

# Working With Sequence Data—Part 2

## Visualization

Christopher G. Wilson, Ph.D.

Loma Linda University  
Dept. of Pediatrics and Center for Perinatal Biology

April 29, 2015

# Outline

1. Review
2. Overview of visualization methods in genomics
3. Tracks
4. Heat Maps
5. Circle Maps
6. DNA Walks
7. Network Maps
8. For today...

## Review from last week

- ▶ DNA Background
- ▶ Sequence tech changes
- ▶ Data handling
- ▶ Flu variant analysis project

## Brief article to read

<http://petridishtalk.com/2012/06/30/gene-visualization/>

## Why is visualization of genomic data important?

- ▶ The genome represents a *LOT* of information so we are dealing with a high information density problem.
- ▶ Being able to understand repetitive patterns is difficult when we just read the base pair sequence.
- ▶ It is easy to miss subtle changes in the sequence without some amplifying visualization tool.
- ▶ Our brains are better able to distinguish changes in a picture than in a text string.

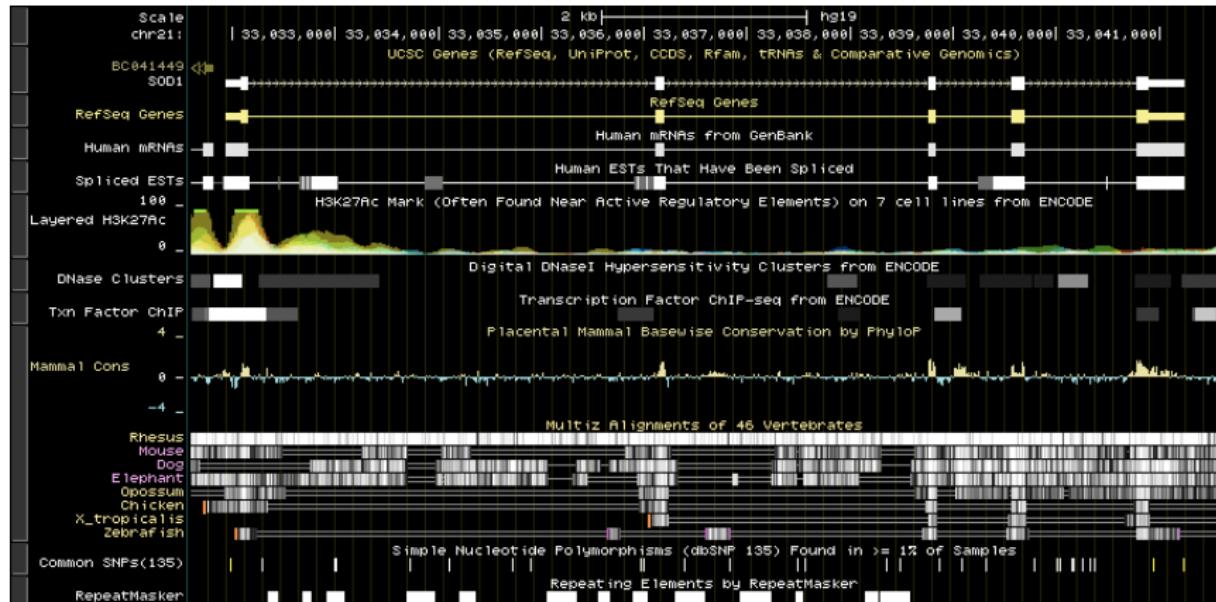
## Five typical methods for genomic sequence visualization

1. **Tracks**—Sequence data is represented in rows of base-pair text (emphUCSC Genome Browser).
2. **Heat Maps**—Rectangles with colored blocks in 2D rows/columns (most viz packages!).
3. **Circular Genome Maps**—Basically tracks arranged in a circle with some other layers of viz/annotation (check out *Genome Projector*).
4. **DNA Walks**—Representation of genomic bases as vectors in a 2D plane where A,T,G,C denote up,down,left,right (or variations thereof). Very similar to gravity analysis (Gerstein).
5. **Network Maps**—Ball and stick representations of genomic data (*Cytoscape*, *Ayasdi*)

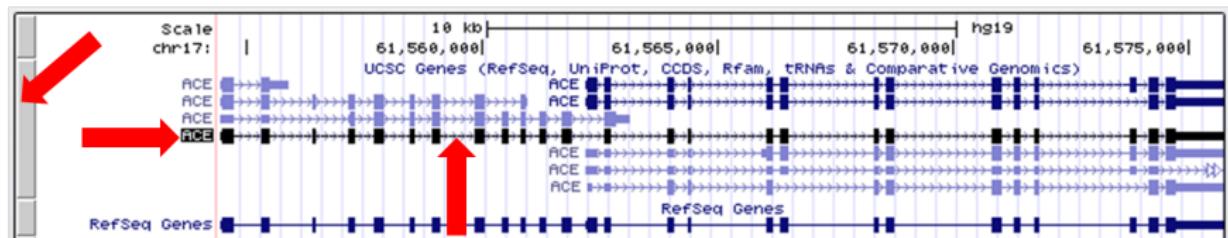
## Tracks

- ▶ Great for very localized changes in base-pairs or for restricted site mutations.
- ▶ Makes it easy to compare across different genomes for base-pair alteration
- ▶ Not great for looking at expansive swaths of genome

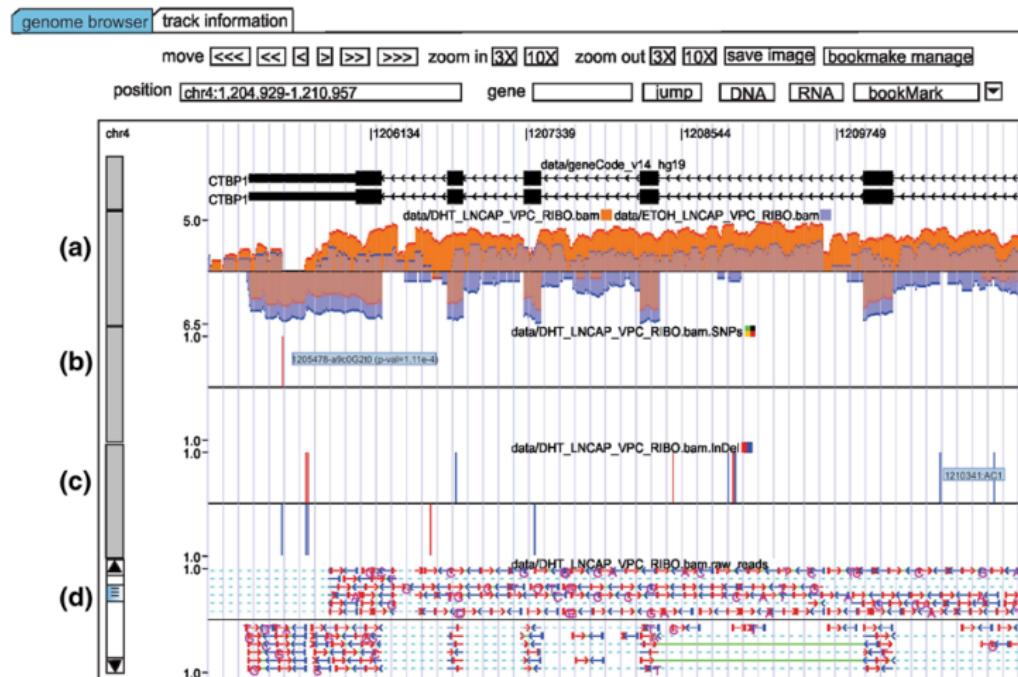
# Track examples



## Track examples (cont.)



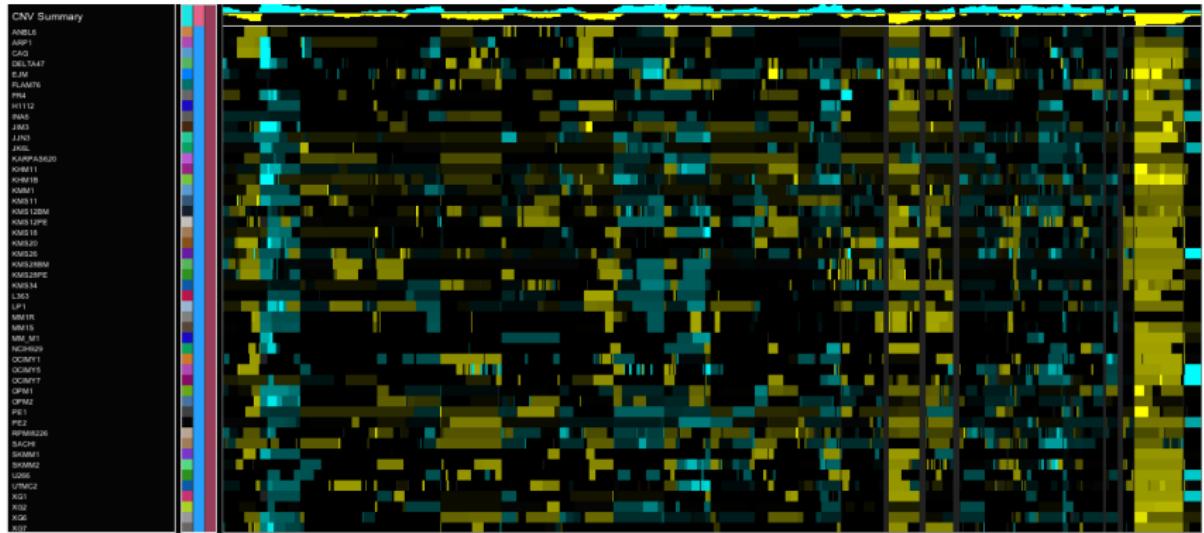
# Track examples (cont.)



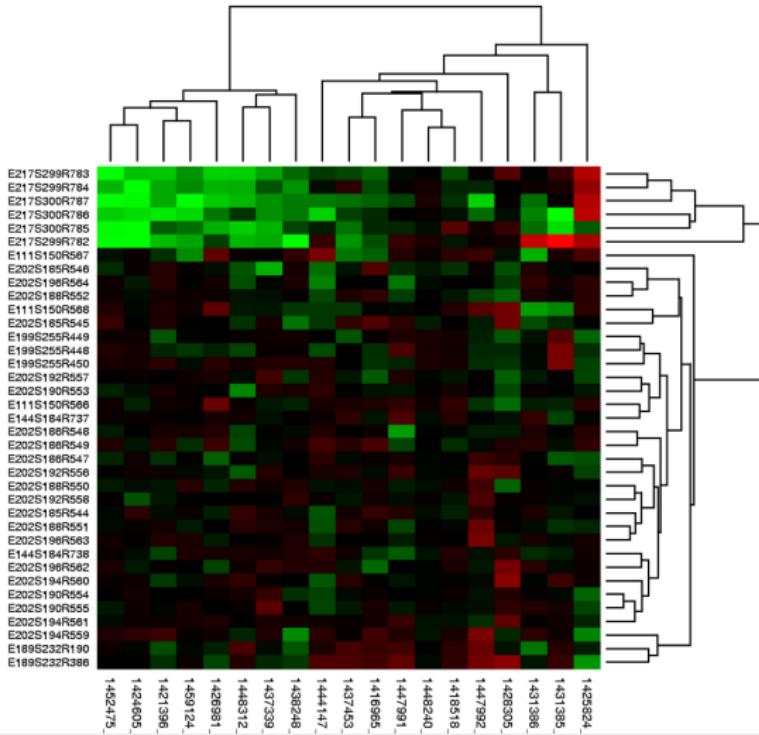
## Heat Maps

- ▶ What most people think of when they think about genome data!
- ▶ Typically a row/column representation of the data with a color scale representing frequency/expression.
- ▶ Available in most genomics tool-suites.

## Heat Map examples



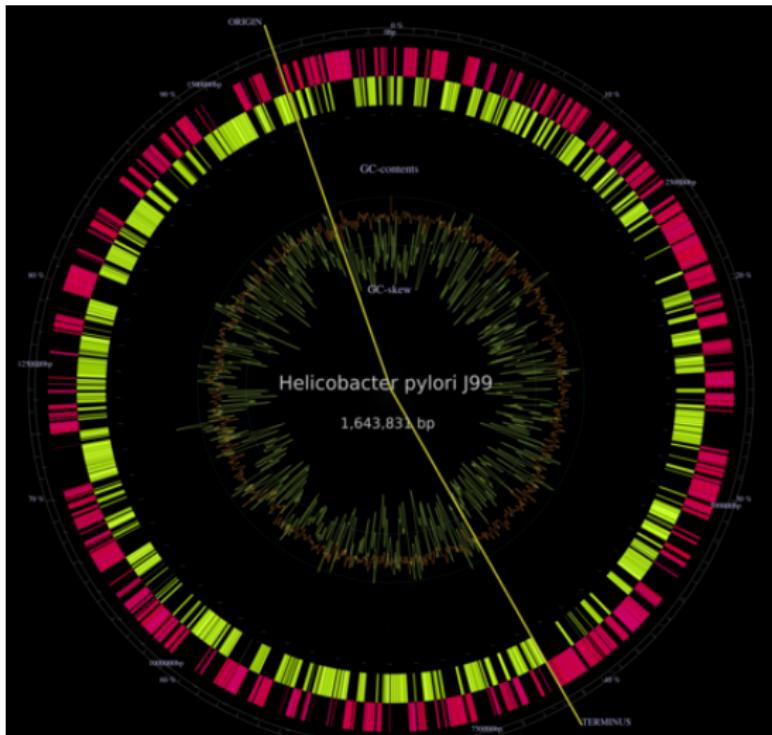
## Heat Map examples (cont.)



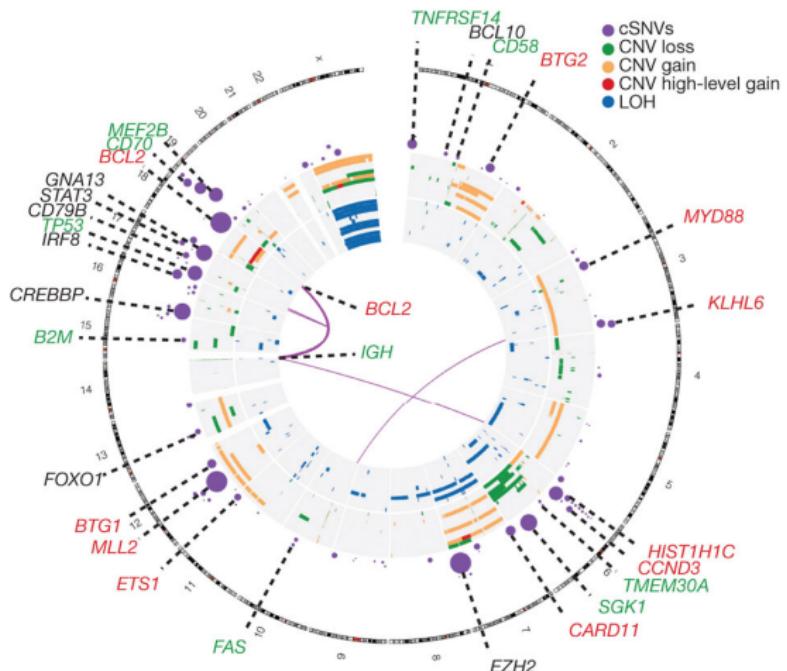
## Circular Genome Maps

- ▶ A method that takes the track maps idea and makes it more visually “pleasing”.
- ▶ The track data is arranged in concentric circles to make it easier to see sequence differences.
- ▶ This allows for rapid identification of potential correlations.

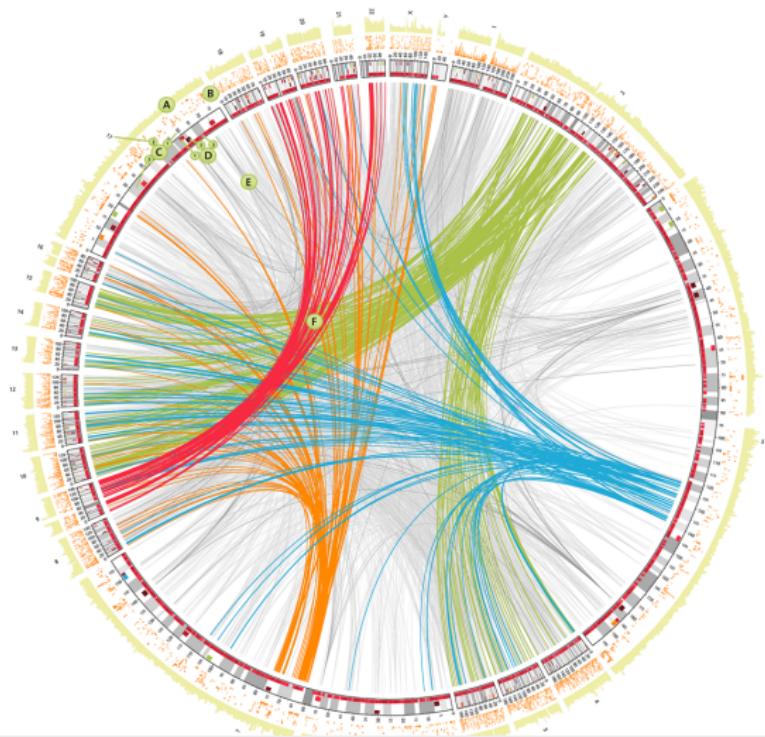
# Circle Map examples



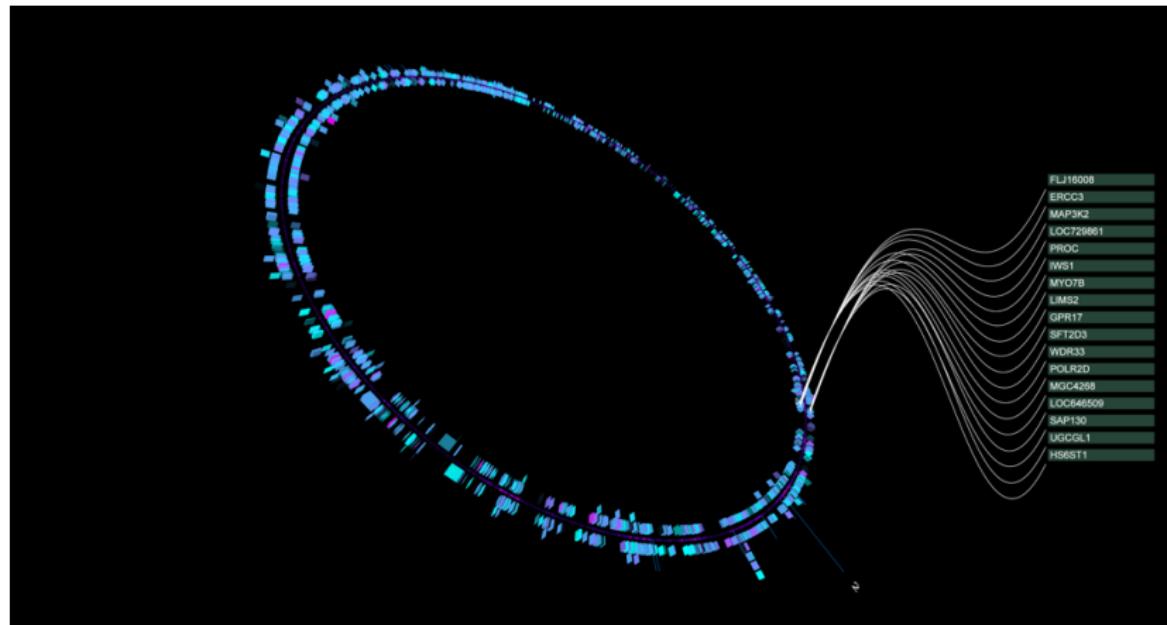
## Circle Map examples (cont.)



## Circle Map examples (cont.)



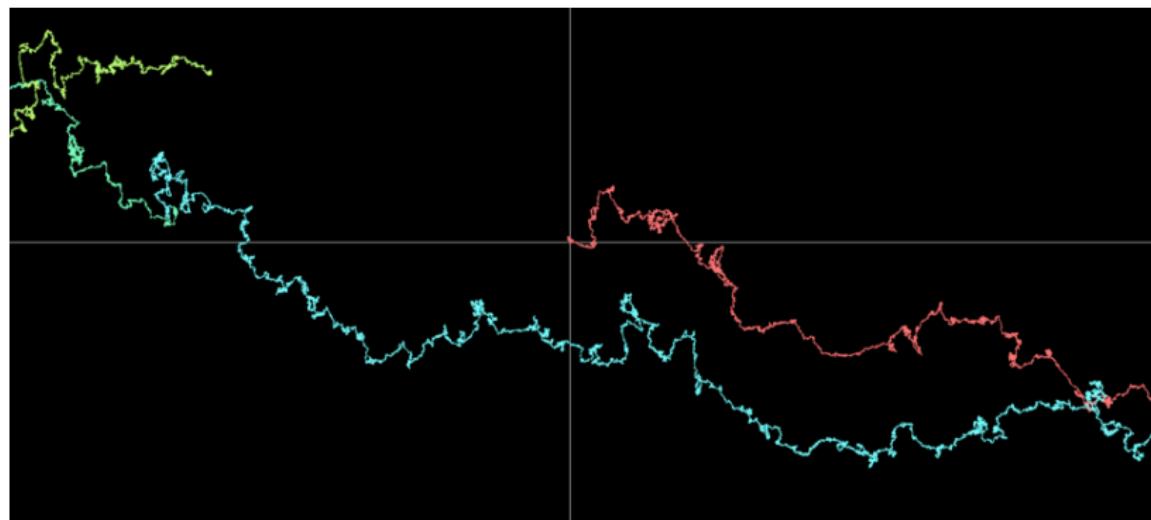
## Circle Map examples (cont.)



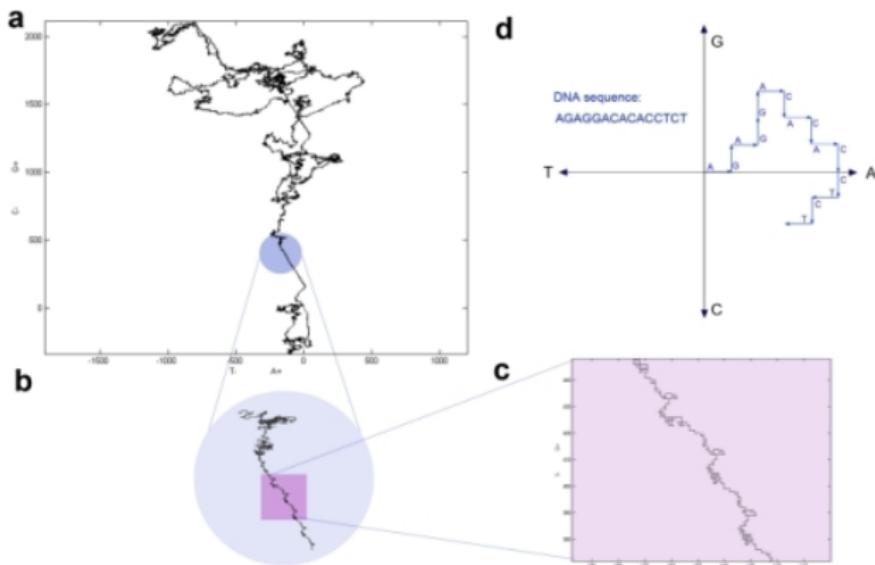
## DNA Walks

- ▶ Relatively new representation of genomic data.
- ▶ Treats each base letter as a “vector”.
- ▶ A = up, T = down, G = left, C = right (or however the encoding is defined!).
- ▶ Very useful for providing visualization of structural changes in the genome or repetitions.

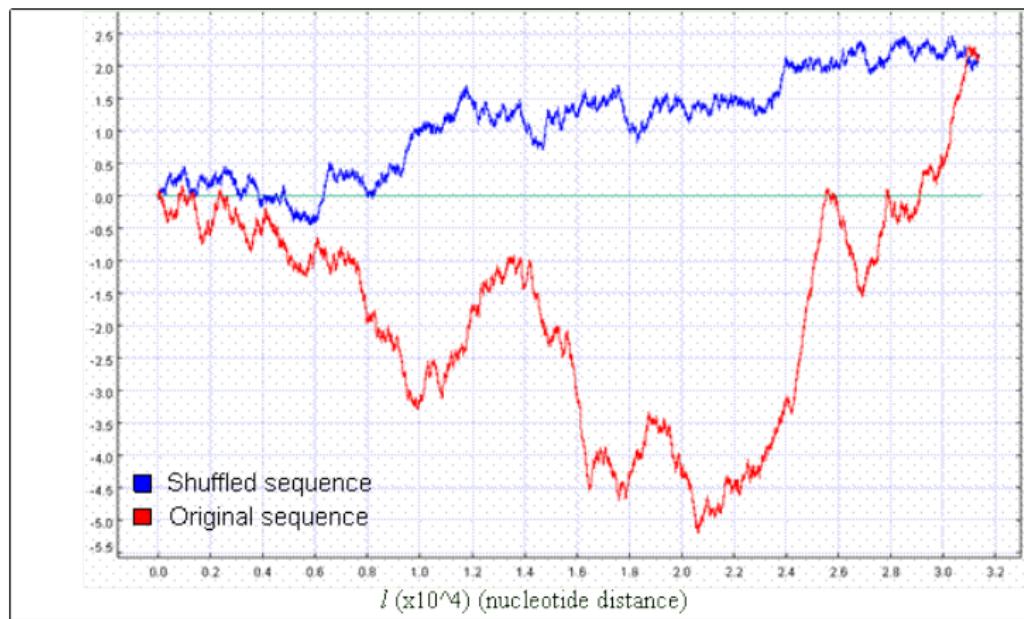
## DNA Walk examples



## DNA Walk examples (cont.)



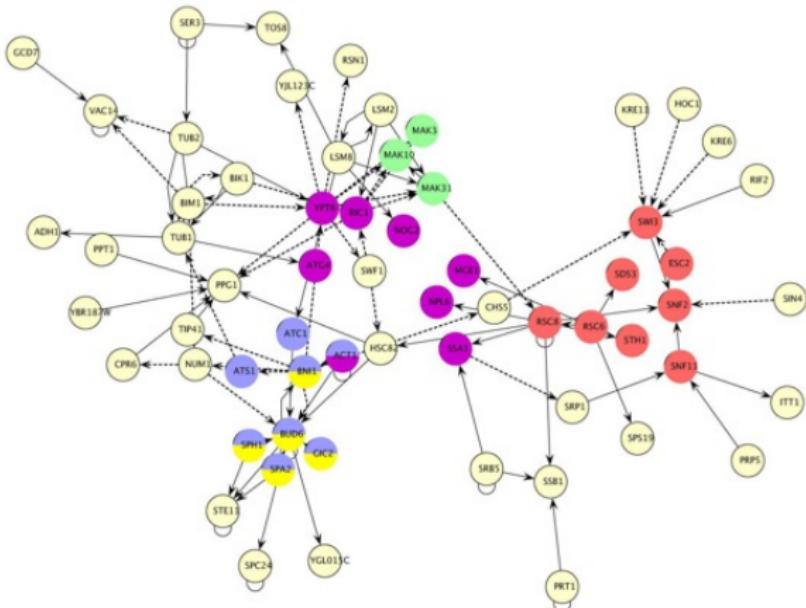
## DNA Walk examples (cont.)



## Network Maps

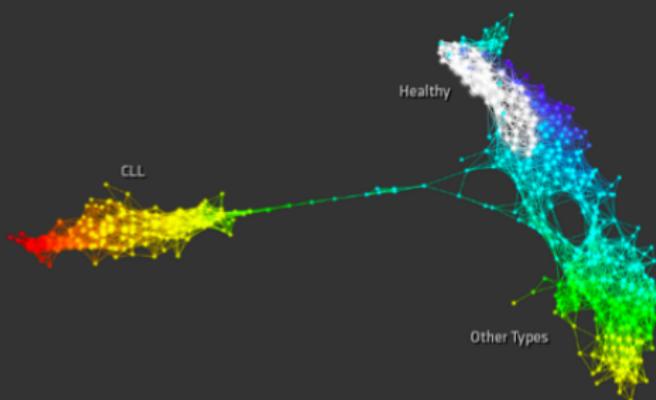
- ▶ Originally used for computer networks, then neural networks, etc.
- ▶ Generalizable to pathways of many kinds (chemical, signaling, etc.)
- ▶ Increasingly used to map gene interactions.
- ▶ Allows “easier” visualization of divisions in the genome.
- ▶ Requires more “annotation” to understand what the interactions mean.

# Network Map examples



## Network Map examples (cont.)

### Portrait of leukemia patient sub-populations



### Gene Expression

#### What is this?

The topological map of leukemia patients as represented by the expression level of different genes in their bone marrow samples.

#### SOURCE

PubMed ID: 8573112

#### DATA SIZE

# Rows: 1905

# Columns: 1500

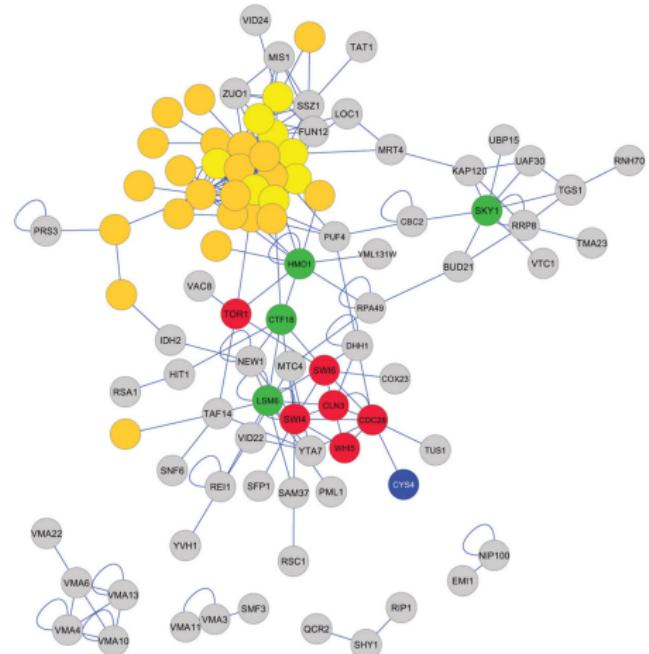
#### THE MAP

Node: Set of patients

Edge: Patients common to both nodes

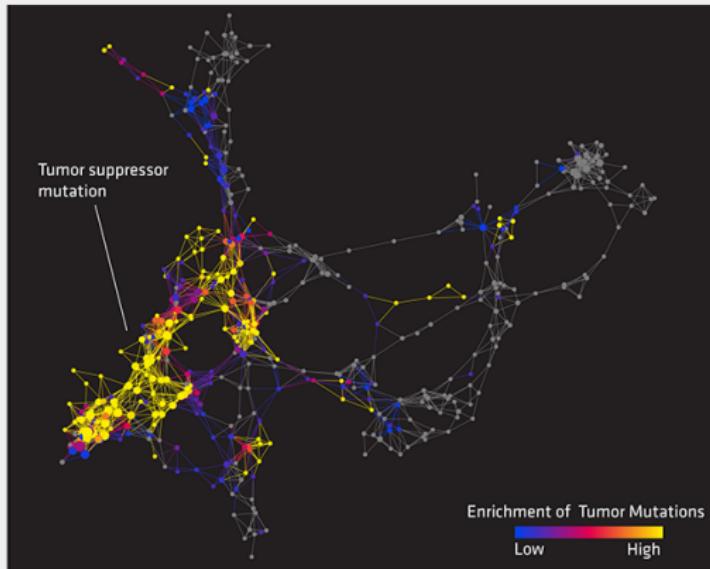
#### PARAMETERS

## Network Map examples (cont.)



# Network Map examples (cont.)

## Topological Network Gallery



### The Cancer Genome Atlas (TCGA)

#### DATA

Cancer tumors characterized by somatic mutations colored by clinical outcomes

#### INSIGHT

Identified cancer subtypes that respond differently to treatments

#### APPLICATION

Development of targeted cancer treatments

## For today

1. Work on your project.
2. Figure out a visualization scheme—see if you can find Python-based ways to create the graphics discussed today.
3. Papers in the class repository for background.
4. Write-up is due on May 7<sup>th</sup>