1	April 2	Course intro
2	April 4	Bayesian Modeling (Data analysis with latent variable models)
3	April 9	Bayesian inference -- MCMC
4	April 11	Bayesian inference -- Variational inference (Variational inference review)
5	April 16	Heritability (Lingyu Zhan)
6	April 18	Phenotype prediction (Ruth Johnson)
7	April 23	Gene expression (Harry Yang)
8	April 25	Time series (Juan de la Hoz)
9	April 30	Latent factor models (Wenshan Li)
10	May 2	Metagenomics (Leah Briscoe)
11	May 7	Guest lecture (Eran Halperin)
12	May 9	Guest lecture
13	May 14	Coalescent (Aaron Zhou)
14	May 16	Sequence models (Jeff Lung)
15	May 21	Chip-Seq and Hi-C (Tommer Schwartz)
16	May 23	Deep Learning (Ruiming Cao). Causality (Vignesh Sairaj)
17	May 28	Project presentations
18	May 30	Project presentations
19	June 4	No class -- work on projects
20	June 6	No class -- work on projects

Project presentation schedule (tentative list)
May 28
Aaron Zhou
Harry Yang, Vignesh Sairaj
Juan de la Hoz Gomez
Leah Briscoe

May 30
Tommer Schwartz
Lingyu Zhan
Ruiming Cao
Ruth Johnson
Wenshan Li
Jeff Lung

Possible topics

Models of genetic variation

- Li and Stephens, Genetics 2003

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
  - Gene expression analysis: PEER
- Other references
  - Automatic differentiation Variational Inference: ADVI

Kernel methods

- Kernel methods in comp. bio
- Applications:
  - Powerful SNP-set analysis for case-control genome-wide association studies
  - 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:
  -

Mixed models and variance components

- Applications
  - Overview of genome-wide association studies (GWAS): Eskin, CACM 2015
  - Mixed models for heritability estimation: Yang et al. Nature Genetics 2010
  - Mixed models for correcting for confounders: Kang et al. Genetics 2008
- Computational issues
  - Loh et al. 2015
  - Lippert et al. 2011

Heritability

- Definitions
- Concepts and challenges: Zuk et al. PNAS 2011
- Mixed models: Yang et al. Nature Genetics 2010

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).

Causality

- Inferring causality from observational data: Didelez et al. 2010

Stochastic process over trees: Coalescent

- Wakeley, chapter 3 (http://dna.ac/filogeografia/PDFs/Wakeley/Wakeley_06_Chapter_3.pdf)

Deep learning

- Methodological papers: Hyperparameter tuning, Sequence models, Permutation-invariance
- Applications
    - Predicting splicing: Alipanahi et al. Nature Biotechnology 2014
    - Predicting function of mutations: Zhou and Troyanskaya 2015
    - Variant calling: Poplin et al. 2016

Interpretability

- SHAP: Lundberg and Lee, 2017
- LIME: Ribero et al. 2016

Sequence models

- Introduction to HMMs
- Applications
    - Li and Stephens, Genetics 2003
    - Li and Durbin 2011

Time series

Methodological papers

- Recurrent neural nets
- Gaussian processes

- State space models

Time series for medical data

- EHR Data and kidney disease prediction
- Recurrent neural nets for time series prediction
- Biases in EHR lab tests

Time series for genomic data

- Ancient DNA: Mathieson et al. 2015
- Gene expression: Lin et al.2008

Multiple testing

- Multiple Testing Correction Intro, Noble 2009
- Paper on GWAS multiple testing: Han et al
- Other references
  - Multiple Hypothesis Testing in Microarray Experiments, Dudoit, et al
  - Section 18.7 of ESL
  - EQTL data: Sul et al. 2015

Metagenomics

- Read assignment
- Association testing

Hidden Markov Models

- Tutorial
- Applications
  - Models of human genetic variation: Li and Stephens, Genetics 2003
  - Inferring demographic history: Li and Durbin 2011
  - Local ancestry inference: Pasaniuc, Sankararaman et al. 2012

Phenotype prediction

- High-dimensional regression
  - Ridge regression
  - LASSO ESL Chapters 3.1-3.4
- LDPRED: http://www.cell.com/ajhg/abstract/S0002-9297(15)00365-1
- Variational inference for high-dimensional regression: Carbonetto and Stephens 2012
- Accurate prediction of human height

Gene expression analysis

- Transcript quantification from RNA-Seq
  - Probabilistic Models

    - Pseudo-alignment
- Correcting for confounders in gene expression: PEER
- Manifold learning for single cell RNA

Genomic privacy

- Erlich and Narayanan, 2014
- Simons et al. Cell Systems 2015
- Dwork et al. 2014

Demographic inference

- HMMs: Li and Durbin 2011
- Gaussian process: Palacios et al. 2016

Radiogenomics

- MRI and gene expression: Stoyanova et al. 2016
- Radiogenomic map: Zhou et al. 2018