Update

- Exercises every week: top 9 (of 14) will be counted towards 30% of the grade
- Exercise 1 and Exercise 2 due at 18.00 tomorrow (5 Oct 2021); now every exercise due the following Tuesday at 18.00
- Exercise 1:
  - 34 students registered
  - 29 Google form responses (github usernames) received —> use https://forms.gle/4dhWdcPa9TfsMhHMA
  - 24 Issues received
- Exercise 2:
  - 6 groups/repos started (Exercise 2 Part a)
  - 5 pull requests received (Exercise 2 Part b)

Journal club

Papers to be selected by 18.00 on 18th October; please discuss it with Hubert and I before submitting PR.

Start: Oct 25

Journal Club schedule to be finalized by 25th October

Given the number of students, groups of 2 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 to the 'material' repo. "First come first served"

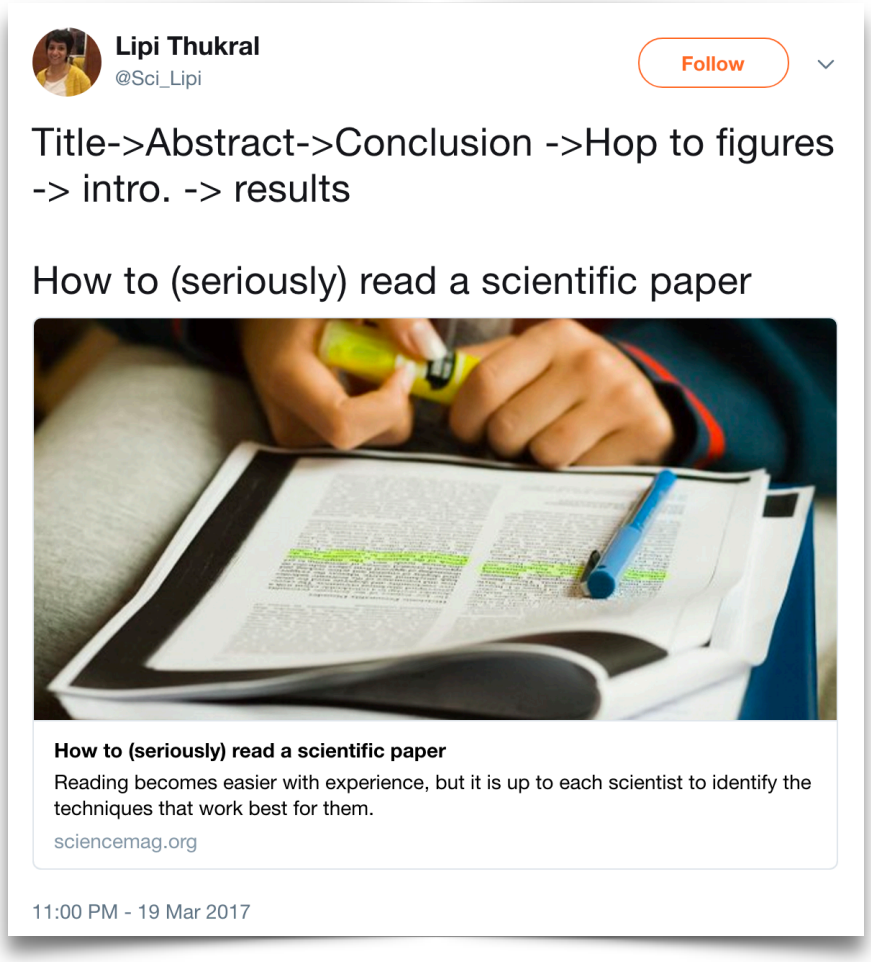
Date	Lecturer	Topic	Exercise	JC1	JC2
20.09.2021	Mark + Hubert	admin; mol. bio. basics	R markdown; git(hub)		
27.09.2021	Mark	interactive technology/statistics session	group exercise: technology pull request		
04.10.2021	Hubert	NGS intro; exploratory data analysis	EDA in R		
11.10.2021	Hubert	mapping	Rsubread		
18.10.2021	Mark	limma + friends	linear model simulation + design matrices		
25.10.2021	Hubert	RNA-seq quantification	RSEM	X	X
01.11.2021	Mark	edgeR+friends 1	basic edgeR/voom	X	X
08.11.2021	Mark	edgeR+friends 2	advanced edgeR/voom	X	X
15.11.2021	TBA	hands-on session #1: RNA-seq	FASTQC/Salmon/etc.	X	X
22.11.2021	Hubert	single-cell 1: preprocessing, dim. reduction, clustering	clustering	X	X
29.11.2021	TBA	hands-on session #2: cytometry	cytof null comparison	X	X
06.12.2021	Mark	single-cell 2: clustering, marker gene DE	marker gene DE	X	X
13.12.2021	Pierre-Luc	hands-on session #3: single-cell RNA-seq (cell type definition, differential state)	full scRNA-seq pipeline	X	X
20.12.2021	Mark	loose ends: HMM, EM, robustness	segmentation, peak finding	X	X

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

Format your signup like this (in markdown, of course):

```
| Date  | Lecturer | Topic | Exercise | JC1 | JC2 |
| --- | --- | --- | --- | --- | --- |
| 20.09.2021  | Mark + Hubert  | admin; mol. bio. basics | R markdown; git(hub) | | |
| 27.09.2021  | Mark | interactive technology/statistics session  | group exercise: technology pull request | | |
| 04.10.2021  | Hubert | NGS intro; exploratory data analysis | EDA in R | | |
| 11.10.2021  | Hubert | mapping  | Rsubread | | |
| 18.10.2021  | Mark | limma + friends | linear model simulation + design matrices | | |
| 25.10.2021  | Hubert | RNA-seq quantification | RSEM  | X | X |
| 01.11.2021  | Mark | edgeR+friends 1 | basic edgeR/voom | X | X |
| 08.11.2021  | Mark | edgeR+friends 2 | advanced edgeR/voom| X | X |
```

How to read a scientific paper



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

limma (= linear models for microarray data) paper

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†

https://doi.org/10.2202/1544-6115.1027

- seminal paper (cited >12,000 times)
- provides the foundation for a lot of (statistical) research in genomics
- we will digest this into a single lecture (**18 Oct**)