Helis Academy

HELIS Academy:
Data Analysis and Stewardship

Courses organized by VIB

---

1. OMICS DATA ANALYSIS THROUGH DATA INTEGRATION - USING BIOLOGICAL PATHWAYS, NETWORKS, AND LINEAR MODELS

Together with Maastricht University

Date and location:

11 – 14 June 2019, Niel (BE)

Participants:

12 participants from BE and NL

Context:

The first Helis Academy Omics Data Analysis course is a collaborative effort between Maastricht University and VIB in which you will learn about transcriptomics data analysis and multi-omics data integration. We will begin the course addressing how you obtain published data from data repositories and then cover basic pre-processing steps. We will continue by introducing you to molecular pathways, from what they are, to how to make your own, to how to use them in advanced omics data analysis (transcriptomics, proteomics, metabolomics) and how this compares Gene Ontology analysis. The course will continue by zooming out to utilize your pathway analysis in a broader network analysis so you can see the bigger picture. Finally, we will finish with a day to show you how to integrate Omics data sets in an unsupervised statistical way using linear models.

Schedule:

11 June 2019:

Helis Academy

11:30 - 12:30 Lecture: Introduction to pathways

12:30 - 13:30 Lunch

13:30 - 13:45 Lecture: Introduction to PathVisio and WikiPathways

13:45 - 15:00 Practical: WikiPathways academy

15:00 - 15:30 Break

15:30 - 17:00 Practical: Draw your own pathway in PathVisio

17:00 - 17:30 Wrap up (Quiz and Q&A)

12 June 2019:

09:30 - 10:30 Lecture: Data repositories and data-preprocessing

10:30 - 12:30 Practical: Pre-processing hands-on with ArrayAnalysis

12:30 - 13:30 Lunch

13:30 - 14:00 Lecture: Pathway analysis

14:00 - 15:00 Practical: Pathway analysis

15:00 - 15:30 Break

15:30 - 16:00 Lecture: Gene Ontology Analysis

16:00 - 17:00 Practical: Gene Ontology Analysis

17:00 - 17:30 Wrap up (Quiz and Q&A)

13 June:

09:30 - 10:30 Lecture: Network Biology

10:30 - 11:45 Practical: Pathways as a source for network biology

11:45 - 12:30 Gene Ontology-analysis in Cytoscape

12:30 - 13:30 Lunch

13:30 - 15:00 Lecture and practical: Protein-Protein Interactions

15:00 - 15:30 Break

15:30 - 16:30 Practical: Automation demo or work on your own dataset

Helis Academy

16:30 - 17:00 Wrap up (Quiz and Q&A)

14 June:

Multi Omics Integration (transcriptomics, proteomics, metabolomics)

09:30 - 11:30 Introduction Multi Omics Integration

11:30 - 12:30 Hands-on MOFA

12:30 - 13:30 Lunch

13:30 - 17:00 Hands-on MixOmics

Trainers:

Lauren Dupuis (Maastricht University)

Martina Summer-Kutmon (Maastricht University)

Susan Steinbusch (Maastricht University)

Friederike Ehrhart (Maastricht University)

Alexander Botzki (VIB & Elixir Belgium)

https://training.vib.be/all-trainings/helis-academy-course-omics-data-analysis-through-data-integration-using-biological

Helis Academy

2. STATISTICAL THINKING, FIRST EDITION

Date and location:

3 March 2020, Ghent (BE)

Participants:

17 participants from BE and NL

Context:

The course mission is to help researchers to see statistics as an intrinsic part of research. Not as a tool that provides standard output when given an input, but a mean to help integrate their knowledge to allow them making objective decisions.

To achieve this, we should get rid of deeply embedded misconceptions about some common statistical tools and replace them by a more intelligent and flexible use of statistics based on understanding of the important principles.

A selection of important topics will be covered in dialogue with the course takers.

Event intended for researchers and research managers interested in improving their research by applying important statistical principles that go beyond applying standard tools without understanding them.

Schedule:

- Introduction
- Case studies to make clear that simple application of statistics lead to problems
    - Discussion in subgroups
    - Plenary discussion
- Place of statistics in research
- Intelligent experimental design
    - Getting the context clear
    - Getting the research question clear
    - Deriving a design from that question
- Case studies on reading and displaying results
    - Discussion in subgroups
    - Plenary discussion

Helis Academy

- More than only data: combining expert knowledge and statistics
  - Models
  - Interpretation
  - Conveying the message
- Recap and putting things in the bigger context again
- Closure

Trainer:

Joris De Wolf, Highwoods

https://training.vib.be/all-trainings/statistical-thinking-o

Helis Academy

3. EXPERIMENTAL DESIGN, FIRST EDITION

Date and location:

1 April, 20 April, 27 April, 11 May, 18 May, 19 May, 21 May & 28 May 2020, online

Participants:

9 participants from BE and NL

Context:

The mission of this course is to explain the underlying principles and concepts of experimental design what will allow the course takers to understand why well-designed experiments are more efficient. We will apply these concepts and use R tools to design or improve concrete and typical biotech research experiments.

We welcome experimental design questions from the audience.

Event intended for researchers with some experience in designing experiments, but that would like to better understand the statistical principles behind it with the aim to improve the relevance and efficiency of their experiments.

A basic knowledge of R is required to follow this course.

Schedule:
- Introduction
- Why do we do experiments?
    o Difference between experiments and observations
    o Causation and correlation
- Think twice before you start: importance of careful design
    o Get the reseach question clear
    o Context and inference space
    o Desired and undesired sources of variability
    o Selecting the measurements
    o Plenty rough or a few accurate observations
- Important concepts:
    o Orthogonality
    o Randomization and blocking

- - Replication and power
  - Independence
  - Fixed and random effects
- First simple case: a design to determine a regression
- Power studies
  - Intro
  - Hands-on in R via traditional methods and via simulations
  - Apply it to the regression problem
- One-factor-at-a-time vs factorial designs
  - Interactions
  - Factor screening vs optimisation
  - Fractional designs (standard designs, optimal designs)
- Finding good design and power study for factorial experiment with R
- Randomisation schemes
  - Completely randomized design
  - Complete block design
  - Incomplete block design
  - Splitplot design
- Case studies
  - Group discussion of existing case
  - Group work: design best experiment based on available material and info
- Recapitulate
- Closure

Trainer:

Joris De Wolf, Highwoods

https://training.vib.be/all-trainings/experimental-design-o

Helis Academy

4. CHALLENGES IN OMICS DATA INTEGRATION

Date:

8 & 9 October 2020 - afternoon sessions, online

Participants:

81 participants from BE and NL (85 participants in total)

Context:

In the era of Big Data, the tsunami of massive 'omics' data is revolutionizing the way we do science. Life science researchers are no longer analyzing one data set at a time but are moving towards multi-disciplinary integrative biology. It has been demonstrated that integration of different 'omics' data types (such as on genomes, transcriptomes, proteomes, epigenomes, etc..), boosts biological discoveries and improves predictions of the underlying interactions and regulation among molecular entities. Integrating different 'omics' datasets is a challenging task that relies heavily on data mining and machine learning algorithms. One must account for the specificities of each data type, solve problems associated with processing data across different platforms, and take into account the variable reliability levels of heterogeneous data.

In this edition of the training on Multi-Omics data integration, four different topics will be addressed: 'Multi-Omics Factor Analysis', 'Single Cell Data Analysis', 'Machine Learning, Deep Learning & Genomics/Proteomics', and 'Machine Learning in Drug Discovery and Disease'. We have invited top level speakers to share their insights with you on the latest developments in the field.

Schedule:

8 October 2020

13h20-13h30   Welcome

13h30-14h10   Michel Dumontier, Institute of Data Science, Maastricht University, NL

Accelerating biomedical discovery science with an Internet of FAIR data and services

14h10-14h50   Ricard Argelaguet, European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK

Multi-Omics Factor Analysis (MOFA): a statistical framework for the unsupervised integration of multi-omics data

14h50-15h30   Edward Marcotte, Molecular Biosciences, College of Natural Sciences, University of Texas at Austin, US

Integrating evolution into proteomics: a case study mapping biochemical machinery across plants and beyond

15h30-16h00   Coffee – networking break

16h00-16h40   Yvan Saeys, VIB-UGent Center for Inflammation Research, BE

Machine learning challenges for multi-modal single-cell data

16h40-17h20   Carl Herrmann, Health Data Science Unit, Medical Faculty University Heidelberg and BioQuant, Heidelberg, DE

17h20-17h50   Closing – networking

9 October 2020

13h25-13h30   Welcome

13h30-14h10   Julien Gagneur, Department of Informatics, Technical University of Munich, DE

14h10-14h50   Avi Ma'ayan, Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, US

The Harmonizome-ML and Drugmonizome-ML Appyters: Web Interfaces to Impute Knowledge about Genes and Drugs with Machine Learning

14h50-15h30   Sushmita Roy, Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, US

Network-based integrative approaches to examine complex systems

15h30-16h00   Coffee – networking break

16h00-16h40   Ernest Fraenkel, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, US

Integrating Multi-Omic and Clinical Data to Understand Neurodegenerative Diseases

16h40-17h20   Stephen MacKinnon, Cyclica, Toronto, CA

Designing Molecules to Satisfy Multiple Predictive Objectives

17h20-17h50   Closing – networking

https://training.vib.be/all-trainings/challenges-omics-data-integration

Helis Academy

5. USING MOFA FOR INTEGRATION OF OMICS DATA, FIRST EDITION

Date:

15 & 16 October 2020, online

Participants:

12 participants from BE and NL (29 participants in total)

Context:

This training provides an introduction to Multi-Omics Factor Analysis (MOFA) for the integration of different omic data sets in an unsupervised fashion. It will enable you to run MOFA on multi-omic data, identify and explore the major drivers of variations across omics and use the inferred factors in various downstream analyses.

Participants can analyze their own data in the course.

- o Questions that will be addressed during the training:
- o What kind of preprocessing of the data is required for MOFA?
- o How to train MOFA on a multi-omic data set?
- o How to interpret the MOFA factors by their loadings, using gene set enrichment or sample ordination?
- o How to use MOFA for downstream analyses including regression, classification or clustering?
- o How to impute missing values with MOFA?
- o How to select the number of factors and compare different MOFA fits?

Familiarity with R and RStudio is essential. If you have no experience with R you should follow the Basic statistics in R training first.

MOFA (Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets) - http://msb.embopress.org/content/14/6/e8124

Schedule:

15 October 2020 14h00 – 16h30: Demo

16 October 2020 14h00 – 16h30: Q&A

Helis Academy

Trainer:

Ricard Argelaguet, European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK

https://training.vib.be/all-trainings/using-mofa-integration-omics-data-online

6. RESEARCH DATA MANAGEMENT IN LIFE SCIENCES

Date:

9 & 10 November 2020, online

Participants:

45 participants from BE and NL

Context:

This in-depth course will help researchers to develop their knowledge and practical skills in handling and managing the research data they collect and use.

This course will guide the attendees through the key aspects on how to manage, document, store and safeguard research data well and how to plan and implement good data management in research projects in accordance to current best practices.

Upon completion of this course, participants should have an understanding of what Research Data Management is, and why it is important in academic research. They should have an understanding of the FAIR data principles, and how they can make data more FAIR. They should be able to successfully manage all types of research data and to document both the research itself, as well as the data in a comprehensive way.

This course targets all researchers active in Life Sciences. There is no prior knowledge or programming skills needed.

Schedule:

9 November 2020

09:30 - 10:30   Thomas Van de Velde (Data Stewards UGent)

Introduction to RDM & Data Lifecycle

10:30 - 11:00   Paula Oset (Data Stewards UGent)

RDM trends and requirements: Funder & journal policies

11:00 - 11:30   Break

11:30 - 12:30	Privacy and GDPR in the research life cycle
12:30 - 13:30	Lunch
13:30 - 14:30	Laura Standaert (Data Stewards UGent)
	DMP and DMPonline
14:30 - 15:00	Flora D'Anna (Data Stewards ELIXIR Belgium)
	FAIRify your data: data documentation and metadata
15:00 - 15:30	Break
15:30 - 16:00	Nele Pauwels (KCGG UGent)
	Organizing your data: structure and versioning
16:00 - 17:00	Alexander Botzki (VIB & ELIXIR Belgium)
	Reusing existing data

10 November 2020

09:30 - 12:30	Tuur Muyldermans (VIB & ELIXIR Belgium)
	Creating and working with a reproducible data analysis environment.
12:30 - 13:30	Lunch
13:30 - 14:30	Alexander Botzki (VIB & ELIXIR Belgium)
	Share and publish your data: repositories and licenses
14:30 - 15:30	Griet Den Herder (VIB)
	Valorisation and Intellectual Property
15:30 - 16:00	Break
16:00 - 17:00	Jan Lammertyn (Data stewards UGent)
	Data security and encryption

https://training.vib.be/all-trainings/research-data-management-life-sciences

7. Upskilling (young) professionals in Data Analysis and Data Stewardship

Together with TU/e, Maastricht University and DTL

Date:

15 March 2021, online

Participants:

68 participants

Context:

In the context of the Helis Academy, and as part of its data analysis and stewardship program, Eindhoven University of Technology (TU/e), Dutch Techcentre for Life Sciences (DTL), Vlaams Instituut voor Biotechnologie (VIB) and Maastricht University (UM) have built up a portfolio of courses and course materials in the past three years. With the end of the Helis project approaching (Summer 2021), we are now organizing this event where we will present what Helis Academy DAS has achieved to help bridge the gap between University education and a career in industry.

The event will feature a group of inspirational speakers from industry, academia and government who will provide their insight on the skills gap in Life Sciences & Health and together with the participants we are going to discuss the perspectives on possible next steps, including the embedding of the developed courses.

The spotlights will be on the challenges that employers face in getting and keeping the digital and data skills of their employees on a sufficient level, on the training solutions that exist, on the needed alignment of the outcomes of 'curricula' and the needs of the 'labor market', and finally we will end with a plea for more collaboration throughout the landscape of public and private providers of education & training.

Schedule:

12:30 - 13:00   Possibility of networking (via the conference platform) before the start of the event

13:00 - 13:10   Welcome (including 5 min tech setup as a starting point)

13:10 - 14:20   Skill gap in Life Sciences & Health in the industry

"Future of Work in Life Sciences & Health sector" by Carmen van Vilsteren (Chair Topsector Life Science & Health, Director Health at TU/e and Chair of e/MTIC Board)

"Prospects for Life Long Learning at TU/e" by Paul Koenraad (Dean Graduate School & Full Professor at TU/e)

"Personalised healthcare: why data skills are crucial to get there" by Mariëlle Gallegos Ruiz (Health Outcomes Data Lead at Roche Nederland)

"Gut-feelings in data-driven decision making" by Uwe Thissen (Digital Strategy Lead at BASF)

14:20 - 14:35   Introduction to Helis Academy

"Helis Academy: Becoming industry-ready in Life Sciences" by Pascale Engelen (Co-General Manager flanders.bio, Project Lead Helis Academy)

14:35 - 15:00   Break & possibility of networking

15:00 - 16:00   Introduction to Helis Data Analysis and Stewardship Portfolio

FAIR Data Stewardship by Celia van Gelder (DTL, Dutch Techcentre for Life Sciences)

Statistics by Alexander Botzki (VIB)

Machine Learning by Harold Weffers (TU/e)

Omics Data Analysis by Lauren Dupuis (UM)

16:00 – 16:20   Helis web-based tools by Rita Neves (TU/e)

RésuMe, a creative orientation tool to guide students/young professionals through their professional path in Life Sciences & Health,

SuitAbility aims to help companies in finding a good match between a vacancy and a potential candidate.

16:20 – 17:00   Outlook and Next steps

Panel discussion moderated by Alexander Botzki (VIB)

Keynote speakers:

Mariëlle Gallegos Ruiz (Health Outcomes Data Lead at Roche Nederland)

Chris Evelo (Head of the Department of Bioinformatics - BiGCaT at Maastricht University)

Jeroen Wynen (Content coordinator VAIA at UHasselt)

17:00   Closing

https://www.eventbrite.nl/e/upskilling-young-professionals-in-data-analysis-and-data-stewardship-tickets-135300612775#

8. STATISTICAL THINKING, SECOND EDITION

Date:

3, 4 and 5 May 2021, each time from 9h00 - 12h00, online

Participants:

7 participants from BE and NL

Context:

The course mission is to help researchers to see statistics as an intrinsic part of research. Not as a tool that provides standard output when given an input, but a mean to help integrate their knowledge to allow them making objective decisions.

To achieve this, we should get rid of deeply embedded misconceptions about some common statistical tools and replace them by a more intelligent and flexible use of statistics based on understanding of the important principles.

A selection of important topics will be covered in dialogue with the course takers.

Event intended for researchers and research managers interested in improving their research by applying important statistical principles that go beyond applying standard tools without understanding them.

Schedule:

- Introduction

- Case studies to make clear that simple application of statistics lead to problems
  - Discussion in subgroups
  - Plenary discussion
- Place of statistics in research
- Intelligent experimental design
  - Getting the context clear
  - Getting the research question clear
  - Deriving a design from that question
- Case studies on reading and displaying results
  - Discussion in subgroups
  - Plenary discussion
- More than only data: combining expert knowledge and statistics
  - Models
  - Interpretation
  - Conveying the message
- Recap and putting things in the bigger context again
- Closure

Trainer:

Joris De Wolf, Highwoods

https://training.vib.be/all-trainings/statistical-thinking-1

9. EXPERIMENTAL DESIGN, SECOND EDITION

Date and location:

17, 18, 20 & 21 May 2021 (morning sessions), online

Participants:

8 participants from BE and NL

Context:

The mission of this course is to explain the underlying principles and concepts of experimental design what will allow the course takers to understand why well-designed experiments are more efficient. We will apply these concepts and use R tools to design or improve concrete and typical biotech research experiments.

We welcome experimental design questions from the audience.

Event intended for researchers with some experience in designing experiments, but that would like to better understand the statistical principles behind it with the aim to improve the relevance and efficiency of their experiments.

A basic knowledge of R is required to follow this course.

Schedule:

- Introduction
- Why do we do experiments?
    o Difference between experiments and observations
    o Causation and correlation
- Think twice before you start: importance of careful design
    o Get the reseach question clear
    o Context and inference space
    o Desired and undesired sources of variability
    o Selecting the measurements
    o Plenty rough or a few accurate observations
- Important concepts:
    o Orthogonality
    o Randomization and blocking

- o Replication and power
- o Independence
- o Fixed and random effects
- First simple case: a design to determine a regression
- Power studies
  - o Intro
  - o Hands-on in R via traditional methods and via simulations
  - o Apply it to the regression problem
- One-factor-at-a-time vs factorial designs
  - o Interactions
  - o Factor screening vs optimisation
  - o Fractional designs (standard designs, optimal designs)
- Finding good design and power study for factorial experiment with R
- Randomisation schemes
  - o Completely randomized design
  - o Complete block design
  - o Incomplete block design
  - o Splitplot design
- Case studies
  - o Group discussion of existing case
  - o Group work: design best experiment based on available material and info
- Recapitulate
- Closure

Trainer:

Joris De Wolf, Highwoods

https://training.vib.be/all-trainings/experimental-design-1

10.   USING **MOFA** FOR INTEGRATION OF OMICS DATA, SECOND EDITION

Date:

25 & 27 May 2021, online

Participants:

16 participants from BE and NL (50 participants in total)

Context:

This training provides an introduction to Multi-Omics Factor Analysis (MOFA) for the integration of different omic data sets in an unsupervised fashion. It will enable you to run MOFA on multi-omic data, identify and explore the major drivers of variations across omics and use the inferred factors in various downstream analyses.

Participants can analyze their own data in the course.

- o   Questions that will be addressed during the training:
- o   What kind of preprocessing of the data is required for MOFA?
- o   How to train MOFA on a multi-omic data set?
- o   How to interpret the MOFA factors by their loadings, using gene set enrichment or sample ordination?
- o   How to use MOFA for downstream analyses including regression, classification or clustering?
- o   How to impute missing values with MOFA?
- o   How to select the number of factors and compare different MOFA fits?

Familiarity with R and RStudio is essential. If you have no experience with R you should follow the Basic statistics in R training first.

MOFA (Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets) - http://msb.embopress.org/content/14/6/e8124

Schedule:

25 May 2021, 14h00 – 16h30: Demo

27 May 2021, 14h00 – 16h30: Q&A

<u>Trainer:</u>

Ricard Argelaguet, European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK

https://training.vib.be/all-trainings/using-mofa-integration-omics-data-online-o

11. WRITING A DATA MANAGMENT PLAN E-LEARNING COURSE

Date:

?

Participants:

?

Context:

Schedule:

- Introduction
    - o What is data?
    - o FAIR data
    - o Research Data Frameworks
    - o Funder Requirements
    - o The Data Management Plan
    - o RDM support
- Preparing a DMP
    - o Tools
    - o DMP Online
    - o VIB forms
    - o Data collection
    - o Copyright and IP right issues
    - o Licenses
    - o Costs involved with managing your data
- Organizing and documenting data
    - o Data documentation
    - o Organizing
    - o Documenting & organizing – References
- Storing data and data security
    - o Data storage
    - o Data security
    - o Privacy
    - o Storing data and data security – References
- Sharing and preserving data

- o Sharing data
- o Preserving data
- o Sharing & preserving data – References
- Roundup
  - o Rounding up
  - o Evaluation
  - o Resources

https://elearning.bits.vib.be/courses/writing-a-data-management-plan/