

# DeCART Summer School 2018

## Data Visualization

*Day 2 - Afternoon*

17 July 2018

Nils Gehlenborg, PhD - Harvard Medical School

# Syllabus

- **Day 1**

- Morning: Introduction to Data Visualization
- Afternoon: Design Process, Evaluation, and Interaction

- **Day 2**

- Morning: Visualization of high-dimensional, network, and temporal data
- **Afternoon: Genomic data visualization techniques  
Introduction to Altair and advanced Altair**

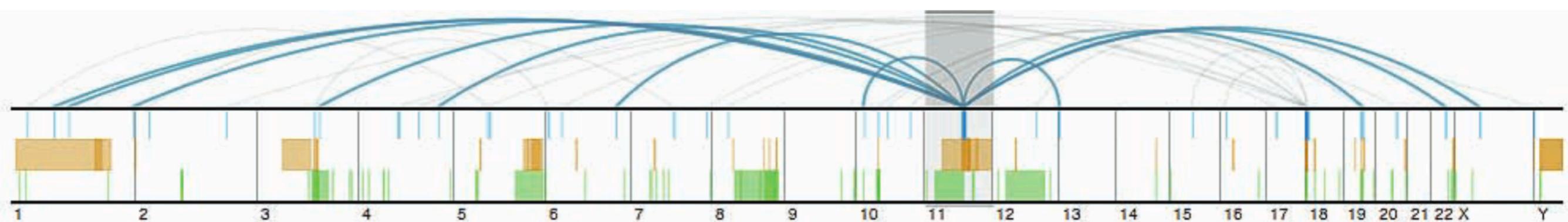
# Syllabus - Day 2 (Afternoon)

- 1:30 - 2:00 | Genomic Data
- 2:00 - 2:30 | Altair Introduction
- 2:30 - 3:00 | Altair Exercise 1
- 3:00 - 3:15 | Break
- 3:15 - 4:45 | Altair Exercise 2
- 4:45 - 5:00 | Questions and Resources

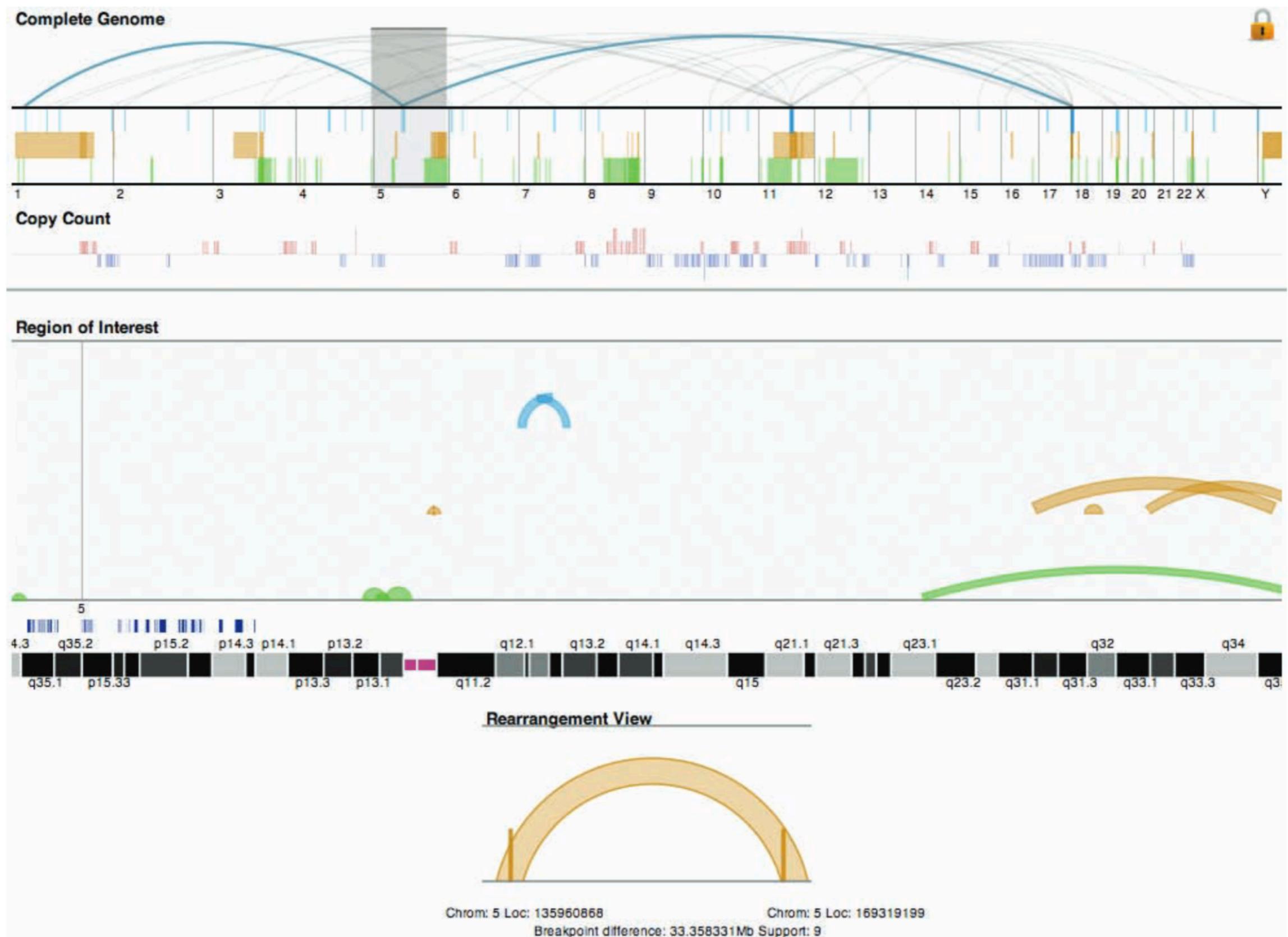
# Genome Interactions

## *Structural Variants*

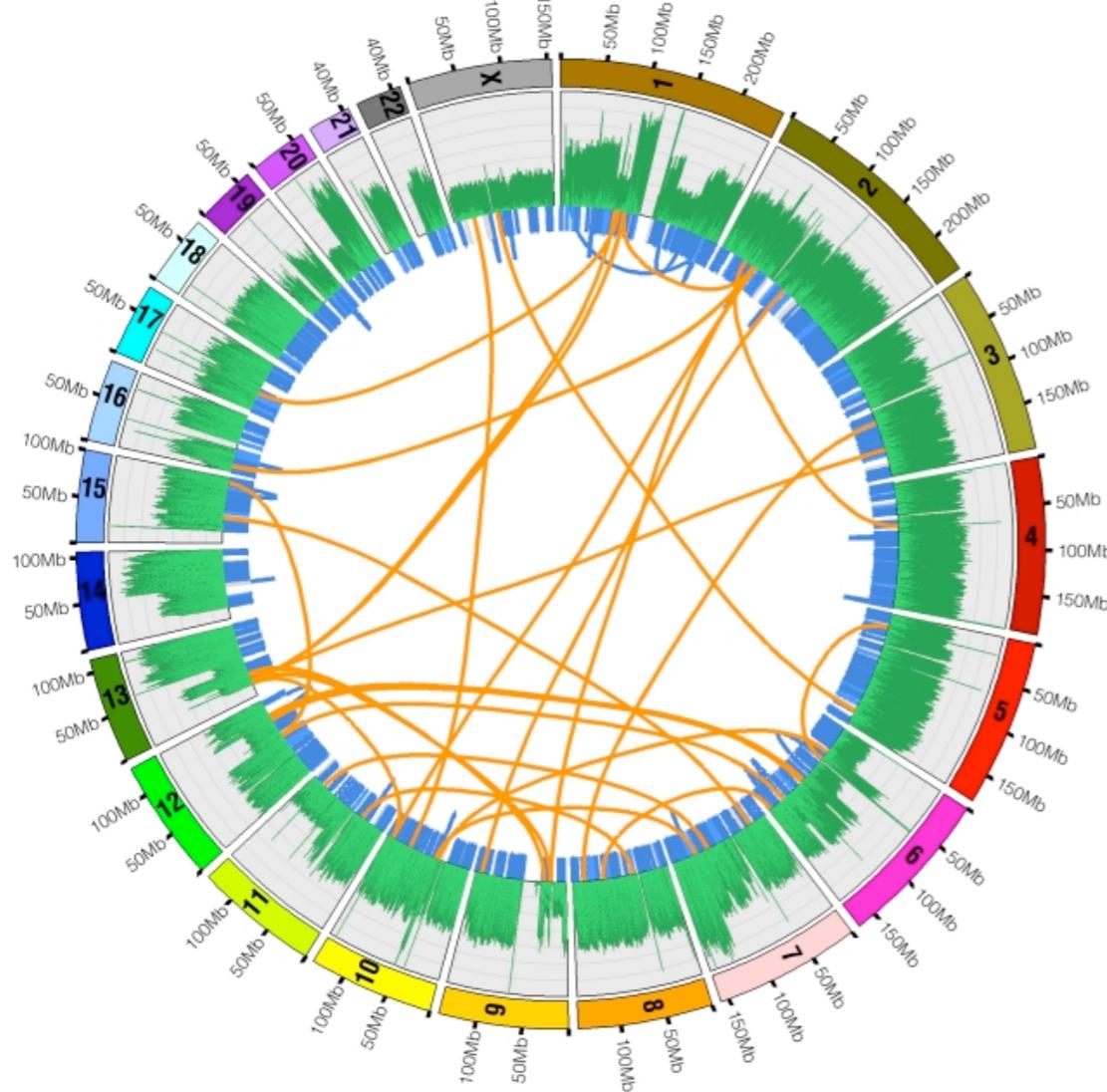
# Gremlin - Structural Variant Visualization



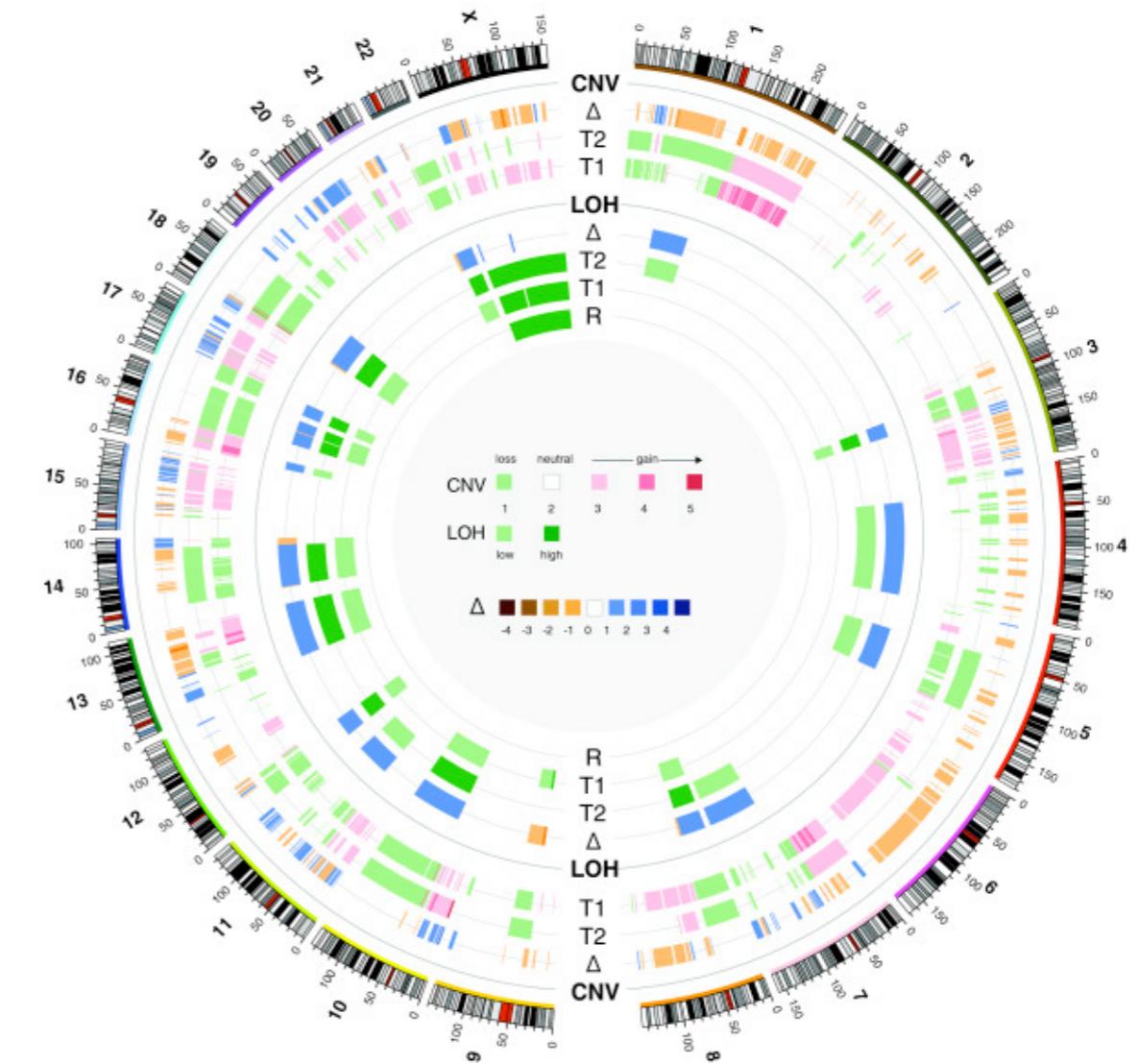
# Gremlin - Structural Variant Visualization



# Sequences: Circos

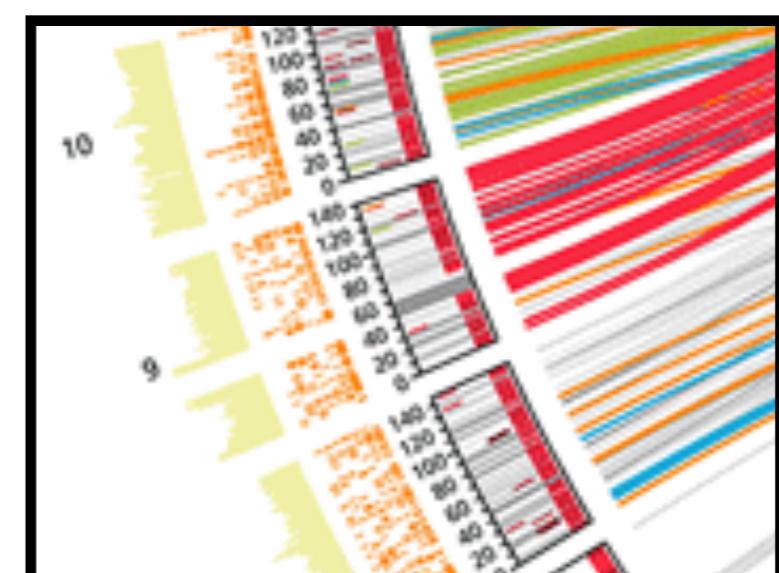
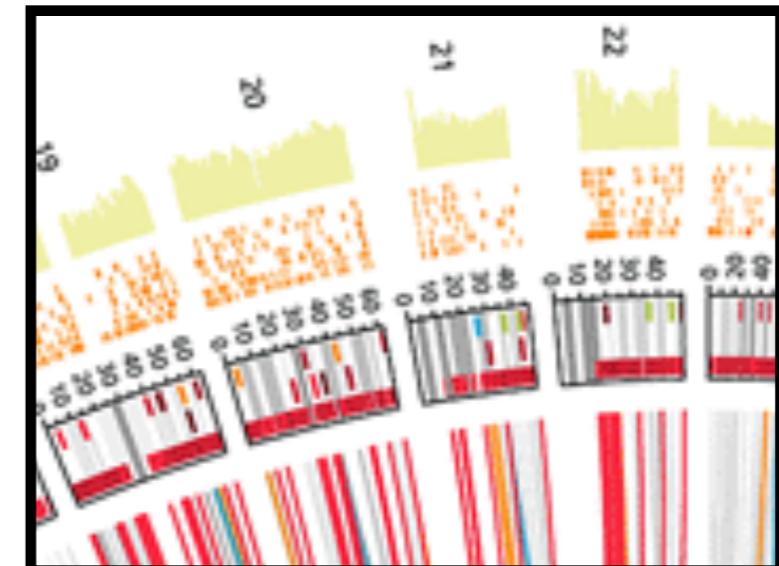


Clark et al. 2009, PLoS Genetics



Jones et al. 2010, Genome Biology

# Sequences: Circos



## HUMAN GENOME (hg18)



## HUMAN GENOME (hg18)

## CLINICAL ATTRIBUTES

## SAMPLE

## SEX

 Male     Female

## TUMOR STAGE

 I     IIa     IIb     IIIa  
 IIIb     IIIc     IVa     IVb

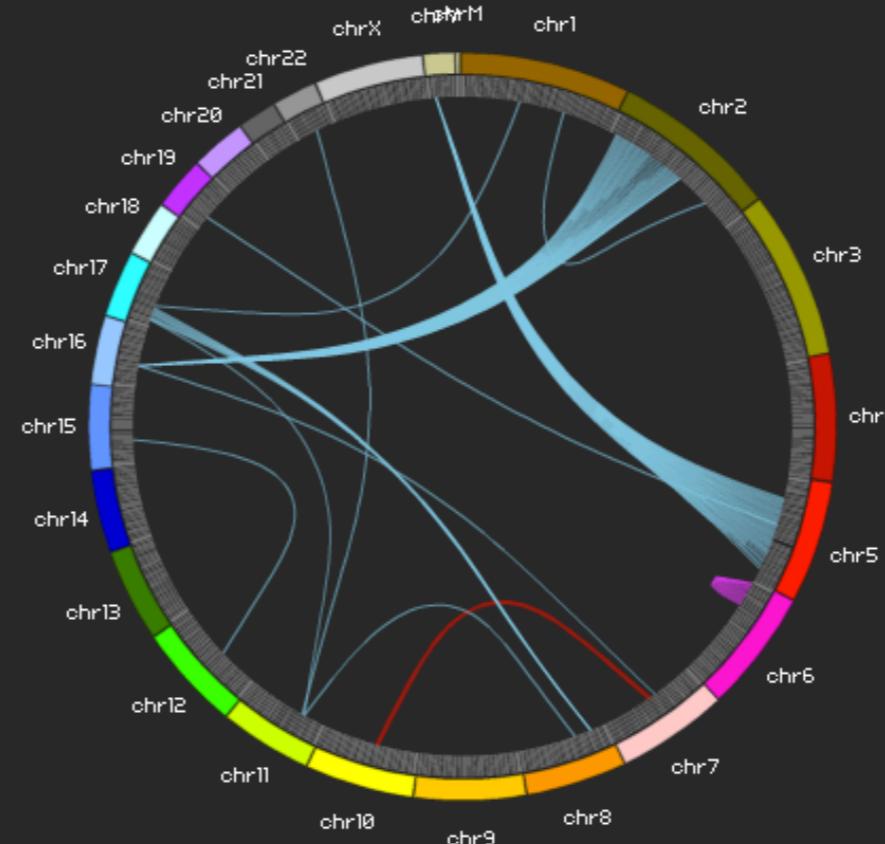
## AGE

 30-44     45-59  
 60-74     75-80

## HISTOLOGY

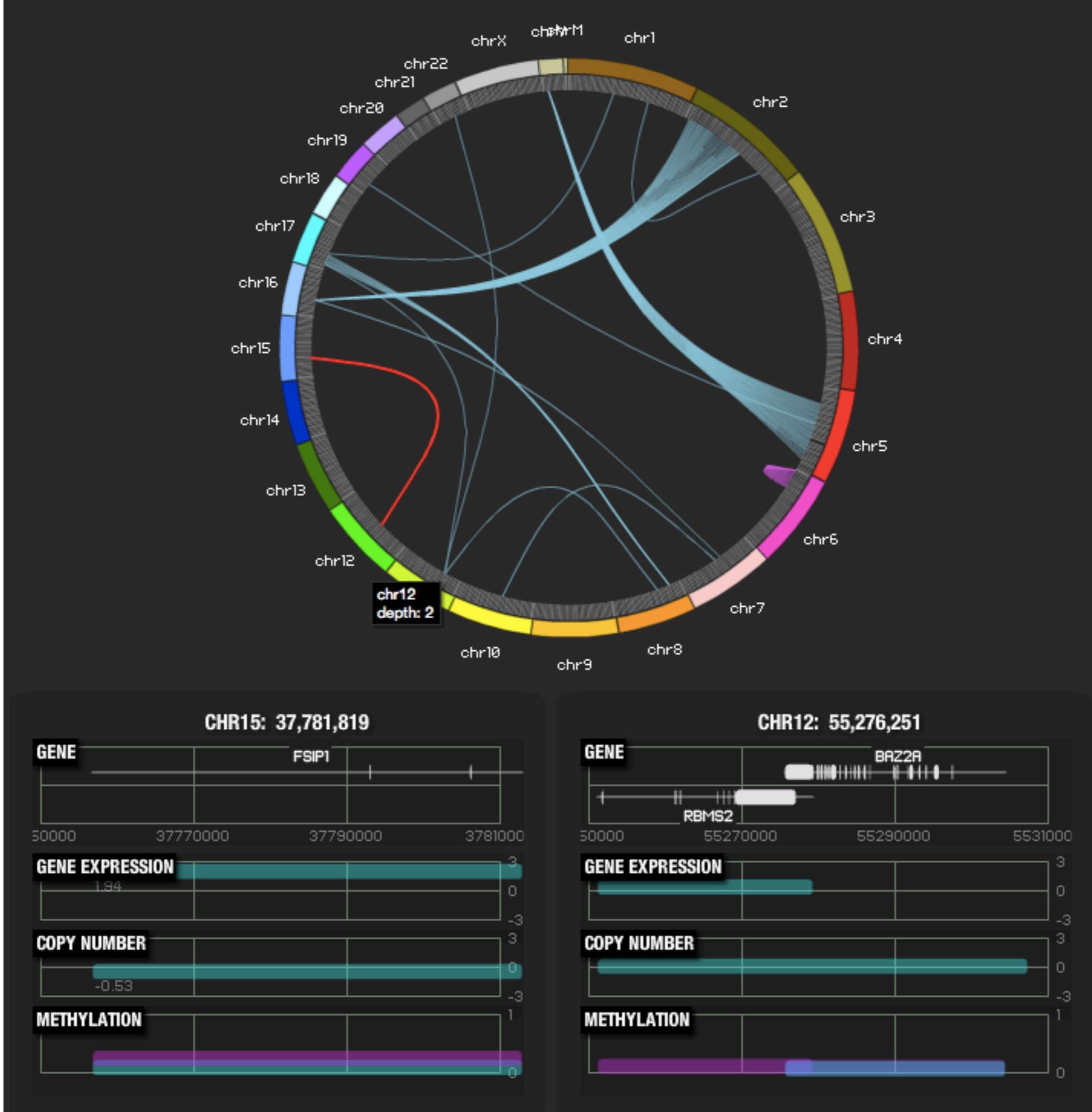
 Colon AD Colon Mucinous AD Rectal AD Rectal Mucinous AD

## FPS:30



ID	TCGA.ID	TYPE	PAIRS_SUPPT	TUMOR_COUNT	NORMAL_COUNT	COPY_RATIO	DEST_CHROM	DEST_BRK	ORG_CHROM	START	END	SIZE	PVALUE_DEPTI
56995	TCGA-RG-3909	ITX	1ITX/3ITX	73	68	1.18	chr7	81627174	chr10	60572278	60572922	645	0.181
56997	TCGA-RG-3909	ITX	1ITX/3ITX	33	48	0.76	chr17	10079441	chr11	37385505	37385892	388	0.911
57000	TCGA-RG-3909	ITX	1ITX/2ITX	11	35	0.35	chr12	55276251	chr15	37781819	37782091	273	1
57001	TCGA-RG-3909	ITX	1ITX/3ITX	1638	1624	1.11	chr2	132739567	chr16	33866699	33867466	768	0.001
57008	TCGA-RG-3909	ITX	1ITX/2ITX	1015027	1445912	0.77	chr8	30265192	chr17	7108460	26629258	19520799	1
57007	TCGA-RG-3909	ITX	1ITX/3ITX	126	149	0.93	chr1	88745410	chr17	29087799	29088923	1125	0.74
57009	TCGA-RG-3909	ITX	1ITX/2ITX	46	57	0.89	chr2	190707698	chr1	162506924	162507345	422	0.755
57012	TCGA-RG-3909	ITX	1ITX/3ITX	49	52	1.04	chr11	43407789	chr22	27395608	27396017	410	0.464
57013	TCGA-RG-3909	ITX	1ITX/3ITX	9200287	9849329	1.03	chr16	33871146	chr2	9608546	132729533	123120988	0
57020	TCGA-RG-3909	ITX	2ITX/1ITX	8899616	9640101	1.02	chrY	10630290	chr5	32406020	153009395	120603376	0
57019	TCGA-RG-3909	ITX	1ITX/4ITX	1059	1148	1.02	chr19	23975735	chr5	71182393	71182777	395	0.359

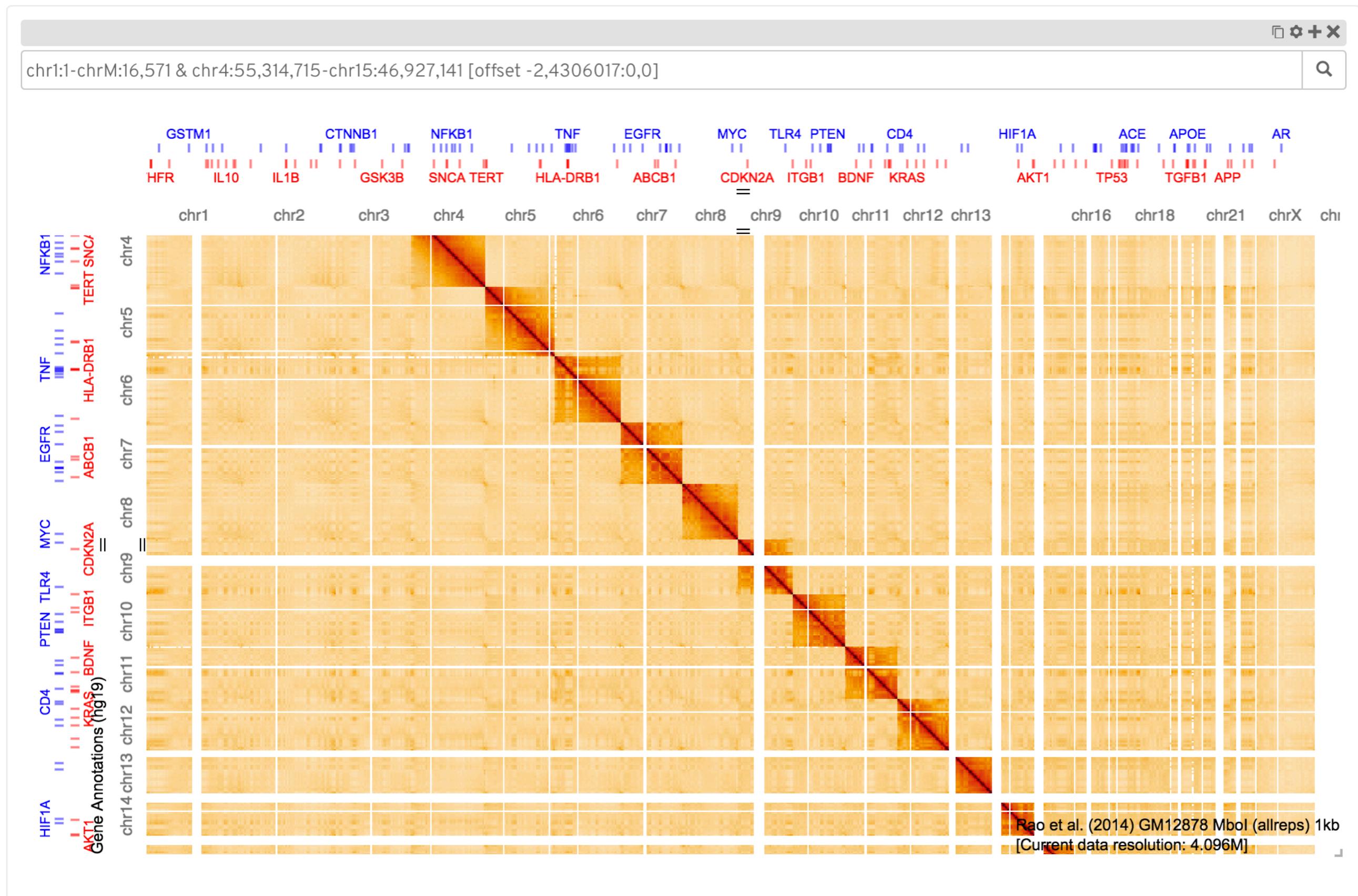




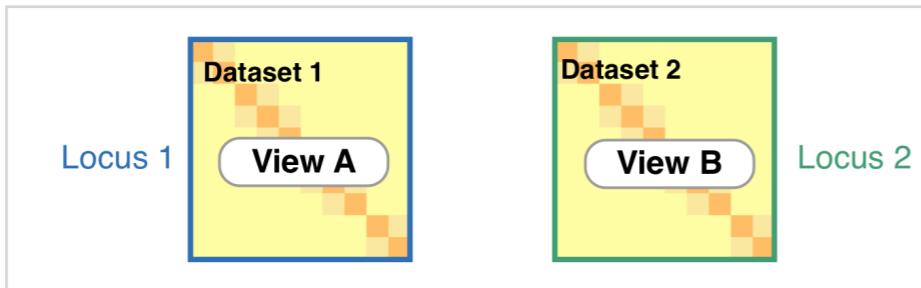
# Genome Interactions

## *3D Interactions*

# HiGlass



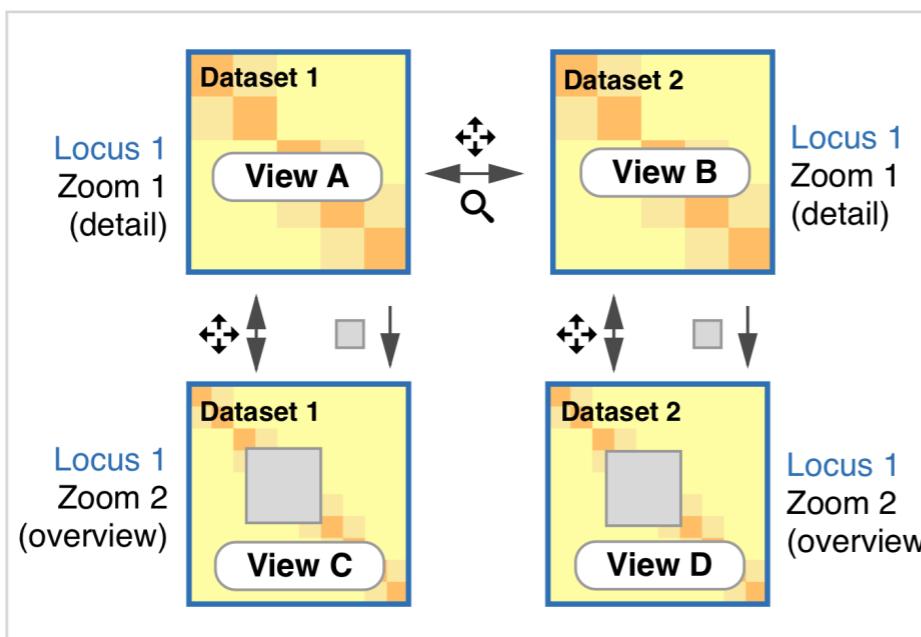
**a Two independent views**



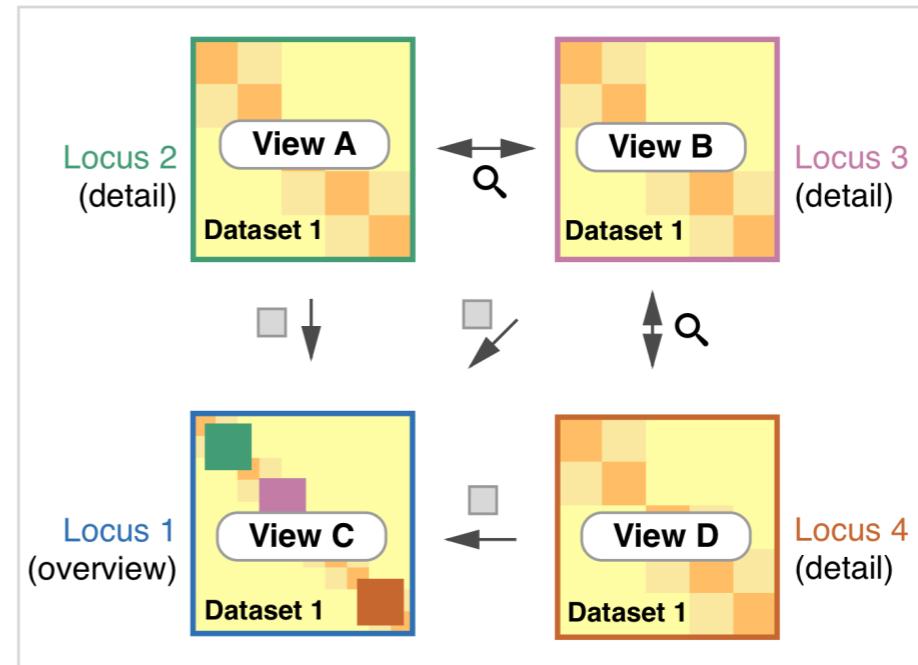
**b Two linked views of the same locus in two datasets**



**c Comparison using overview and detail views**



**d Overview and three linked detail views**

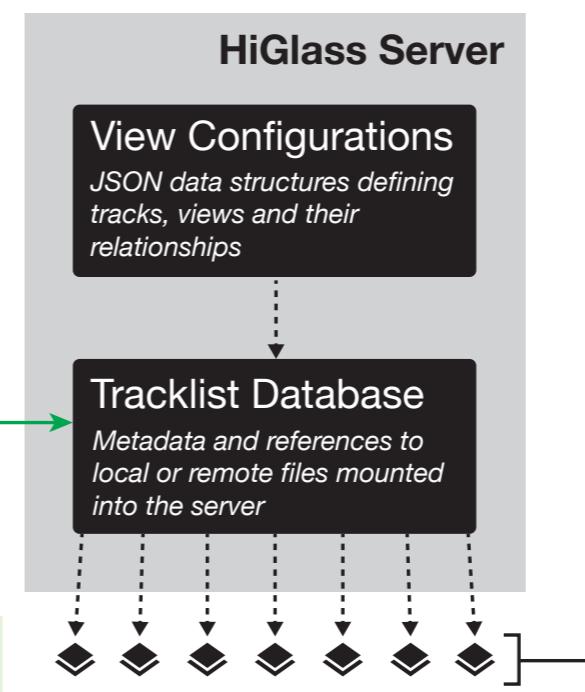
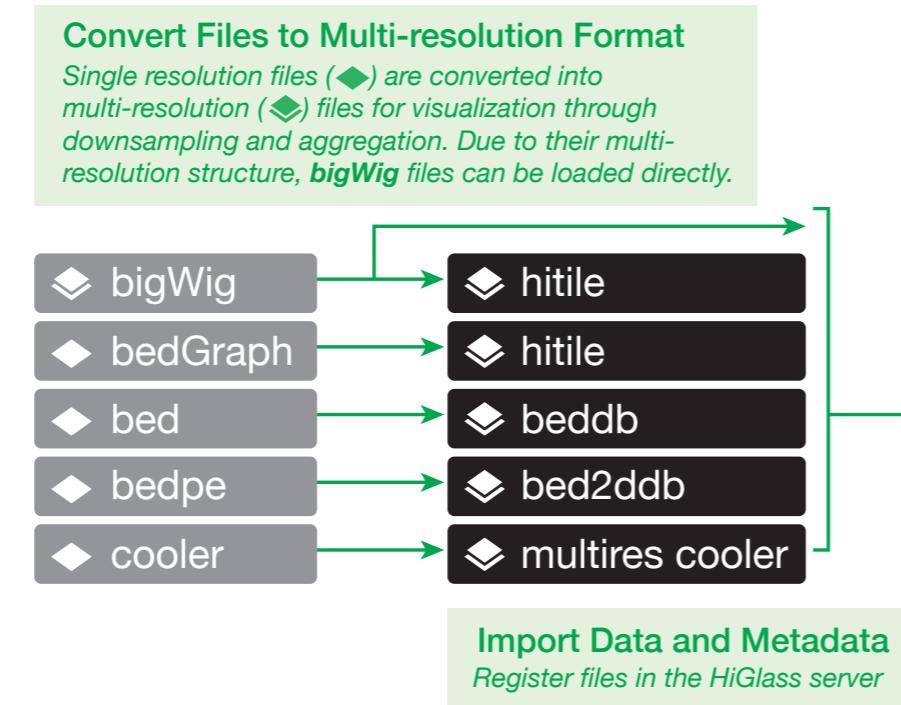
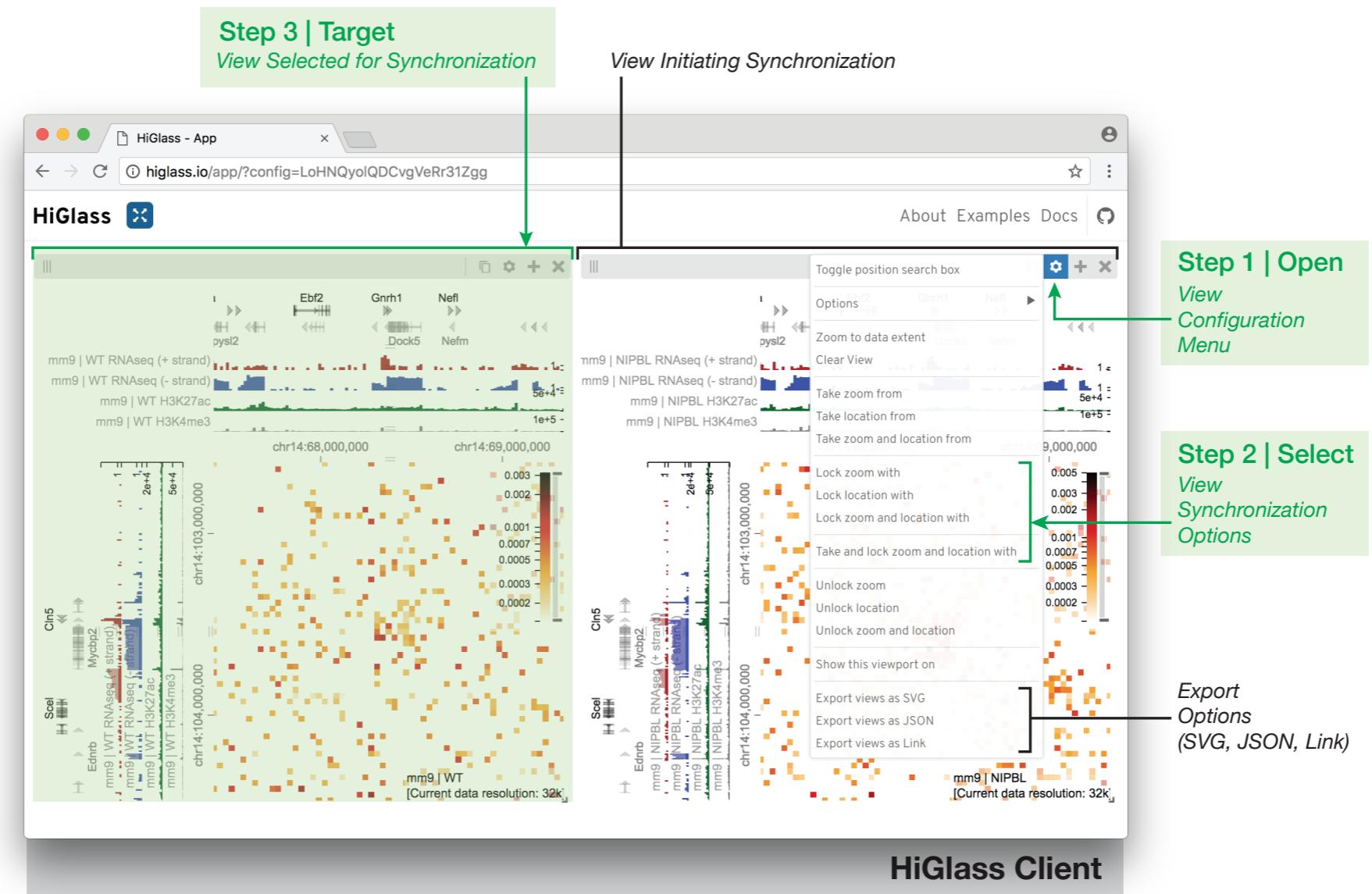


↔ Q Views linked by **zoom level and location**

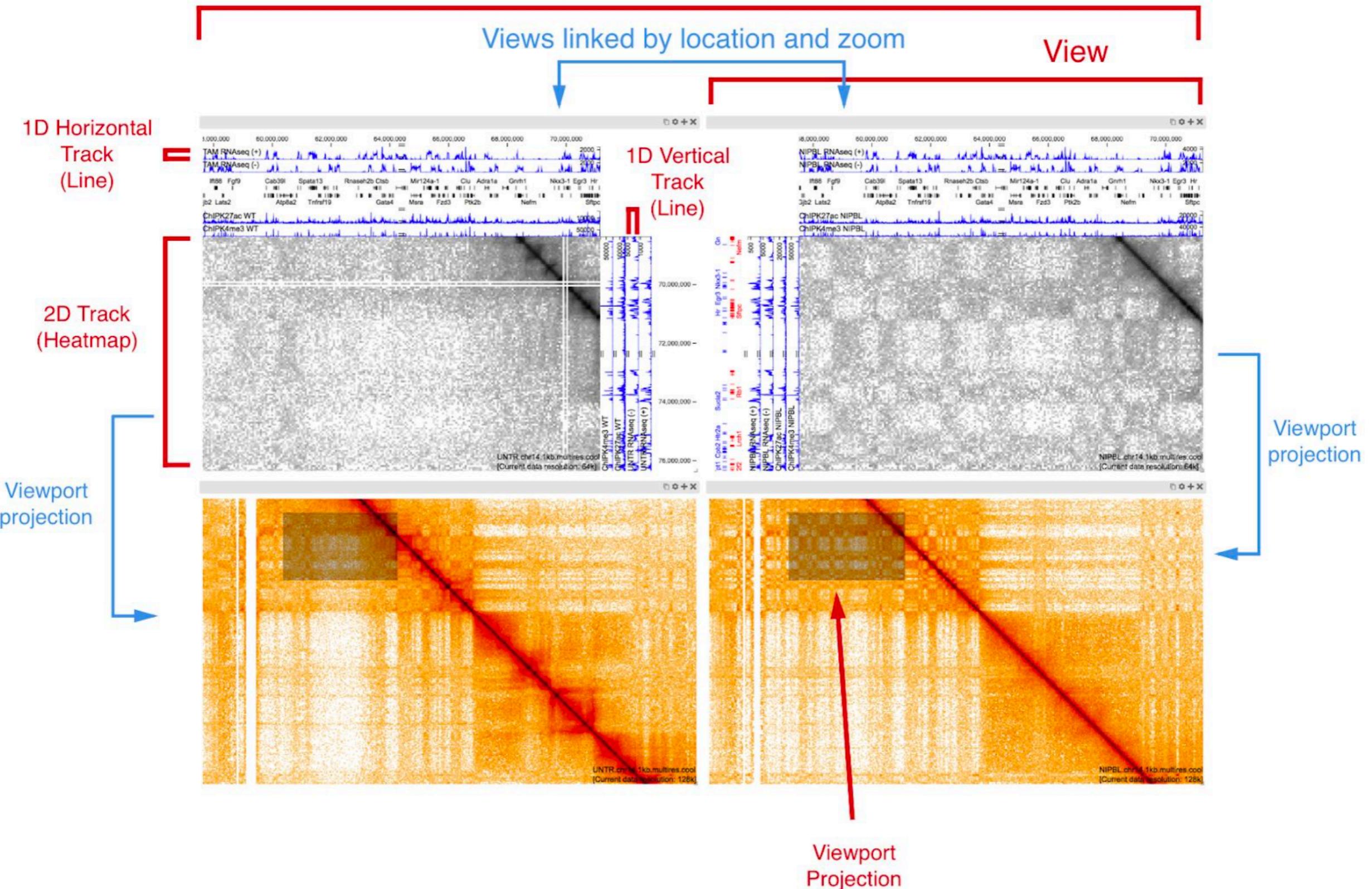
↔ Views linked by **location only**

↔ Q Views linked by **zoom only**

→ Bounds of source view shown on target  
**(viewport projection)**

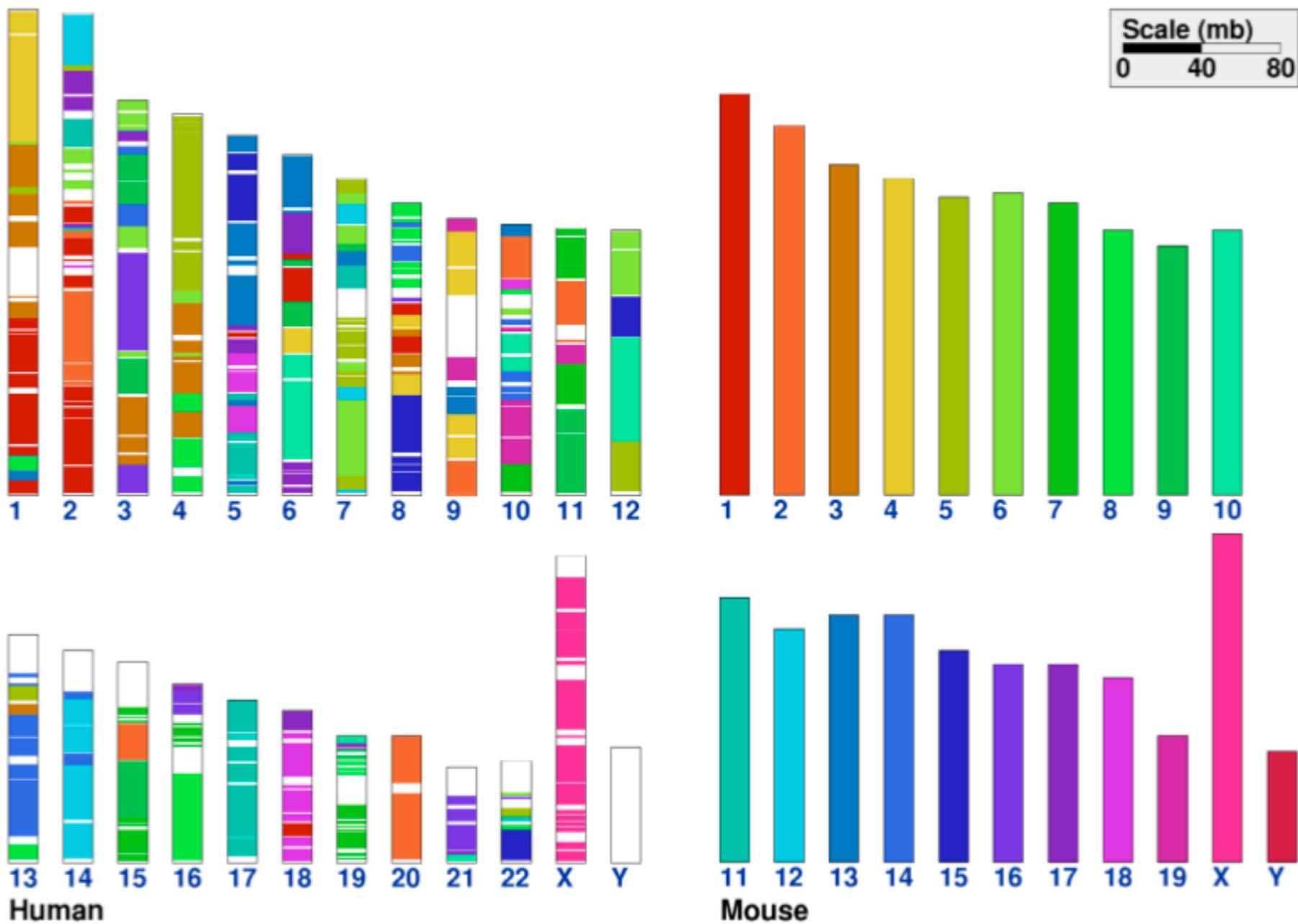


## View Composition

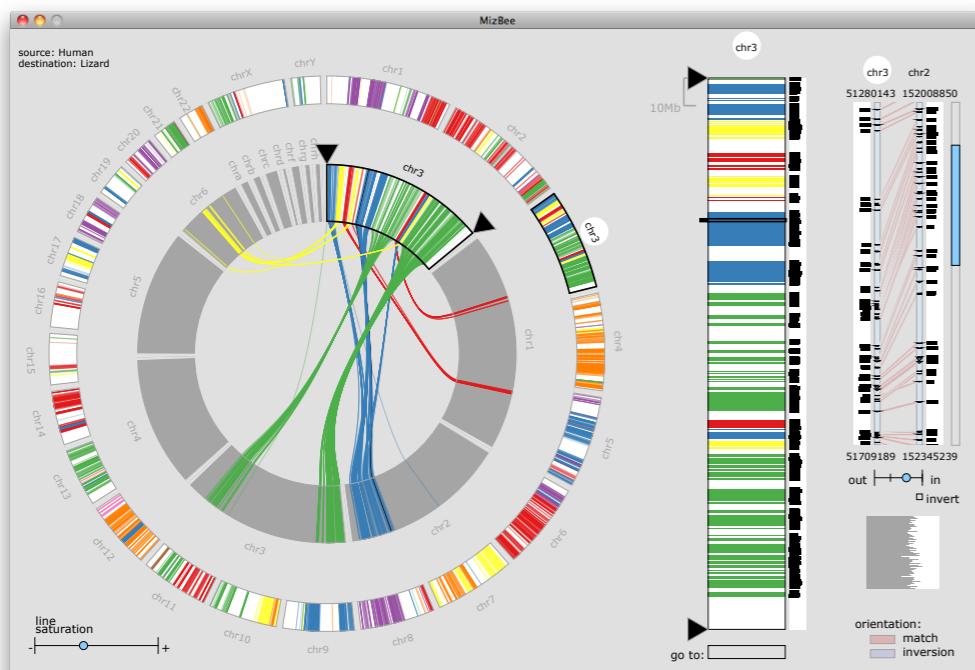


# Synteny

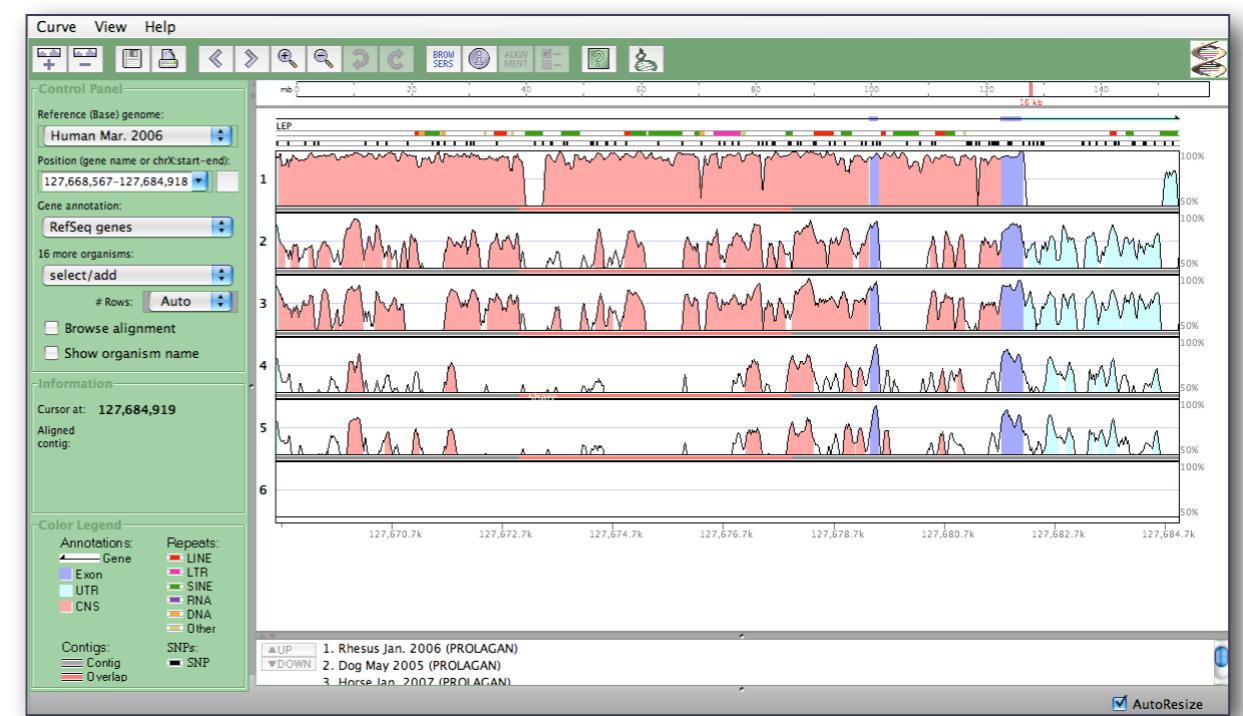
# Sequences: Comparative Genomics



# Sequences: Comparative Genomics

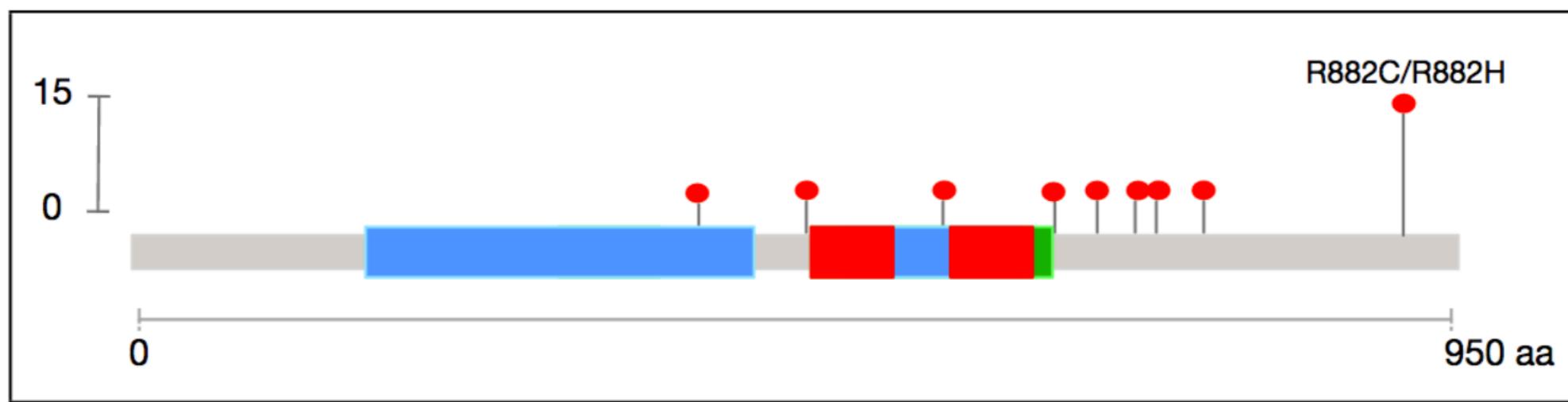


<http://www.mizbee.org>

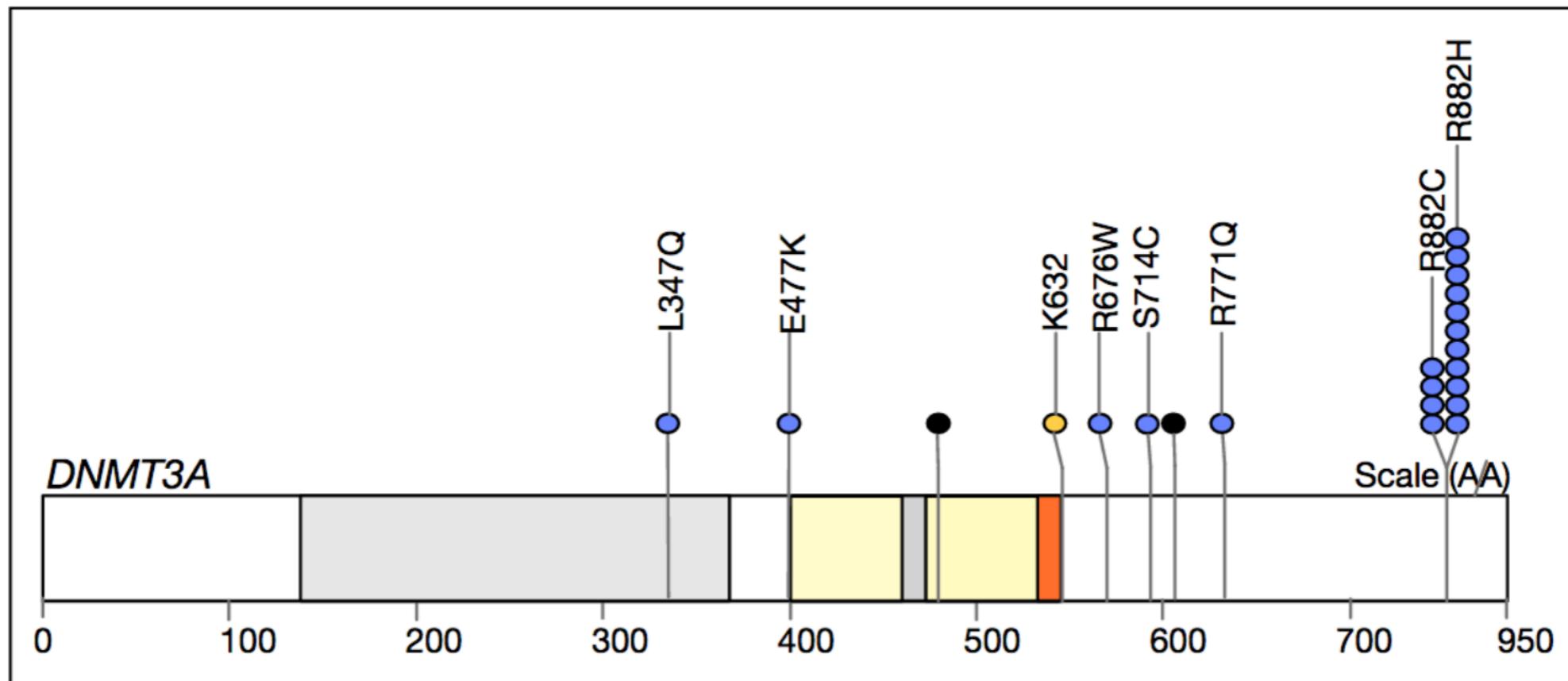


<http://genome.lbl.gov/vista>

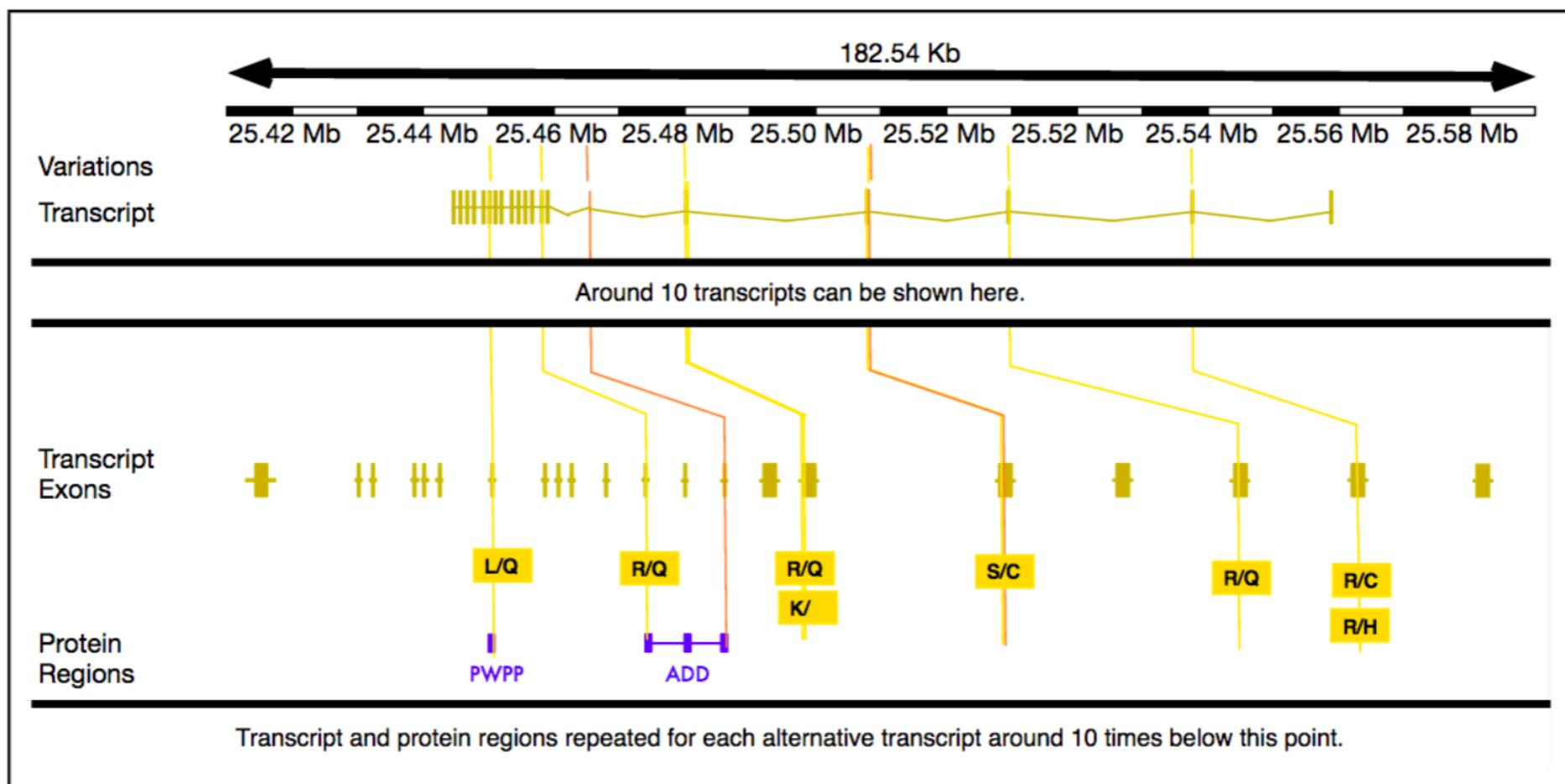
# **SNPs and Indels**



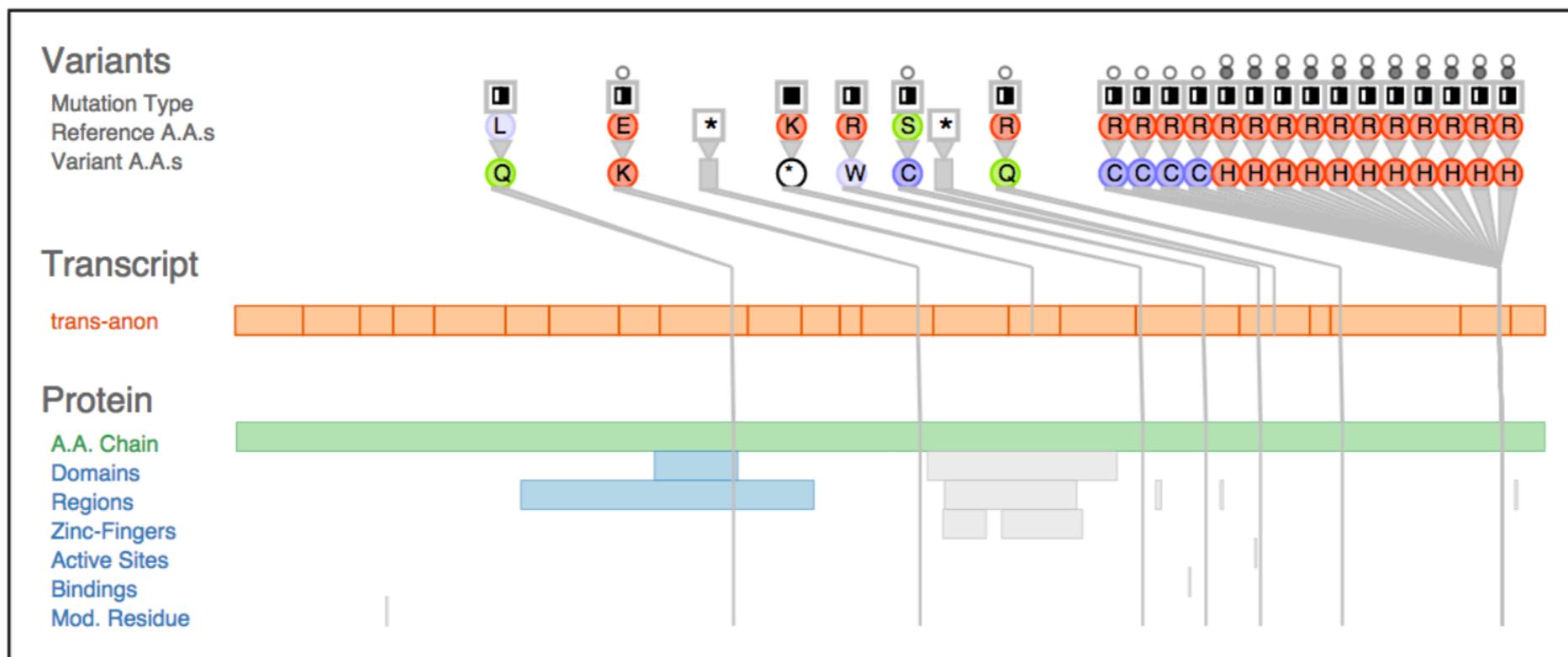
(a)



(b)



(c)

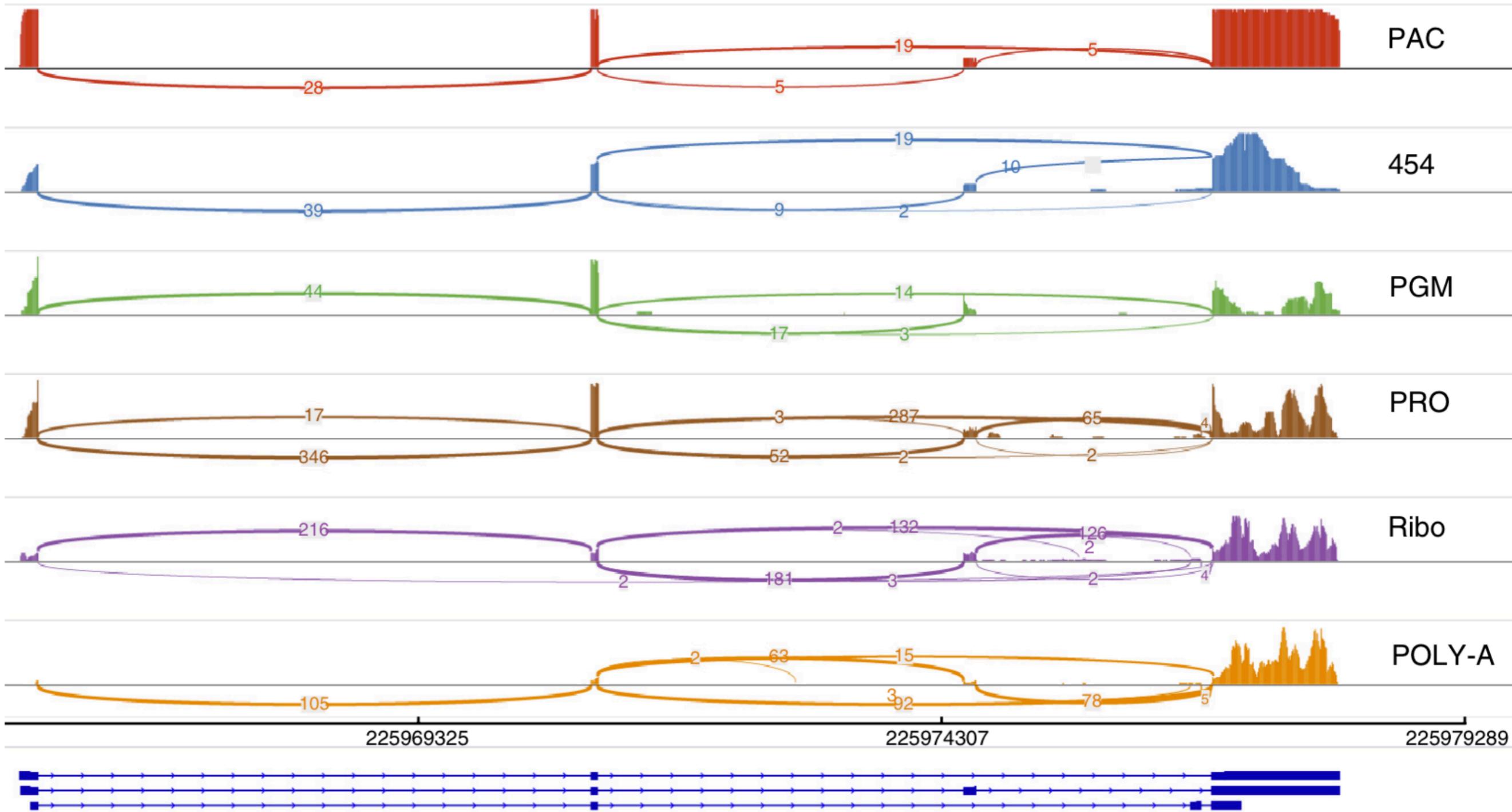


(d)

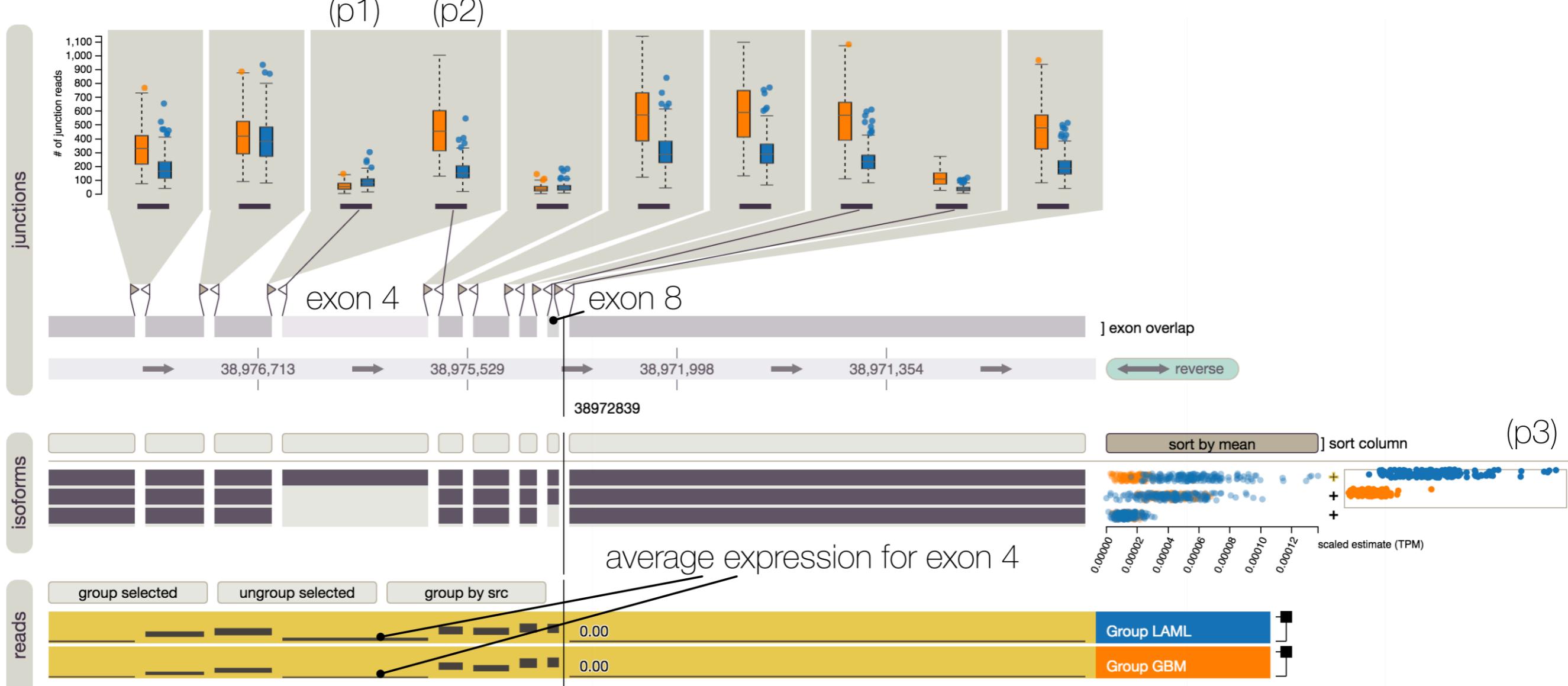
VariantView, Ferstay et al.

# Alternative Splicing

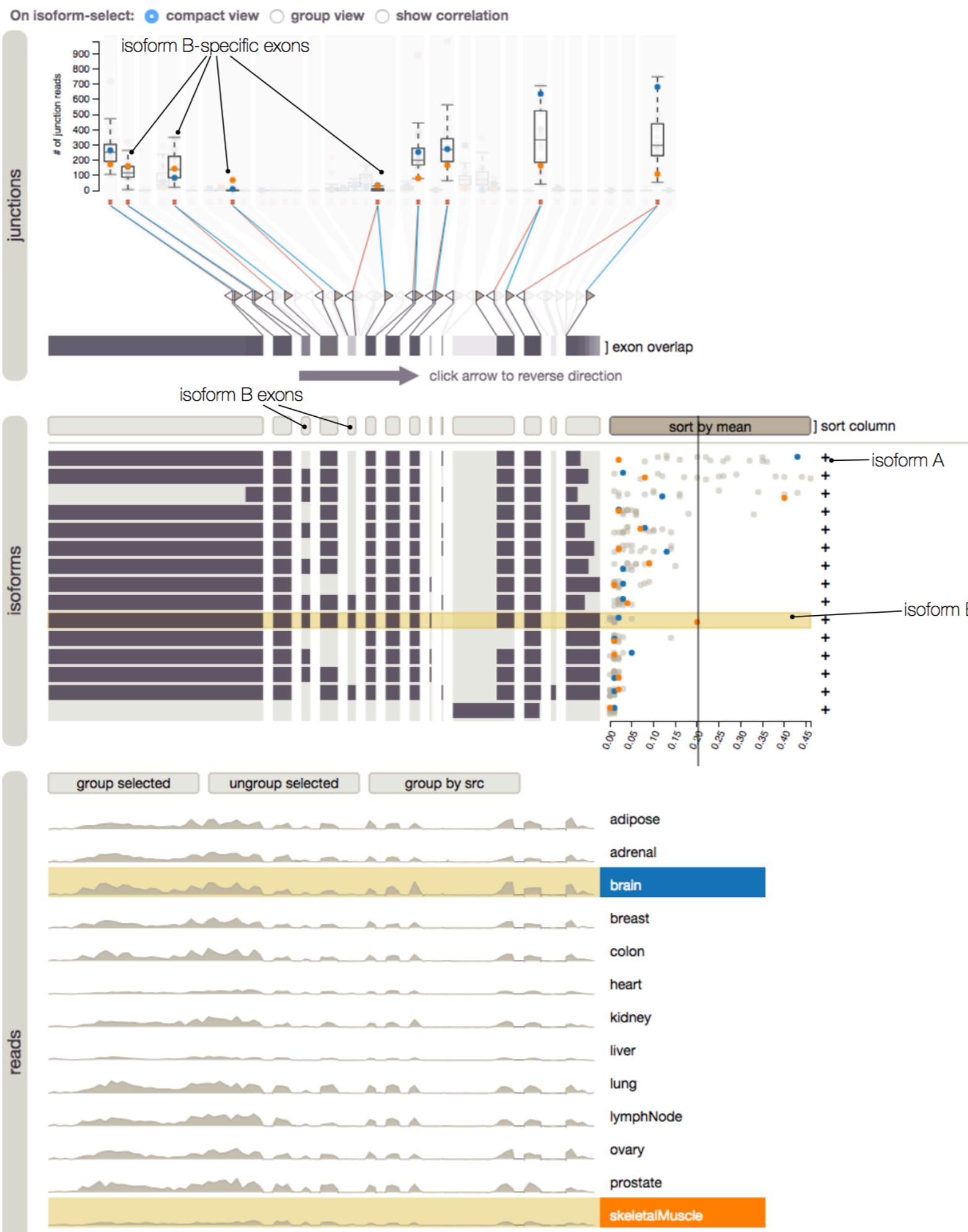
# Sashimi Plots



# Vials

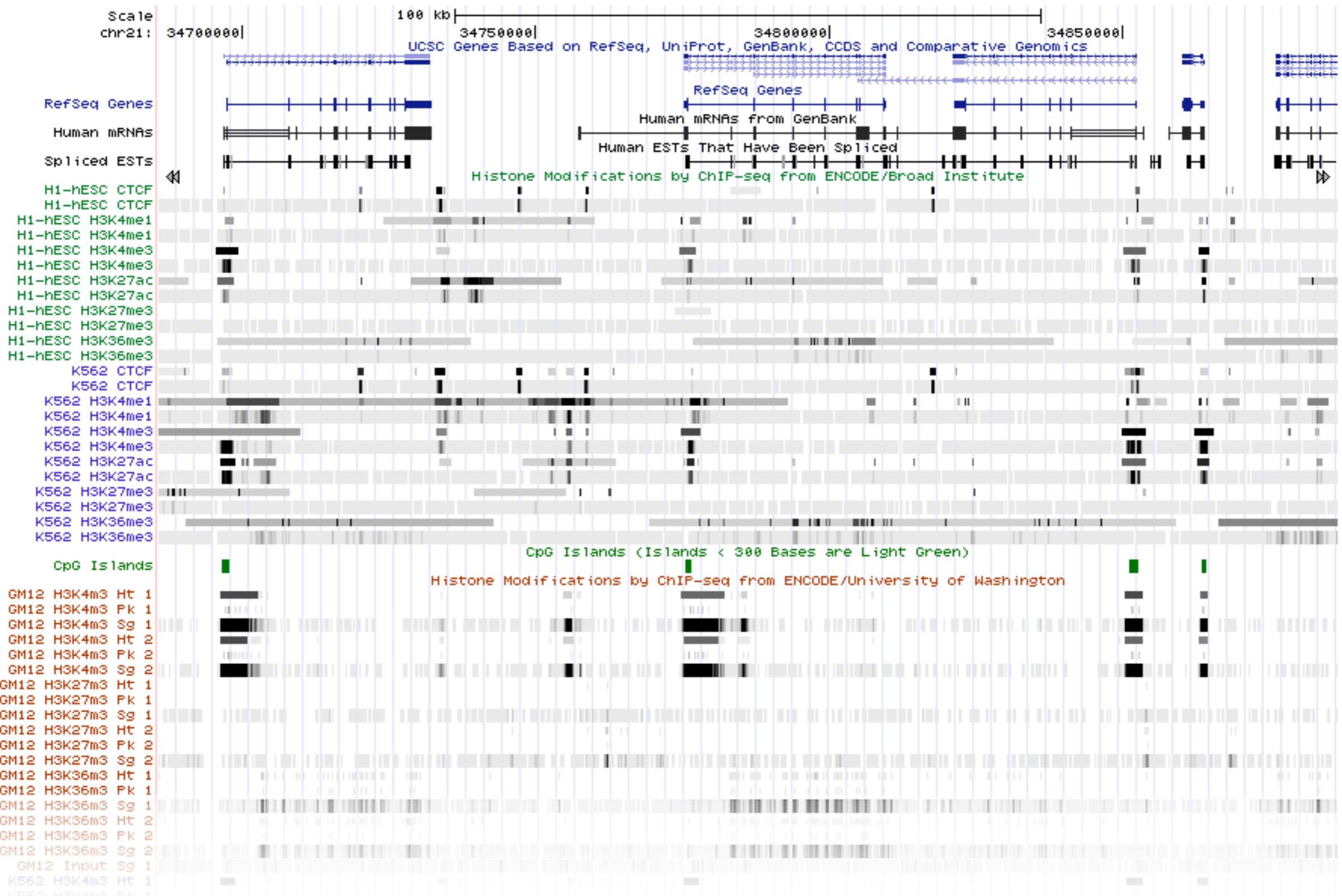


# Vials



# Genome Sequence

# Sequences: UCSC Genome Browser



# Sequences: UCSC Genome Browser

“squished”



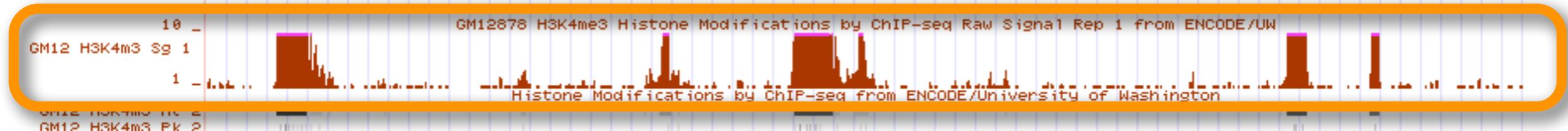
“dense”



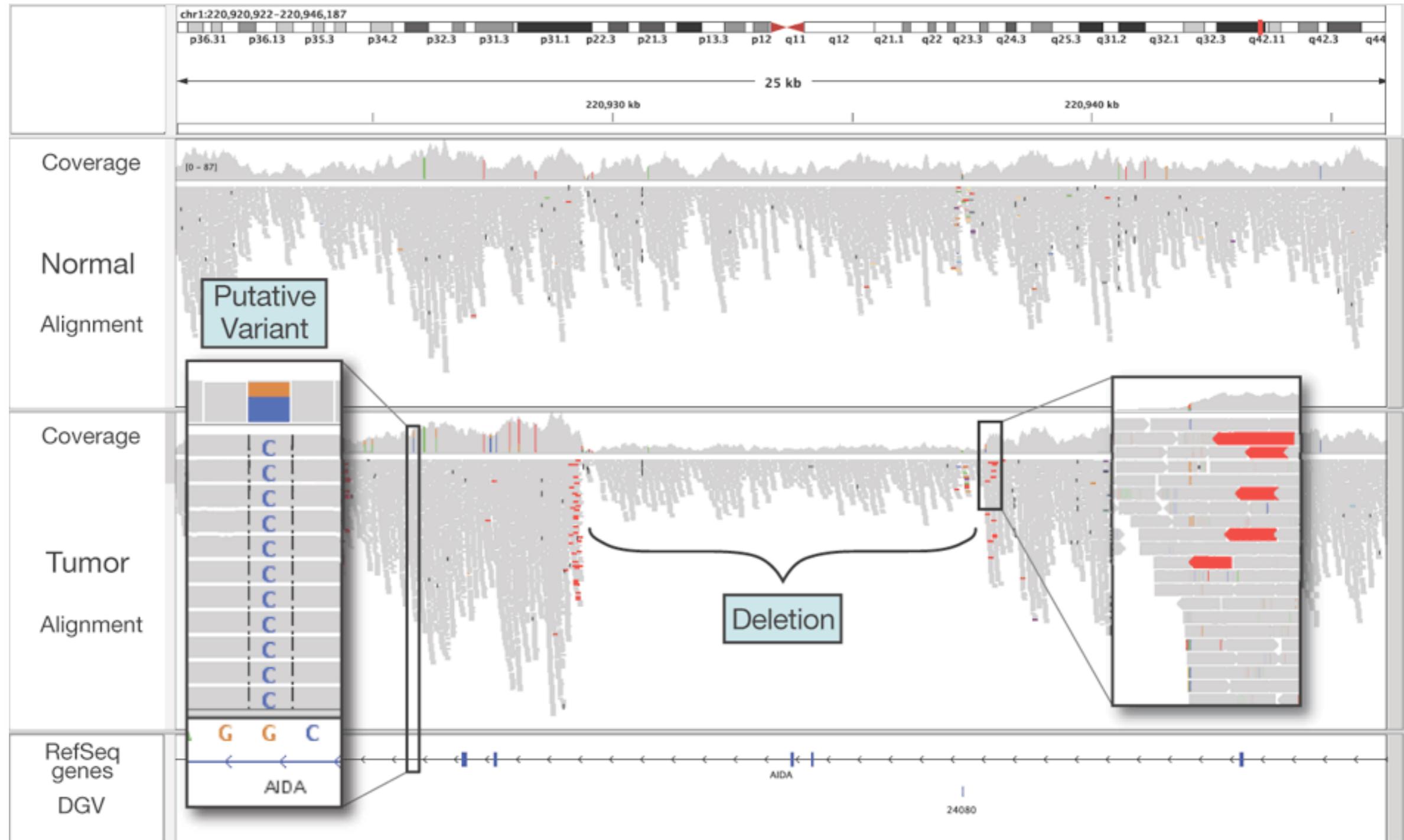
“packed”



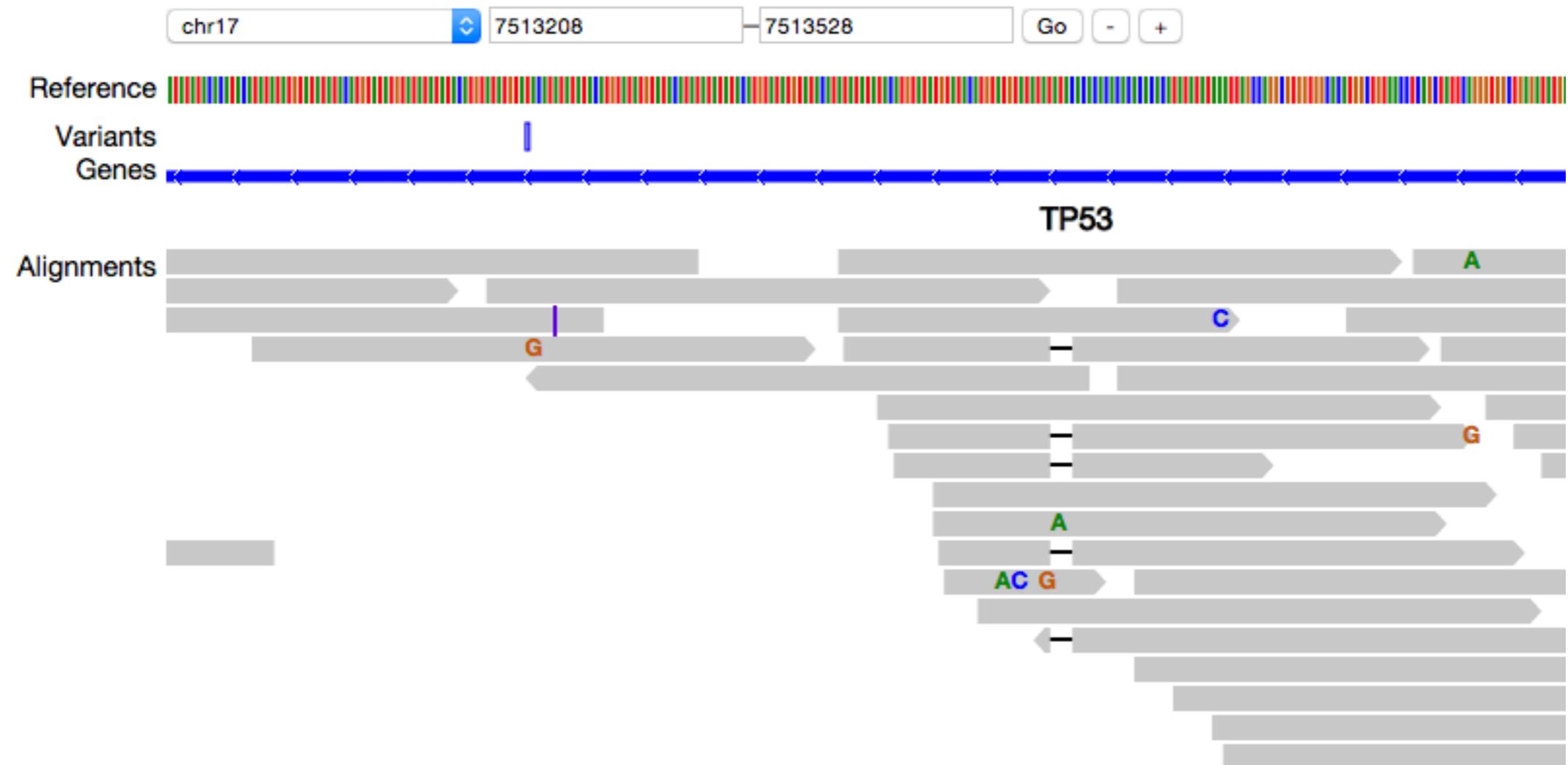
“full”



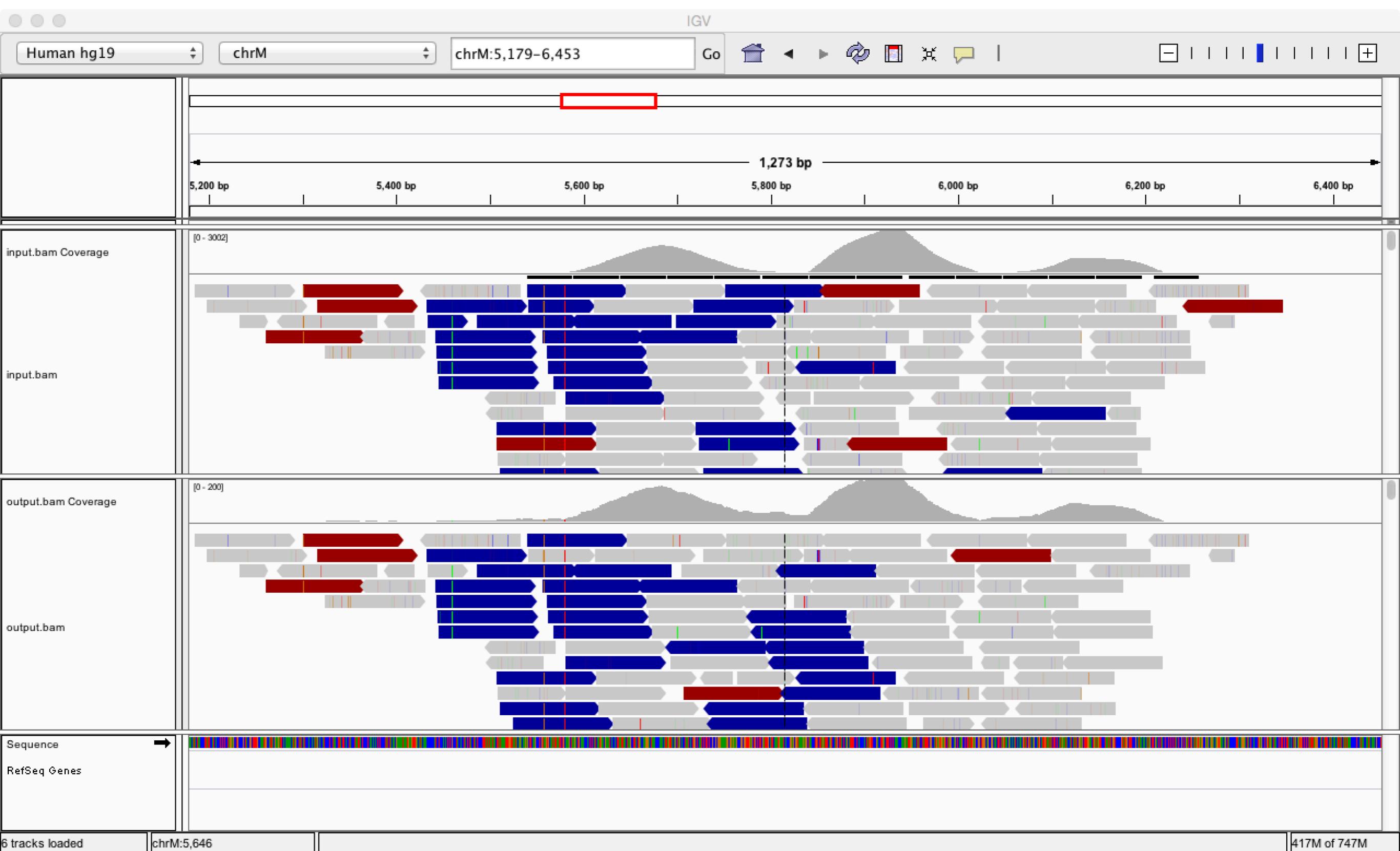
# Sequences: Integrative Genomics Viewer (IGV)



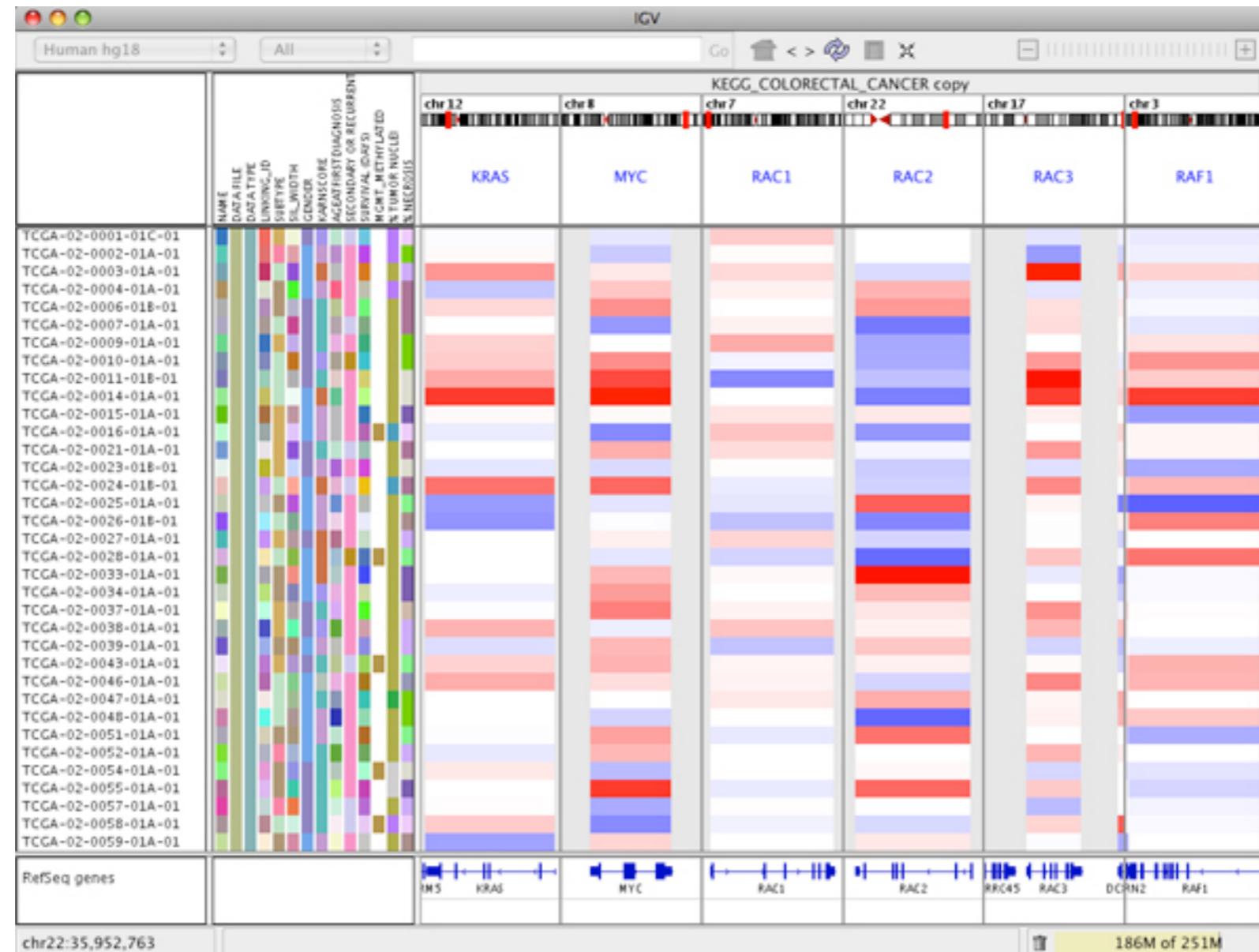
# PileUp Visualization



# PileUp Visualization

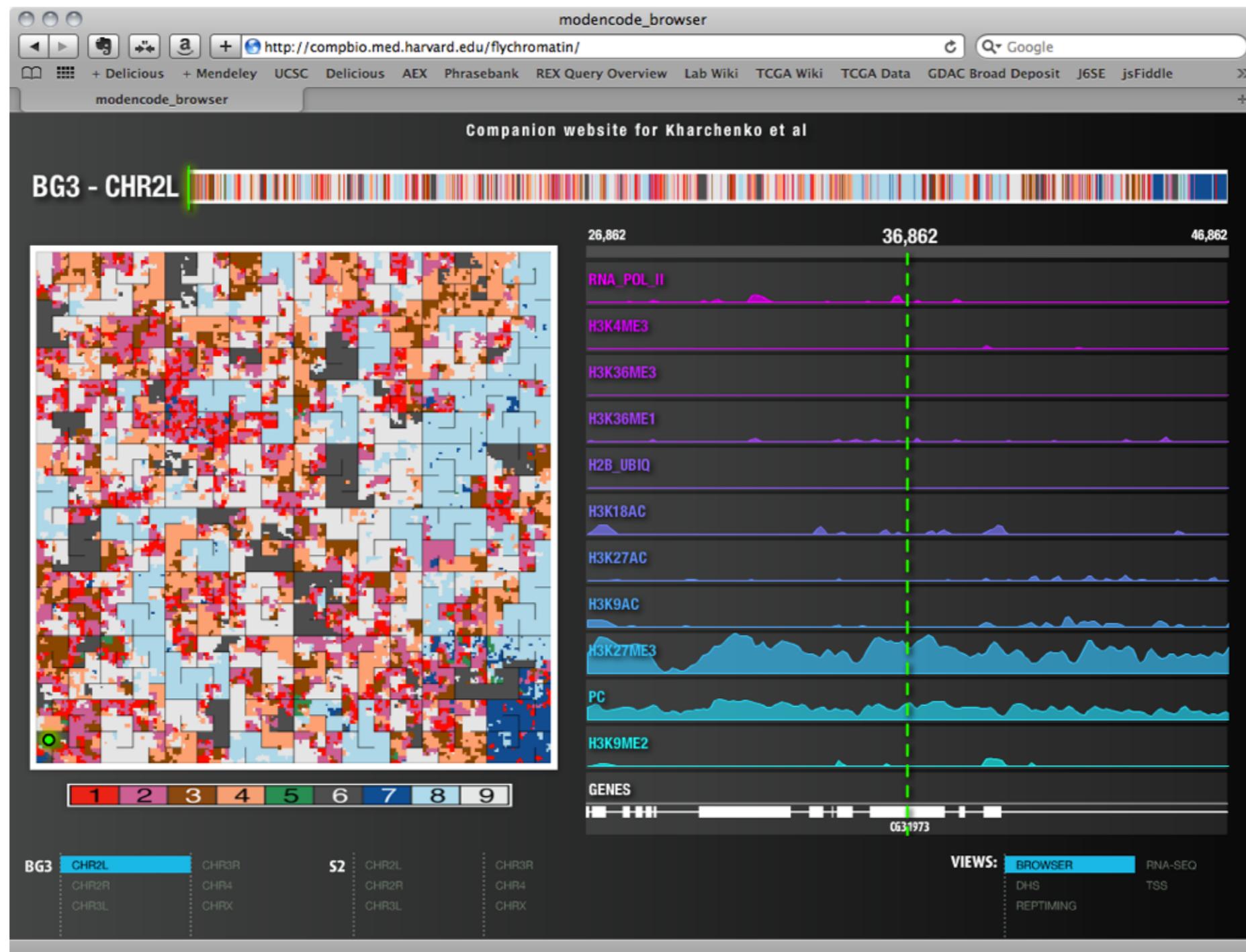


# Sequences: Integrative Genomics Viewer (IGV)



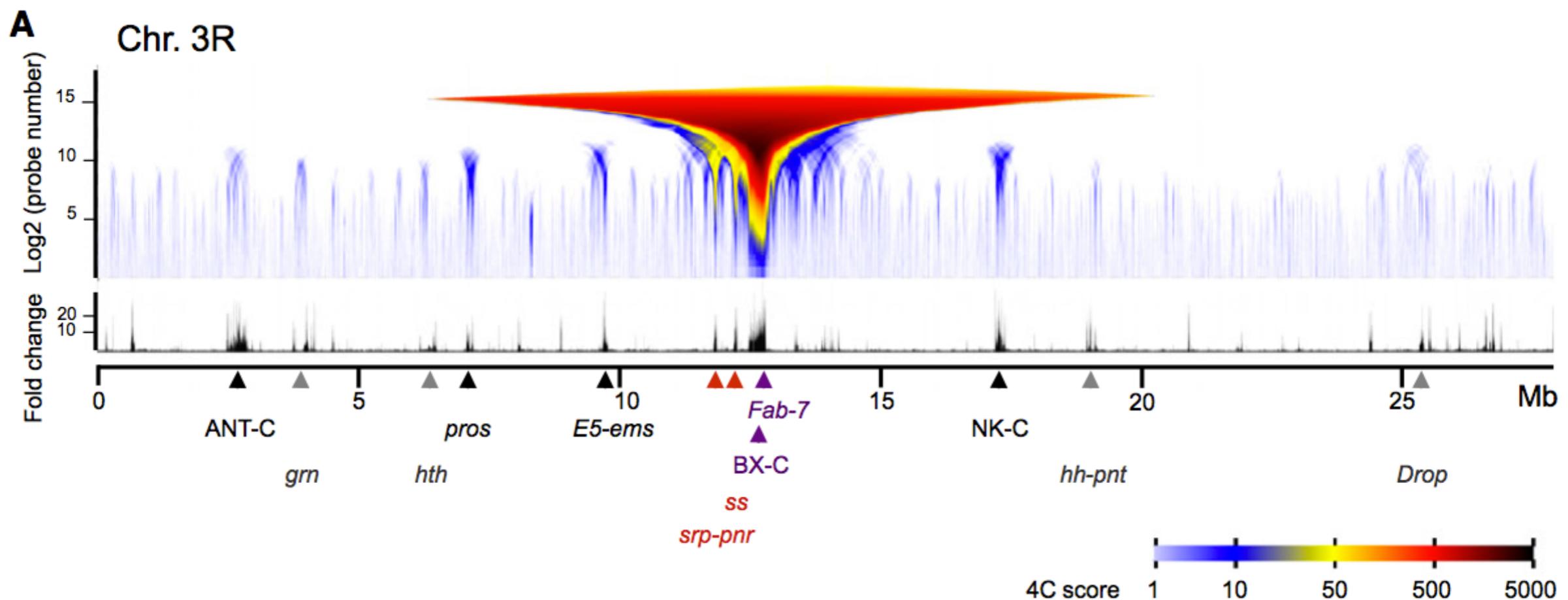
# Epigenomic Profiles

# Sequences: Epigenomics

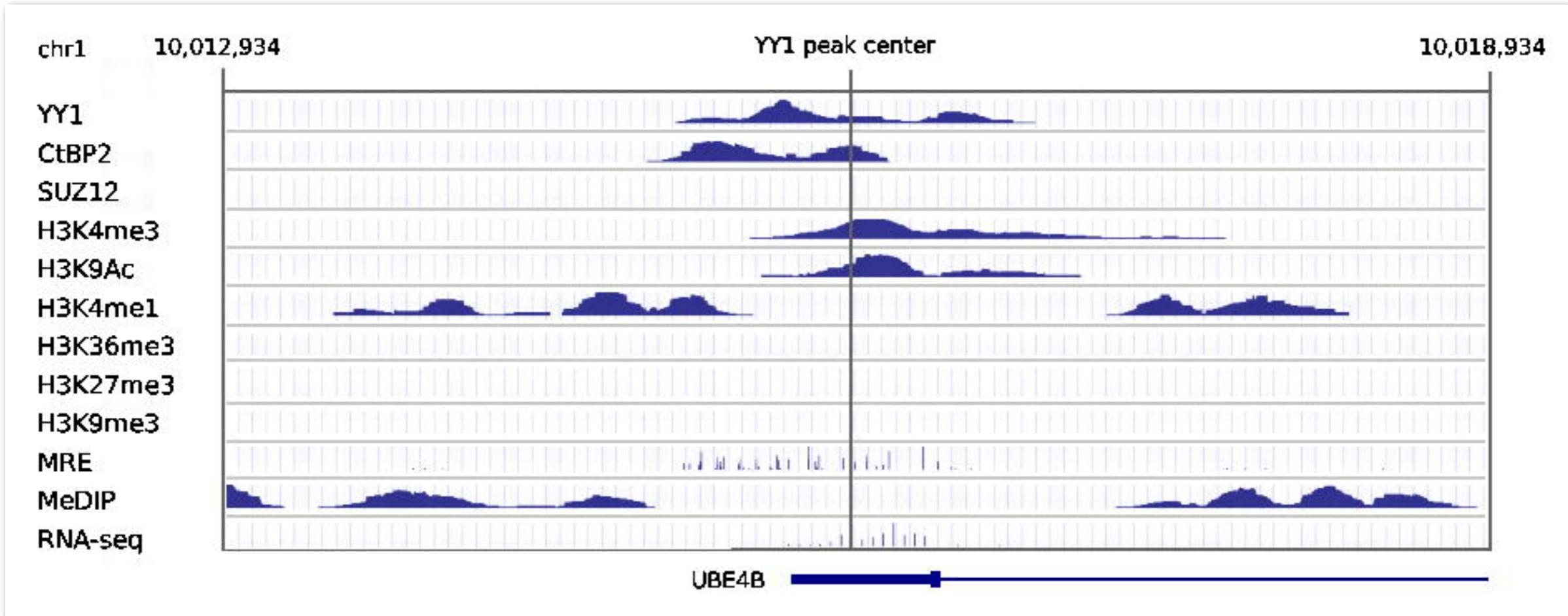


# Sequences: Epigenomics

Domainogram



# Sequences: Epigenomics



# Sequences: Epigenomics

## Spark Profile Summaries

