

# Microbiome data science

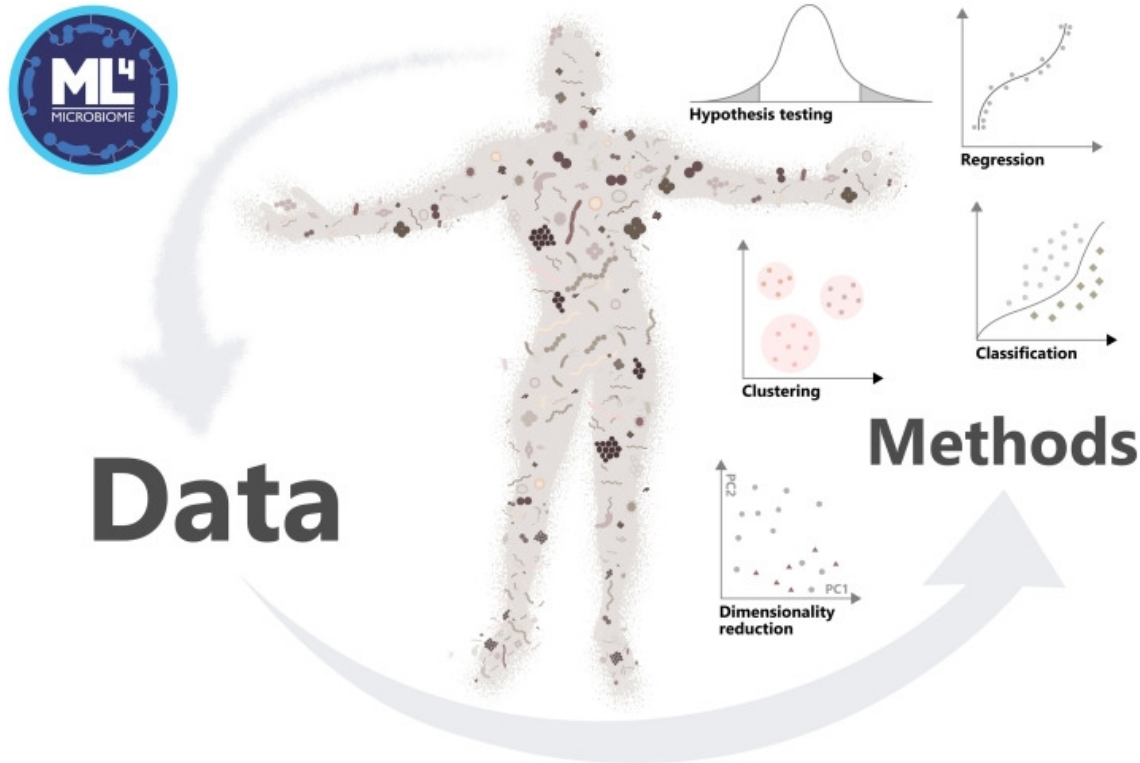


Figure source: Moreno-Indias et al. (2021)  
Statistical and Machine Learning Techniques  
in Human Microbiome Studies: Contemporary  
Challenges and Solutions. URL:  
<https://doi.org/10.3389/fmicb.2021.635781>.  
Frontiers in Microbiology 12:11.

**Associate Prof. Leo Lahti** | [datascience.utu.fi](mailto:datascience.utu.fi)  
Department of Computing, University of Turku, Finland

 @antagomir

# Acknowledgments

## Course organizers

- Mirjam Bloemendaal

## Course assistants

- Henrik Eckermann
- Tuomas Borman
- Chouaib Benchraoui

## Miaverse collective

- Sudarshan Shetty
- Felix Ernst

# Learning goals

Key concepts in microbiome bioinformatics

Common data science tools & best practices

# Overview of the week

## Key data analytical concepts

- Intro to microbiome data science
- Alpha diversity
- Beta diversity
- Differential abundance analysis (Henrik)

## Conclusion

- Course work presentations

# Monday: from raw sequences to ecological data analysis

## Lectures (by Tom)

Microbiota analysis: association studies vs. causality; microbiota sequencing methods (16S, shotgun, metagenomics)

DNA isolation and 16S rRNA gene sequencing: from raw sequences to OTU table in a biom file

## Getting started

Prerequisites

Data generation and import

Open science methods

## Practical

Importing microbiome data to R for interactive data analysis

Initializing reproducible report with Rmarkdown

# Getting started

- Checklist
- R, R studio, R packages installation ok?
- additional reading
- online chat (Gitter & Slack)
- course assignment

# Course assignment – case study options

**Option 1** ADHD-associated changes in gut microbiota and brain in a mouse model

**Option 2** Western diet intervention study in serotonin transporter modified mouse model

Mice received three weeks of Western (high fat, high sugar) or Control diet in three Genotype groups, a wild-type group and 2 groups with either partial or complete knock-out of the serotonin transporter (SERT), leading to an excess of serotonin in the gut, brain and blood circulation. Mice were measured before and after the intervention. In total there are three factors in this design, Time (pre-,post-intervention), Diet (Control, Western diet) and Genotype group (WT, HET, KO). There are about 6-8 mice per group.

**Option 3** Open data set of your own choice, see e.g. Bioconductor ExperimentHub

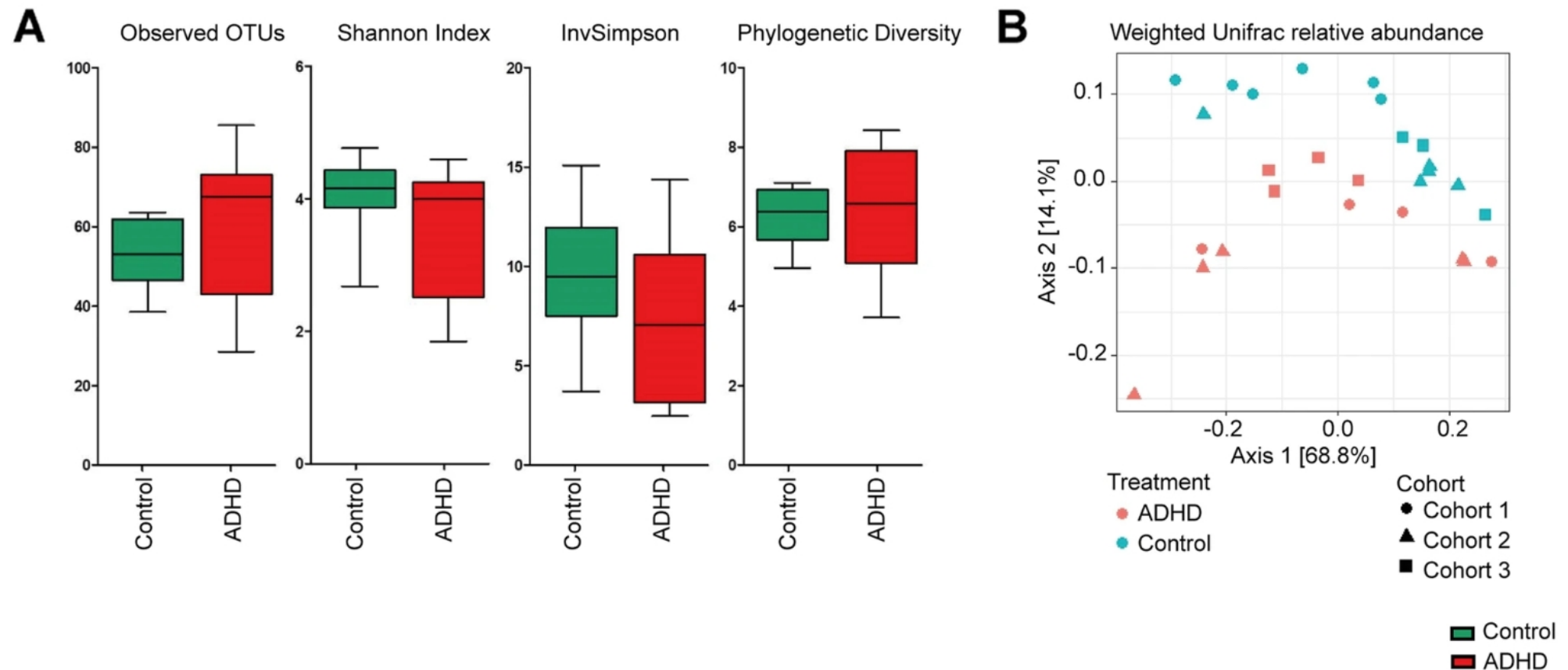
- microbiomeDataSets
- curatedMetagenomicData

# Gut microbiota from persons with attention-deficit/hyperactivity disorder affects the brain in mice

Anouk C. Tengeler, Sarita A. Dam, Maximilian Wiesmann, Jilly Naaijen, Miranda van Bodegom, Clara Belzer, Pieter J. Dederen, Vivienne Verweij, Barbara Franke, Tamas Kozicz, Alejandro Arias Vasquez & Amanda J. Kiliaan ✉

*Microbiome* 8, Article number: 44 (2020) | [Cite this article](#)

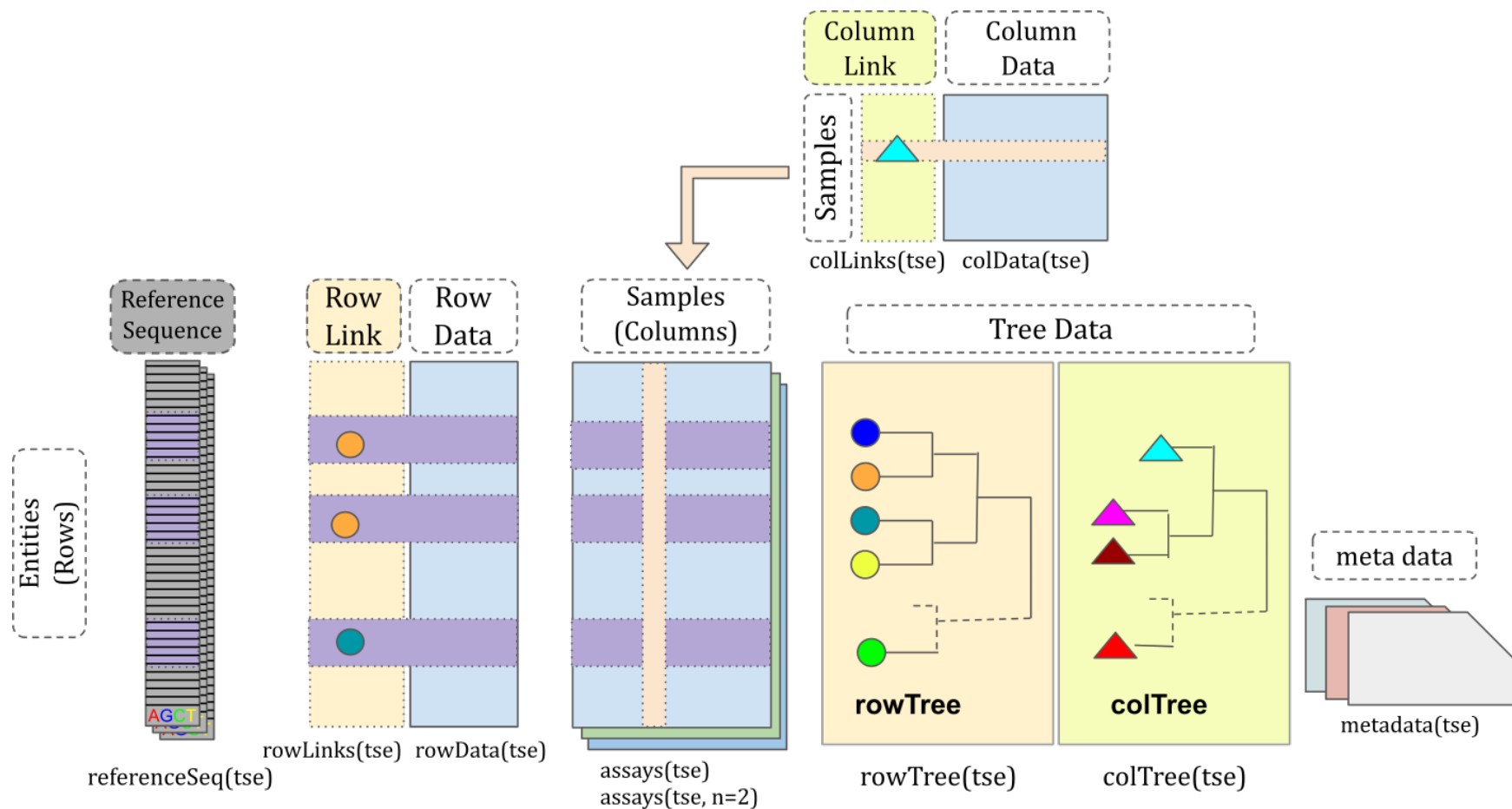
From: [Gut microbiota from persons with attention-deficit/hyperactivity disorder affects the brain in mice](#)





# Importing microbiome data to R

**Task:** import biom data files for your chosen case study into a specific data container (structure) in R, TreeSummarizedExperiment (TSE) Huang et al. (2020). This provides the basis for downstream data analysis.



# Initializing reproducible report

**Task:** create Rmarkdown document that imports the data

## 1 Welcome!

### 1.1 Installation Instructions

## 2 Amazing Resources

## 3 Dynamic Documents

### 3.1 Reproducible Research

## 4 Markdown

## 5 RMarkdown

### 5.1 Why R Markdown?

### 5.2 Simple Workflow

### 5.3 Creating a .Rmd File

### 5.4 YAML Headers

### 5.5 Markdown Basics

### 5.6 Embed Code

#### 5.6.1 Inline R Code

#### 5.6.2 R Code Chunks

## 6 Knitr

### 6.1 Chunk Labels

### 6.2 Chunk Options

### 6.3 Global Options

## Creating Dynamic Documents with RMarkdown and Knitr

Code ▾

*By: Marian L. Schmidt, @micro\_marian, marschmi at umich.edu*

*May 11th, 2016*

## 1 Welcome!

This workshop was hosted on **May 11th, 2016** and has 2 supplemental materials available:

1. **Class notes** can be found on the workshop [etherpad](#).
2. This workshop was **recorded** live and is available on [YouTube](#). Please be welcome to tune in on YouTube!

This tutorial was constructed as a part of Dr. C Titus Brown's [Data Intensive Biology \(DIB\)](#) training program at the University of California, Davis. The DIB training program hosts local + remote workshops covering topics in bioinformatic tools and data analysis.

There were three remote classrooms that tuned in:

1. University of California, Davis
2. Simon Fraser University
3. Ontario Institute for Cancer Research

The **Github repository** for this lesson can [be found here](#).

# Data exploration

**Task:** see the online material

