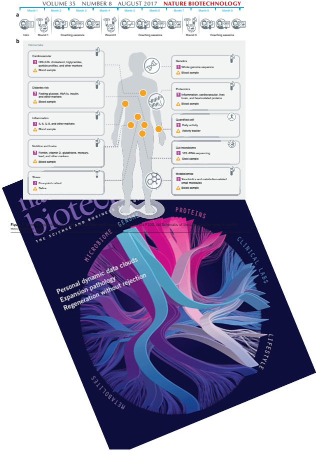
Statistical Genomics: Master of Science in Bioinformatics and Master of Science in Statistical Data Analysis

Lieven Clement
Ghent University, Belgium

Scientific Integrity and Reproducible Research

Bio-informatics research is based on empirical data



NATURE METHODS | VOL.12 NO.7 | JULY 2015 | 605

Mass spectrometrists should search only for peptides they care about

William Stafford Noble

WEBINAR
Proteogenomic strategies to advance drug development and precision medicine
Recorded live on September 6, 2017
Click to view
Science AAAS
Sponsored by Olink Proteomics
Log in | My account | Contact us
member
Renew my subscription
Sign up for newsletters
big data
Shift
SHARE
73
Advertisement
ONE APP... THOUSANDS OF JOBS

PHOTO: MAKSIM KABAKOU/SHUTTERSTOCK.COM

This special feature is brought to you by the Science/AAAS Custom Publishing Office

TECHNOLOGY FEATURE

Big biological impacts from big data

By Mike May | Jun. 13, 2014 , 5:45 PM

Scientific Integrity and Reproducible Research

Bio-informatics research is based on empirical data

VOLUME 35 NUMBER 8 AUGUST 2017 NATURE BIOTECHNOLOGY

NATURE METHODS | VOL.12 NO.7 | JULY 2015 | 605

Mass spectrometrists should search only for peptides they care about

William Stafford Noble

Mass spectrometrists should search for all peptides, but assess only the ones they care about

NATURE METHODS | VOL.14 NO.7 | JULY 2017 | 643

Adriaan Sticker1–4, Lennart Martens2–5 & Lieven Clement1,4,5

COMMENT

A tragedy of errors

Mistakes in peer-reviewed papers are easy to find but hard to fix, report David B. Allison and colleagues.

Personal dynamic data clouds
Expansion pathology
Regeneration without rejection

This special feature is brought to you by the Science/AAAS Custom Publishing Office

TECHNOLOGY FEATURE

Big biological impacts from big data

By Mike May | Jun. 13, 2014, 5:45 PM

→ Number of observations <<< number of features

→ Need for statistics to distinguish real patterns from random patterns in high dimensional data

Topics

Module I: Quantitative Proteomics

1. Identification and quantification of peptides and proteins
2. Data exploration and quality control using plots
3. Preprocessing: log-transformation, Filtering, Normalization, Summarization
4. Dealing with batch effects and other confounders
5. Statistical Concepts
    1. Linear models/Linear mixed models
    2. Trade-off between biological relevance/effect size vs statistical significance
    3. Empirical Bayes Methods
    4. Multiple testing

Module II: Next generation sequencing (NGS, Transcriptomics)

1. NGS Data exploration
2. Preprocessing/normalization
3. Additional Statistical Concepts
    1. Generalized linear models (GLM) for binary data
    2. GLM for count data
    3. Overdispersion

Organisation

1. Theory and Tutorials are blended
   - Module I: week 1-5
   - Module II: week 6-10
   - Project: week 1-10 via small assignments + week 11-12

2. Communication and submission of projects via Ufora
3. All tutorials from week 2 onwards are based on R/Bioconductor
   - via R-studio
   - Scripts are made in R/markdown: a file format to combine text, R code and R output.

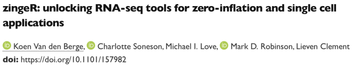
   → This makes it very easy to document your analysis and to distribute them in a way which is reproducible.

Organisation

4. Project
   - Projects: 10/20
   - Written Exam: 10/20.
     - Open book
     - Deep insight expected
     - Critical assessment of R-output,

Projects + Master thesis

- Project 201415, Master thesis 201516:

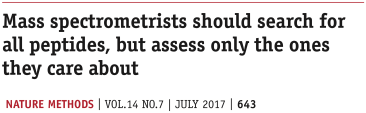


zingeR: unlocking RNA-seq tools for zero-inflation and single cell applications

Koen Van den Berge, Charlotte Soneson, Michael I. Love, Mark D. Robinson, Lieven Clement

doi: https://doi.org/10.1101/157982

- Project 201516: Neurogenomic profiling reveals distinct gene expression profiles between brain parts that are consistent across cichlid species of the genus Ophthalmotilapia. Derycke et al. 2018.
- Project 201516: Manuscript in preparation. A leap of the hurdle in mass spectrometry based proteomics. (Presentation at HUPO conference 2017).



Mass spectrometrists should search for all peptides, but assess only the ones they care about

NATURE METHODS | VOL.14 NO.7 | JULY 2017 | **643**

- Master thesis 201516:

Adriaan Sticker[1–4], Lennart Martens[2–5] & Lieven Clement[1,4,5]

- Design Project 201718: paper in preparation.
- Continuing on statistical genomics project for thesis is possible.