Week 5 notes:

- Projects: some ideas
- Journal club
    - signups via pull request
    - Feedback forms
    - "Benchmarking"

Project ideas

- n.b.: 50% of mark
- As always, reproducing analyses from a paper or designing your own simulation to evaluate some methods is always a possibility
- Not urgent; I will put pressure on in a few weeks
- “Consulting” type possibilities:
  - Comparing fixed effects and mixed effects models for the paired comparison problem
  - Time course differential gene expression

33 signed up

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
05.11.2018	Mark	edgeR+friends 1	basic edgeR/voom	Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data (RB, RG)	
12.11.2018	Mark	edgeR+friends 2	GLM/DEXSeq	A general and flexible method for signal extraction from single-cell RNA-seq data (AL, VL)	Integrating single-cell transcriptomic data across different conditions, technologies, and species (PV, FN, ES)
19.11.2018	Mark	single-cell dim. reduction + clustering; FDR	conquer	Normalization of RNA-seq data using factor analysis of control genes or samples (RM, JD, CV)	Diffusion maps for high-dimensional single-cell analysis of differentiation data (SP, GK)
26.11.2018	Lukas	hands-on session #2: cytometry	cytof null comparison	Epigenome-wide association studies without the need for cell-type composition (RL, SG)	X
03.12.2018	Hubert	classification	MLInterfaces	Predicting cell types in single cell mass cytometry data (CM, SS)	Bayesian Trees for Automated Cytometry Data Analysis (CB, XC)
10.12.2018	Mark	loose ends: HMM, EM, robustness	segmentation, peak finding	Differential expression analysis for sequence count data (AA, PS)	Visualizing Data using t-SNE (MJT, TB, MP)
17.12.2018	Mark	hands-on session #3: single-cell RNA-seq	full scRNA-seq pipeline	tba (SB,ST)	X

Journal Club procedure

- Your job during journal clubs: give the presenters some constructive feedback

- n.b.: Please put your GitHub names on the first slide of journal club and/or add them to the Slack on the day.

STA426 Journal Club Feedback Form

Please give your classmates some constructive feedback on their presentations. Note that several of the questions were taken from https://neurograd.ucsf.edu/sites/neurograd.ucsf.edu/files/journal_club_evaluation_form.pdf

What is my github username? *

Short answer text

What are the github usernames of the presenters? *

Long answer text

What is the topic of today's Journal Club? (i.e., not the title, but a quick description of what the method is/does) *

Long answer text

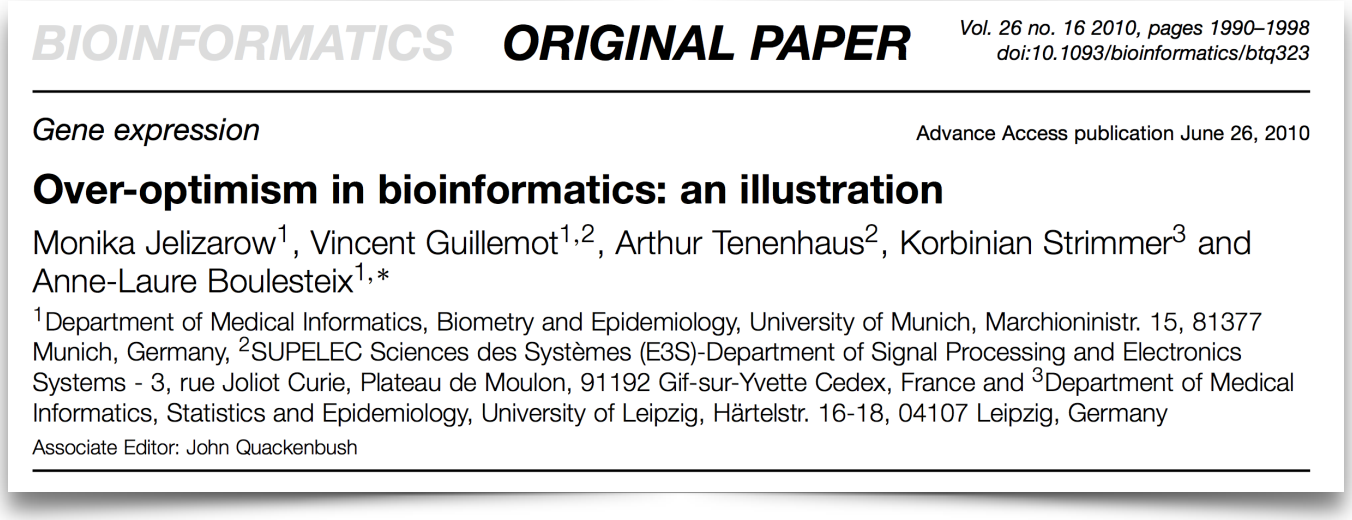
Paper Selection (importance, interest): *

- ○ Poor
- ○ Good
- ○ Very Good
- ○ Outstanding

From the feed: "Over-optimism" + Terry's IMS Bulletin

We will see a lot of methods in this course – **how do we evaluate what works well in practice ?**

http://bulletin.imstat.org/2012/11/terences-stuff-does-it-work-in-practice/



BIOINFORMATICS ORIGINAL PAPER
Vol. 26 no. 16 2010, pages 1990–1998
doi:10.1093/bioinformatics/btq323

Gene expression

Advance Access publication June 26, 2010

Over-optimism in bioinformatics: an illustration

Monika Jelizarow[1], Vincent Guillemot[1,2], Arthur Tenenhaus[2], Korbinian Strimmer[3] and Anne-Laure Boulesteix[1,*]

[1]Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Marchioninistr. 15, 81377 Munich, Germany, [2]SUPELEC Sciences des Systèmes (E3S)-Department of Signal Processing and Electronics Systems - 3, rue Joliot Curie, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France and [3]Department of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany

Associate Editor: John Quackenbush

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

Ten Simple Rules for Reducing Overoptimistic Reporting in Methodological Computational Research

Anne-Laure Boulesteix*

Institute for Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany

* boulesteix@ibe.med.uni-muenchen.de

"if the improvement of a quantitative criterion such as the error rate is the main contribution of a paper, the superiority of new algorithms should always be demonstrated on independent validation data."

In class exercise + discussion

- (5 minutes) Read the excerpt from “Terence’s Stuff” column
- (5-10 minutes; discuss with your neighbour) Answer the following 4 questions:
  1. How do we tell what works in practice?
  2. What problems arise using simulated (synthetic) data?
  3. What problems arise using real data?
  4. What are positive/negative controls?
- Discuss
- If simulation: what metrics could/should we use?
- **n.b. include this method comparison context in your Journal Club talks**