Multi-omic data science with R/Bioconductor

University of Oulu & University of Turku

2022-04-23

Contents

Chapter 1

Overview

Welcome to Oulu Summer School, June 2022

Venue University of Oulu. June 20-23, 2022. Organized together with University of Turku, Finland.

Figure source: Moreno-Indias *et al.* (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

1.1 Introduction

This course is based on data science with R, a popular open source environment for scientific data analysis. It provides general capabilities for the analysis, integration, and visualization of multi-omic data from biomedical studies.

The course is based on *miaverse* (mia = **MI**crobiome **A**nalysis), an R/Bioconductor framework for microbiome data science.

The data science framework consists of the following elements:

- efficient multi-omic data container
- a package ecosystem, providing algorithmic data analysis methods
- demonstration data sets
- open documentation

The framework is explained in a greater depth in the online book Orchestrating Microbiome Analysis. The book is currently a development version.

1.2 Learning goals

The course aims to provide the basic understanding and skills for biomedical data analysis with R and Bioconductor. The course provides an overview of reproducible data analysis and reporting workflow in multi-omic studies, with recent example data sets published microbiome studies.

The participants become familiar with standard concepts and methods in multi-omic data analysis, in particular in human microbiome research. This includes better understanding of the specific statistical challenges, practical hands-on experience with the commonly used methods, and reproducible research with R.

After the course you will know how to approach new tasks in microbiome data science by utilizing available documentation and R tools.

Target audience Advanced MSc and PhD students, Postdocs, and biomedical researchers who wish to develop their skills in scientific programming and biomedical data analysis.

1.3 Schedule

The course takes place daily between 9am – 5pm, including coffee and lunch breaks.

The course will be organized in a live format but the material will be openly available online during and after the course.

A priority will given for local students from Oulu. Participants from other higher education institutions are welcome to apply.

The mornings will start with lectures, and afternoons are mainly dedicated to hands-on sessions that consist of practical tasks and example data from recent research literature. Students will solve the exercises based on available online examples and resources that are pointed out in the study material. There is often more than one way to solve a given task, and the teachers will be available for assistance.

Monday:

- Orientation
- Best practices in reproducible reporting and open science
- Hands-on: Introduction to R

Tuesday:

- Key concepts and challenges in biomedical data analysis
- Hands-on: Biomedical data exploration

Wednesday:

- Key concepts in biomedical data visualization
- Hands-on: Biomedical data visualization

Thursday:

- Advanced topics: common machine learning techniques
- Hands-on: multi-omic data integration and reproducible workflows

Friday:

- Student presentations
- Summary & Conclusions

1.4 Material

The teaching material follows open online documentation created by the course teachers, extending the online book Orchestrating Microbiome Analysis (https://microbiome.github.io/OMA). We will teach generic data analytical skills that are applicable to common data analysis tasks encountered in modern omics research.

The training material walks through example workflows that go through standard steps of biomedical data analysis covering data access, exploration, analysis, visualization, reproducible reporting, and best practices in open science. The teaching format allows adaptations according to the student's learning speed.

You can run the workflow by simply copy-pasting the examples. For further, advanced material, you can test and modify further examples from the OMA book, or try to apply the techniques to your own data.

We expect that the students will install the necessary software in advance. Online support will be available.

Chapter 2

Organizers

Jointly organized by:

- Health and Biosciences Doctoral Programme, University of Oulu Graduate School (HBS-DP)
- Department of Computing, University of Turku

Supported by:

- IT Center for Science (CSC), Finland

Teachers

- Leo Lahti is the main teacher and Associate Professor in Data Science at the University of Turku, with specialization on biomedical data analysis.

- Tuomas Borman is research assistant and one of the main developers of the open training material covered by the course.

- Jenni Hekkala is a local PhD researchers from Oulu who will coordinate the local arrangements and contribute to course teaching.

2.1 Acknowledgments

Citation "Introduction to microbiome data science (2021). URL: https://microbiome.github.io".

Borman et al. (2022)

We thank all miaverse developers and contributors who have contributed open resources that supported the development of the training material.

Contact Leo Lahti, University of Turku, Finland

License All material is released under the open CC BY-NC-SA 3.0 License and available online during and after the course, following the recommendations on open teaching materials of the national open science coordination in Finland**.

Source code

The source code of this repository is fully reproducible and contains the Rmd files with executable code. All files can be rendered at one go by running the file main.R. You can check the file for details on how to clone the repository and convert it into a gitbook, although this is not necessary for the training.

- Source code (github): miaverse teaching material
- Course page (html): miaverse teaching material

Chapter 3

Program

The course takes place on each working day from 9am – 1pm (CEST). Short breaks will be scheduled between sessions.

The hands-on sessions consist of a set of questions and example data. Solve the exercises by taking advantage of the online examples and resources that are pointed out in the study material. There is often more than one way to solve a given task. It is assumed that you have already installed the required software. Do not hesitate to ask support from the course assistants.

3.1 Day 1: from raw sequences to ecological data analysis

Lectures

- Microbiota analysis: association studies vs. causality; microbiota sequencing methods (16S, shotgun, metagenomics) - by dr. Tom Ederveen (Radboud UMC Nijmegen, The Netherlands)

- DNA isolation and 16S rRNA gene sequencing; bioinformatics step 1: from raw sequences to OTU table in a biom file – by Tom Ederveen (Radboudumc Nijmegen, The Netherlands)

Demo & Practical

- Importing data to R for interactive data analysis

- Task: initialize reproducible report

3.2 Day 2 - Alpha diversity

Demo

- Microbiome data exploration

Lecture

- Key concepts in microbiome data science

Practical

- Alpha diversity: estimation, analysis, and visualization

3.3 Day 3 - Beta diversity

Demo

- Community similarity

Practical

- Beta diversity: estimation, analysis, and visualization

3.4 Day 4- Differential abundance

Lecture

- Differential abundance analysis methods

Practical

- Differential abundance in practice

Lecture

- Overview of microbiota data science methods & concepts

3.5 Day 5 : Presentations & closing

Student presentations on microbiome data analytics

Bibliography

Borman, T., Hekkala, J., Reunanen, J., and Lahti, L. (2022). *Multi-omic data science with R/Bioconductor.* Oulu Summer School.