

Tentative syllabus

- 1 03/29 Course intro
- 2 03/31 Bayesian modeling for genomic data ([Data analysis with latent variable models](#))
- 3 04/05 Bayesian inference (MCMC)
- 4 04/07 Bayesian inference (Variational inference)
- 5 04/12 [Latent variable models for scRNA-seq](#) (Presenter: Sriram)
- 6 04/14 Modeling LD (Shaviv)
- 7 04/19 Gene expression from imaging (Jessica)
- 8 04/21 Gene expression from single-cell RNA-seq (Jason)
- 9 04/26 Representation learning for gene expression (Taykhoom). AlphaFold (Derek)
- 10 04/28 Visualization methods (Ella). Pseudotime/trajectory analysis (Connor)
- 11 05/03 RNA velocity (Samuel). Single-cell transcriptomics for cancer (Emily).
- 12 05/05 ~~Predicting function of mutations (Aishwarya)~~. Kernel methods (Wenye)
- 13 05/09 Bayesian nonparametrics (Jingyou). ~~Latent factor models (Dennis)~~
- 14 05/11 ~~Interpretable ML (Da)~~. ~~Predicting function of mutations (Aishwarya)~~.
- 15 05/16 Experiment design (Brendon). Interpretable ML (Da).
- 16 05/18 Inferring causality from observational data (Alex). Latent factor models (Dennis)
- 17 05/23 Predicting function of mutations (Aishwarya). Project presentations
- 18 05/25 Project presentations
- 19 05/30 No class -- Memorial day
- 20 06/01 No class -- work on projects

TOPICS:

1. Visualization methods (tSNE, UMAP): Ella Peter
2. Single-cell transcriptomics for cancer: Emily Kao
3. Gene expression from imaging: Jessica Ho
4. Bayesian nonparametrics: Jingyou Rao
5. AlphaFold (protein folding): Derek Ma
6. Pseudotime/trajectory analysis: Connor Razma
7. Representation learning for gene expression: Taykhoom Dalal
8. Interpretable ML: Da Yin
9. Experiment design: Brendon Ng
10. Inferring causality from observational data: Alex Chen
11. RNA velocity: Samuel Lee
12. Gene expression from single-cell RNA-seq: Jason Panelli
13. Modeling LD: Shaviv Hoffman-Lowitz
14. Predicting function of mutations: Aishwarya Dev

Single-cell genomics

Experiment design

- [Zhang et al. Nature Communications 2020](#)
- [Mandric et al. Nature Communications 2020](#)

Differential expression analysis

- [MAST: Finak et al. Genome Biology 2015](#)
- [Ntranos et al. Nature Methods 2019](#)
 - Also see matters arising: [Bech et al. Nature Methods 2020](#)
- Also papers for DE analysis for bulk RNA-seq data
 - [Love et al. Genome Biology 2014](#)
 - [LIMMA](#)

Pseudotime/trajectory analysis

- [Monocle: Trapnell et al. Nature Biotechnology 2014](#)
- [tradeseq: Berge et al. Nature Communications 2020](#)
- [Optimal transport analysis: Schiebinger et al. Cell 2019](#)

Representation learning

- [Convolutional Neural Network for Coexpression: Yuan and Joseph PNAS 2019](#)

Allele-specific gene expression

- [SCALE](#)

RNA velocity

- [Manno et al. Nature 2018](#)
- [Bergen et al. Nature Biotechnology 2020](#)

Visualization

- [UMAP: McInnes et al. 2018](#)
- [tSNE: van der Maaten JMLR 2008](#)
- [PHATE: Moon et al. Nature Biotech. 2019](#)
- [Narayan et al. Biorxiv 2020](#)

Spatial transcriptomics

- [Gene expression cartography: Nitzan et al. 2019](#)
- [Moncada et al. Nature Biotech 2020](#)
- [Graph-based convolutional networks: Song and Su Briefings in Bioinformatics 2021](#)

Denoising/Imputation/Prediction

- [SAVER: Huang et al. Nature Methods 2018](#)
- [SAVER-X: Wang et al. Nature Methods 2019](#)
- [NetNMF-SC: Elyanow et al. Genome Research 2020](#)

Data integration/Identifying batch effects/shared variation

- [scLVM: Buettner et al. Nature Biotech. 2015](#)
- [SEURAT: Buttlar et al. Nature Biotech 2018](#)
- [Korunsky et al. Nature Methods 2019](#)

Cell-type specific gene expression

- [csSAM: Shen-Orr et al. Nature Methods 2010](#)

Single-cell transcriptomics for cancer

- [Fan et al. 2020](#)
- [Fan et al. Genome Research 2018](#)

Single-cell multiomics

- [Surface protein imputation from scRNA-seq: Zhou et al .Nature communications 2020](#)
- [Gene expression prediction from imaging: Schumauch et al. Nature Communications2020](#)

Manifold learning

- [Moon et al. 2018](#)

Lineage reconstruction

-

Foundational topics

The practice of probabilistic modeling

- [David Blei](#)(Chapters 1,2,3.1,3.2,5)
- Probabilistic programming
 - [STAN](#)

MCMC

- [Geyer](#)
- Other references
 - [Andrieu, deFreitas et al.](#)
 - [Neal](#)

Variational Inference

- [Variational Inference: A Review for Statisticians](#)
- [Scalable variational inference](#)
- [Black box variational inference](#)
- Applications:
 - Ancestry analysis (aka topic models): [FastStructure](#), [TeraStructure](#)
 - Gene expression analysis: [PEER](#)
- Other references
 - Automatic differentiation Variational Inference: [ADVI](#)

Kernel methods

- [Kernel methods in comp. bio](#)
- Applications:
 - [Multikernel linear mixed models for complex phenotype prediction](#)
 - [Multiple kernel learning for single cell analysis](#)

Bayesian nonparametrics

- Dirichlet process: [Tutorial](#)
- [Gaussian processes, chapter 2 and 4](#)
- Applications:
 - Predicting complex traits using latent dirichlet process regression: [LDPR](#)
 - Predicting gene expression using dirichlet processes: [TIGAR](#)

Deep learning

- Deep learning frameworks for genomics: [Janggu](#)
- Applications
 - [Cancer prediction](#)
 - [Gene expression coevolution](#)
 - [Gene expression prediction from imaging](#)
 - [scRNA](#)
 - Predicting splicing: [Alipanahi et al. Nature Biotechnology 2014](#)
 - Predicting function of mutations: [Zhou and Troyanskaya 2015](#)
 - Variant calling: [Poplin et al. 2016](#)

Optimal transport analysis

- [Schiebinger et al. Cell 2019](#)

- [Gene expression cartography: Nitzan et al. 2019](#)

Latent factor models

- Probabilistic PCA: [Tipping and Bishop 2002](#)
- PCA for correcting for population stratification: [Price et al. 2006](#)
- PCA and relation to population structure: [Patterson et al. 2006](#)
- PCA and factor analysis: [Engelhardt and Stephens 2012](#)
- Other reference
 - PCA for large-scale genetic data (Galinsky et al. 2015).

Interpretability

- SHAP: [Lundberg and Lee. 2017](#)
- LIME: [Ribero et al. 2016](#)
- ELDR:

Causality

- Inferring causality from observational data: [Didelez et al. 2010](#)