Lecture 9: Introduction to Genomics
Announcements

- Upcoming deadlines:
  - Project proposal was due last Friday
  - A2 due next Wed Oct 21
Some biology basics: starting from DNA

Figure credit: virtualmedicalcentre.com
Some biology basics: starting from DNA

Figure credit: virtualmedicalcentre.com
Some biology basics: starting from DNA

~ 37 trillion cells in the human body

Figure credit: virtualmedicalcentre.com
Some biology basics: starting from DNA

Nucleus: “brain of the cell”. Contains genetic material in the form of DNA.
Some biology basics: starting from DNA
Some biology basics: starting from DNA

Figure credit: https://en.wikipedia.org/wiki/Nucleobase#/media/File:DNA_chemical_structure.svg

Figure credit: virtualmedicalcentre.com
Chromosomes and genes

U.S. National Library of Medicine

Figure credit: https://gr.nlm.nih.gov/primer/illustrations/chromosomes.jpg

Figure credit: https://www.ncbi.nlm.nih.gov/books/NBK22266/bin/a01chr.jpg
Chromosomes and genes

23 pairs of chromosomes (22 autosomes + sex chromosomes)
Chromosomes and genes

Genes: segments of DNA within chromosomes

Figure credit: https://ghr.nlm.nih.gov/primer/illustrations/chromosomes.jpg

Figure credit: https://www.ncbi.nlm.nih.gov/books/NBK22266/bin/a01chr.jpg
Chromosomes and genes

Genes: segments of DNA within chromosomes

Genes provide code for proteins

Figure credit: https://ghr.nlm.nih.gov/primer/illustrations/chromosomes.jpg

Figure credit: https://www.ncbi.nlm.nih.gov/books/NBK22266/bin/a01chr.jpg

Genes: segments of DNA within chromosomes

Genes provide code for proteins
Chromosomes and genes

Genes: segments of DNA within chromosomes

Genes provide code for proteins
But 99% of genes are “non-coding!”

Figure credit: https://www.ncbi.nlm.nih.gov/books/NBK22266/bin/a01chr.jpg

Figure credit: https://ghr.nlm.nih.gov/primer/illustrations/chromosomes.jpg
DNA replication and transcription

Replication

Transcription and Translation

Figure credit:
https://www.bosterbio.com/media/images/MB_Replication_and_Transcription.png
DNA replication and transcription

Mitosis

Replication

Meiosis

Transcription and Translation

Figure credit:
https://en.wikipedia.org/wiki/Mitosis#/media/File:Major_events_in_mitosis.svg
https://en.wikipedia.org/wiki/Meiosis#/media/File:Meiosis_Overview_new.svg

https://www.bosterbio.com/media/images/MB_Replication_and_Transcription.png
DNA replication and transcription

Replication

Transcription and Translation

Figure credit:
https://www.bosterbio.com/media/images/MB_Replication_and_Transcription.png
Transcription and translation

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
Transcription and translation

Transcription: DNA -> RNA

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
Transcription and translation

Transcription: DNA -> RNA

Translation: RNA -> Protein

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
DNA -> Pre-mRNA

Figure credit: http://u18439936.onlinehome-server.com/craig.milgrim/Bio230/Outline/ECBFigures_Tables/Chapter_7/FigureJPGs/figure_07_09.jpg

Serena Yeung  BIODS 220: AI in Healthcare  Lecture 9 - 20
DNA -> Pre-mRNA

Gene to transcribe

Figure 7-9 Essential Cell Biology 3/e (© Garland Science 2010)

Figure credit:
http://u18439936.onlinehome-server.com/craig.milgrim/Bio230/Outline/ECBFigures_Tables/Chapter_7/FigureJPGs/figure_07_09.jpg

Serena Yeung  BIODS 220: AI in Healthcare  Lecture 9 - 21
DNA -> Pre-mRNA

RNA polymerase: enzyme that binds to promoter region and uses DNA template to synthesize complementary RNA

Figure credit:
http://u18439936.onlinehome-server.com/craig.milgrim/Bio230/Outline/ECBFigures_Tables/Chapter_7/FigureJPGs/figure_07_09.jpg
DNA -> Pre-mRNA

Figure credit:
Pre-mRNA -> mRNA

Figure credit: http://academic.pgcc.edu/~kroberts/Lecture/Chapter%207/transcription.html
mRNA splicing: remove introns (non-coding regions), splice together exons (coding regions)
mRNA -> Proteins

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
mRNA -> Proteins

Ribosome: cell organelle that synthesizes proteins

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
mRNA -> Proteins

tRNA: molecule carrying amino acids corresponding to each 3-nucleotide codon

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
mRNA -> Proteins

Figure credit: https://www.cancer.gov/images/cdr/live/CDR761782-571.jpg
mRNA -> Proteins

Figure credit:
https://philschatz.com/biology-concepts-book/resources/Figure_09_04_02.jpg
mRNA -> Proteins

Codon -> amino acid mapping

Figure credit:
https://philschatz.com/biology-concepts-book/resources/Figure_09_04_02.jpg
Epigenomics

Study of processes that regulate how and when genes are turned on and off ("gene expression")
Epigenomics

Study of processes that regulate how and when genes are turned on and off (“gene expression”)

- E.g. **transcription factors**: proteins that bind to the promoter and other noncoding regions, can enhance or repress transcription
Epigenomics

Study of processes that regulate how and when genes are turned on and off (“gene expression”)

- **E.g. transcription factors**: proteins that bind to the promoter and other noncoding regions, can enhance or repress transcription

- **E.g. DNA methylation**: addition of large methyl group to promoter region makes it difficult for proteins to bind -> represses transcription
Epigenomics

Study of processes that regulate how and when genes are turned on and off (“gene expression”)

- E.g. **transcription factors**: proteins that bind to the promoter and other noncoding regions, can enhance or repress transcription

- E.g. **DNA methylation**: addition of large methyl group to promoter region makes it difficult for proteins to bind -> represses transcription

- E.g. **Histone modification**: addition or removal of acetyl groups affects charge interaction to relax or tighten chromatin structure (easier for proteins to bind)
Transcriptomics

- Study of the transcriptome (the RNA of a cell)
- One reason of interest: Harder to measure proteins (the functional molecules!), but we can sequence RNA as a (highly imperfect) proxy for proteins to quantify cell state
Transcriptomics

- Study of the transcriptome (the RNA of a cell)
- One reason of interest: Harder to measure proteins (the functional molecules!), but we can sequence RNA as a (highly imperfect) proxy for proteins to quantify cell state

Proteomics

- Study of the proteins in a cell
Data: genomic sequencing

Produces readout of DNA template strands

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Add some special (and fluorescently labeled) nucleotides that cause a chain being synthesized to terminate

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Random interaction of nucleotides with template strand lead to chains of different early-terminated lengths

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Sorting by length (e.g. electrophoresis) gives sequence readout

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

Arrange many short templates on an array

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

Now all added nucleotides are chain-terminating

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

All templates get next sequence element attached (and terminated), then read

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

Apply process to “restore” the chain-terminating nucleotides to be normal, then repeat to extend synthesizing sequence by one more nucleotide

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: genomic sequencing

Produces readout of DNA template strands

Sanger sequencing: Invented in 1977, based on “chain termination”

Next-generation sequencing (NGS): Used since 2000s, based on massively parallelized sequencing of short sequences

Set of read-out images at every step gives sequences of all template strands. Then analyze data to reconstruct longer sequences.

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig1_HTML.jpg
Data: DNA microarray

Produces relative expression of genes in normal vs disease tissue samples

Figure credit: http://www.vce.bioninja.com.au/_Media/microarray_med.jpeg
Data: DNA microarray

Produces relative expression of genes in normal vs disease tissue samples

Isolate mRNA ("expressed genes") from tissue samples and synthesize complementary DNA (cDNA).

Figure credit: http://www.vce.bioninja.com.au/_Media/microarray_med.jpeg
Data: DNA microarray

- Produces relative expression of genes in normal vs disease tissue samples

- Isolate mRNA (“expressed genes”) from tissue samples and synthesize complementary DNA (cDNA).

- Use fluorescent tags to label cDNA from normal tissue green, and from disease tissue red.

Figure credit: http://www.vce.bioninja.com.au/_Media/microarray_med.jpeg
Data: DNA microarray

Produces relative expression of genes in normal vs disease tissue samples

Each spot of DNA microarray contains single-stranded DNA corresponding to a gene

Figure credit: http://www.vce.bioninja.com.au/_Media/microarray_med.jpeg
Data: DNA microarray

Produces relative expression of genes in normal vs disease tissue samples.

cDNA will bind to the corresponding DNA strands on microarray. Color indicates ratio of cDNA (relative gene expression) in the normal vs disease tissue.

Figure credit: http://www.vce.bioninja.com.au/_Media/microarray_med.jpeg
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Isolate RNA and generate cDNA

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Use NGS to sequence cDNA

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Map back to reference genome for analysis

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Map back to reference genome for analysis

Now standard approach for transcriptomics study

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: RNA-seq

Produces readout of mRNA content in a tissue sample

Map back to reference genome for analysis

Now standard approach for transcriptomics study

More recently in 2010s, single-cell RNA-seq!

Figure credit: https://cdn.technologynetworks.com/tn/images/body/dnasequencinga1529596208892.png
Data: ChIP-seq

Produces reads of DNA sequences where a protein binds

Use formaledehyde treatment to cross-link (fix) proteins to their bound DNA

Figure credit:
Data: ChIP-seq

Produces reads of DNA sequences where a protein binds

Disintegrate non-bound DNA -> what is left is DNA segments bound to protein

Figure credit:
Data: ChIP-seq

Produces reads of DNA sequences where a protein binds

Treat sample to remove proteins
Data: ChIP-seq

Produces reads of DNA sequences where a protein binds.

Use NGS to read-out remaining DNA sequences

Figure credit: https://www.france-genomique.org/wp-content/uploads/2019/08/CHIP-selon-Park-1-e1566900408602.jpg
Data: ChIP-seq

Produces reads of DNA sequences where a protein binds

Visualize distribution of locations on DNA where protein binds

Figure credit:

Figure credit:
https://www.researchgate.net/publication/262150050/figure/fig2/AS:27256850559751@1441996433141/Chromatin-domain-containing-VDR-binding-sites-The-IGV-browser-was-used-to-display-the.png
ENCODE: identifying and analyzing all functional elements in the human genome

- Launched by US National Human Genome Research Institute in 2003
- Contributions from worldwide consortium of research groups
ENCODE data
ENCODE data

- **long read RNA-seq of left lung**
  - *Homo sapiens* left lung male adult (40 years)
  - Lab: Ali Mortazavi, UCI
  - Project: ENCODE

- **long read RNA-seq of left lung**
  - *Homo sapiens* left lung female child (16 years)
  - Lab: Ali Mortazavi, UCI
  - Project: ENCODE

- **long read RNA-seq of ovary**
  - *Homo sapiens* ovary female adult (41 years)
  - Lab: Ali Mortazavi, UCI
  - Project: ENCODE

- **long read RNA-seq of mucosa of descending colon**
  - *Homo sapiens* mucosa of descending colon female adult (51 years)
  - Lab: Ali Mortazavi, UCI
  - Project: ENCODE
### ENCODE data

**Common Cell Types: Tier 1 and Tier 2**

<table>
<thead>
<tr>
<th>Cell, tissue or DNA sample</th>
<th>Description</th>
<th>Lineage</th>
<th>Tissue</th>
<th>Karyotype</th>
<th>Sex</th>
<th>Documents</th>
<th>Vendor ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM12878</td>
<td>B-lymphocyte, lymphoblastoid, International HapMap Project - CEPH/Utah - European Caucasian, Epstein-Barr Virus</td>
<td>mesoderm</td>
<td>blood</td>
<td>normal</td>
<td>F</td>
<td>ENCODE</td>
<td>Coriell GM12878</td>
</tr>
<tr>
<td>H1-hESC</td>
<td>embryonic stem cells</td>
<td>inner cell mass</td>
<td>embryonic stem cell</td>
<td>normal</td>
<td>M</td>
<td>ENCODE</td>
<td>WiCell Research Institute WA01</td>
</tr>
<tr>
<td>K562</td>
<td>leukemia, &quot;The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises.&quot; - ATCC</td>
<td>mesoderm</td>
<td>blood</td>
<td>cancer</td>
<td>F</td>
<td>ENCODE</td>
<td>ATCC CCL-243</td>
</tr>
</tbody>
</table>

Total = 3

**Cell, tissue or DNA sample:** Cell line or tissue used as the source of experimental material.

<table>
<thead>
<tr>
<th>cell</th>
<th>Tier</th>
<th>Description</th>
<th>Lineage</th>
<th>Tissue</th>
<th>Karyotype</th>
<th>Sex</th>
<th>Documents</th>
<th>Vendor ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>A549</td>
<td>2</td>
<td>epithelial cell line derived from a lung carcinoma tissue. (PMID: 175022), &quot;This line was initiated in 1972 by D.J. Giard, et al. through explant culture of lung carcinomatous tissue from a 58-year-old caucasian male.&quot; - ATCC, newly promoted to tier 2: not in 2011 analysis</td>
<td>endoderm</td>
<td>epithelium</td>
<td>cancer</td>
<td>M</td>
<td>Myers Crawford Stam</td>
<td>ATCC CCL-185</td>
</tr>
</tbody>
</table>
Other datasets

https://www.ga4gh.org/community/catalogue
Other datasets

https://www.ga4gh.org/community/catalogue

Catalogue of Genomic Data Initiatives

The purpose of the catalogue is to identify and aggregate global resources for sharing clinical and genomic data.

1000 Genomes

CATEGORY: eHealth
INITIATIVE TYPE: Database

The 1000 Genomes Project set out to catalogue common human genetic variation, publishing a set of variations based on sequencing of 2504 individuals from 26 populations. Additional work was done to investigate structural variations in the human genome. Variant calls, sequence data, high-density genotyping chip calls and cell lines from the Project are all available. Data from the 1000 Genomes Project is now housed in the International Genome Sample Resource (IGSR), which is realigning sequence data from the 1000 Genomes Project to the updated GRCh38 human genome assembly and also expanding the data resources produced by 1000 Genomes to include new samples with similarly open consent, new populations and a wider range of data types. Further information, access to data and user support are available.

Antigenic Variation Database (VarDB)

CATEGORY: 
INITIATIVE TYPE:

Serena Yeung

BIODS 220: AI in Healthcare

Lecture 9 - 70
Remember from Lecture 1

1953 - Watson and Crick discover double helix structures of DNA

1977 - Fred Sanger sequences first full genome of a virus

1990 - 2003: Human Genome Project sequences full human genome

2003: ENCODE project launched to identify and characterize genes in human genome

1000 Genomes Project: 2008 - 2015

UK100,000 Genomes Project: 2012 - 2018
DeepBind

Input: DNA sequence
Output: Score of whether a particular protein will bind to the sequence or not

DeepBind

Input: DNA sequence
Output: Score of whether a particular protein will bind to the sequence or not

**DeepBind**

Input: DNA sequence  
Output: Score of whether a particular protein will bind to the sequence or not

- Processing to handle different sources of experimental (training) data and input / output data formats
- Trained on 12 TB of sequence data; learned 927 DeepBind models representing 538 transcription factor (TF) proteins and 194 RNA-binding proteins (RBPs)

DeepBind

Input: DNA sequence
Output: Score of whether a particular protein will bind to the sequence or not

- Processing to handle different sources of experimental (training) data and input / output data formats
- Trained on 12 TB of sequence data; learned 927 DeepBind models representing 538 transcription factor (TF) proteins and 194 RNA-binding proteins (RBPs)

DeepBind

Learned DeepBind motifs

DeepBind

Predicted effect of sequence mutations

SP1 loss in LDL-R promoter

WT

Mutant: C>G

G>A would increase binding

chr19:11,200,089 (c.–136)

G>C would decrease binding

GATA1 gain in α-globin cluster

WT chr16:209,709

Mutant: A>G

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

Input: DNA sequence pair with SNP
Output: Predicted chromatin effects (919 total)
- 690 transcription factor profiles
- 125 DNase I hypersensitive sites (DHS) profiles (looser chromatin structure, easier protein binding)
- 104 histone-mark profiles (histone modifications)

Multi-task training!

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

Input: DNA sequence pair with SNP
Output: Predicted chromatin effects (919 total)
- 690 transcription factor profiles
- 125 DNase I hypersensitive sites (DHS) profiles (looser chromatin structure, easier protein binding)
- 104 histone-mark profiles (histone modifications)

Multi-task training!

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

Input: DNA sequence pair with SNP
Output: Predicted chromatin effects (919 total)
  - 690 transcription factor profiles
  - 125 DNase I hypersensitive sites (DHS) profiles (looser chromatin structure, easier protein binding)
  - 104 histone-mark profiles (histone modifications)

Multi-task training!

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

Input: DNA sequence pair with SNP
Output: Predicted chromatin effects (919 total)
  - 690 transcription factor profiles
  - 125 DNase I hypersensitive sites (DHS) profiles (looser chromatin structure, easier protein binding)
  - 104 histone-mark profiles (histone modifications)

Multi-task training!

DeepSea

Predict chromatin effects of (non-coding) sequence alterations with single-nucleotide sensitivity (SNPs: single nucleotide polymorphism)

Input: DNA sequence pair with SNP
Output: Predicted chromatin effects (919 total)
- 690 transcription factor profiles
- 125 DNase I hypersensitive sites (DHS) profiles (looser chromatin structure, easier protein binding)
- 104 histone-mark profiles (histone modifications)

Multi-task training!

DeepSea

Model Architecture:

2. Pooling layer (Window size: 4. Step size: 4.)
4. Pooling layer (Window size: 4. Step size: 4.)
6. Fully connected layer (925 neurons)
7. Sigmoid output layer

DeepVariant

Variant calling: identifying variants from reference genome (SNPs, small indels, etc.)

Figure credit: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4633438/bin/40142_2015_76_Fig2_HTML.jpg

DeepVariant

Variant calling: identifying variants from reference genome (SNPs, small indels, etc.)

A) GENOME REFERENCE

Challenge with short, errorful sequence reads from NGS!

B) Homozygous reference  Heterozygous  Homozygous alternate

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

Used an Inception v3 CNN

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

Used an Inception v3 CNN

DeepVariant

Input: “Pileup images” of reference sequence + NGS reads, + other features

Output: Categorical prediction of variant type (hom-ref, het, hom-alt), or no variant

Used an Inception v3 CNN

Won highest performance for SNPs in the 2016 FDA variant calling Truth Challenge

More examples of deep learning in genomics

**Epigenomics:**
- Predicting methylation states, gene expression from histone modifications, etc.

**Transcriptomics:**
- Predicting phenotypes from transcriptome, identifying genes associated with transcriptomic data, etc.

**Proteomics:**
- Predicting secondary structure of proteins, protein-protein interactions, etc.
Summary

Today we covered:

- Biology basics for genomics
- Epigenomics, transcriptomics, proteomics
- Genomics data
- Examples of deep learning for genomics