

# Data Visualization for 3D Genome Analysis

Nils Gehlenborg, PhD

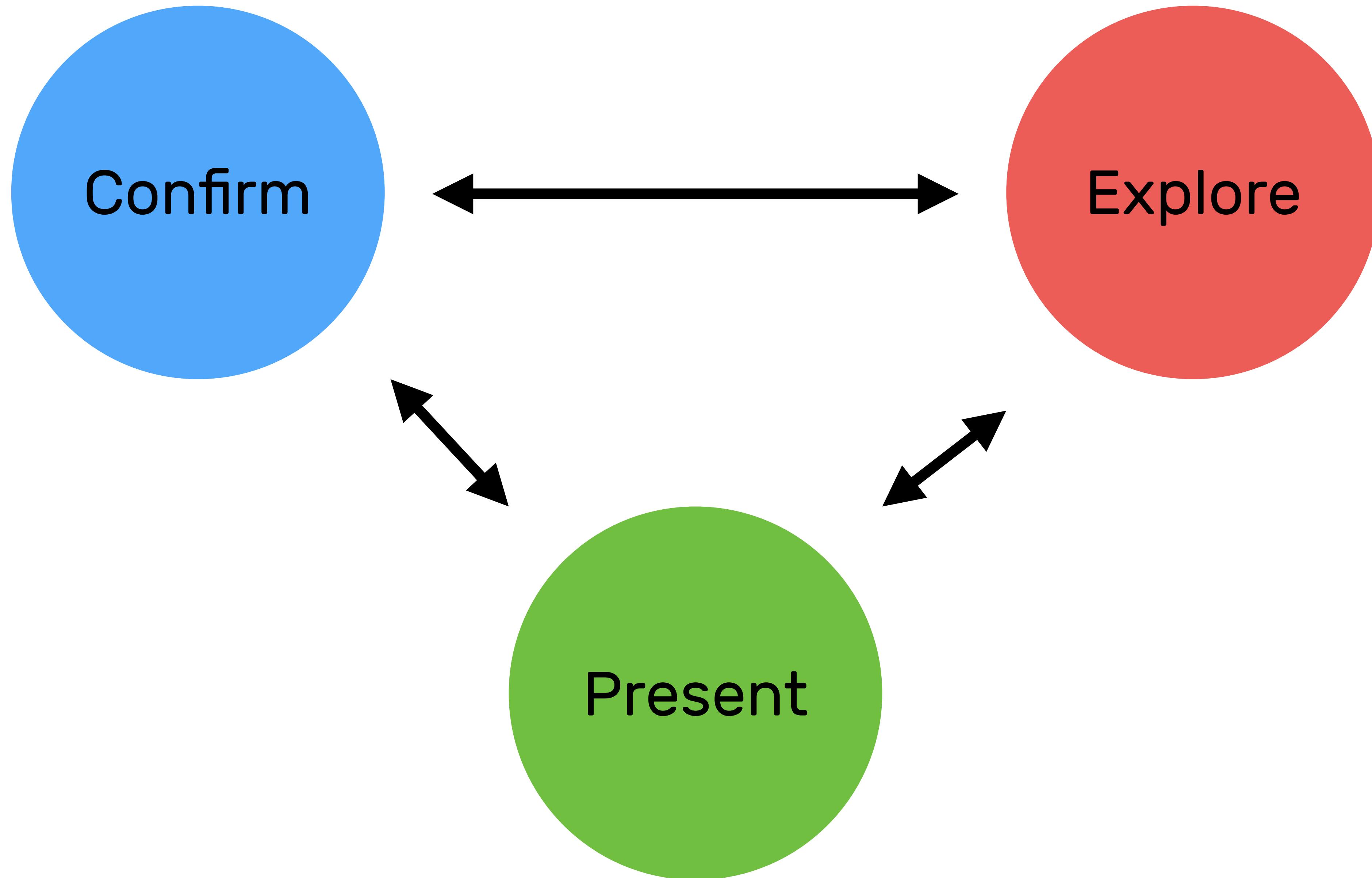
**HARVARD MEDICAL SCHOOL**  
DEPARTMENT OF BIOMEDICAL INFORMATICS

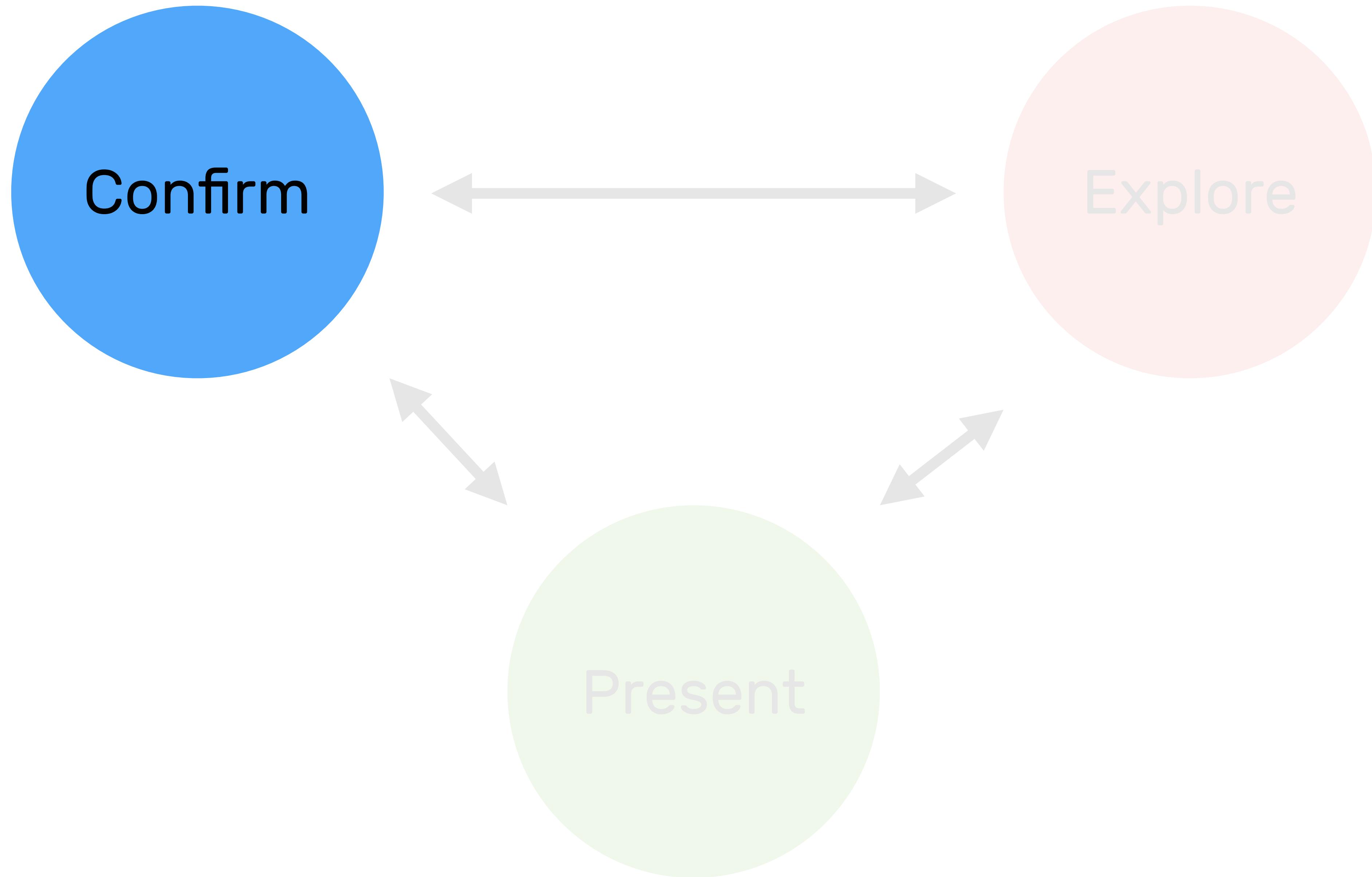
<http://gehlenborglab.org>



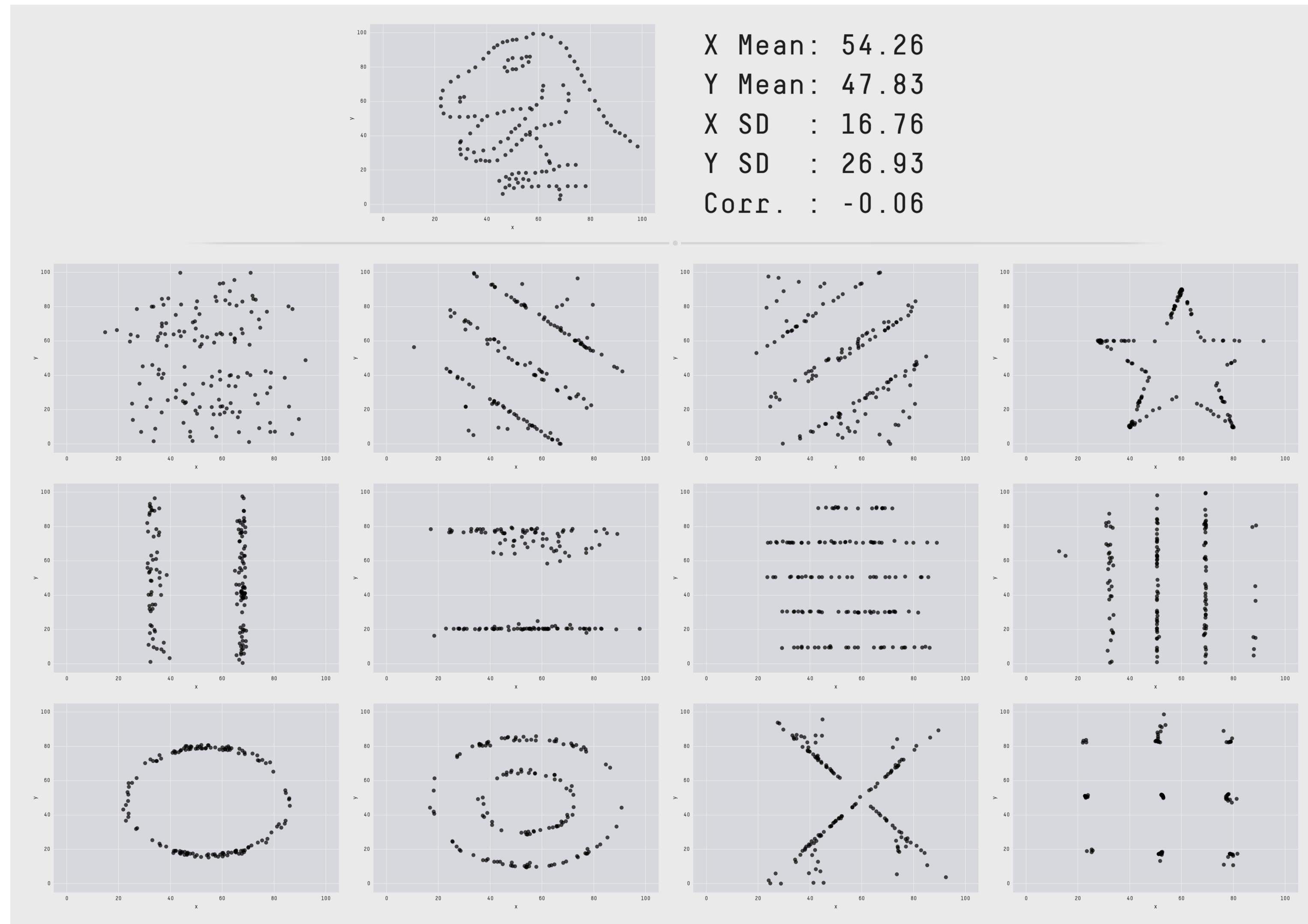
@nils\_gehlenborg

Why do we need to visualize data?





# Confirm



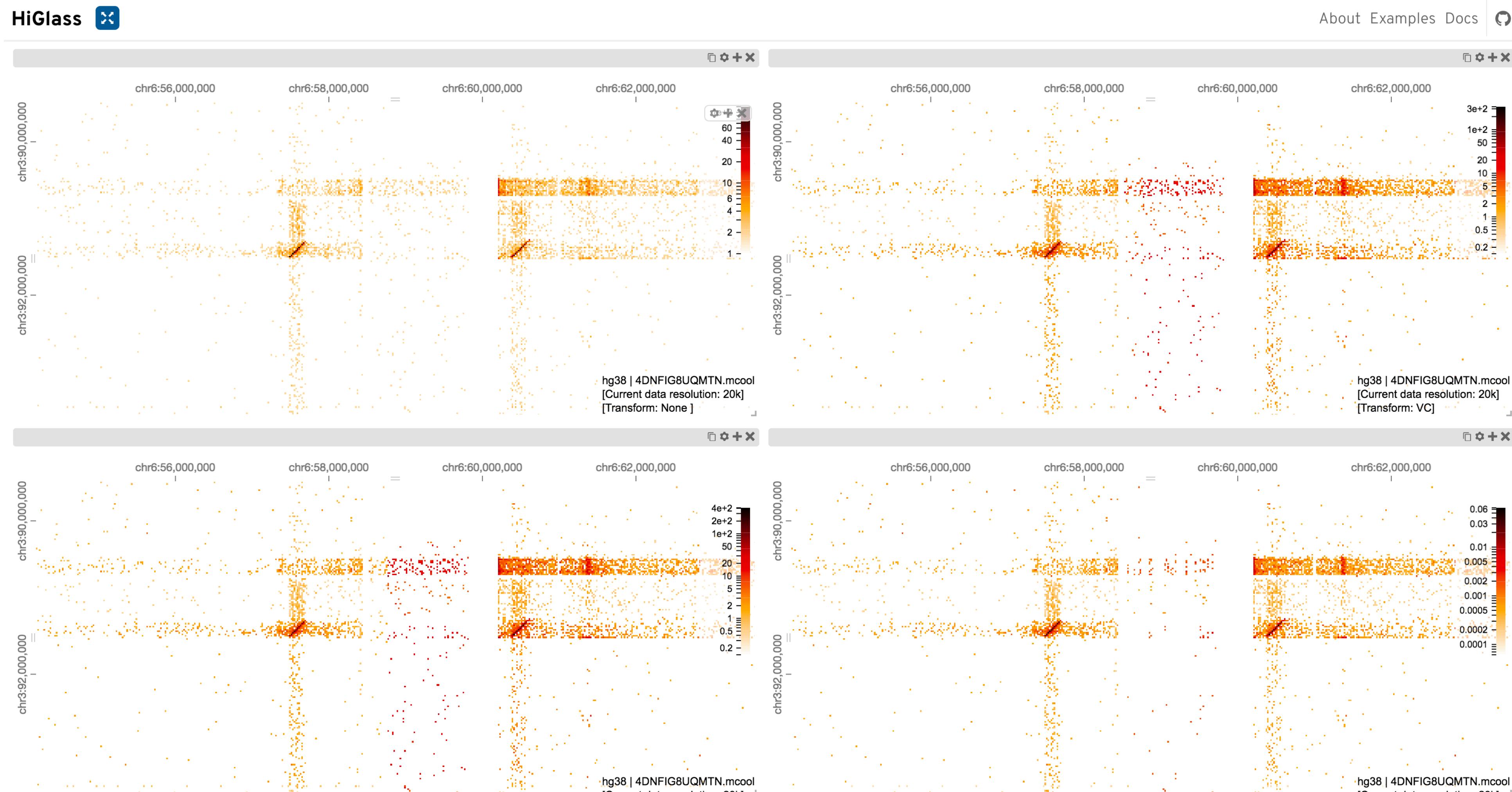
# Confirm

1. Does it look like a Hi-C matrix?
2. Did the filtering work?
3. Did we normalize properly?
4. ...

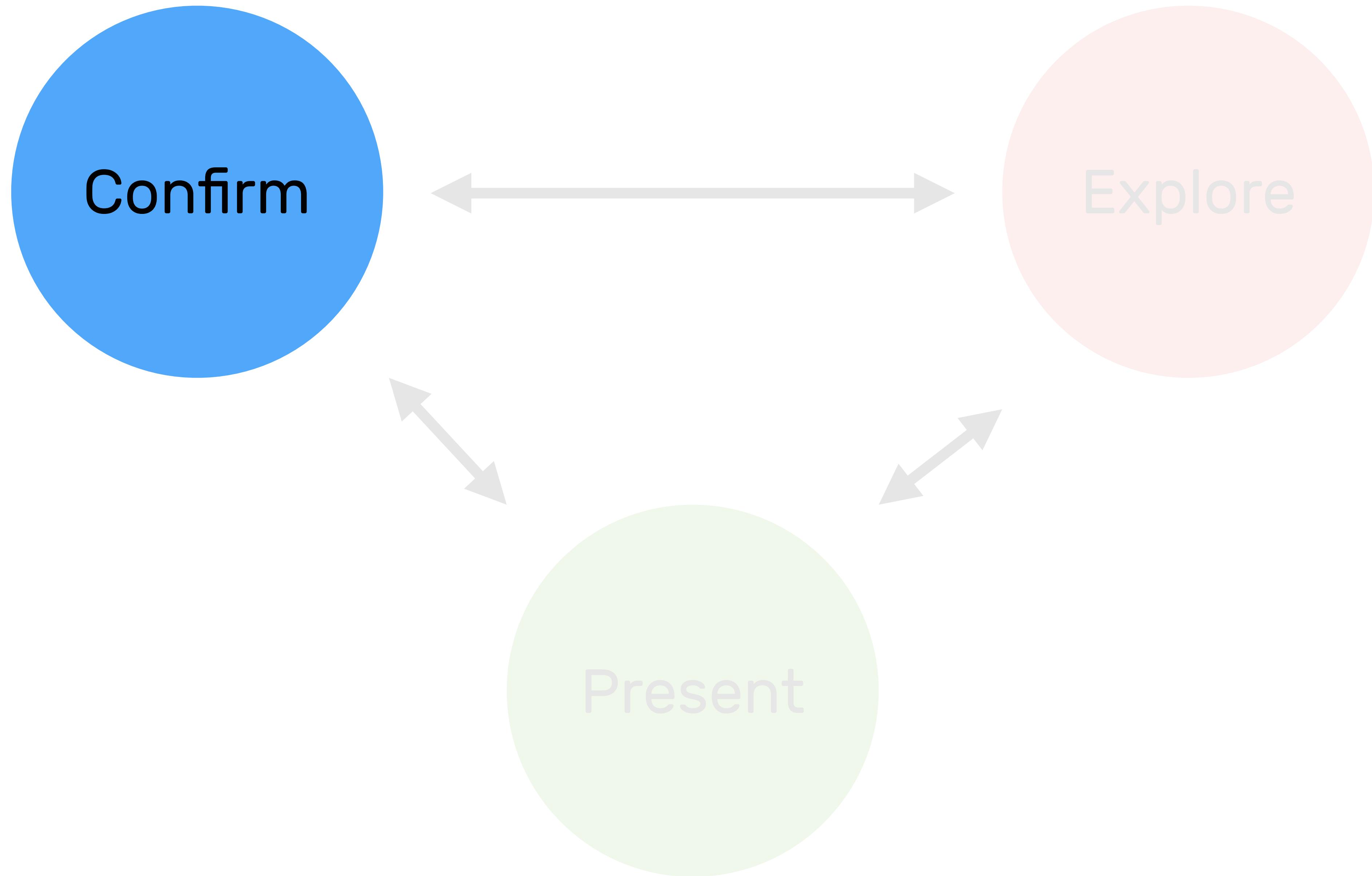
# Confirm

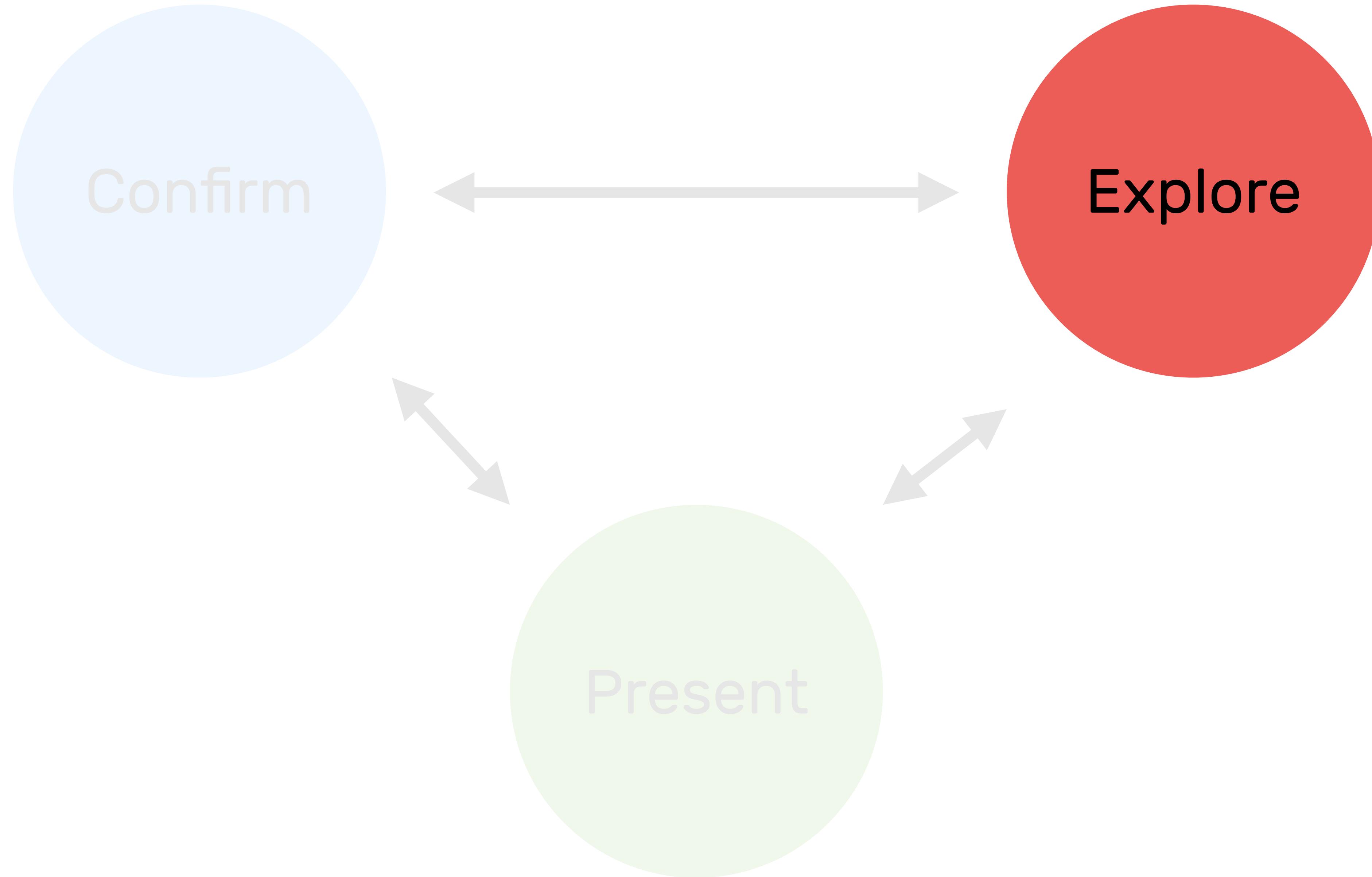
# Issue

## Multi-mapped reads not removed!



<http://higlass.io/app/?config=UDtCT0OjSqvNS-sSHvzgPw>





# Explore

# Explore

1. View interactions at different scales from genome to individual bins -> **Zoom, Pan**

# Explore

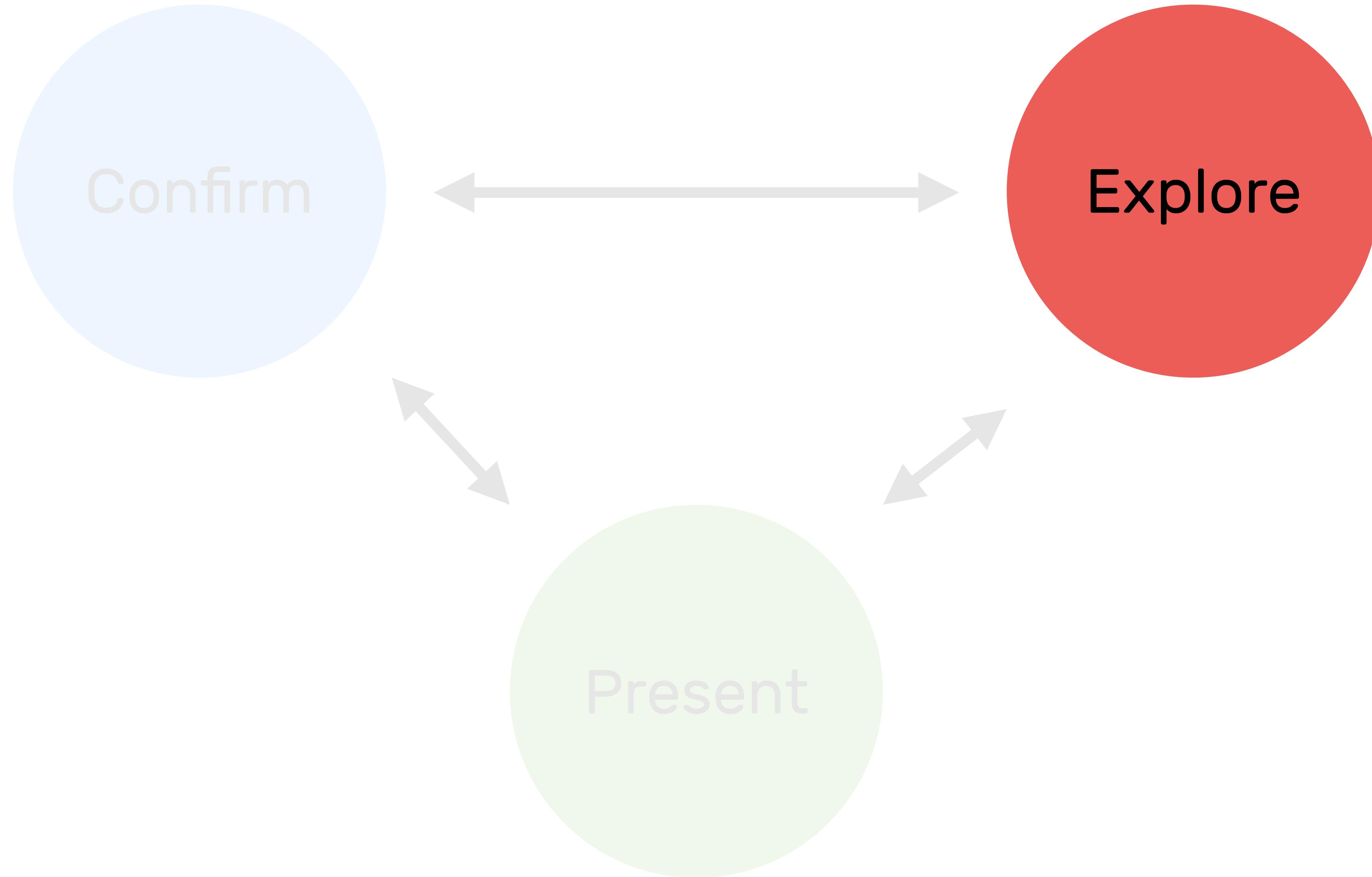
1. View interactions at different scales from genome to individual bins -> **Zoom, Pan**
2. Compare maps across two or more conditions  
-> **Juxtapose or Overlay Maps**

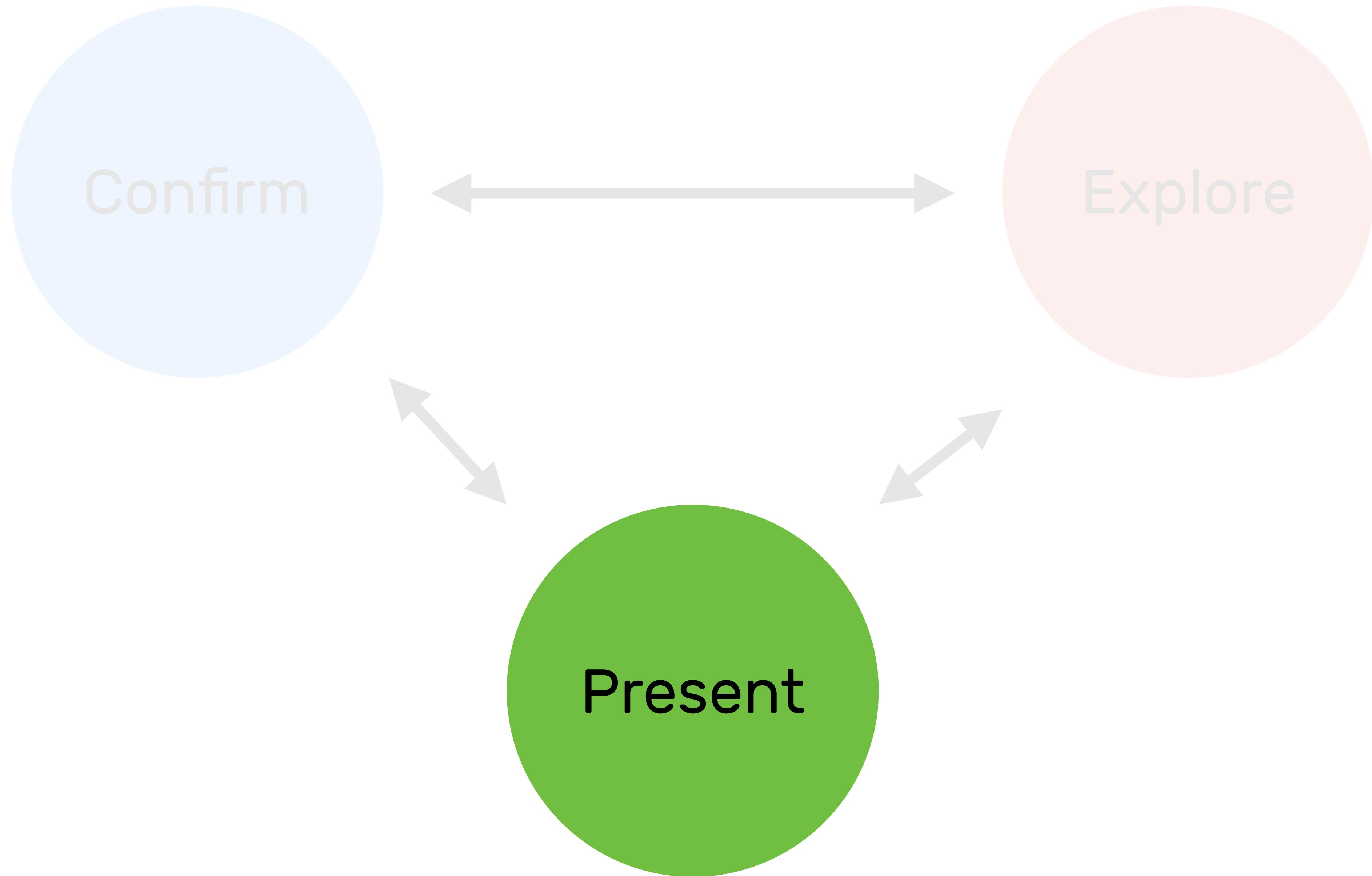
# Explore

1. View interactions at different scales from genome to individual bins -> **Zoom, Pan**
2. Compare maps across two or more conditions  
-> **Juxtapose or Overlay Maps**
3. View and compare features within and across maps  
-> **Link Maps, Extract Features**

# Explore

1. View interactions at different scales from genome to individual bins -> **Zoom, Pan**
2. Compare maps across two or more conditions  
-> **Juxtapose or Overlay Maps**
3. View and compare features within and across maps  
-> **Link Maps, Extract Features**
4. Correlate interaction patterns with other (epi)genomic data and genomic annotations -> **Add Genomic Tracks**





# Present

# Present

1. Highlight key observations & annotate features  
-> Label features

# Present

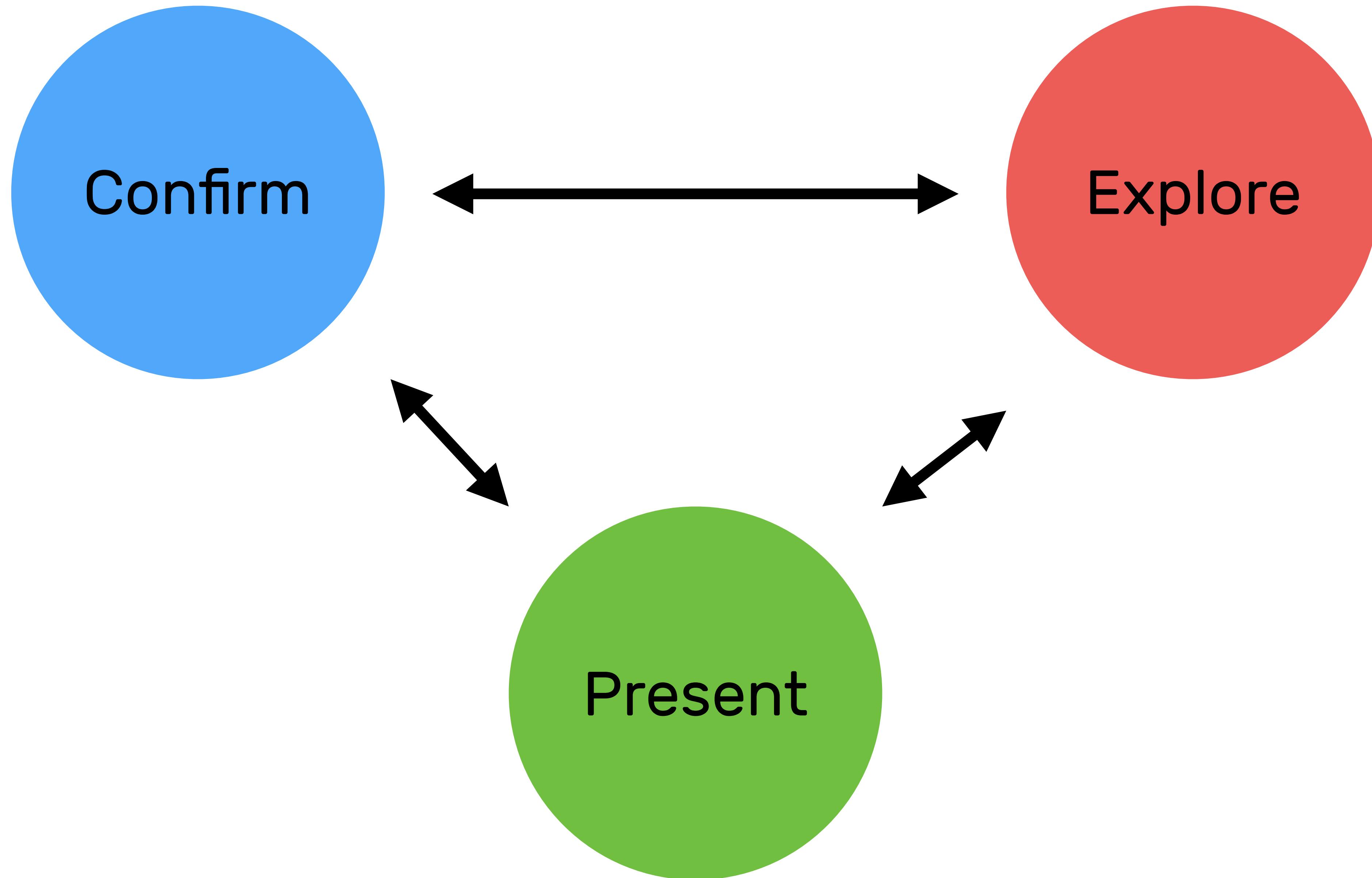
1. Highlight key observations & annotate features  
-> Label features
2. Communicate process of exploratory analysis  
-> Track visualization provenance\*

\* see <http://vistories.org>

# Present

1. Highlight key observations & annotate features  
-> Label features
2. Communicate process of exploratory analysis  
-> Track visualization provenance\*
3. Share findings with collaborators -> Save state and share with collaborators
4. ...

\* see <http://vistories.org>



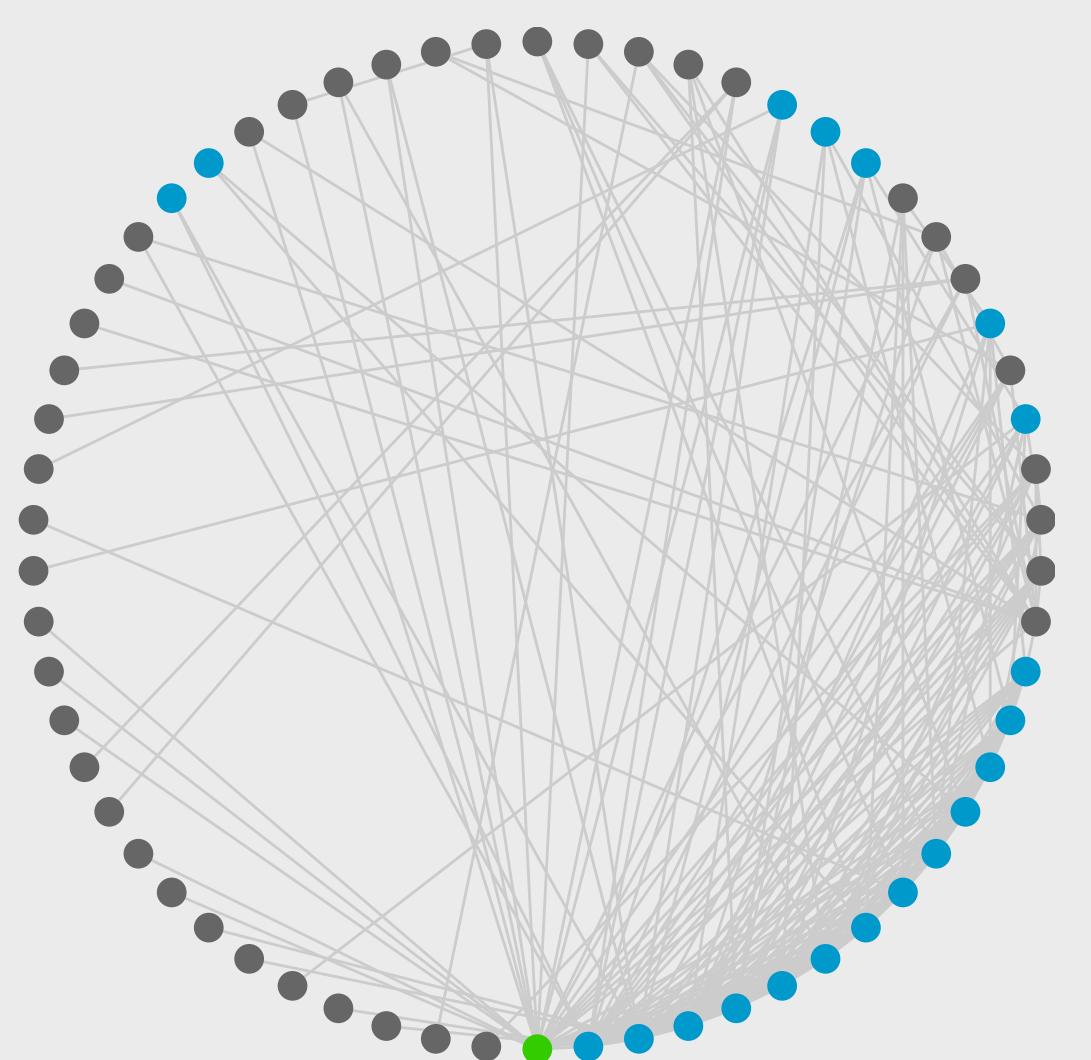
How can we visualize the data?

# Network Visualization

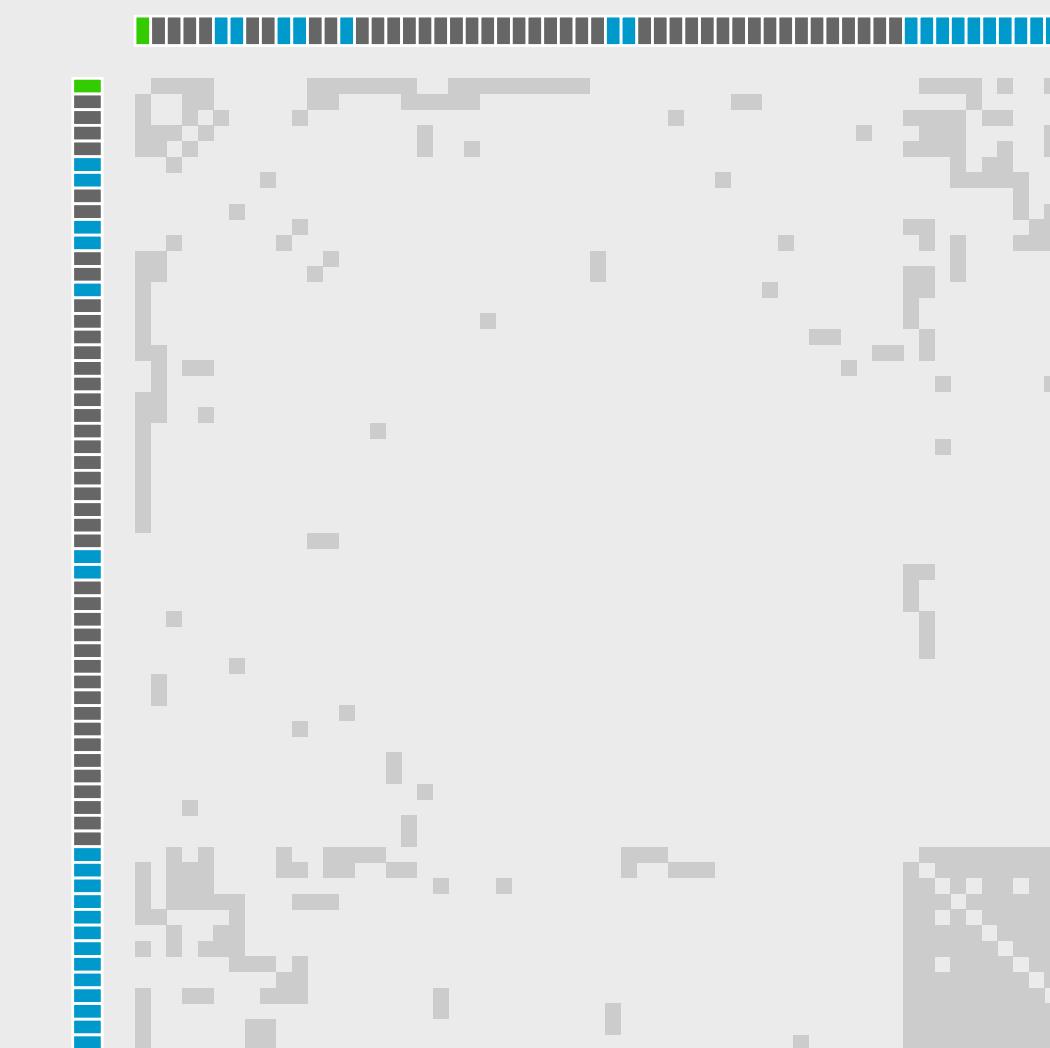
1.  $(3M \times 3M)/2$  interactions
2. Weight for each interaction
3. Constraint: nodes in sequence order

# Network Visualization

1.  $(3M \times 3M)/2$  interactions
2. Weight for each interaction
3. Constraint: nodes in sequence order



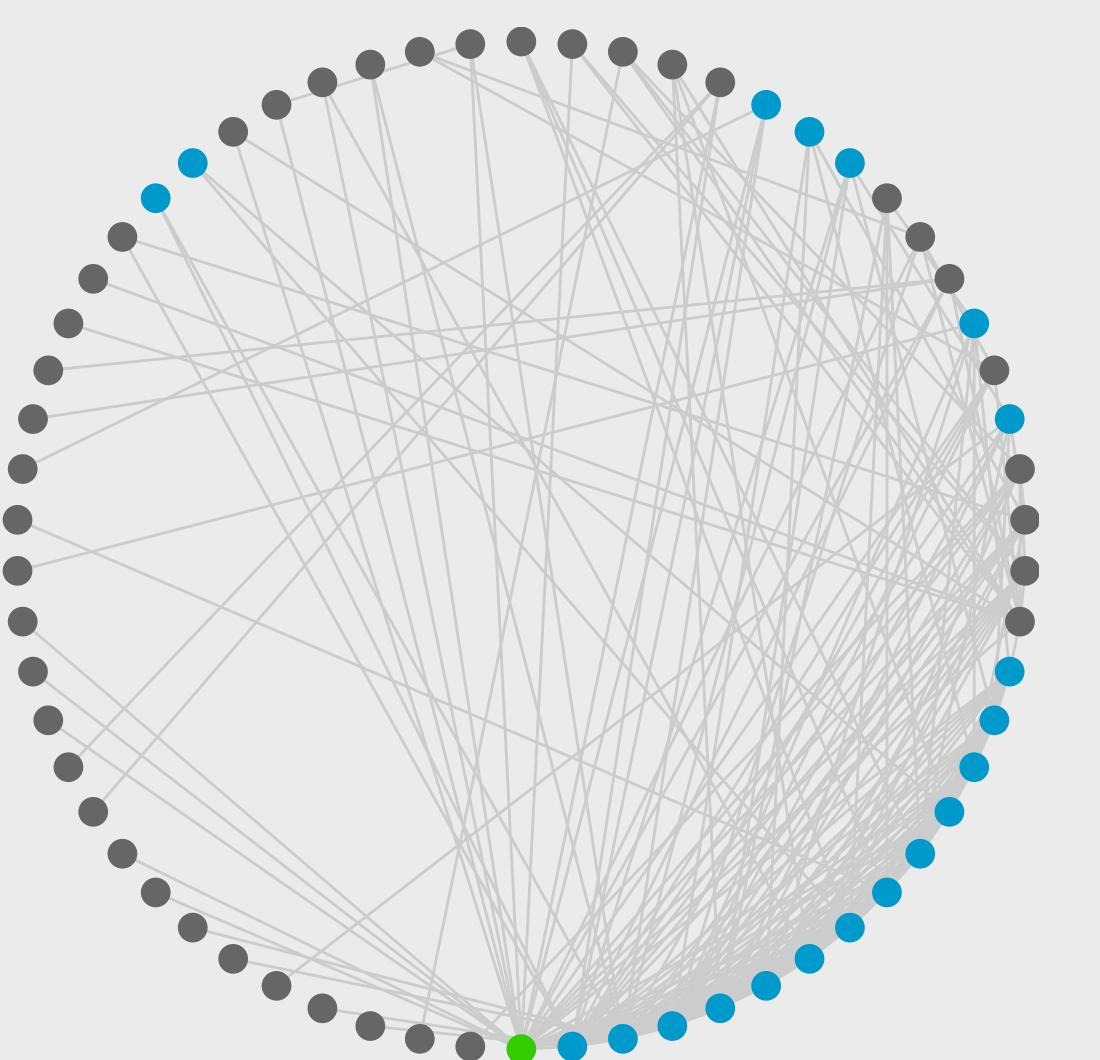
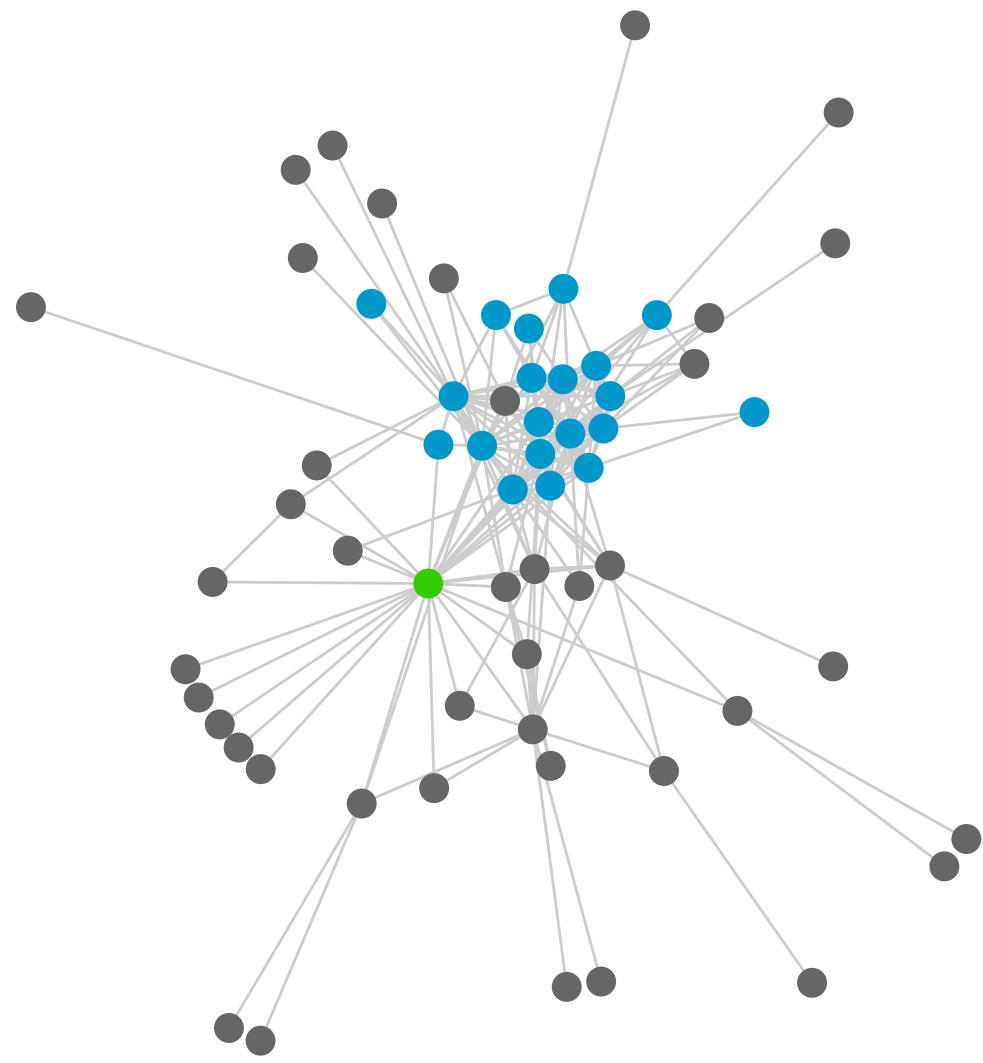
Node-Link  
Diagram



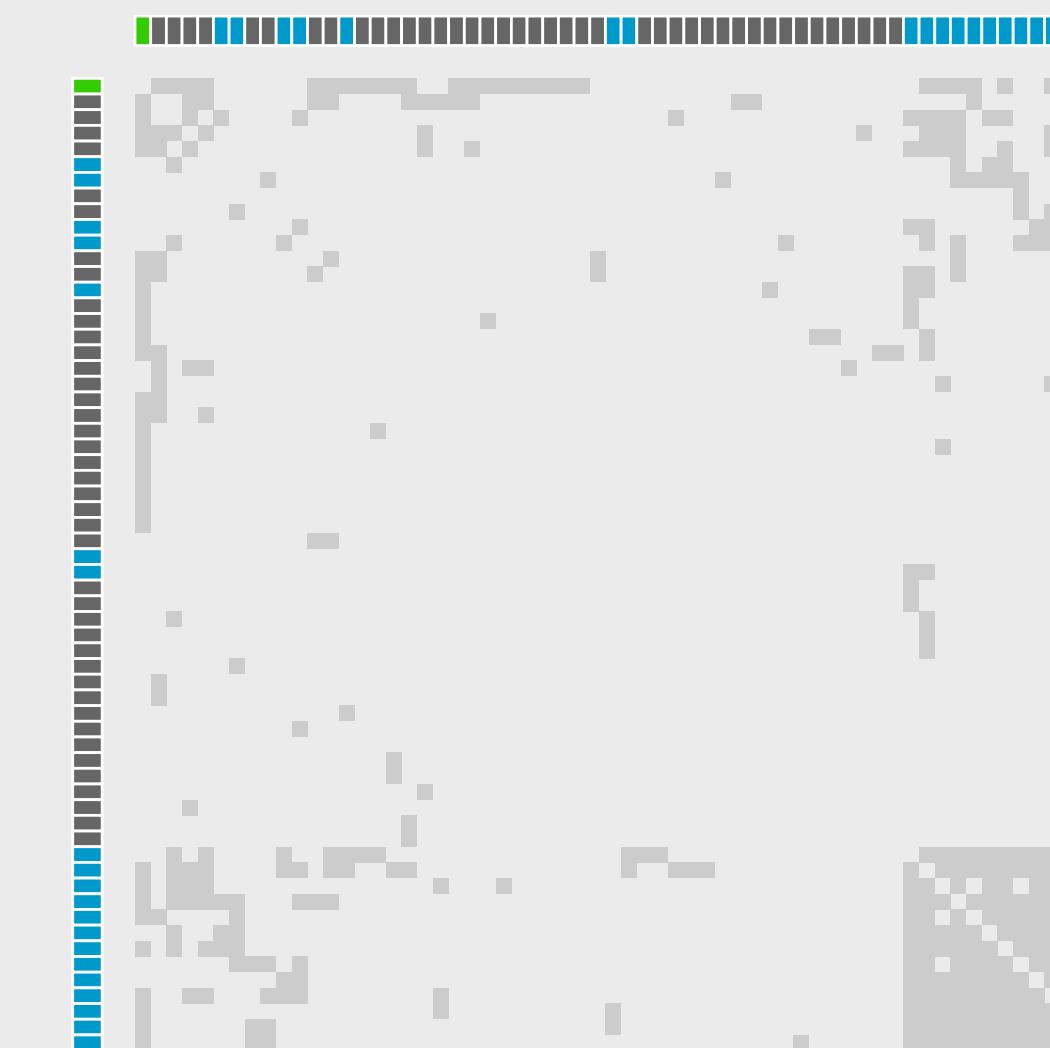
Matrix

# Network Visualization

1.  $(3M \times 3M)/2$  interactions
2. Weight for each interaction
3. Constraint: nodes in sequence order



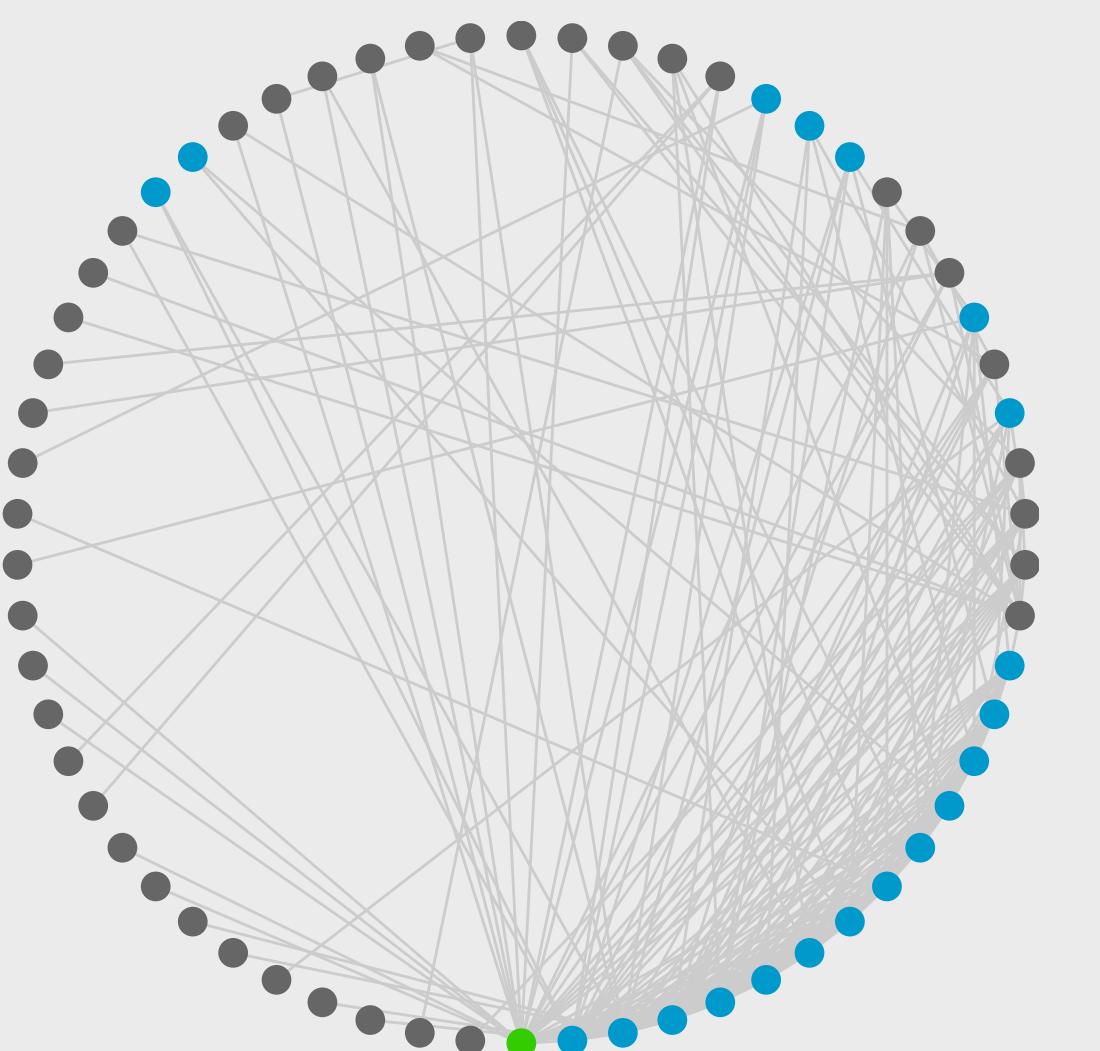
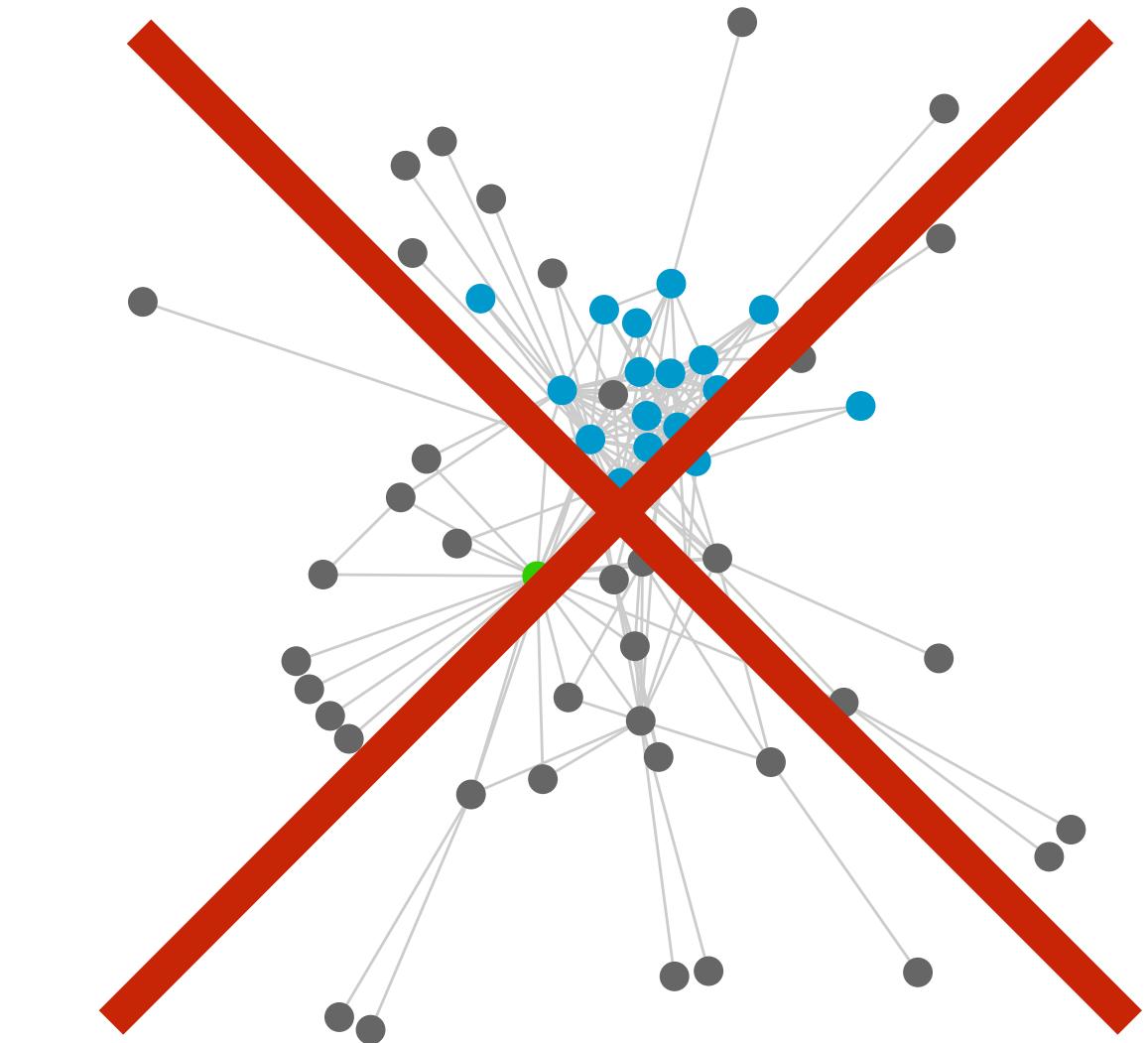
Node-Link  
Diagram



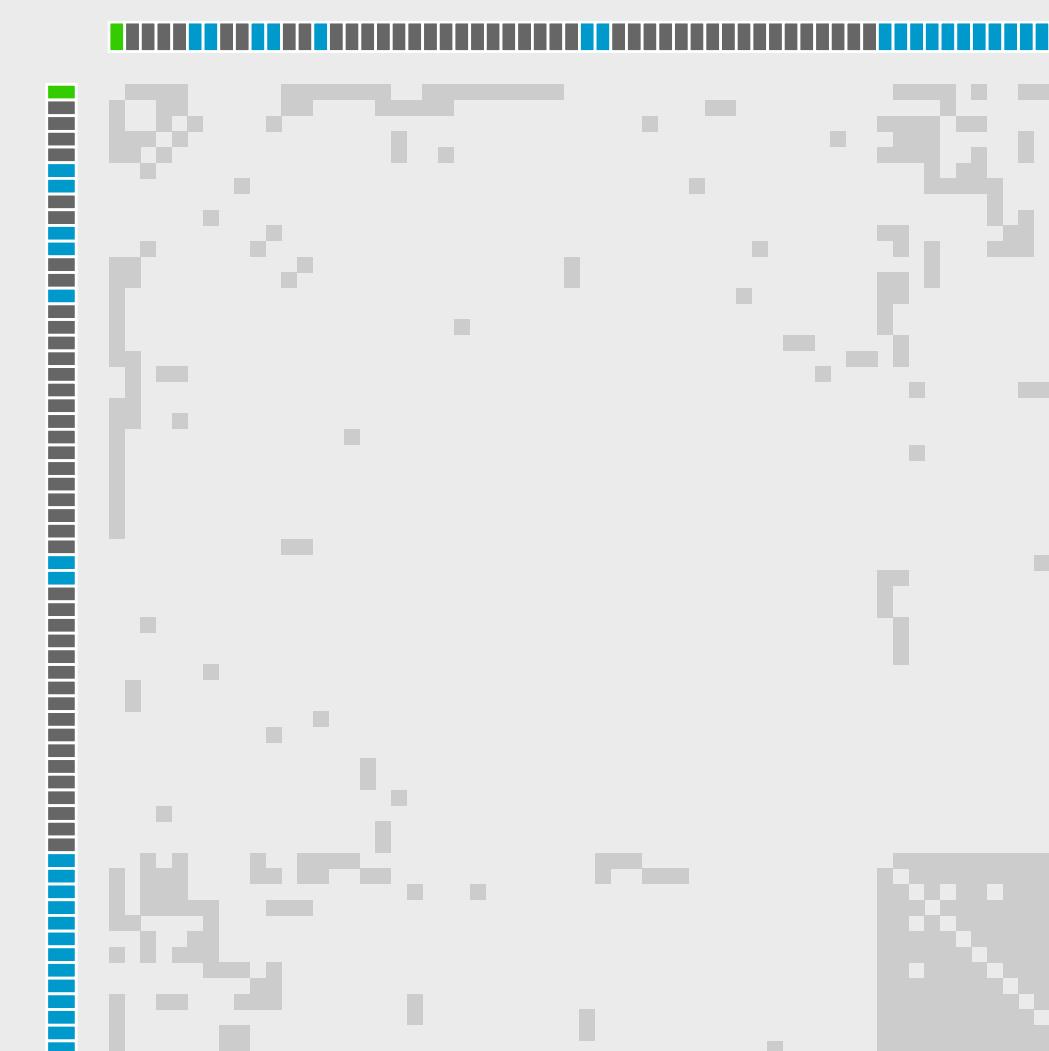
Matrix

# Network Visualization

1.  $(3M \times 3M)/2$  interactions
2. Weight for each interaction
3. Constraint: nodes in sequence order



Node-Link  
Diagram



Matrix

# Genome Interaction Data Visualization

## Encoding

1. Heatmap
2. Node-Link Diagram (here: Arc Diagram)

## Scale

1. Global Interactions (whole chromosome or genome)
2. Local Interactions (immediate feature neighborhood)
3. Individual Features

# Genome Interaction Data Visualization

Yardimci and Noble *Genome Biology* (2017) 18:26  
DOI 10.1186/s13059-017-1161-y

Genome Biology

REVIEW

Open Access

## Software tools for visualizing Hi-C data



CrossMark

Galip Gürkan Yardimci<sup>1</sup> and William Stafford Noble<sup>2\*</sup>

### Abstract

High-throughput assays for measuring the three-dimensional (3D) configuration of DNA have provided unprecedented insights into the relationship between DNA 3D configuration and function. Data interpretation from assays such as ChIA-PET and Hi-C is challenging because the data is large and cannot be easily rendered using standard genome browsers. An effective Hi-C visualization tool must provide several visualization modes and be capable of viewing the data in conjunction with existing, complementary data. We review five software tools that do not require programming expertise. We summarize their complementary functionalities, and highlight which tool is best equipped for specific tasks.

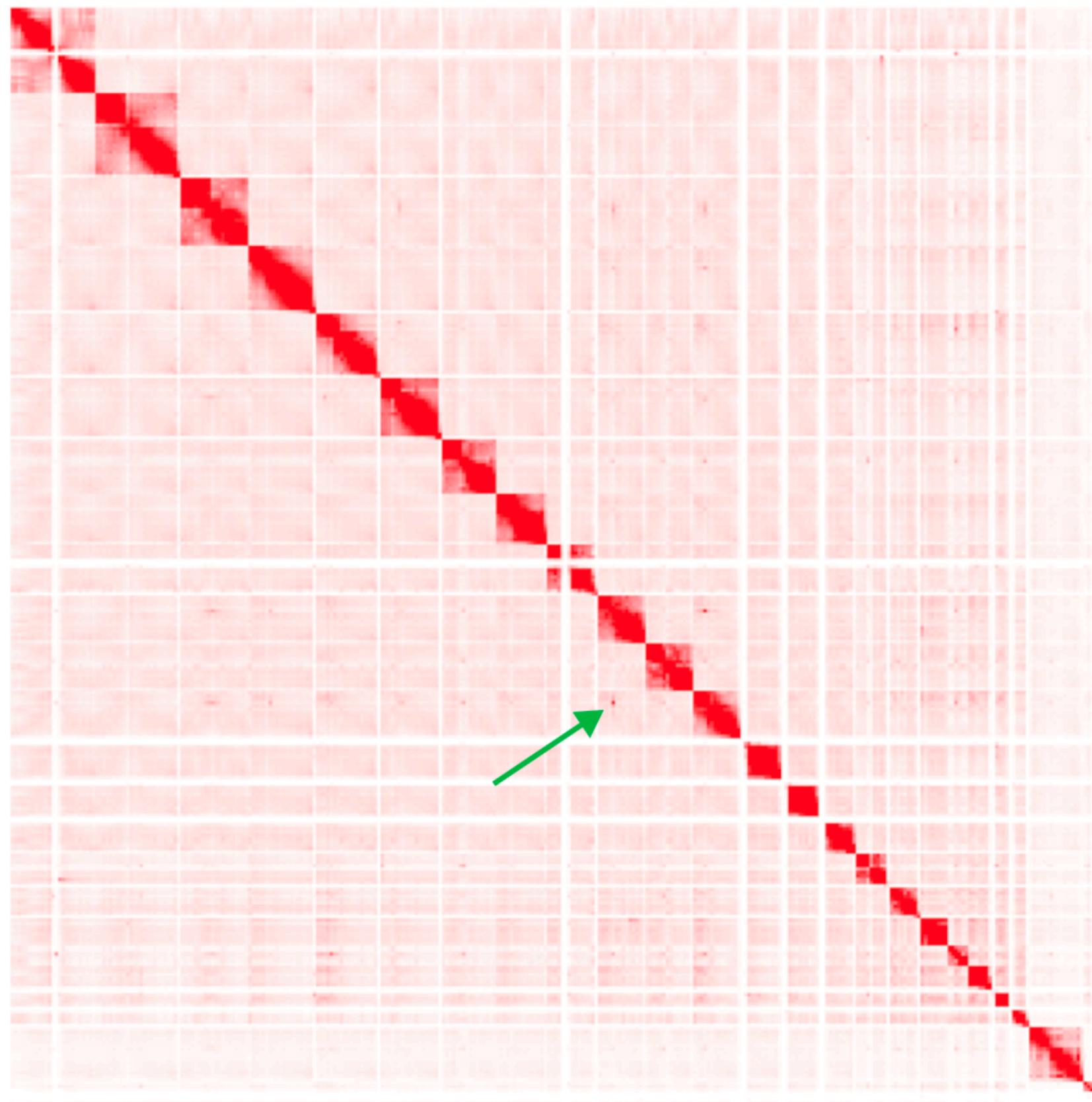
fragments to form a circular molecule marked with biotin, (4) shearing circular DNA and pulling down fragments marked with biotin, and (5) paired-end sequencing of the pulled-down fragments. A pair of sequence reads from a single ligated molecule map to two distinct regions of the genome, and the abundance of such fragments provides a measure of how frequently, within a population of cells, the two loci are in contact. Thus, by contrast with assays such as DNase-seq and chromatin immunoprecipitation sequencing (ChIP-seq) [7, 8], which yield a one-dimensional count vector across the genome, the output of Hi-C is a two-dimensional matrix of counts, with one entry for each pair of genomic loci. Production of this matrix involves a series of filtering and normalization steps (reviewed in [9] and [10]).

A critical parameter in Hi-C analysis pipelines is the effective resolution at which the data is analyzed [10, 11]. In this context, “resolution” simply refers to the size of the loci for which Hi-C counts are aggregated. At present, deep sequencing to achieve very high resolution data for

### Introduction

The three-dimensional (3D) conformation of the genome in the nucleus influences many key biological processes.

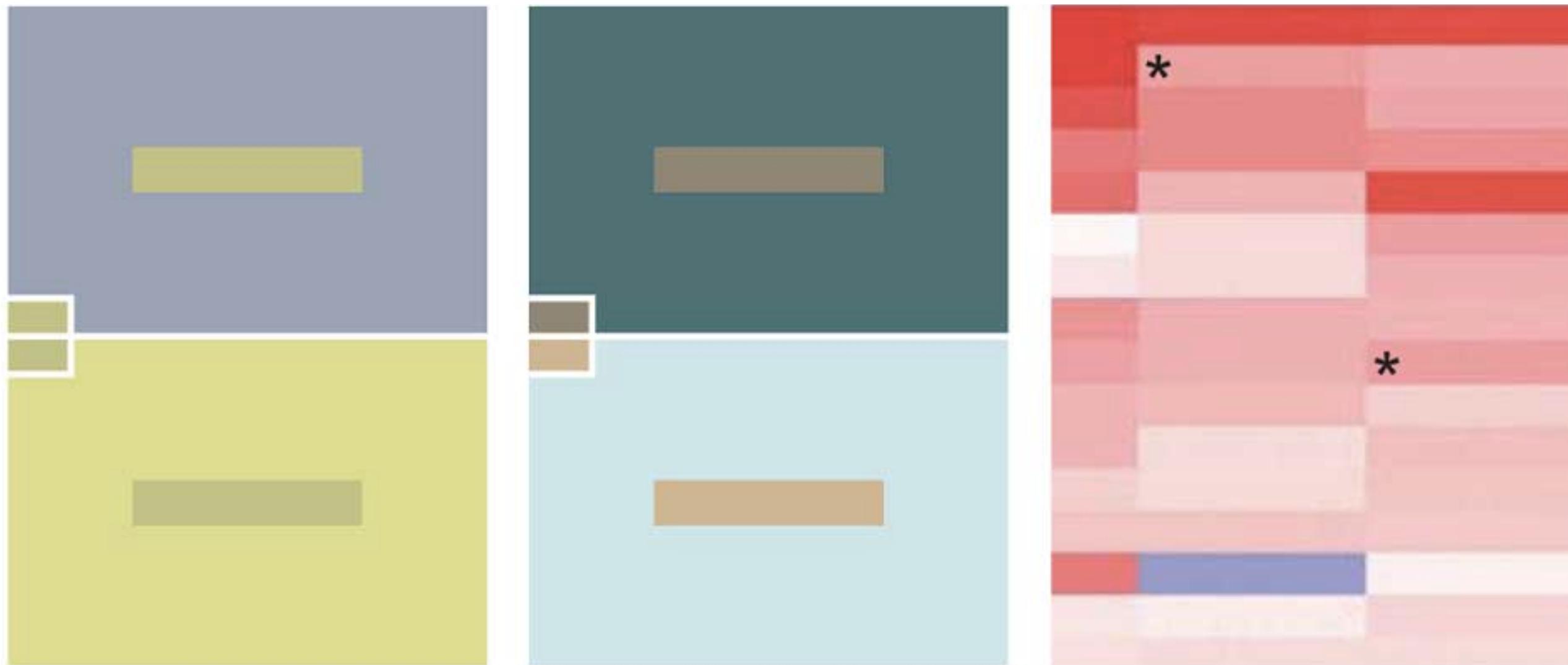
# Global Interactions



Juicebox  
**HEATMAP**

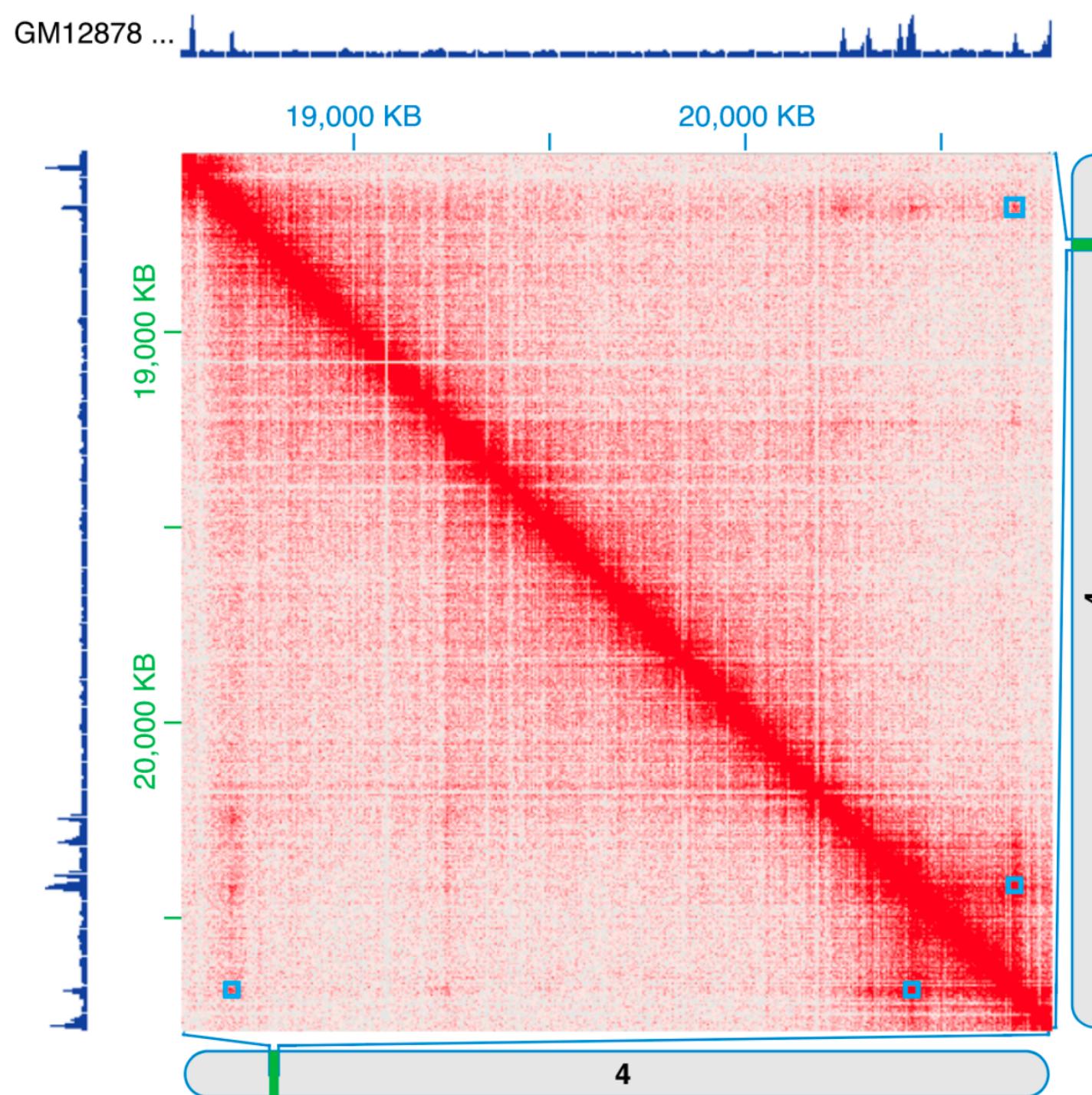
**Caveat**  
only qualitative interpretation of color map possible

# Mini Excursion: Color



Color is a relative medium!

# Global Interactions

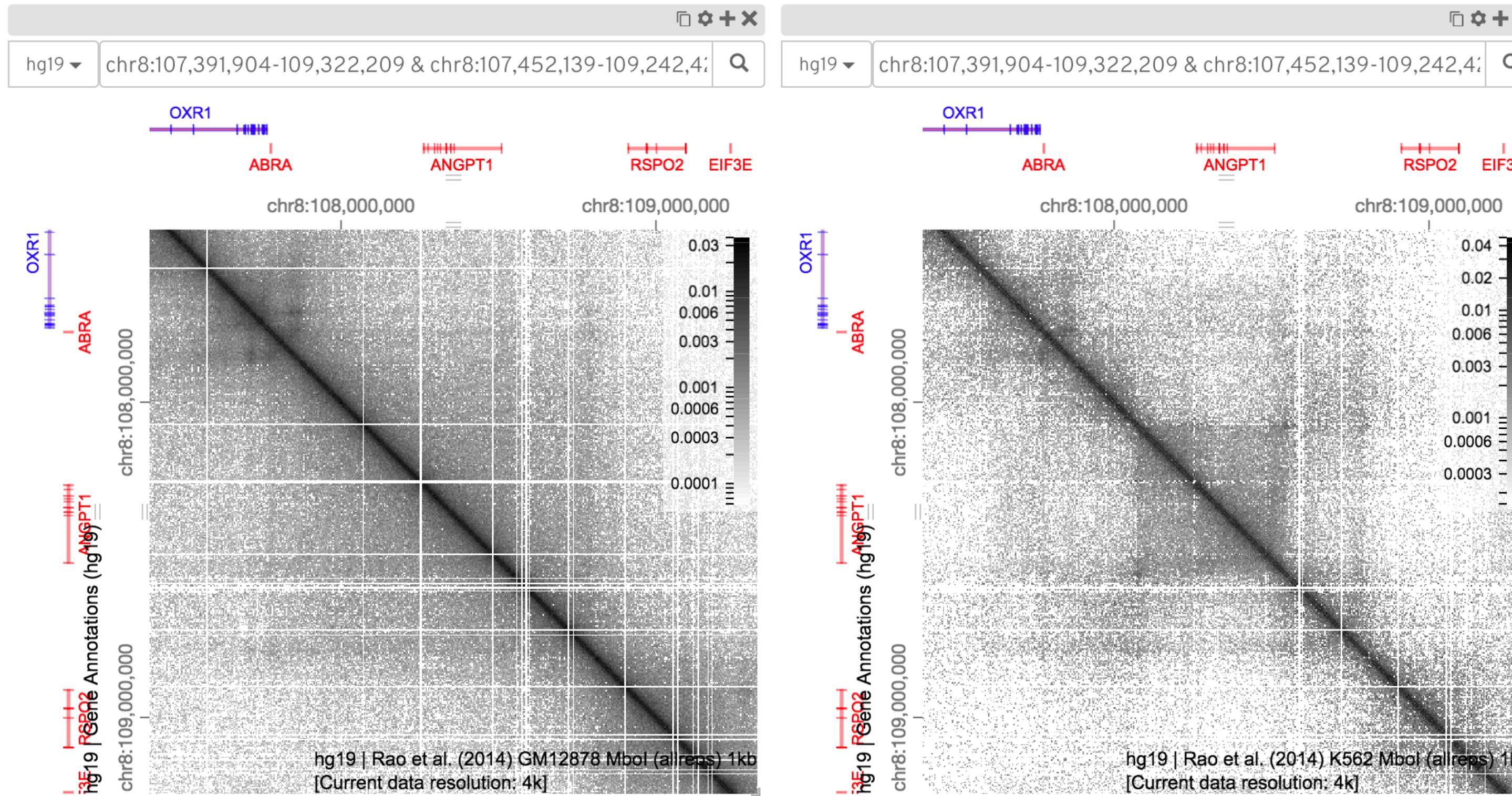


**Juicebox**  
**HEATMAP**

## Caveat

only qualitative interpretation of color map possible

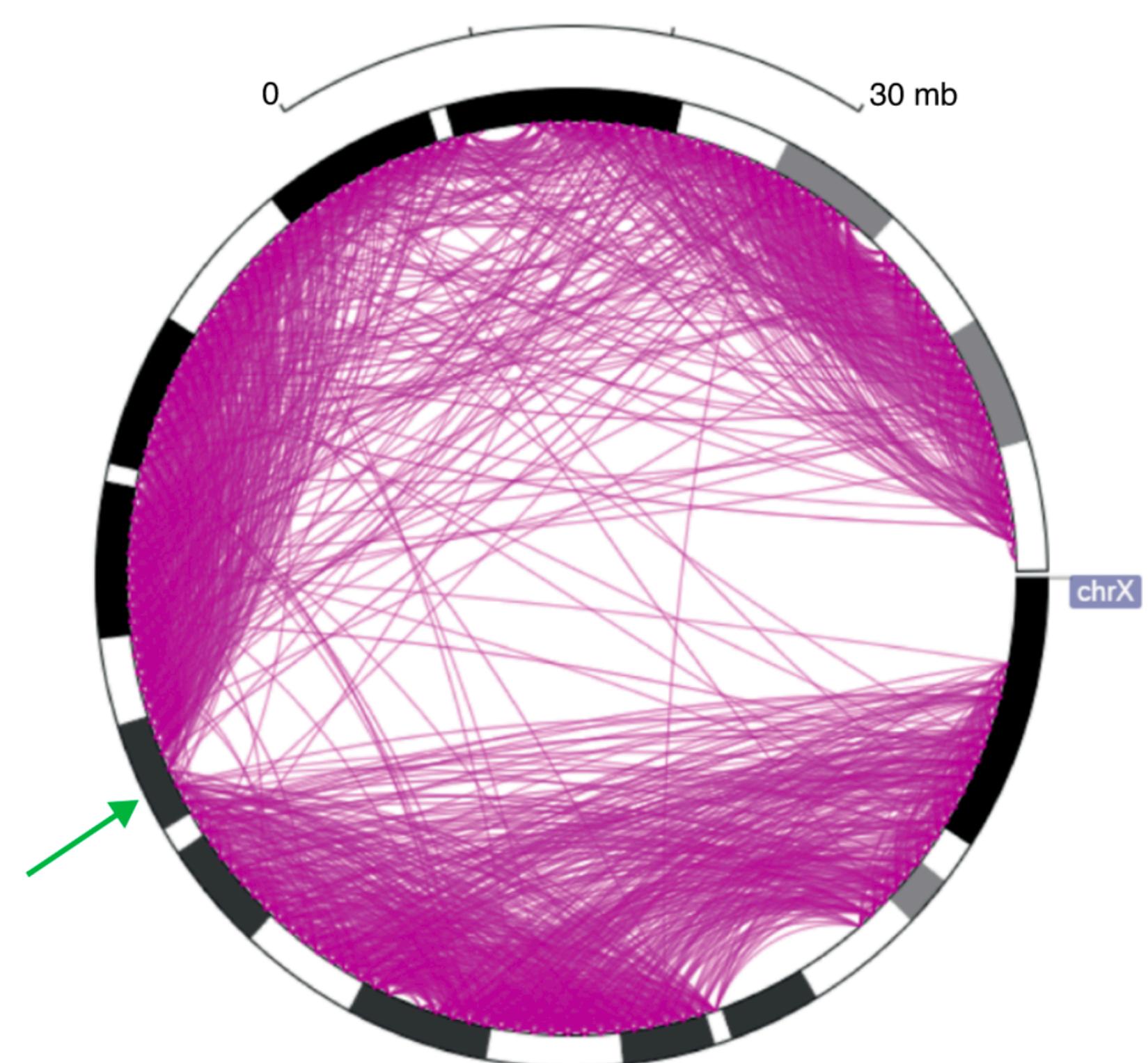
# Global Interactions



HiGlass  
HEATMAP

**Caveat**  
only qualitative interpretation of color map possible

# Global Interactions

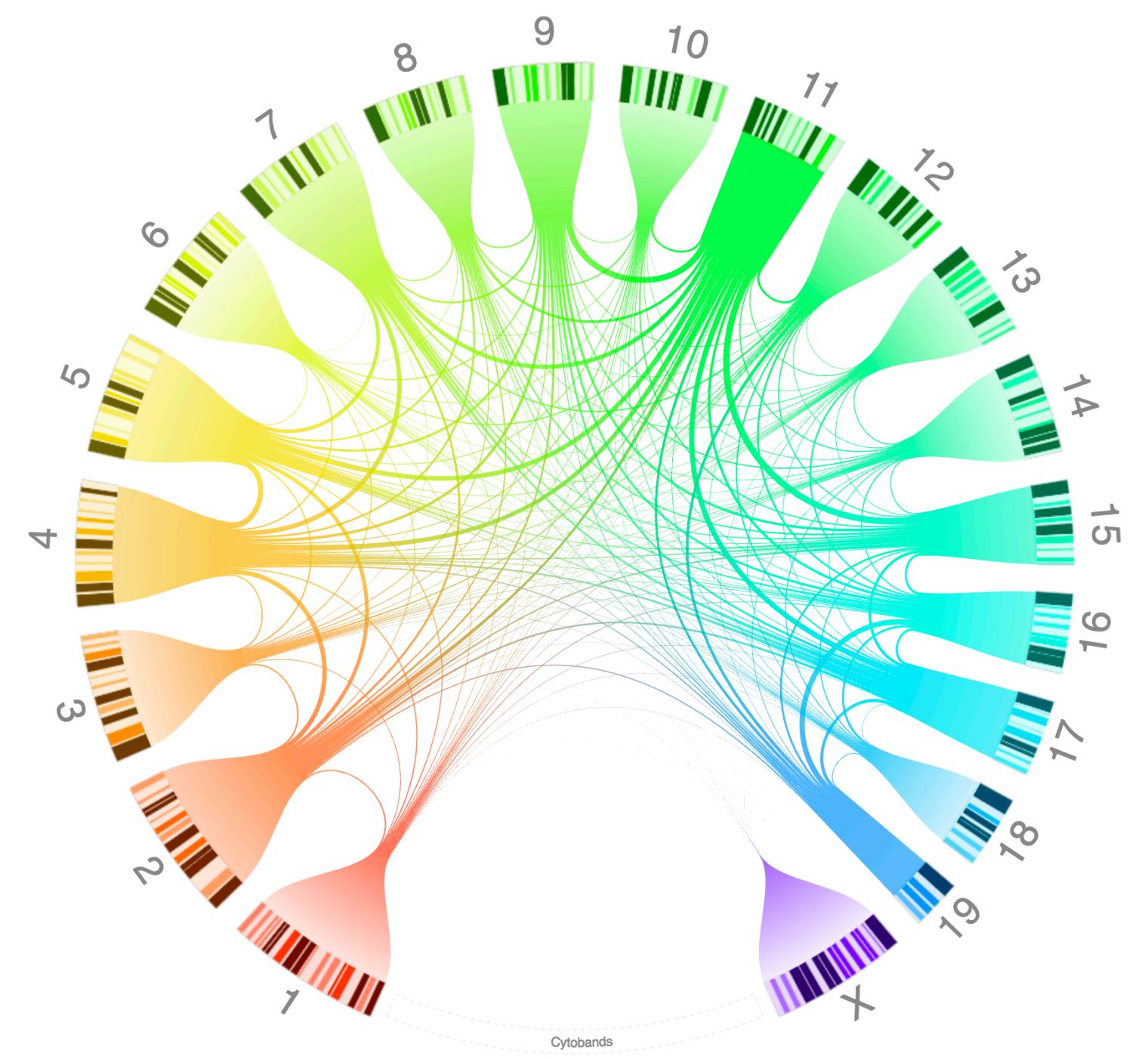


Washington University  
Epigenome Browser  
**ARC DIAGRAM**

## Caveats

line crossings, very limited dynamic range, zooming complex

# Global Interactions

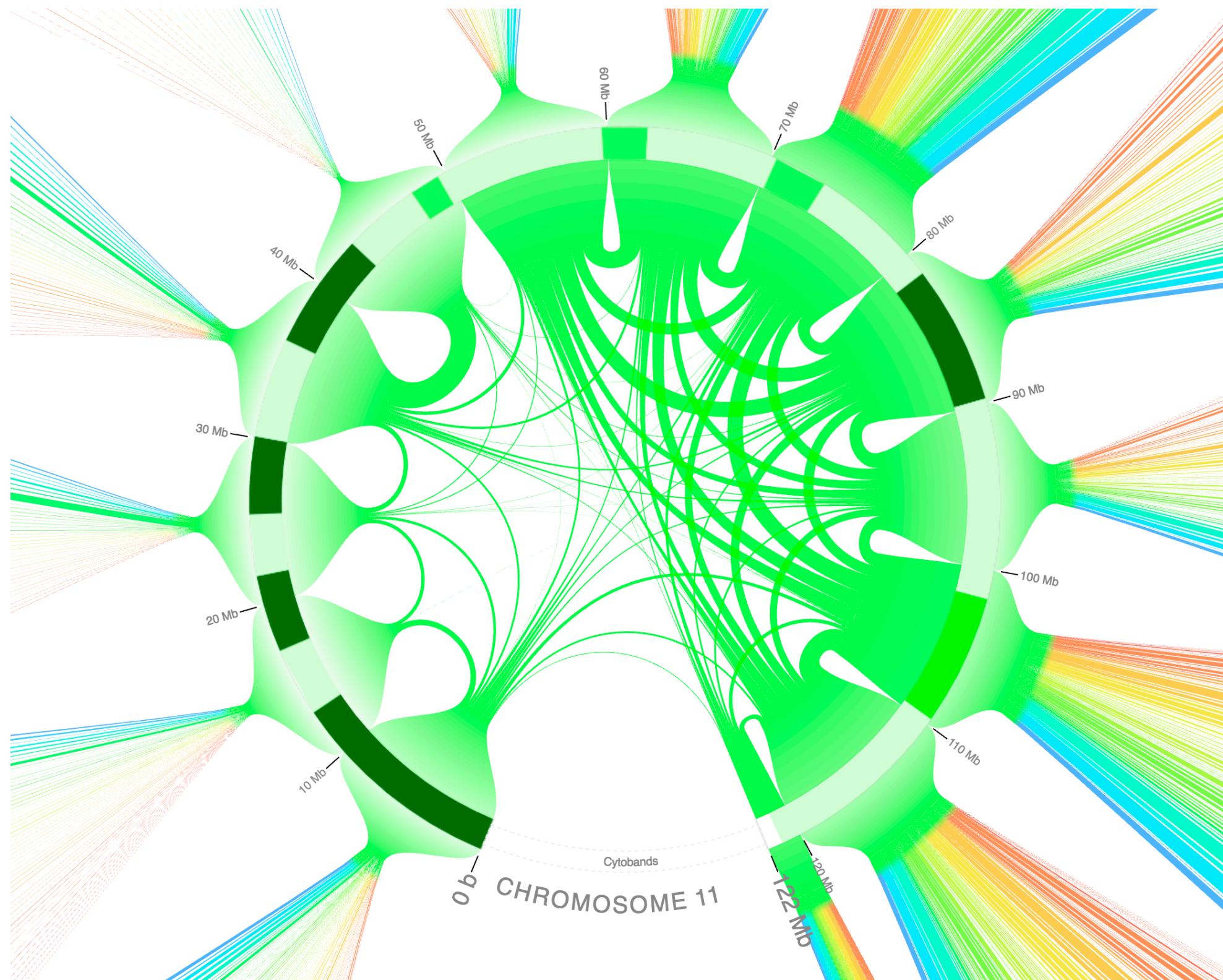


Rondo  
**ARC DIAGRAM**

## Caveats

line crossings, very limited dynamic range

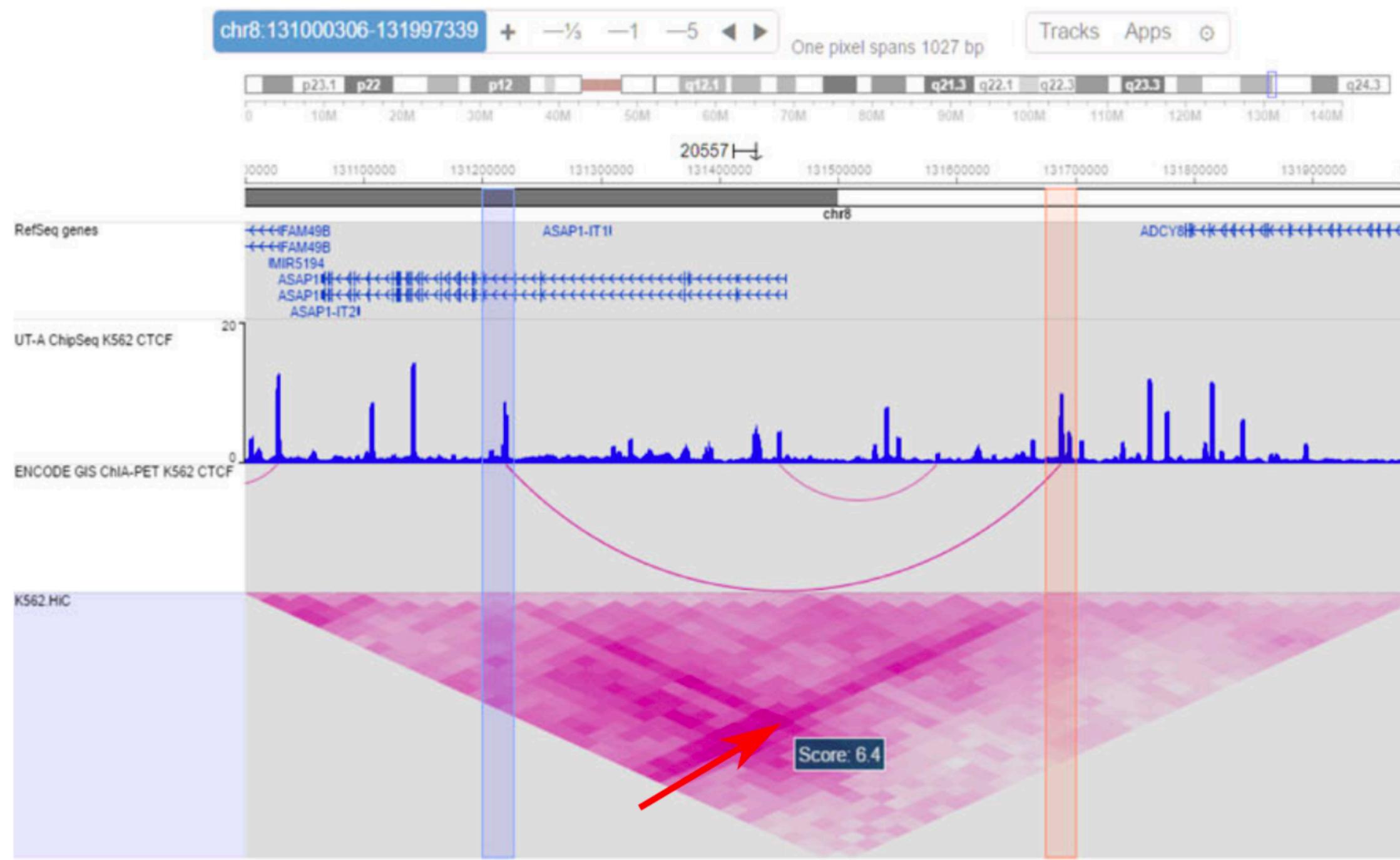
# Global Interactions



Rondo  
**ARC DIAGRAM**

**Caveat**  
line crossings, colors unmappable

# Local Interactions



3D Genome Browser

**HEATSTRIP**

## Caveat

height of triangle grows with  
distance of interaction

# Local Interactions



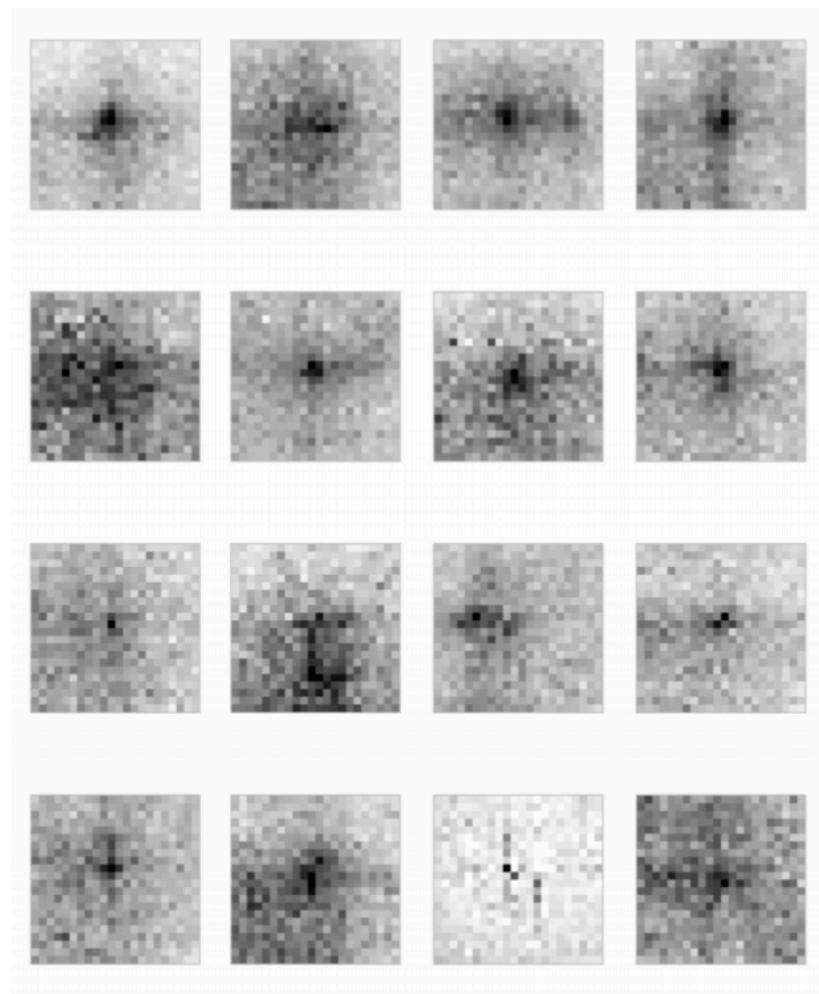
Washington University  
Epigenome Browser  
**ARC DIAGRAM**

## Caveats

zooming is problematic, no context

# Individual Features

Grid



Projection (here: MDS)

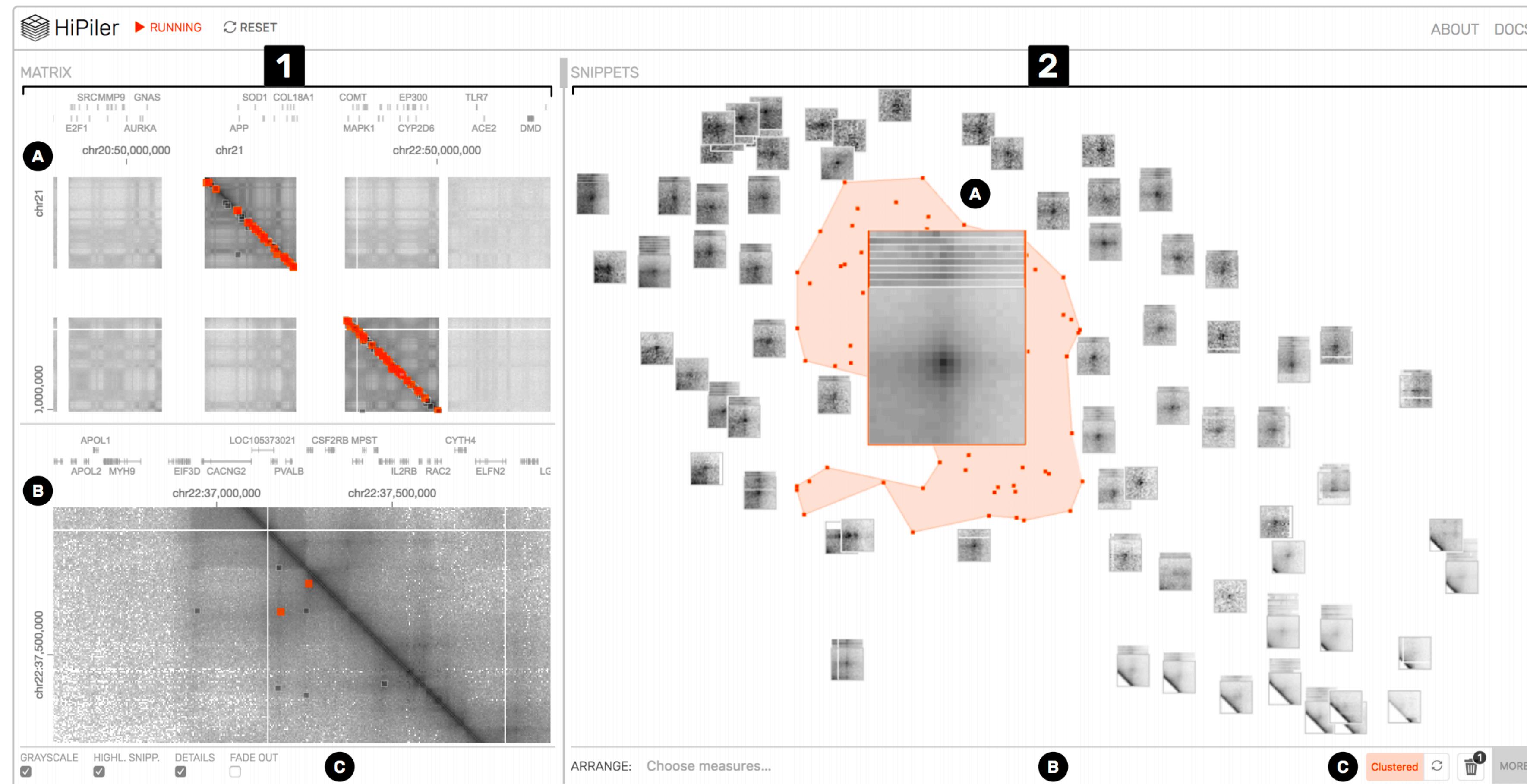


HiPiler

**HEATMAP SNIPPETS**

# Individual Features

HiPiler



# My lab is hiring postdocs!

Data visualization, analysis, and management for:

- exploration tools for data repositories
- provenance graphs
- genomic structural variants
- dynamics of the 3D genome
- cancer subtypes in patient cohorts

<http://gehlenborglab.org>