

Summer Institutes of Statistical Genetics, 2023

Module 2: INTRODUCTION TO GENETICS AND GENOMICS

Greg Gibson and Joe Lachance

Georgia Institute of Technology

Lecture 7: GENE EXPRESSION STUDIES

Rationale for Gene Expression Profiling

Question: What makes a muscle cell different from a skin cell different from a liver cell?

Analogy: What makes a living room different from a bathroom different from the kitchen?

Answer: What furniture and appliances and windows are placed where and when

What genes are expressed where and when and how much

Strategy: Measure the abundance of mRNA transcripts in a bunch of samples, and use statistically rigorous approaches to identify differential expression

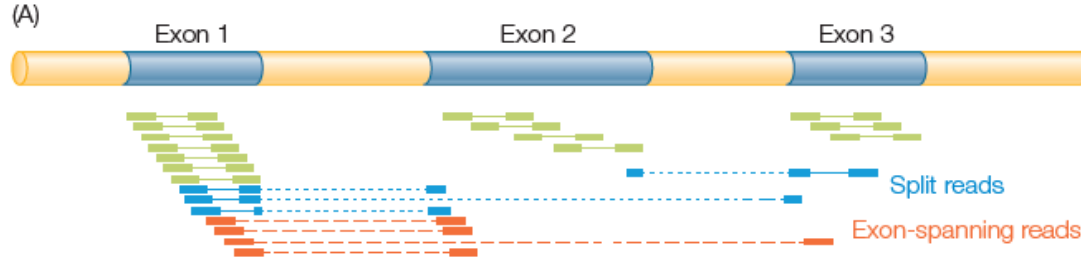
Methods: Prior to 1995: One gene at a time qPCR

1995-2015: Microarrays

Since 2015: RNAseq

Since 2019: single cell RNAseq

Principle of RNA-seq



Gene / Exon / Transcript Abundance

(B)

Reference:

```

ACGGCATTCATCCTACGGGCCATCCACTACGGCTGCTAAGCCACACCCATATAACGGGC
  GGCATTCATCCTACGGGCCATCCACTACGACTGCTAAG
    TTCATCCTACGGGCCATCCACTACGGCTGCTAAGC
      CATCCTACGGGCCATCCACTACGACTGCTAAGCCA
        CTACGGGCCATCCACTACGACTGCTAAGCCACAC
          TACGGGCCATCCACTACGGCTGCTAAGCCACACCCAT
            GCGCCATCCACTACGGCTGCTAAGCCACACCCAT
              GCGCCATCCACTACGACTGCTAAGCCACACCCAGAT
                CGCCATCCACTACGGCTGCTAAGCCACACCCATATAC
                  CGCCATCCACTACGGCTGCTAAGCCACACCCATATAC
                    ACTACGACTGCTAAGCCACACCCAGATACCG
                      TACGACTCCTAAGCCACACCCAGATACCGG
                        GGCTGCTAAGCCACACCCATATAACGGGC
                                ↑   ↑                               ↑

```

Allele-specific Expression

(C)

| | | | | | |
|------|-----|---|----|-----------------------------------|-----------------------|
| seq1 | 272 | G | 17 | , . \$ | <<<+; <<<<<<<<; <&&< |
| seq2 | 273 | G | 18 | . , A , . a . a . a , T , a T . T | <<; ; <<<<<<&3; <=&<< |
| seq3 | 274 | C | 18 | \$. , | <<; ; <<&<<&<; <<&<=6 |
| seq4 | 275 | T | 16 | | <<<+<<=<<&7; <<<<; |
| seq5 | 276 | G | 17 | , C , | <6 <<<<<<<<; <<<<&4< |
| seq6 | 277 | C | 15 | | <; ; <<<<<<<<<; <<<< |

FastQ format stores sequence and quality

Gene Ontology Analysis: Annotating Gene Function

Biological process

All [716478]
GO:0008150 : biological_process [563081]
GO:0032502 : developmental process [51920]
GO:0048856 : anatomical structure development [52222]
GO:0001568 : blood vessel development [4180]
GO:0060840 : artery development [423]
GO:0035904 : aorta development [160]
GO:0035905 : ascending aorta development [20]



Expression in a tissue generally implies the gene is active there

Molecular function

All [716478]
GO:0003674 : molecular_function [577197]
GO:0001071 : nucleic acid binding transcription factor activity [18771]
GO:0003700 : sequence-specific DNA binding transcription factor activity [18743]
GO:0000981 : sequence-specific DNA binding RNA polymerase transcription factor activity [4244]
GO:0000982 : RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity [1208]
GO:0001078 : RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription [337]



Co-expression with similar types of gene may imply “guilt by association”

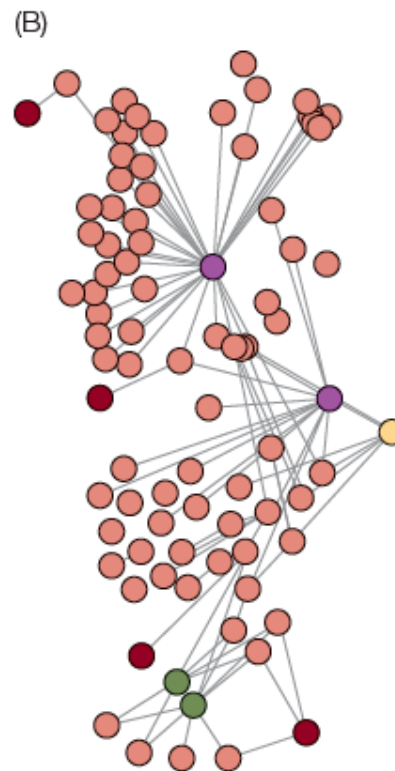
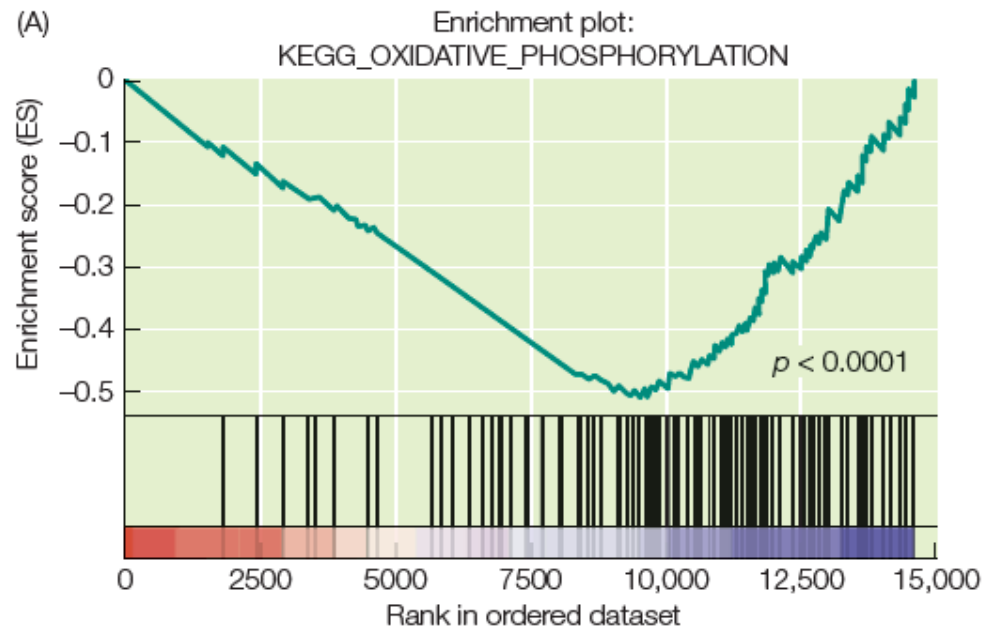
Cellular component

All [716478]
GO:0005575 : cellular_component [505276]
GO:0005623 : cell [298439]
GO:0005622 : intracellular [262296]
GO:0043299 : intracellular organelle [209322]
GO:0043231 : intracellular membrane-bounded organelle [175805]
GO:0005634 : nucleus [83664]



FISH and sub-cellular imaging of proteins tells us where they act

Gene Enrichment Analysis: Hypothesis generation



2 Approaches to Enrichment:

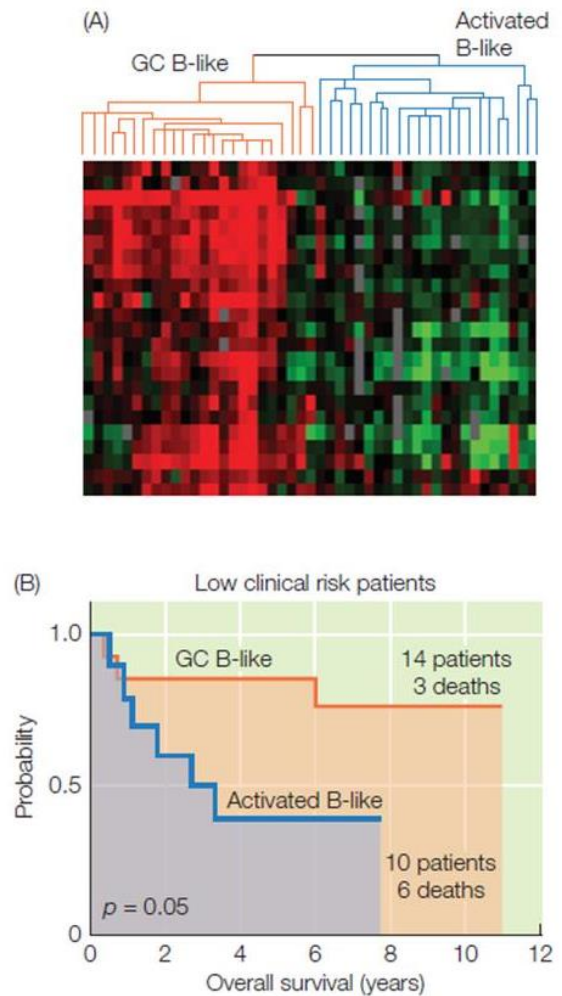
GSEA asks whether the rank of expression of genes in known pathways is skewed up or down

GO asks whether there is an excess of genes in a DE (differentially expressed) set in a GO term relative to the whole transcriptome

Network analysis places genes as hubs and nodes

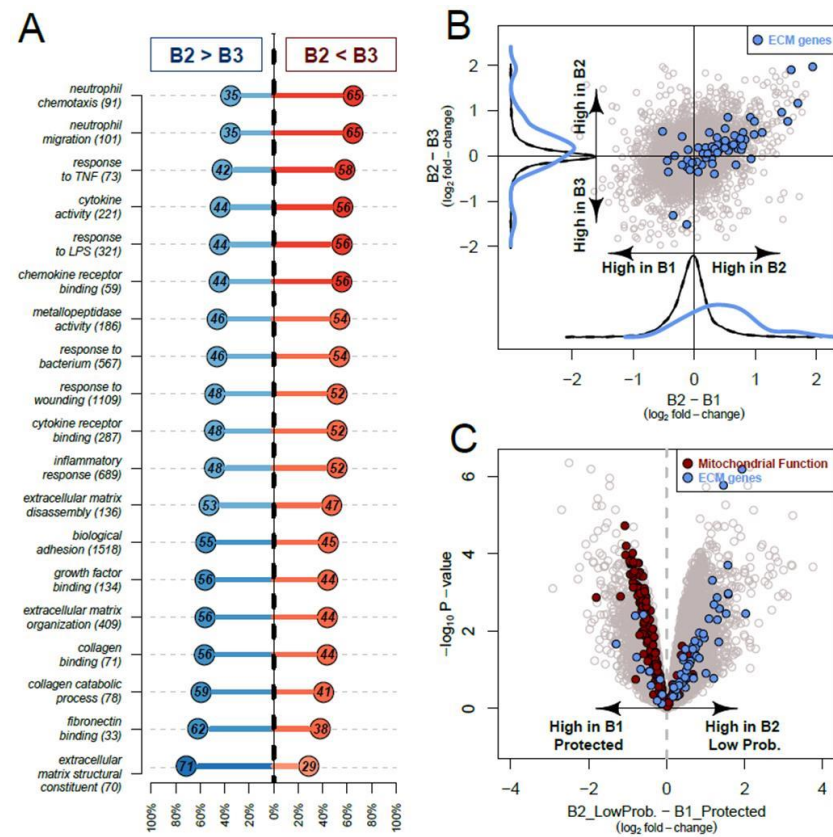
Differential Gene Expression

Cancer



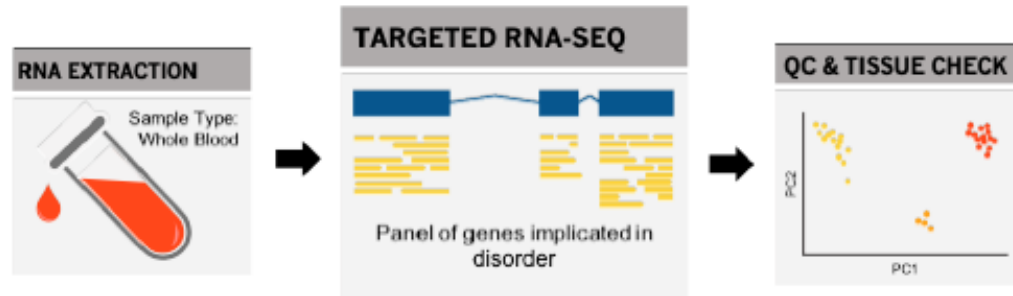
Alizadeh *et al* (2000) *Nature* **403**: 503-511

Inflammatory Bowel Disease



Kugathasan *et al* (2017) *Lancet* **389**: 1710-1718

Clinical Diagnostics



RNA VARIANT CALLING

Confirmation of specific variants from DNA sequencing

Identification of novel variants

Observation of transcript structure effects of DNA variants

AGCTACAGACAA

ALLELE EXPRESSION IMBALANCE

Nonsense mediated decay
Non-stop mediated decay
Allele specific expression

lesser allele read count/ Total

ctrl case

ABERRANT SPLICING

exon skip
pseudoexon gain
exon extension
exonic splice gain

ISOFORM ABUNDANCE

usage

Differential exon usage
Percent Spliced In

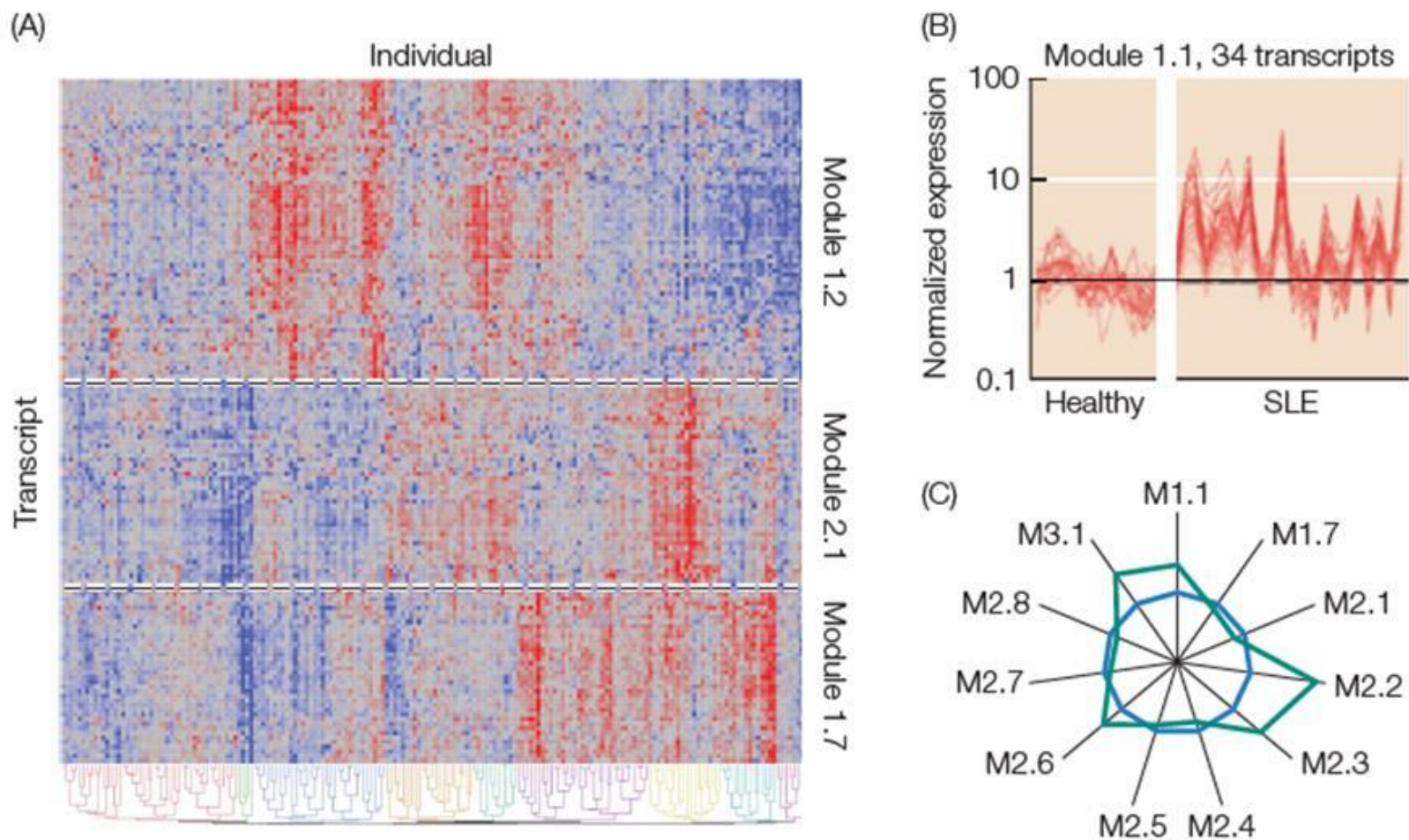
exonic part 1 2 3 4 5

gene model

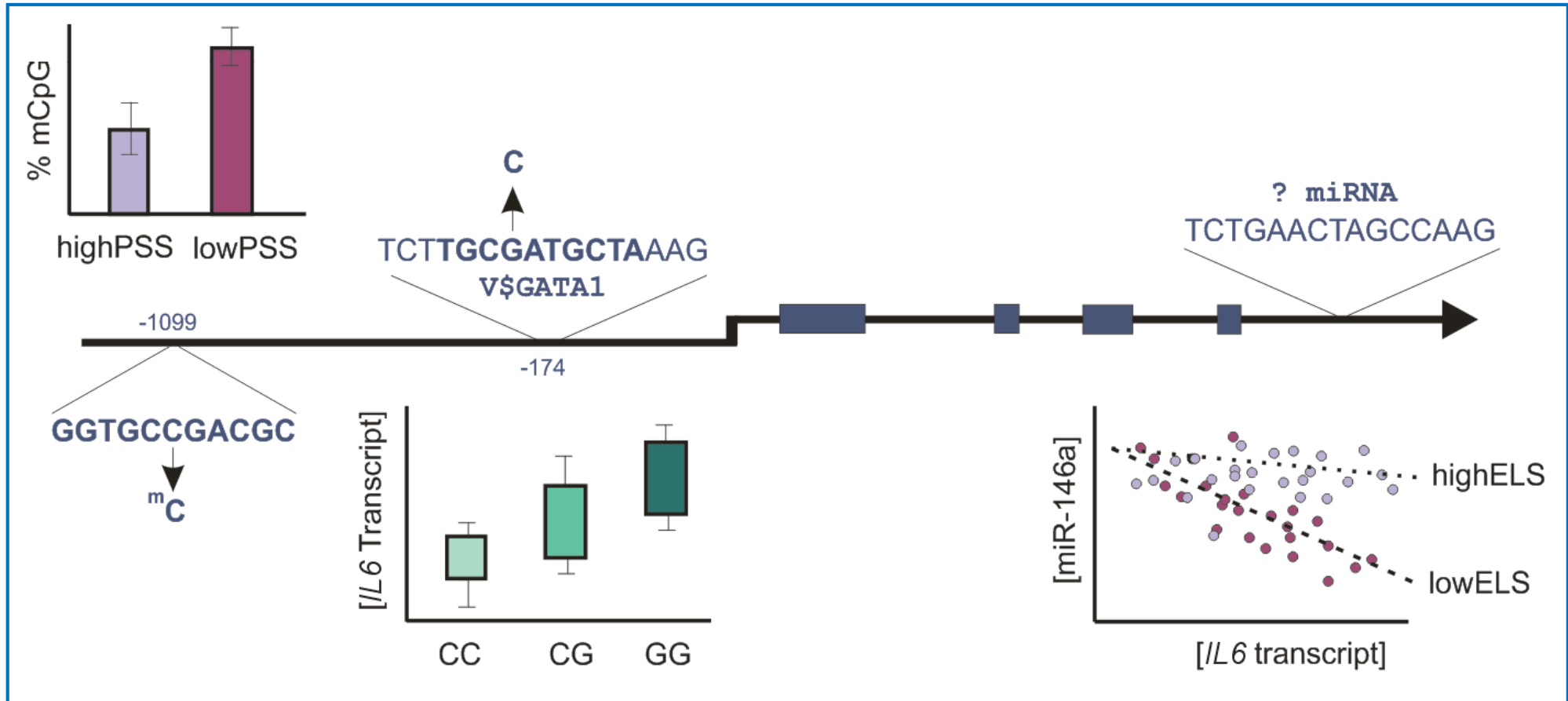
isoform A

isoform B

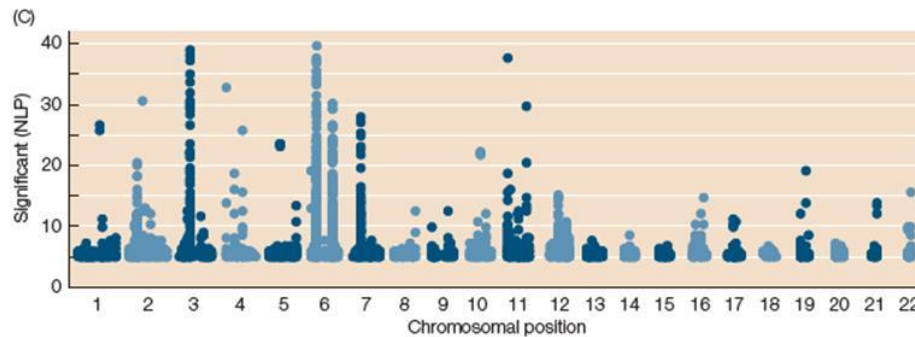
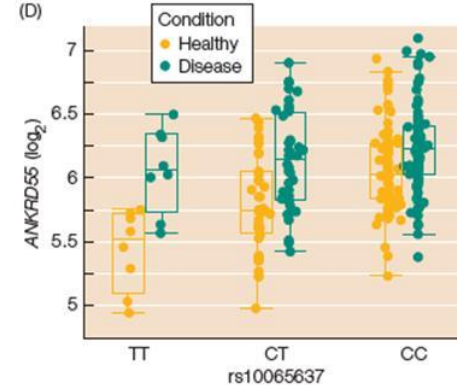
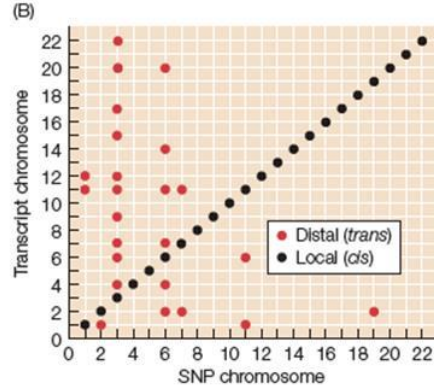
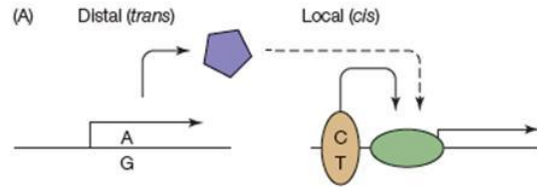
Cohort Studies



Genetics of Gene Regulation

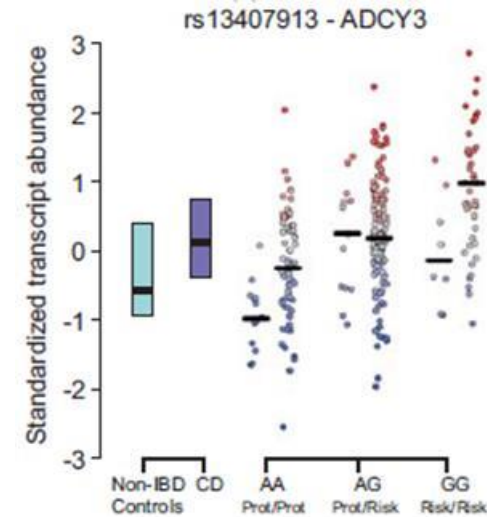


eQTL: Variants regulating gene expression

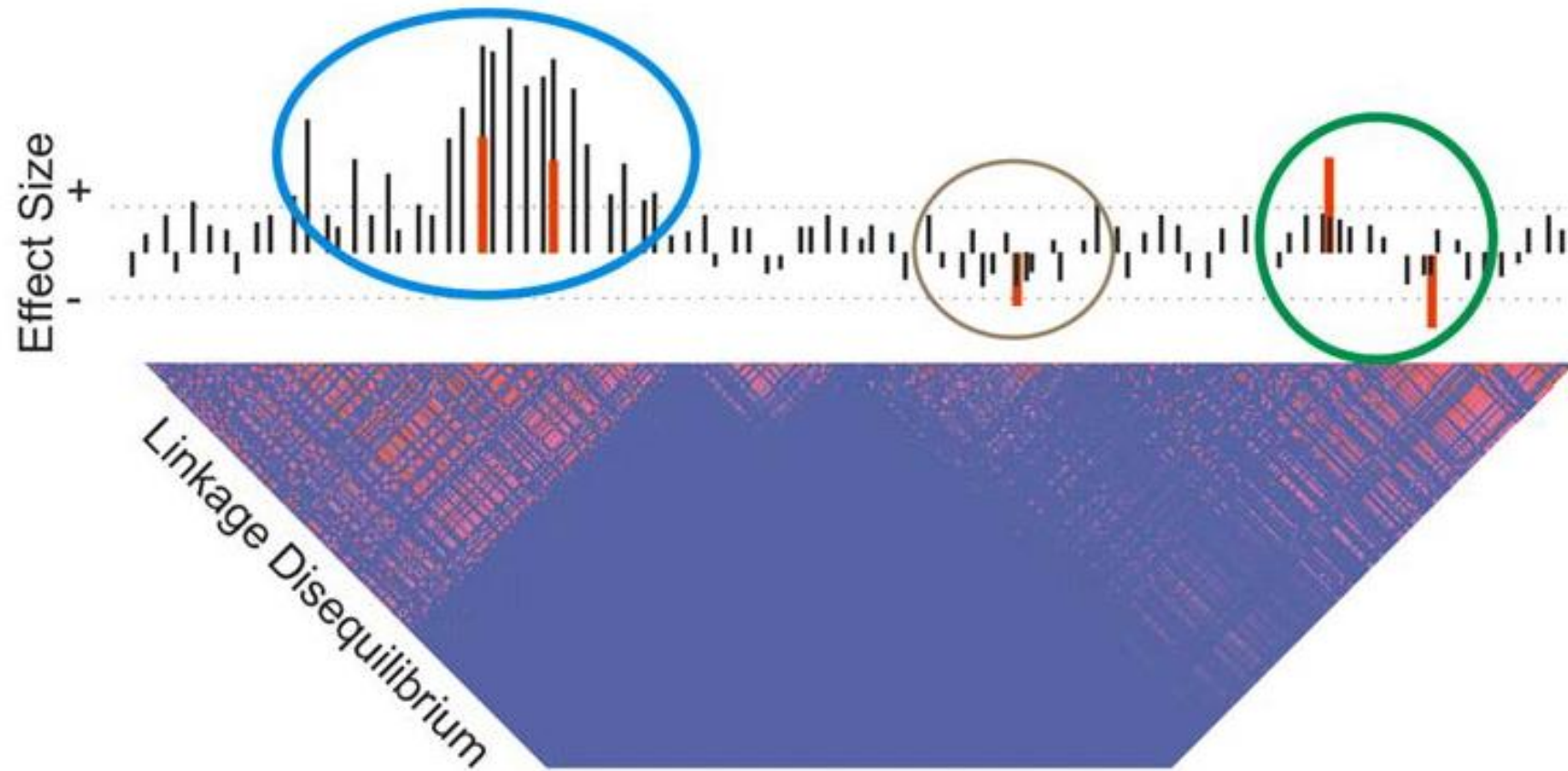


Expression QTL are polymorphisms that affect the expression of a gene.

They are particularly interesting when they overlap with GWAS hits.

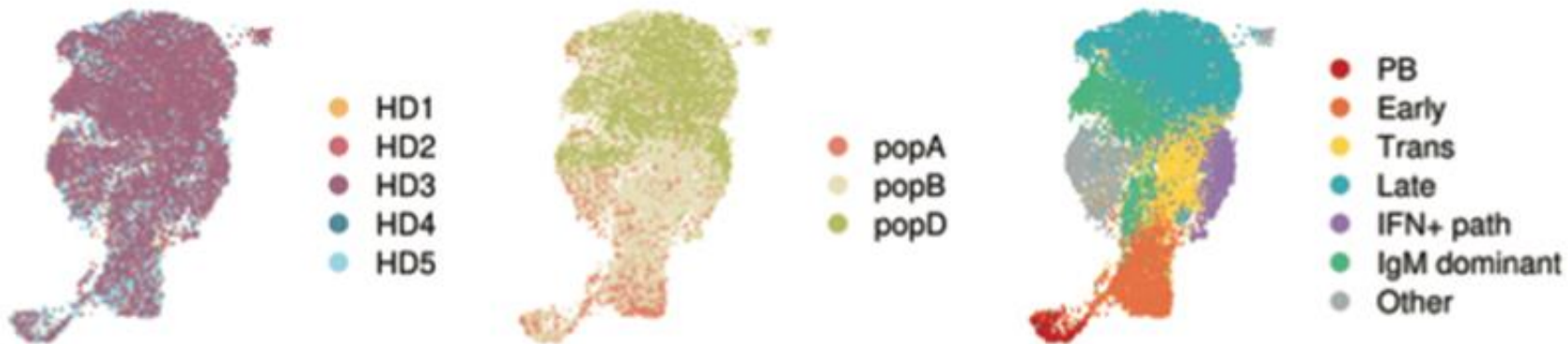
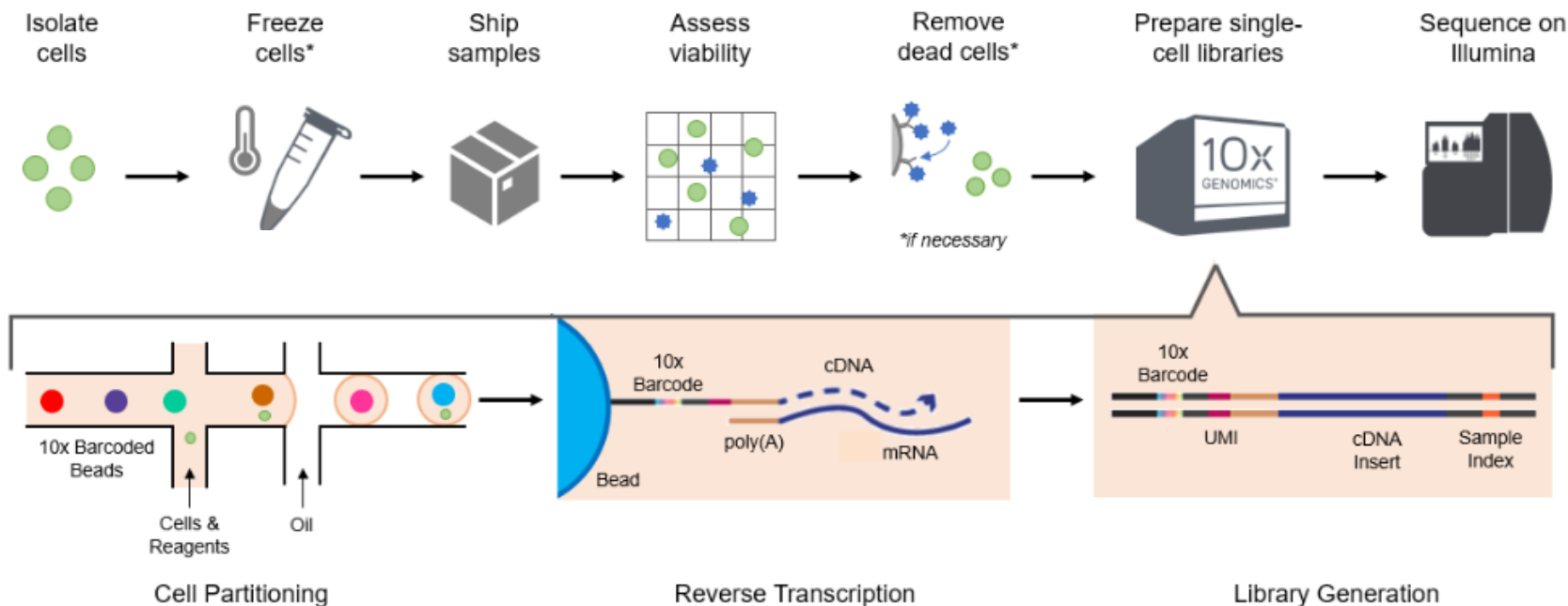


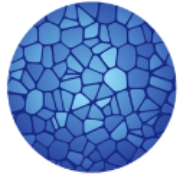
Causal Regulatory Variant Fine-Mapping is Difficult



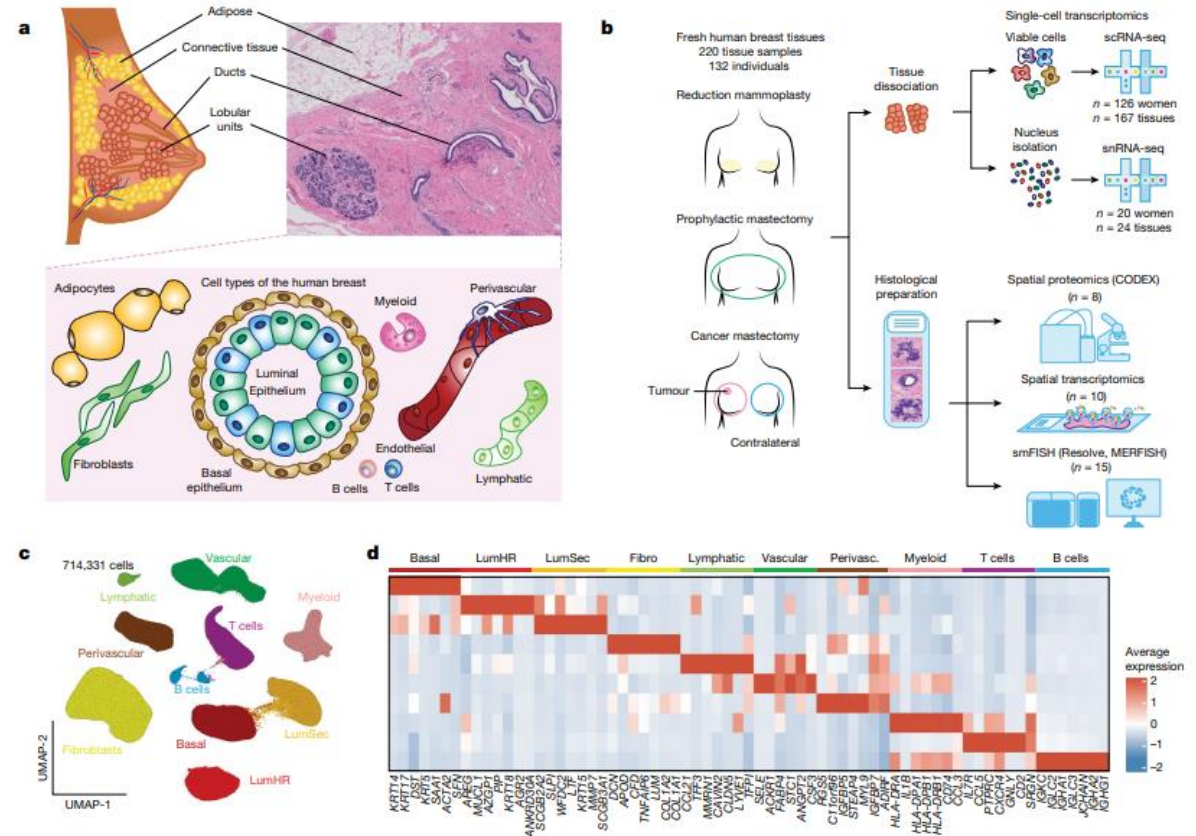
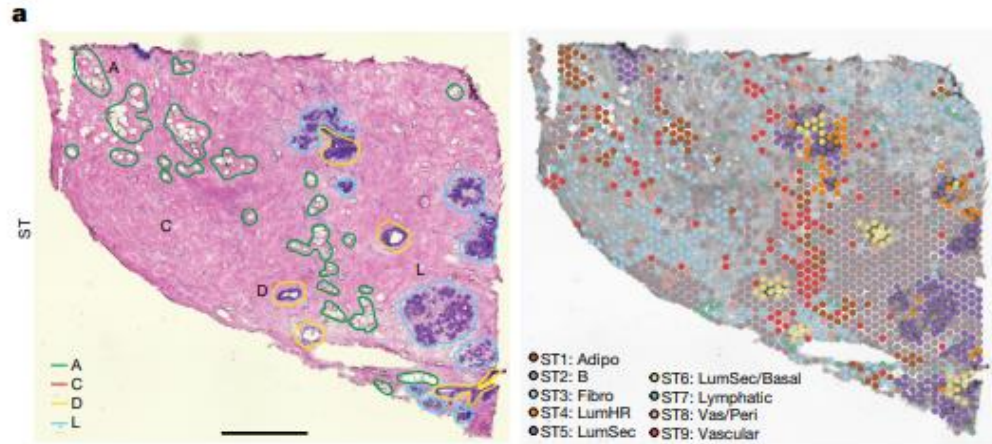
Droplet Single Cell Genomics: scRNAseq

GeneWiz website





The Human Cell Atlas (HCA) is an international group of researchers using a combination of these new technologies to create cellular reference maps with the position, function and characteristics of every cell type in the human body. We believe the work of the HCA is immensely beneficial to all of humanity, further information can be read [here](#).



Epigenomics

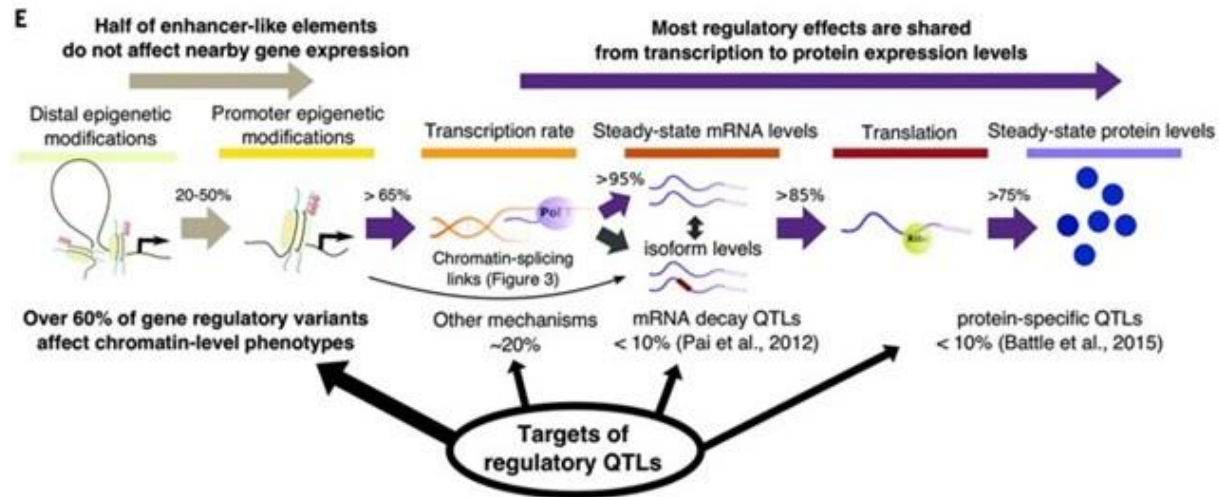
ENCODE Data Encyclopedia Materials & Methods Help

ENCODE: Encyclopedia of DNA Elements

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

[Get Started](#)

(Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI))



Integrative genomics / Systems Biology

