Seminar
Statistical Bioinformatics for Omics Data

Preliminaries

- English /Deutsch
- Du / Sie
- Time / Location
- Participants / Waiting List
- Campus Management
- Website: renard.it/seminar

Goal

- Hard skills
  - Know several state-of-the-art statistical methods in bioinformatics beyond the standard course level
  - Know how to come from a biological problem to a statistical solution and to a complete pipeline
- Soft skills
  - Know how to read a research paper
  - Know how to give a scientific talk (not on your own work)
  - Know how to stay in time
  - Know how to express yourself in English
  - Know how to master a discussion

Goal

- The Today Show says:
- http://www.youtube.com/watch?v=dPbxYsWxp8A

Requirements

- Attendance
  - 85% of all meetings, PLEASE be respectful and be on time!
- Active Participation
  - rule of thumb: at least one (non-random) question/remark per meeting
- Give a talk
  - 30-40 minutes length
    (roughly 10-15 minutes introduction of the biological problem and workflow, 15-20 minutes on the main statistical concept, 0-10 minutes evaluation, results, discussion)
  - stay in time
  - slides (ppt, latex, prezi, etc. but no media-show) and/or blackboard
  - answer questions in a 15-20 minutes discussion
  - in English
- Write a summary
  - 2-5 pages in English
  - focusing on the major statistical concepts
  - understandable and meant for your classmates (no c&p from wikipedia or the paper)
  - distributed at the meeting after your talk

Help

- **Do not expect to understand your paper at once!**
  Start reading early (ideally 4 weeks before your talk) and take your time
- **Collect your questions and set up a meeting with me**
  (not mandatory), please send me an email at least 1 week earlier to make sure I am available
- **Send me a complete outline of your slides (mandatory!)**
  at least 1 week before your talk (so that we can have several iterations if necessary)
- **Send me the complete summary**
  at least 48 hours before handing it out (so that I can find bugs before you share them)
- **Help each other!**
  We are far more people than I expected or can fully handle!
- **Get feedback**
  from me (and your classmates?)

Help

- **How to read a paper**
  http://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/07/paper-reading.pdf
  http://www.biochem.arizona.edu/classes/bioc568/papers.htm

- **How to give a talk**
  http://www.mpi-inf.mpg.de/~sofronie/seminar-decproc08/slides/giving-talks.pdf
  https://www.st.cs.uni-saarland.de/zeller/GoodTalk.pdf (excellent adivise, although for this specific seminar, please do not leave out all formulats)
  http://pharmacy.ucsd.edu/labs/gilson/seminar1a.html
  http://www.win.tue.nl/~hermanh/teaching/2IL35/givingatalk.html

- **How to write a summary**
  http://rwc.hunter.cuny.edu/reading-writing/on-line/summary.pdf
  http://www.writingcenter.uconn.edu/pdf/How_to_Summarize_a_Research_Article.pdf
  http://chiron.valdosta.edu/dtwasieleski/artisumm.htm

Participants

14 Teilnehmer	Montag, 8.00	Montag, 10.00	Montag, 14.00	Donnerstag, 08.00	Freitag, 16.00
Valentin	✓		(✓)	(✓)	✓
Lam-Ha	(✓)	(✓)	✓	(✓)	(✓)
Lisa	(✓)	✓	✓	(✓)	(✓)
Timo	✓		✓	✓	
Katja	(✓)			✓	✓
Adrian	(✓)	(✓)	✓	(✓)	✓
Alexander		✓	✓	(✓)	(✓)
Ria	✓		(✓)	✓	✓
Stefan	✓		✓	✓	
Herbert	✓		(✓)	✓	
Rukeia	✓			✓	
Johanna			(✓)	✓	(✓)
Vishalini	(✓)			✓	
Benjamin	✓			(✓)	
Ihr Name	Ja (Ja) Nein	Ja (Ja) Nein	Ja (Ja) Nein	Ja (Ja) Nein	Ja (Ja) Nein
	Montag, 8.00	Montag, 10.00	Montag, 14.00	Donnerstag, 08.00	Freitag, 16.00

X
X

Waiting list:

Sebastian
Jakob
~~Annkatrin~~

Topics

- Choice is always subjective
  - You will note some authors who I enjoy reading
  - I tried to stay away from my own work
  - There are plenty of important papers left out
- Idea
  - Important statistical concept, not taught in previous statistics courses
  - Relevant application in omics bioinformatics

Topics

- differential expression in RNAseq - statistical testing on distributions beyond the normal
- SNP calling in samtools - likelihood ratio testing and independence
- experimental design and reproducibility - conjugate priors and bayesian hierarchical modeling
- analyzing microarrays - normalization matters
- RNAi screening - nested effect models and causality
- phenotype screening and observational data - stability selection and causality
- viral quasispecies - expectation maximization
- de novo protein identification - hidden markov models
- database protein identification and confidence - false discovery rate
- protein inference - markovian graphical models
- tumor grading from MSI data - active learning
- brain image classification - semi-supervised learning
- biomarker classification - missing data
- predicting chemosensitivity - reproducible research

Time Slots

Day	First Talk	Second Talk
April 24	Talk 5: RNAi Screen (Jakob)	Talk 2: SNP calling (Valentin)
~~May 1st (Holiday)~~		
May 8	Talk 1: RNAseq (Lisa)	Talk 4: Microarrays (Katja)
~~May 15 (B in Izmir)~~		
May 22	Fritz (Reproduc./Genom-wide)	Talk 6: Obs. Data (Adrian)
~~May 29 (Holiday)~~		
June 5	Talk 7: Quasispecies (Ria)	Talk 8: De Novo ID (Timo)
June 12	Talk 9: Prot FDR (Alexander)	Talk 12: brain class (Benjamin)
~~June 19 (B in Baltimore?)~~		
June 26	Talk 11: MSI Grading (Sebastian)	Talk 10: Prot. Infer. (Vishalini)
~~July 3 (B back from Sheffield?)~~	Talk 13: Biomarker (Lam-Ha)	Talk 14: Bayes (Johanna)
~~July 10 (B in Boston)~~		
July 17	Back-Up	

Questions