Workshop
R For Genomics

Organizer

CUBiDA (Core Unit Bioinformatik, Datenintegration und Analyse) at the Universitätsklinikum Erlangen

Instructors

Manoj Kumar Selvaraju and Leila Taher

Outline

This workshop will introduce participants to the fundamentals of using R for genomics data analysis. The focus will be on good coding practices, data wrangling with dplyr and tidyr, and data visualization with ggplot2. Participants will learn to work effectively with common genomics data formats and develop the skills necessary for independent analysis within the R environment.

Outcome

Upon successful completion of this workshop, participants will be able to:

- Understand and apply good coding practices for reproducibility and maintainability.
- Efficiently load and handle various genomic data formats.
- Master data manipulation techniques using the dplyr package.
- Effectively reshape and tidy data using the tidyr package.
- Create informative and visually appealing plots using the ggplot2 package.
- Customize plots with titles, labels, legends, colors, themes, and scales.
- Generate various plot types relevant to genomics.
- Develop an understanding of common data structures and analyses in genomics research.
- Gain the confidence and skills to independently conduct basic genomics data analysis in R.

Number of participants

5-15

Target audience

This workshop is designed for biomedical researchers with an interest in analyzing genomic data in R.

Prerequisites

This workshop assumes a foundational understanding of R programming, including familiarity with R syntax and data structures, such as vectors and data frames. Participants should have experience with basic R functions as well as a basic understanding of how to use control flow structures and define custom functions. Participants who are completely new to R will find it beneficial to complete an introductory R programming course before attending this workshop.

Teaching format

This workshop will combine interactive lectures with hands-on exercises. All participants will have access to dedicated workstations equipped with the necessary software and datasets. We will also provide you with all teaching materials, including presentations, scripts, and data files, for your future reference.

Provisional workshop schedule

DAY 1	
8:30 am-4:30 pm	Setting up your R environment (package installation and management, project setup and organization). Writing clean, readable, and reproducible code. Commenting and documenting code effectively. Review of Basic R Programming concepts (data structures, control flow, functions). Introduction to common genomics data and file formats.

	Loading data into R. Working with data frames. Data Wrangling with dplyr. Tibbles. Data manipulation verbs (filter, select, mutate, arrange, rename, group_by, and summarize). The pipe operator. Two-table verbs (inner_join, left_join, right_join, full_join). Data Tidying with tidyr. Splitting cells (separate, unite). Reshaping data (pivot_longer, pivot_wider).
	DAY 2
8:30 am-4:30 pm	Introduction to ggplot2. The grammar of graphics: Layers, aesthetics, and geoms. Creating basic plots. Customizing plots. Creating publication-ready plots. Generating complex visualizations (e.g., heatmaps, volcano plots). Case Study: Applying data wrangling, visualization, and analysis techniques to a real-world genomics dataset.

Questions

Please, contact us at mik-cubida@uk-erlangen.de.