



# Update

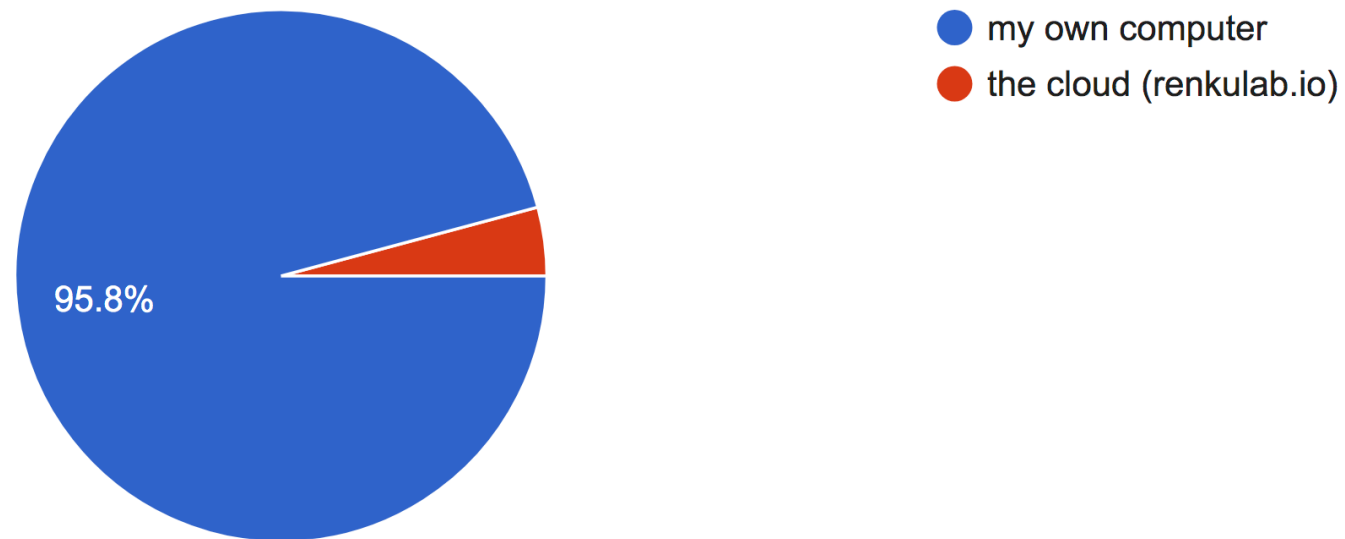
- Exercise 1 and Exercise 2 due at 18.00 tomorrow; every following exercise due the following Tuesday at 18.00
- Exercise 1:
  - 50 students registered (48 in UZH system)
  - 48 Google form responses received (Exercise 1 Part a) —> 43 unique GitHub usernames received
- Exercise 2:
  - 8 groups/repos started (Exercise 2 Part a)
  - 6 pull requests received (Exercise 2 Part b)



# Computing

What is your preferred computing option to do exercises / project?

48 responses





## Journal club

**Papers to be selected by 18.00 on 13th October;** discuss it with Hubert and I.

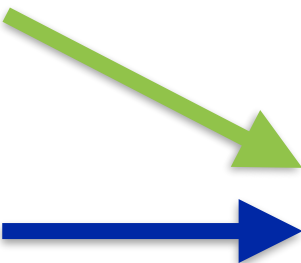
Start: Oct 12 or Oct 19

Journal Club schedule to be finalized by 19th October

Given the number of students, groups of 3 are recommended.

Use the #journal-clubs channel (e.g., to find a group member). I will put some suggestions there.

**Sign up by pull request (give a link to the paper, give initials of group members)**



### Schedule

Date	Lecturer	Topic	Exercise	JC1	JC2
14.09.2020	Mark + Hubert	admin; mol. bio. basics	R markdown; git(hub)		
21.09.2020	Mark	interactive technology/statistics session	group exercise: technology pull request		
28.09.2020	Hubert	NGS intro; exploratory data analysis	EDA in R		
05.10.2020	Hubert	mapping	Rsubread		
12.10.2020	Mark	limma + friends	linear model simulation + design matrices		
19.10.2020	Hubert	RNA-seq quantification	RSEM		
26.10.2020	Mark	edgeR+friends 1	basic edgeR/voom		
02.11.2020	Mark	edgeR+friends 2	GLM/DEXSeq		
09.11.2020	Kathi	hands-on session #1: RNA-seq	FASTQC/Salmon/etc.	X	X
16.11.2020	Hubert	single-cell 1: preprocessing, dim. reduction, clustering		X	X
23.11.2020	Helena	hands-on session #2: cytometry	cytof null comparison	X	X
30.11.2020	Mark	single-cell 2: cell type definition, differential state	scRNA exercise 2	X	X
07.12.2020	Pierre-Luc	hands-on session #3: single-cell RNA-seq	full scRNA-seq pipeline	X	X
14.12.2020	Mark	loose ends: HMM, EM, robustness	segmentation, peak finding	X	X



## limma (= linear models for microarray data) paper

### Linear models and empirical bayes methods for assessing differential expression in microarray experiments

[GK Smyth](#) - Statistical applications in genetics and molecular ..., 2004 - [degruyter.com](#)

... This section describes how **gene-wise linear models** arise from experimental designs and states the distributional assumptions about the ... There may be more or fewer contrasts than coefficients for the **linear model**, although if more than the contrasts will be **linearly** dependent ...

☆ Cited by 10031 Related articles All 39 versions Import into BibTeX

### Limma: linear models for microarray data

[GK Smyth](#) - ... and computational biology solutions using R and ..., 2005 - [Springer](#)

... **limma** uses **linear models** to analyze designed microarray experiments (Yang and Speed, 2003; Smyth, 2004) ... reference microarray data, **linear modeling** is much the same as ordinary ANOVA or multiple **regression** except that a **model** is fitted for ... **Linear Models** Case Study 401 ...

☆ Cited by 4799 Related articles All 18 versions Import into BibTeX

### limma powers differential expression analyses for RNA-sequencing and microarray studies

[ME Ritchie](#), [B Phipson](#), [D Wu](#), [Y Hu](#)... - Nucleic acids ..., 2015 - [academic.oup.com](#)

... **limma** powers **differential expression** analyses for RNA-sequencing and microarray studies. Matthew E. Ritchie. 1. Molecular Medicine Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia ...

☆ Cited by 3437 Related articles All 14 versions Import into BibTeX

## Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments\*

Gordon K. Smyth

Walter and Eliza Hall Institute of Medical Research  
Melbourne, Vic 3050, Australia

January 2004<sup>†</sup>

- seminal paper (cited >11,000 times)
- provides the foundation for a lot of (statistical) research in genomics
- we will digest this into a single lecture (**12 Oct**)



## Format your signup like this:

22.10.2018	Mark	limma + friends	linear model simulation + design matrices	Averaged gene expressions for regression (AS, LB, MK)	Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing (DT, HP)
29.10.2018	Charlotte	hands-on session #1: RNA-seq	FASTQC/Salmon/etc.	Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis (MS, CR)	X

## How to read a scientific paper



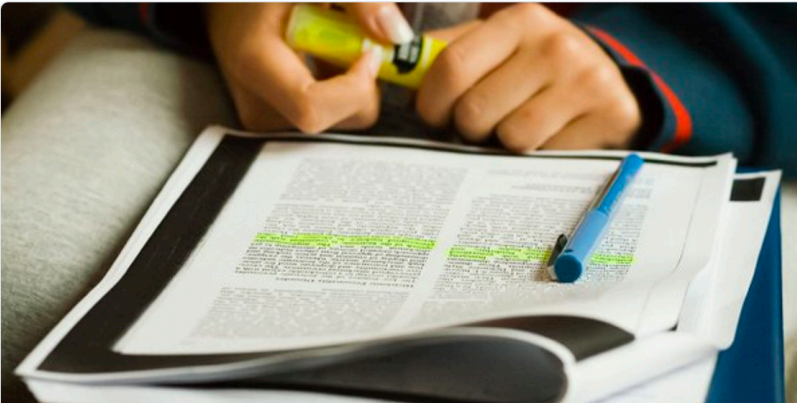
Lipi Thukral  
@Sci\_Lipi

Follow



Title->Abstract->Conclusion ->Hop to figures  
-> intro. -> results

### How to (seriously) read a scientific paper



#### How to (seriously) read a scientific paper

Reading becomes easier with experience, but it is up to each scientist to identify the techniques that work best for them.

sciencemag.org

11:00 PM - 19 Mar 2017

<https://github.com/jtleek/readingpapers>

### How much should you read?

Academic papers come out all the time. Thousands are published every year, including hundreds in any given specific area. Unless you devote yourself full time to reading academic papers you won't be able to keep up with them all. I believe in the idea that you should read papers that you find interesting. Science is awesome and you shouldn't waste your time on the boring parts if you can avoid it.

In general there are two main ways to find papers that I like. The way I used to do it was set up an aggregator with the RSS feeds from journals that I like, then I use the following (approximate) rates of reading parts of papers.

- 100% - read the title
- 20-50% - read the abstract
- 5-10% - look at the figures/captions
- 1-3% - read the whole paper

The new way that I do it is follow bioRxiv and a bunch of other people who have similar interests on Twitter. I use the above percentages for papers tweeted from aggregators and if I see a paper tweeted by 2-3 people I trust I usually end up reading that paper.



## Expectations: **journal club** presentation

- 20-25 minutes (+5 minutes discussion)
- MUST:
  - ➔ be a paper about a **statistical** method in genomics
  - ➔ be approved by Mark/Hubert
- Should:
  - ➔ describe the biological context and/or data collected
  - ➔ describe the (new) model used
  - ➔ describe comparisons to existing methods
- Should not:
  - ➔ be one of the papers discussed in detail in lectures: limma, edgeR, DEXSeq, etc.
- (since 2017) feedback forms from fellow students