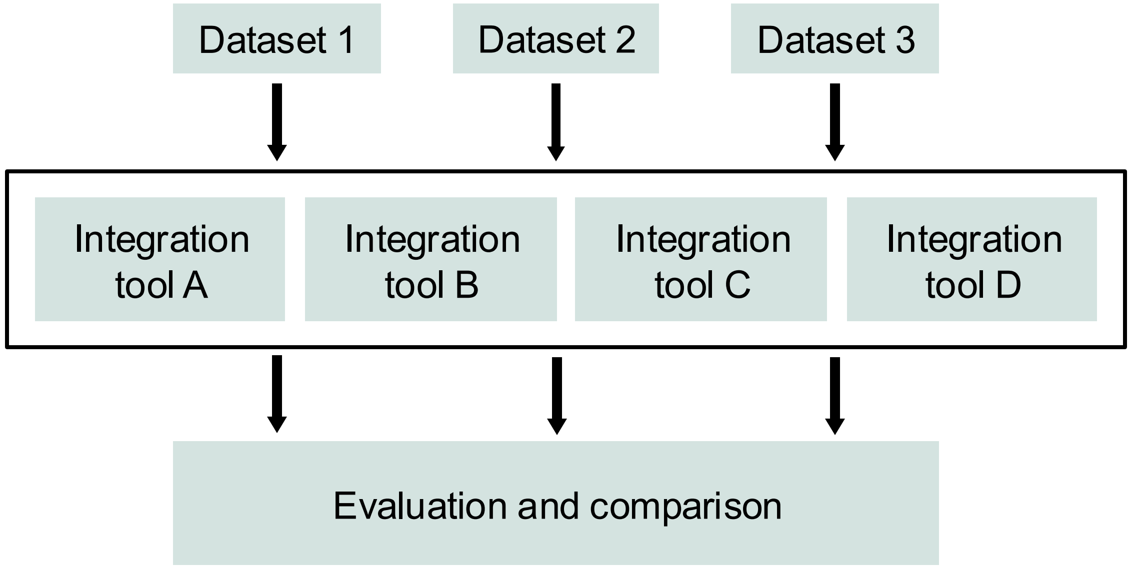
Different requirements for input data format

Different formatting of results

Unequal visualisation capabilities

Multi-omics integration comparison project

- Goal: develop a framework to enable the comparison of integration tools
- Lots of data wrangling and repetitive tasks
- Can be formalised as an R package!



Dataset 1

Dataset 2

Dataset 3

Integration tool A

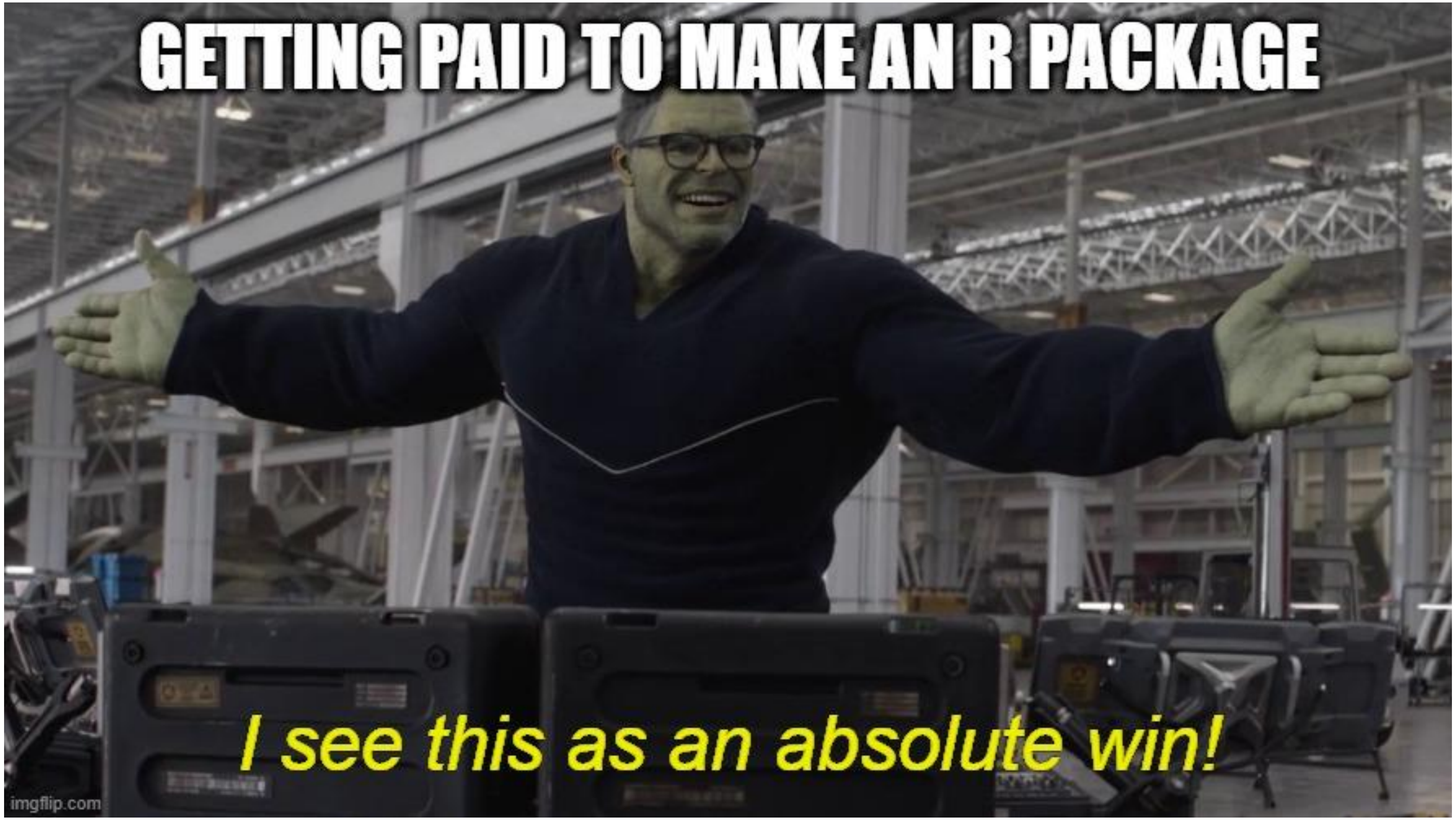
Integration tool B

Integration tool C

Integration tool D

Evaluation and comparison

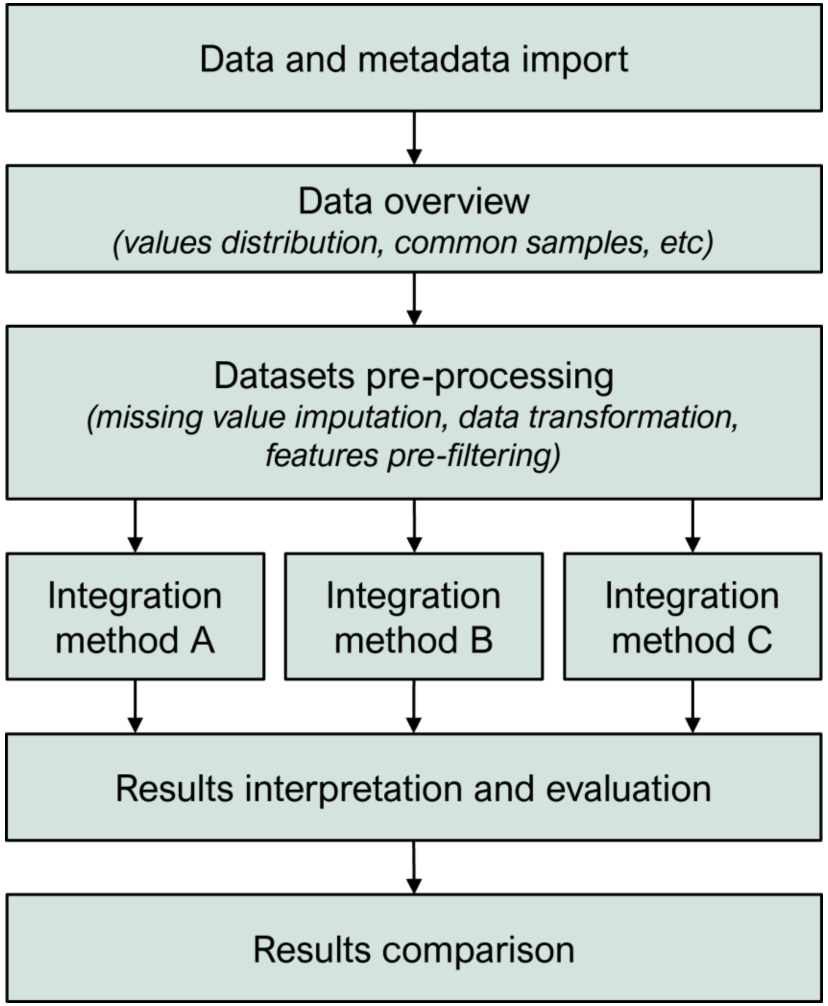
Multi-omics integration comparison project



The {moiraine} package

Construction of reproducible pipelines for testing and comparing multi-omics integration methods

https://github.com/Plant-Food-Research-Open/moiraine



Data and metadata import

Data overview
(values distribution, common samples, etc)

Datasets pre-processing
(missing value imputation, data transformation, features pre-filtering)

Integration method A

Integration method B

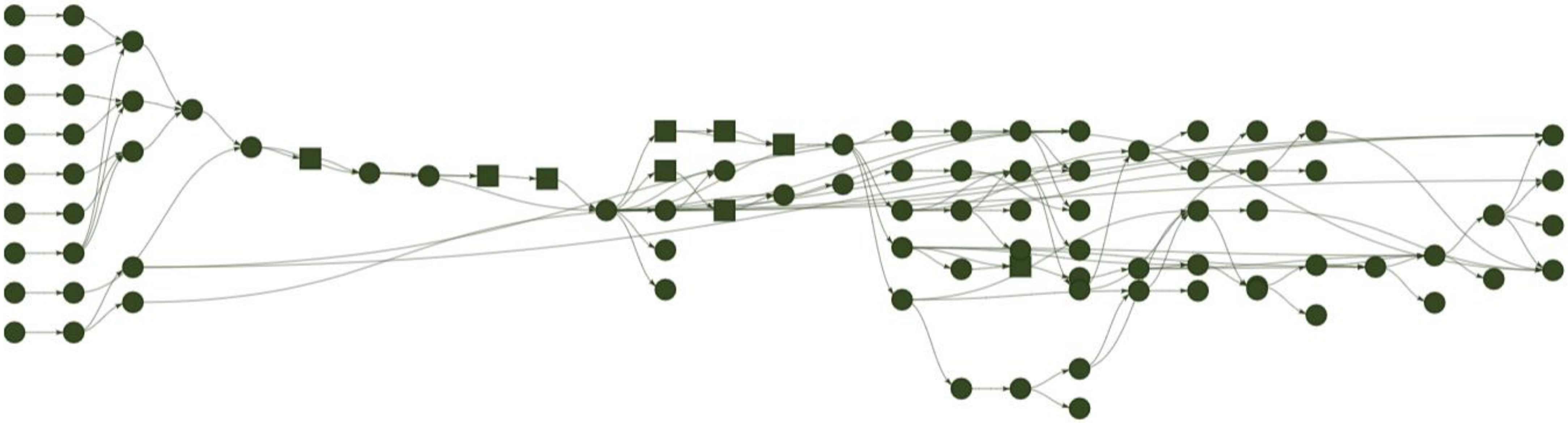
Integration method C

Results interpretation and evaluation

Results comparison

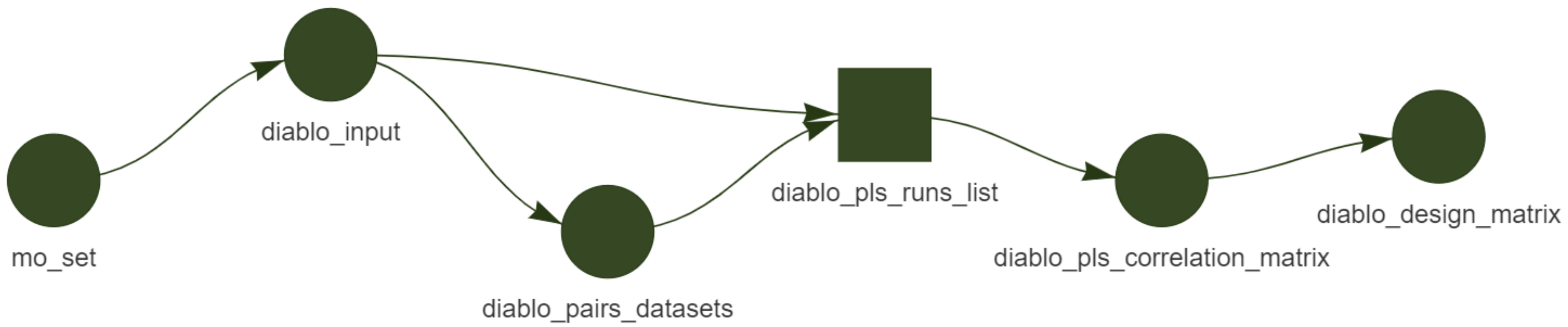
Reproducible pipelines with {targets}

Uses the {targets} package to build reproducible analysis pipelines

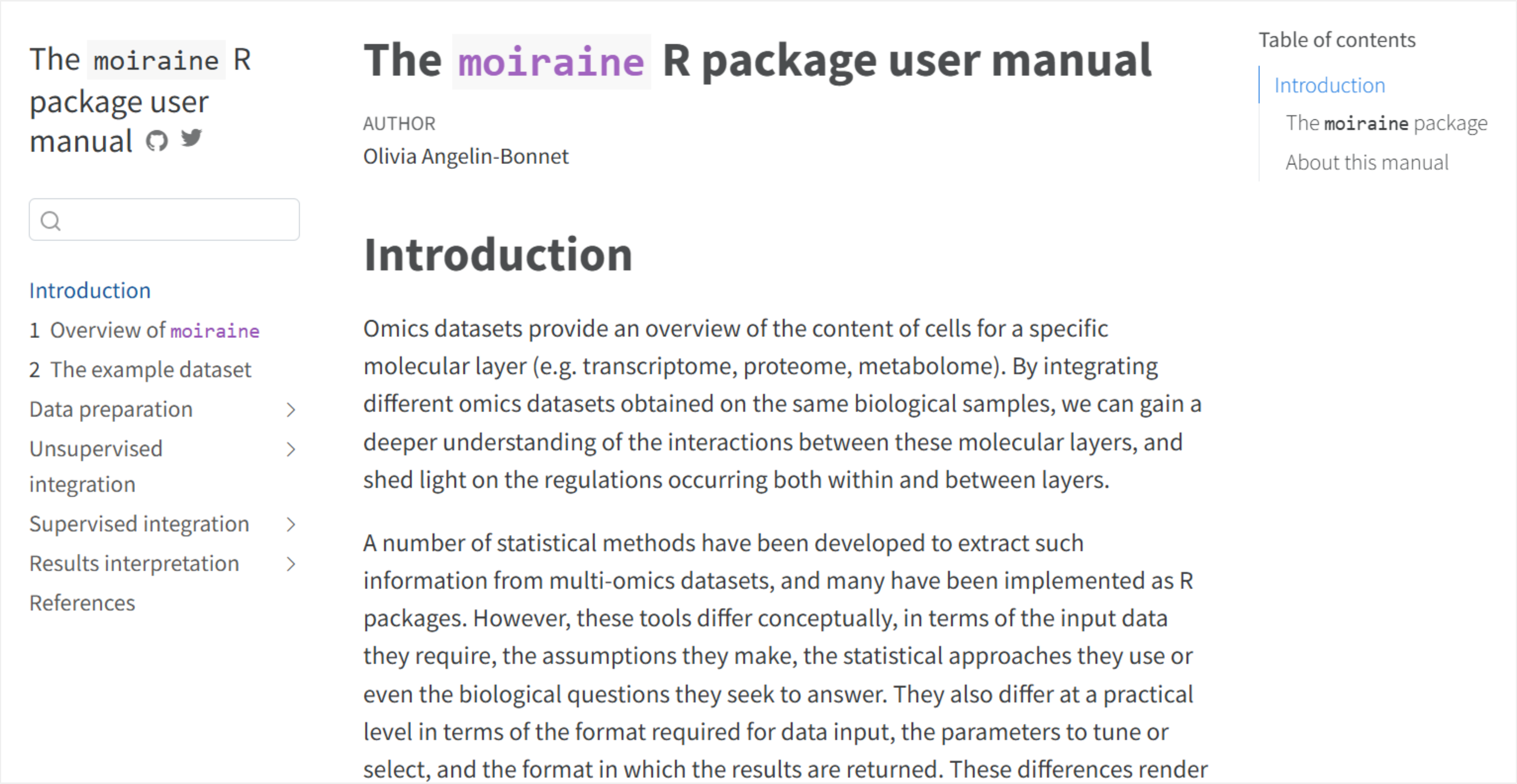


Reproducible pipelines with {targets}

Target factories to automate multi-step processes (e.g. estimation of a parameter of interest):



Documentation to facilitate users' journey



The `moiraine` R package user manual

AUTHOR
Olivia Angelin-Bonnet

Introduction

Omics datasets provide an overview of the content of cells for a specific molecular layer (e.g. transcriptome, proteome, metabolome). By integrating different omics datasets obtained on the same biological samples, we can gain a deeper understanding of the interactions between these molecular layers, and shed light on the regulations occurring both within and between layers.

A number of statistical methods have been developed to extract such information from multi-omics datasets, and many have been implemented as R packages. However, these tools differ conceptually, in terms of the input data they require, the assumptions they make, the statistical approaches they use or even the biological questions they seek to answer. They also differ at a practical level in terms of the format required for data input, the parameters to tune or select, and the format in which the results are returned. These differences render

https://plant-food-research-open.github.io/moiraine-manual/

Documentation to facilitate users’ journey

moiraine

`moiraine` is a package for facilitating the construction of a reproducible analysis pipeline for multi-omics data integration. It provides functions to automate data import, pre-processing, transformation, integration through several tools, as well as interpretation and comparison of the integration results. It relies on the targets package to generate reproducible workflows.

Overview

The workflow for a typical multi-omics integration analysis handled with `moiraine` includes the following steps:

- Data import: this covers the import of omics measurements as well as associated metadata (i.e. information about the omics features and samples) – moiraine relies on the `MultiDataSet` package to store this information in a consistent format;
- Inspection of the omics datasets: including checking values density distribution, samples overlap between omics datasets, or presence of missing values;
- Preprocessing of the omics datasets: missing values imputation, transformation, and pre-filtering of samples and omics features;
- Integration of the omics datasets by one or more of the supported tools; currently, the following integration methods are covered in `moiraine`:
  - sPLS and DIABLO from the `mixOmics` package
  - sO2PLS from the `OmicsPLS` package
  - MOFA and MEFISTO from the `MOFA2` package
- Interpretation of the integration results using standardised visualisations enriched with features and samples metadata;
- Comparison of the integration results obtained by different methods or pre-processing approaches.

An overview of the capabilities of the package is available here.

Installation

You can install the development version of moiraine from GitHub with:

Links

Browse source code

Report a bug

License

Full license

MIT + file LICENSE

Citation

Citing moiraine

Developers

Olivia Angelin-Bonnet
Author, maintainer

Dev status

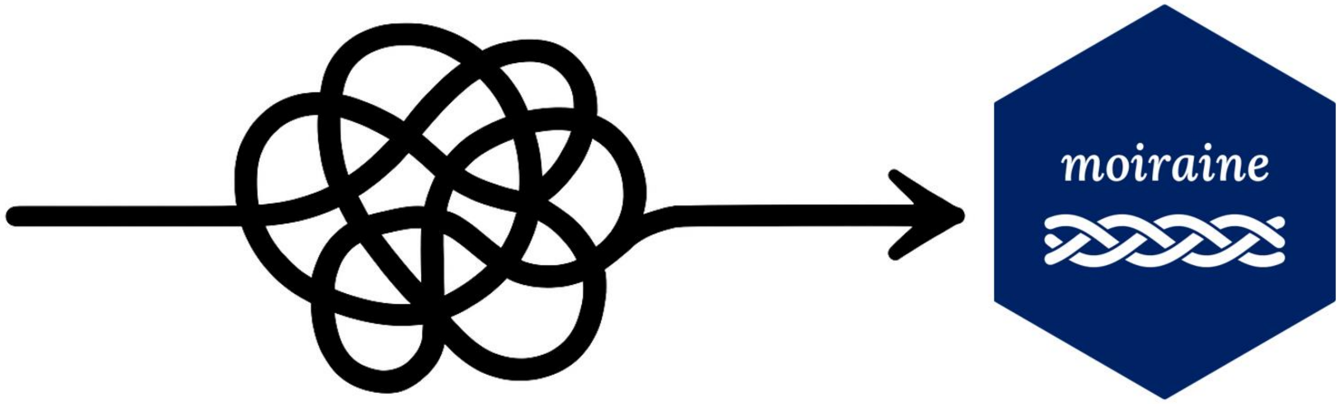
- Additional layer of test for the package
- Highlights gaps and frictions
- Makes you think like a user
- Advertisement for your package

Next steps

- {moiraine} currently available on GitHub
- Publication as a software article
- Submission to CRAN, Bioconductor or R-Universe… but:

```
Imports:
    Biobase,
    circlize,
    ComplexHeatmap,
    dplyr,
    GGally,
    ggforce,
    ggnewscale,
    ggplot2,
    ggpubr,
    ggrepel,
    grid,
    methods,
    mixOmics,
    MOFA2,
    MultiDataSet,
    OmicsPLS,
    patchwork,
    pcaMethods,
    purrr,
    RColorBrewer,
    readr,
    rlang,
    scales,
    stats,
    stringi,
    stringr,
    tarchetypes,
    targets,
    textclean,
    tibble,
    tidyr,
    tidyselect,
    utils
Suggests:
    bestNormalize,
    BiocParallel,
    cluster,
    corrplot,
    DESeq2,
    DT,
    gage,
    GenomicFeatures,
    ggbeeswarm,
    grDevices,
    hexbin,
    htmltools,
    kableExtra,
    knitr,
    plotly,
    rmarkdown,
    rtracklayer,
    testthat (>= 3.0.0),
    UpSetR,
    vsn
```

Conclusion



- Improved my R programming skills a lot! From statistician-who-codes to software engineer-ish
- Important for managers to recognise the need for projects dedicated to tool development
- {targets} = 😍

Resources

- R Packages (2nd edition) by Hadley Wickham and Jennifer Bryan
- Code smells and feels talk by Jenny Bryan at useR! 2018
- Beautiful Code, Because We're Worth It! blog post by Maëlle Salmon and Yanina Bellini Saibene for ROpenSci
- Simple techniques for writing readable code talk by Joachim Wuttke at LMU Open Science Center Summer school 2023

Acknowledgements

- Susan Thomson
- Roy Storey
- Lindy Guo
- Nigel Joyce
- Blake List
- Leonardo Salgado

moiraine

Plant & Food™
Research
Rangahau Ahumāra Kai

Thank you

olivia.angelin-bonnet@plantandfood.co.nz

plantandfood.co.nz

A smart
green
future.
Together.

Presentation disclaimer

Presentation for
Rexchange 2025, Wellington, New Zealand, 15 May 2025

Publication data:
Angelin-Bonnet Olivia May 2025. Reproducible multi-omics integration with the {moiraine} R package. A Plant & Food Research PowerPoint presentation.
SPTS No. 27088.

Presentation prepared by:
Olivia Angelin-Bonnet
Statistical Scientist, Statistical Science
May 2025

Presentation approved by:
Mark Wohlers
Science Group Leader, Data Science
May 2025

For more information contact:
Olivia Angelin-Bonnet
DDI: +64 6 355 6156
Email: olivia.angelin-bonnet@plantandfood.co.nz

DISCLAIMER

The New Zealand Institute for Plant and Food Research Limited does not give any prediction, warranty or assurance in relation to the accuracy of or fitness for any particular use or application of, any information or scientific or other result contained in this presentation. Neither The New Zealand Institute for Plant and Food Research Limited nor any of its employees, students, contractors, subcontractors or agents shall be liable for any cost (including legal costs), claim, liability, loss, damage, injury or the like, which may be suffered or incurred as a direct or indirect result of the reliance by any person on any information contained in this presentation.

COPYRIGHT

© COPYRIGHT (2025) The New Zealand Institute for Plant and Food Research Limited. All Rights Reserved. No part of this report may be reproduced, stored in a retrieval system, transmitted, reported, or copied in any form or by any means electronic, mechanical or otherwise, without the prior written permission of The New Zealand Institute for Plant and Food Research Limited. Information contained in this report is confidential and is not to be disclosed in any form to any party without the prior approval in writing of The New Zealand Institute for Plant and Food Research Limited. To request permission, write to: The Science Publication Office, The New Zealand Institute for Plant and Food Research Limited – Postal Address: Private Bag 92169, Victoria Street West, Auckland 1142, New Zealand; Email: SPO-Team@plantandfood.co.nz.

This report has been prepared by The New Zealand Institute for Plant and Food Research Limited (Plant & Food Research).
Head Office: 120 Mt Albert Road, Sandringham, Auckland 1025, New Zealand, Tel: +64 9 925 7000, Fax: +64 9 925 7001.
www.plantandfood.co.nz