Update

- Reminder: exercises every week; top 9 (of 14) will be counted towards 30% of the grade
- Mark + feedback sent via Slack
- Exercise 1 and Exercise 2 due at 18.00 tomorrow (4 Oct 2022); from now on every exercise due the following Tuesday at 18.00
- Exercise 1:
  - 34 students registered
  - 30 Google form ("Computing options for STA 426 (2022)") responses (github usernames) received —> use https://forms.gle/sc7ci6jPFweBE8xKA
  - 26 Issues received
- Exercise 2:
  - 10 groups/repos started (Exercise 2 Part a)
  - 6 pull requests received (Exercise 2 Part b)

Journal club

Papers to be selected by 18.00 on 17th October; please discuss it with Hubert and I before submitting pull request.

Start: Oct 24

Journal Club schedule to be finalized by 24th 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, README.md. "First come first served"

Date	Lecturer	Topic	Exercise	JC1	JC2
19.09.2022	Mark + Hubert	admin; mol. bio. basics	quarto; git(hub)		
26.09.2022	Mark	interactive technology/statistics session	group exercise: technology pull request		
03.10.2022	Hubert	NGS intro; exploratory data analysis	EDA in R		
10.10.2022	Mark	limma + friends	linear model simulation + design matrices		
17.10.2022	Hubert	mapping	Rsubread		
24.10.2022	Hubert	RNA-seq quantification	RSEM	X	X
31.10.2022	Mark	edgeR+friends 1	basic edgeR/voom	X	X
07.11.2022	Mark	edgeR+friends 2	advanced edgeR/voom	X	X
14.11.2022	YYY	hands-on session #1: RNA-seq	FASTQC/Salmon/etc.	X	X
21.11.2022	Hubert	single-cell 1: preprocessing, dim. reduction, clustering	clustering	X	X
28.11.2022	YYY	hands-on session #2: cytometry	cytof null comparison	X	X
05.12.2022	Mark	single-cell 2: clustering, marker gene DE	marker gene DE	X	X
12.12.2022	YYY	hands-on session #3: single-cell RNA-seq (cell type definition, differential state)	full scRNA-seq pipeline	X	X
19.12.2022	Mark	loose ends: HMM, EM, robustness	segmentation, peak finding	X	X

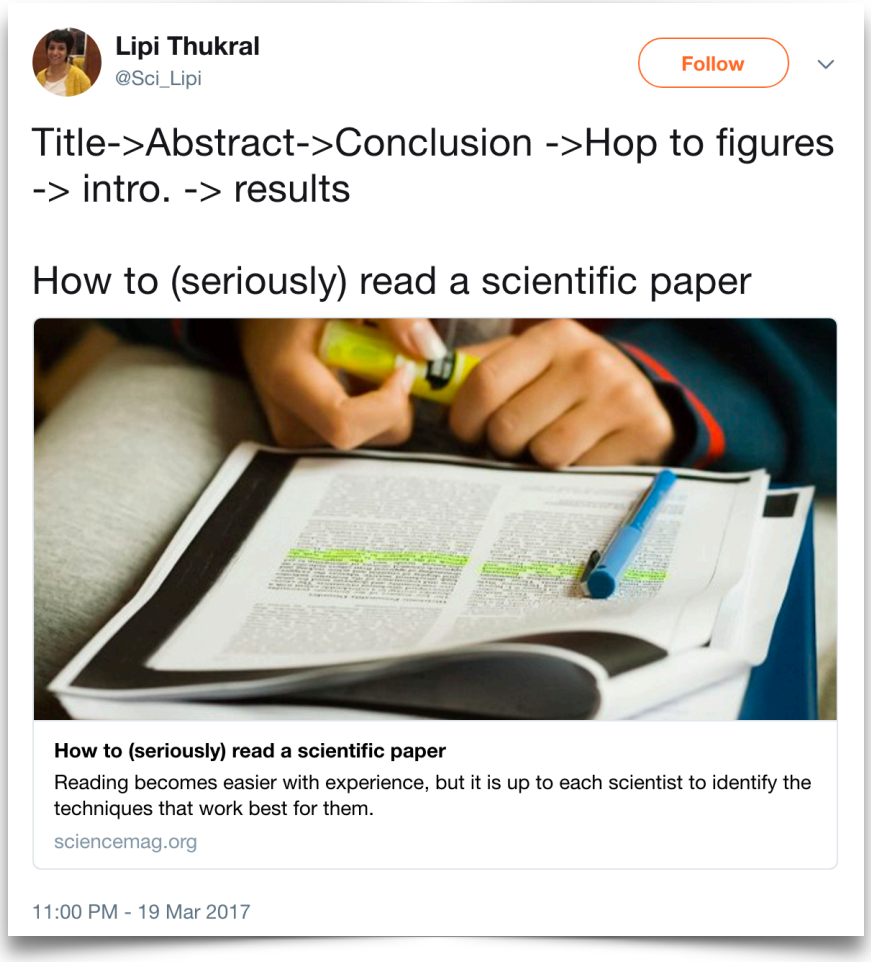
Format your signup like this (in markdown, of course):

README.md: paper title as link to website, initials of group members

01.11.2021	Mark	edgeR+friends 1	basic edgeR/voom	OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data (BT, KN)	Powerful and robust non-parametric association testing for microbiome data via a zero-inflated quantile approach (ZINQ) (RM, DS)
08.11.2021	Mark	edgeR+friends 2	advanced edgeR/voom	ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system (TB, OF)	X

```
| 01.11.2021  | Mark | edgeR+friends 1 | basic edgeR/voom | [OUTRIDER: A Statistical
Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data](https://
www.cell.com/ajhg/fulltext/S0002-9297(18)30401-4) (BT, KN)  | [Powerful and robust non-
parametric association testing for microbiome data via a zero-inflated quantile approach
(ZINQ)](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-021-01129-3)
(RM, DS) |
| 08.11.2021  | Mark | edgeR+friends 2 | advanced edgeR/voom| [ZeitZeiger: supervised
learning for high-dimensional data from an oscillatory system](https://dx.doi.org/
10.1093%2Fnar%2Fgkw030) (TB, OF) | 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 almost 13,000 times)
- provides the foundation for a lot of (statistical) research in genomics
- we will digest this into a single lecture (**10 Oct**)