

Data Visualization for Biomedical Applications

Lecture 5

BMI706 - 19 April 2018

Presented by: Peter Kerpeljiev, PhD
Slides by: Nils Gehlenborg, PhD

Homework 5

- Start implementing visualizations for your data sets
- Starting point should be your sketches from HW3
- It is ok to add additional data sets at any point
 - discuss rationale, data set characteristics, and how you are integrating it with your existing data in the process book
- **Submit:** state of your repository including an updated process book

Review of Session 4

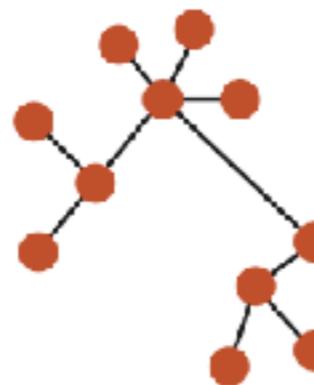
Arrange Networks and Trees

→ Node–Link Diagrams

Connection Marks

NETWORKS

TREES

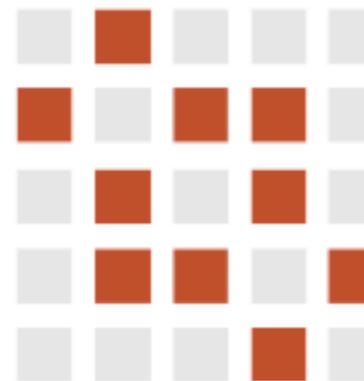


→ Adjacency Matrix

Derived Table

NETWORKS

TREES



→ Enclosure

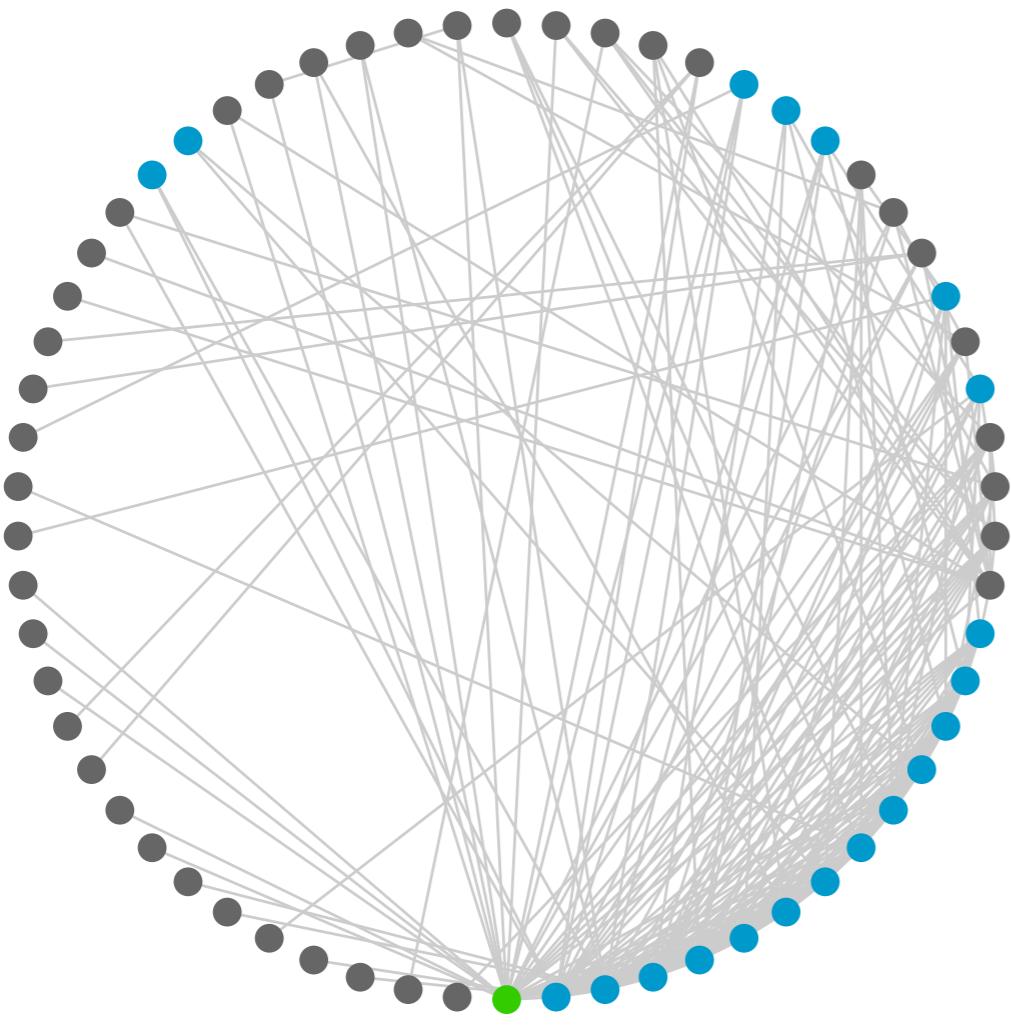
Containment Marks

NETWORKS

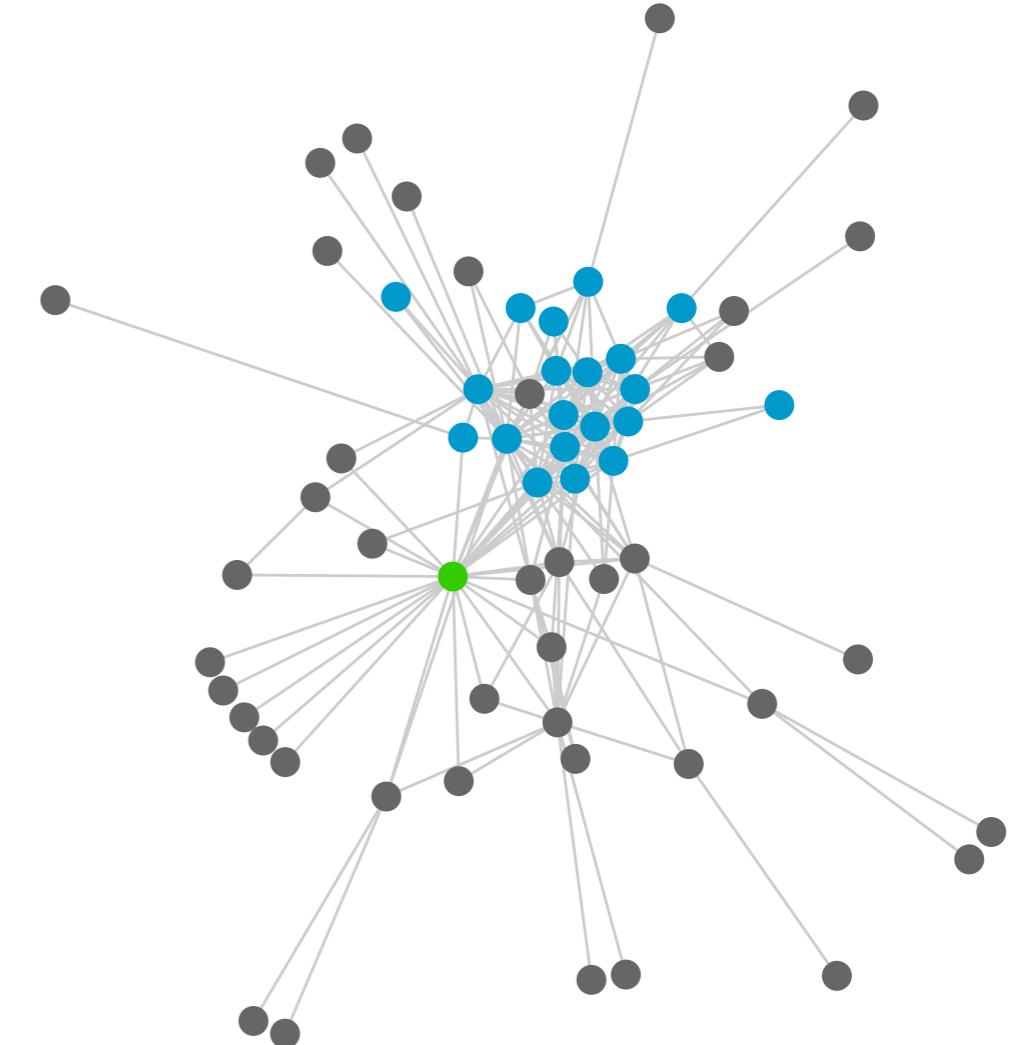
TREES



Graph Layouts

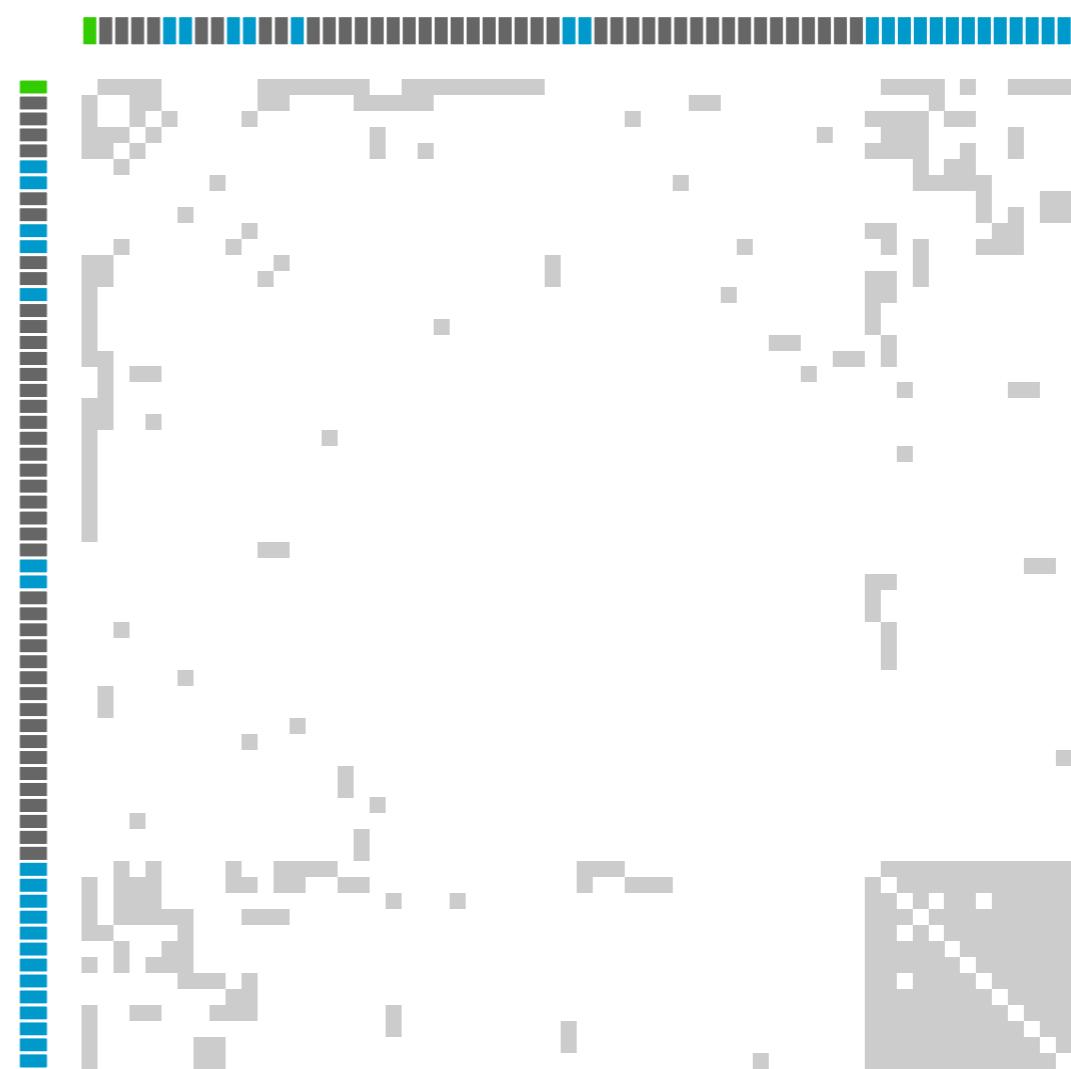
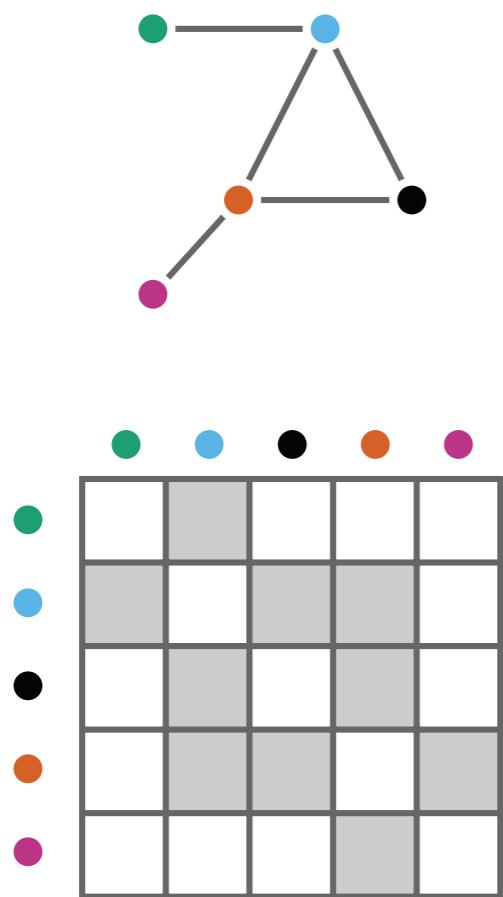


Circular Layout



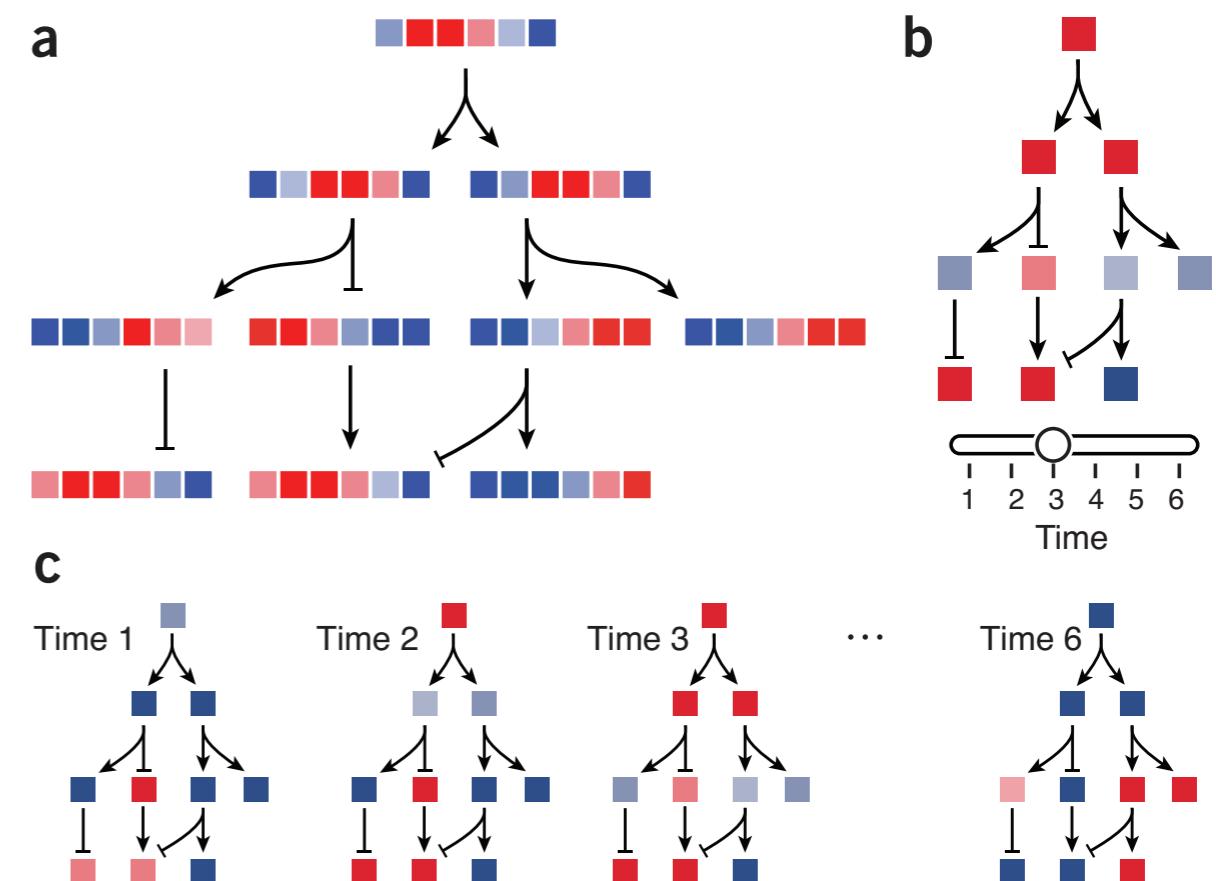
Force-directed Layout

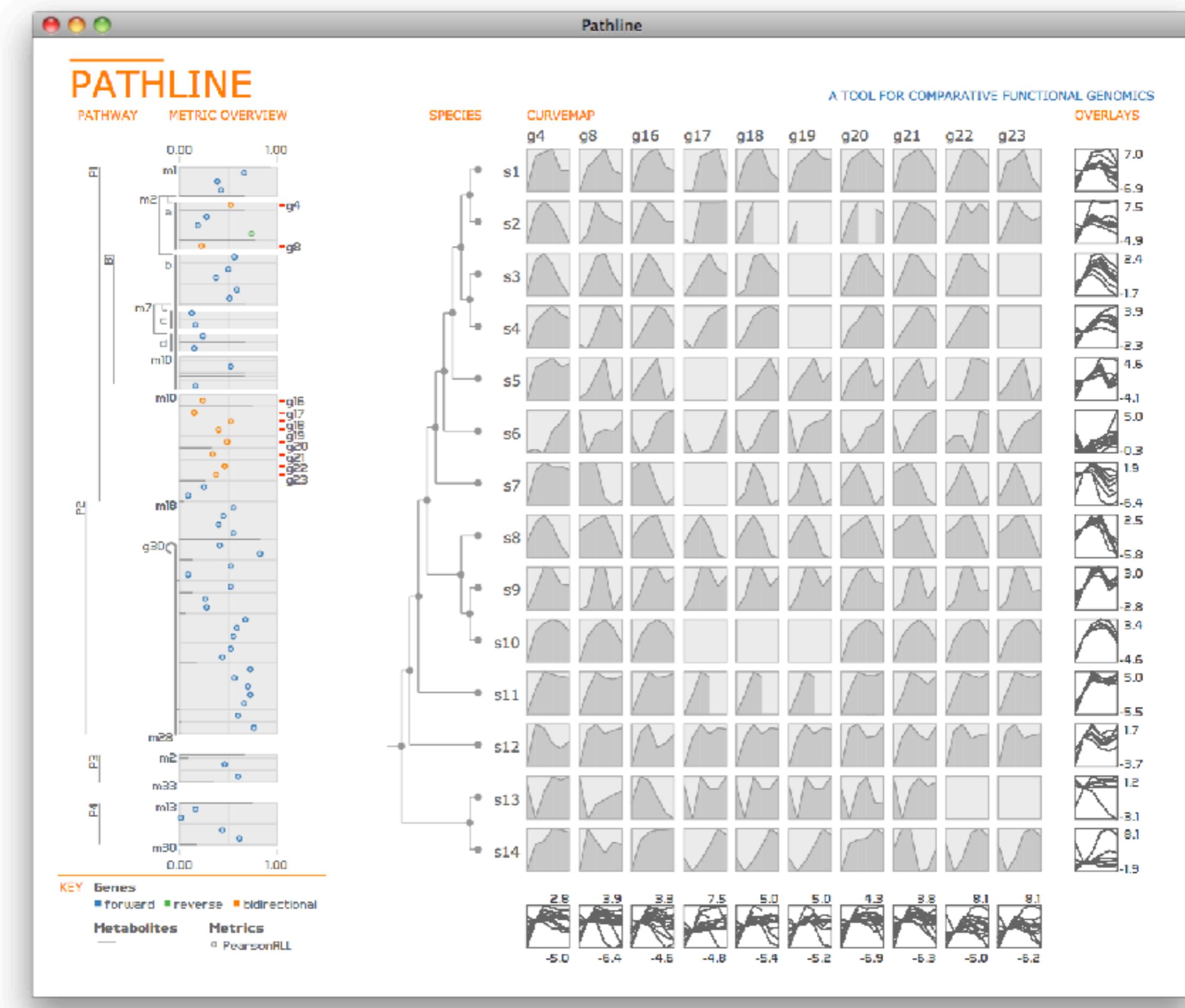
Adjacency Matrix



Interactions: And Multivariate Data!?

- (a) **Complex glyphs** = multiple values per node?
- (b) **Animation** = one value per node, one network shown at a time?
- (c) **Small multiples** = one value per node, all networks shown simultaneously?





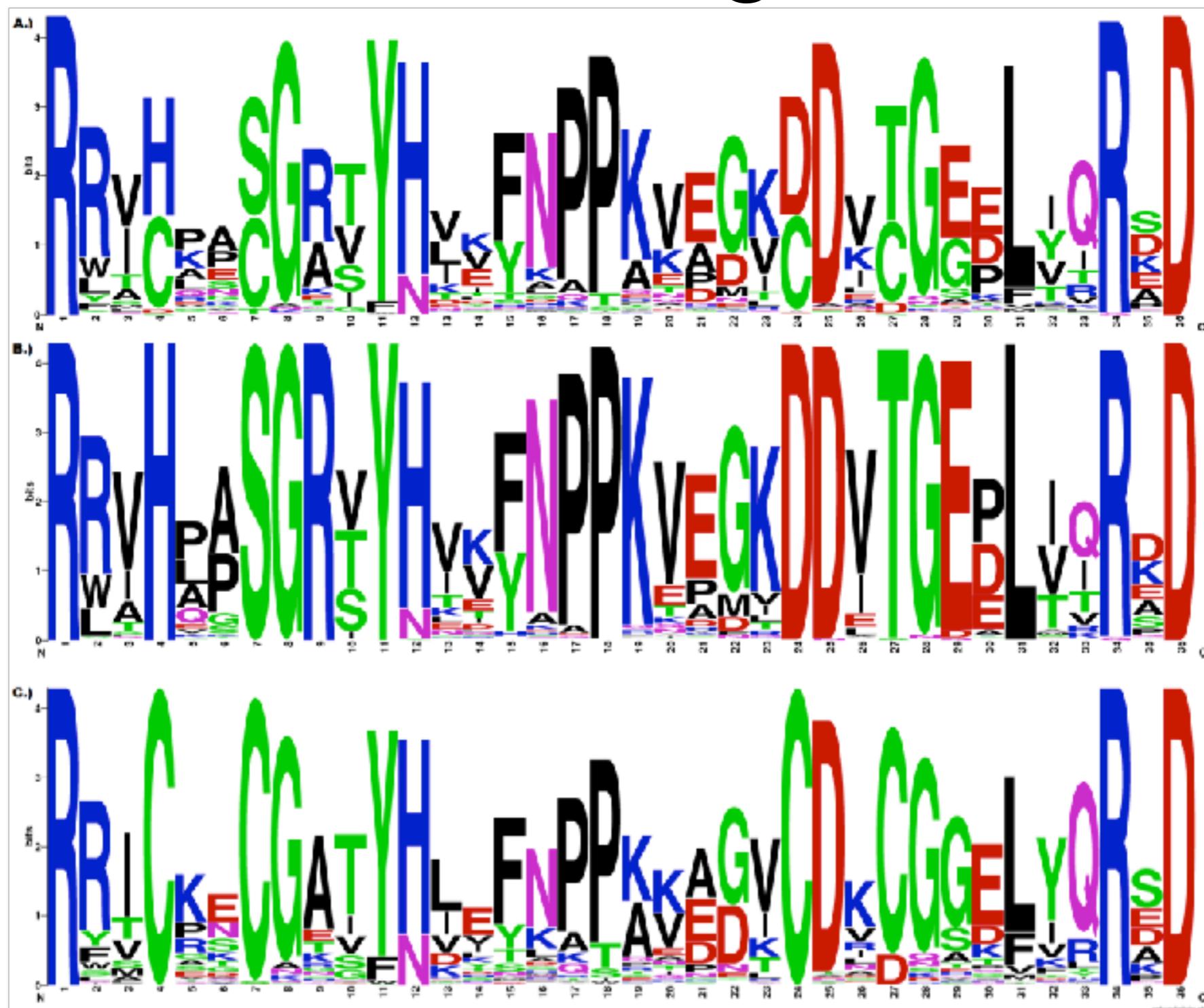
Lecture 5: Learning Goals

- What kind of genomic data do we need to visualize?
- What are common approaches for visualization of genomic data?
- What approaches exist to deal with the multi-scale nature of genomic data?

Genomic Data Types

- Genome Sequence: nucleotide sequence, mapped reads (“pile up”)
- Genomic Features (point, blocks, complex, values associated): gene model, binding site, SNPs, indels, inversion
- Continuous Signal: epigenomic profiles, copy number profiles
- Genomic Interactions: 3D interactions (Hi-C), structural variants
- Species Interactions: synteny

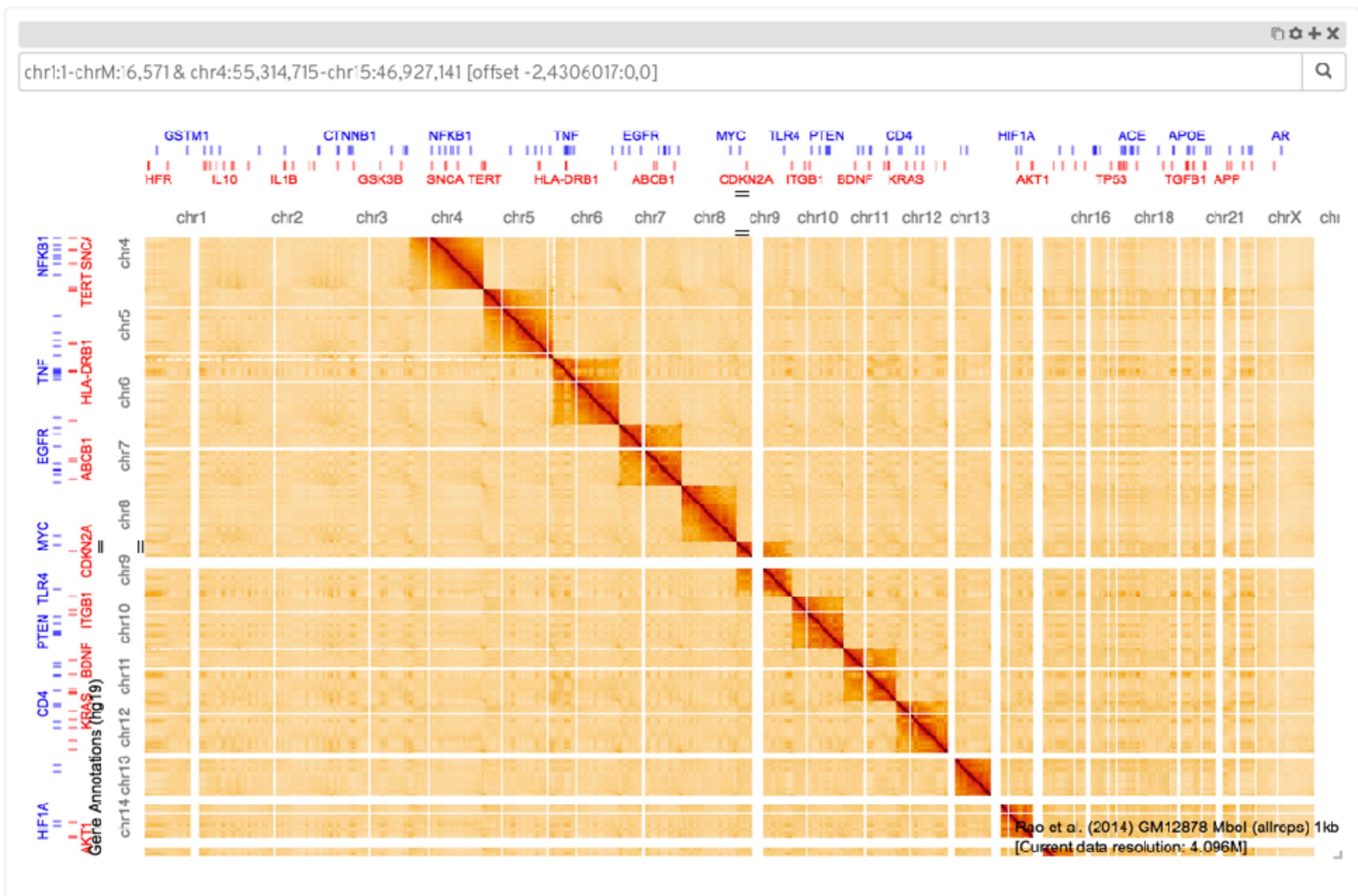
Redesign



Genome Interactions

3D Interactions

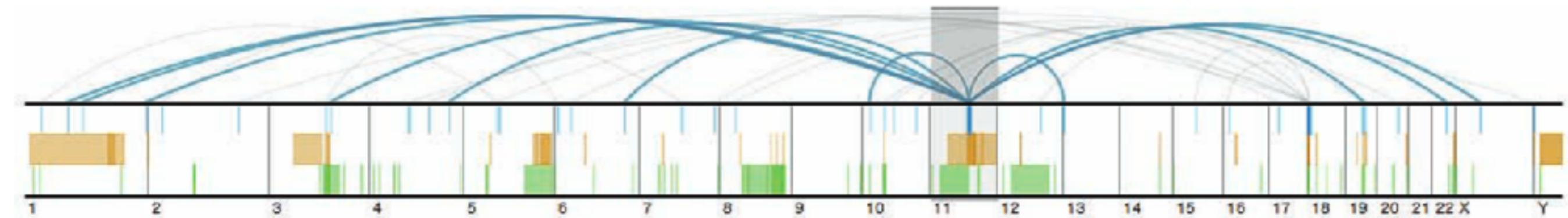
See Peter Kerpeljiev's Slides



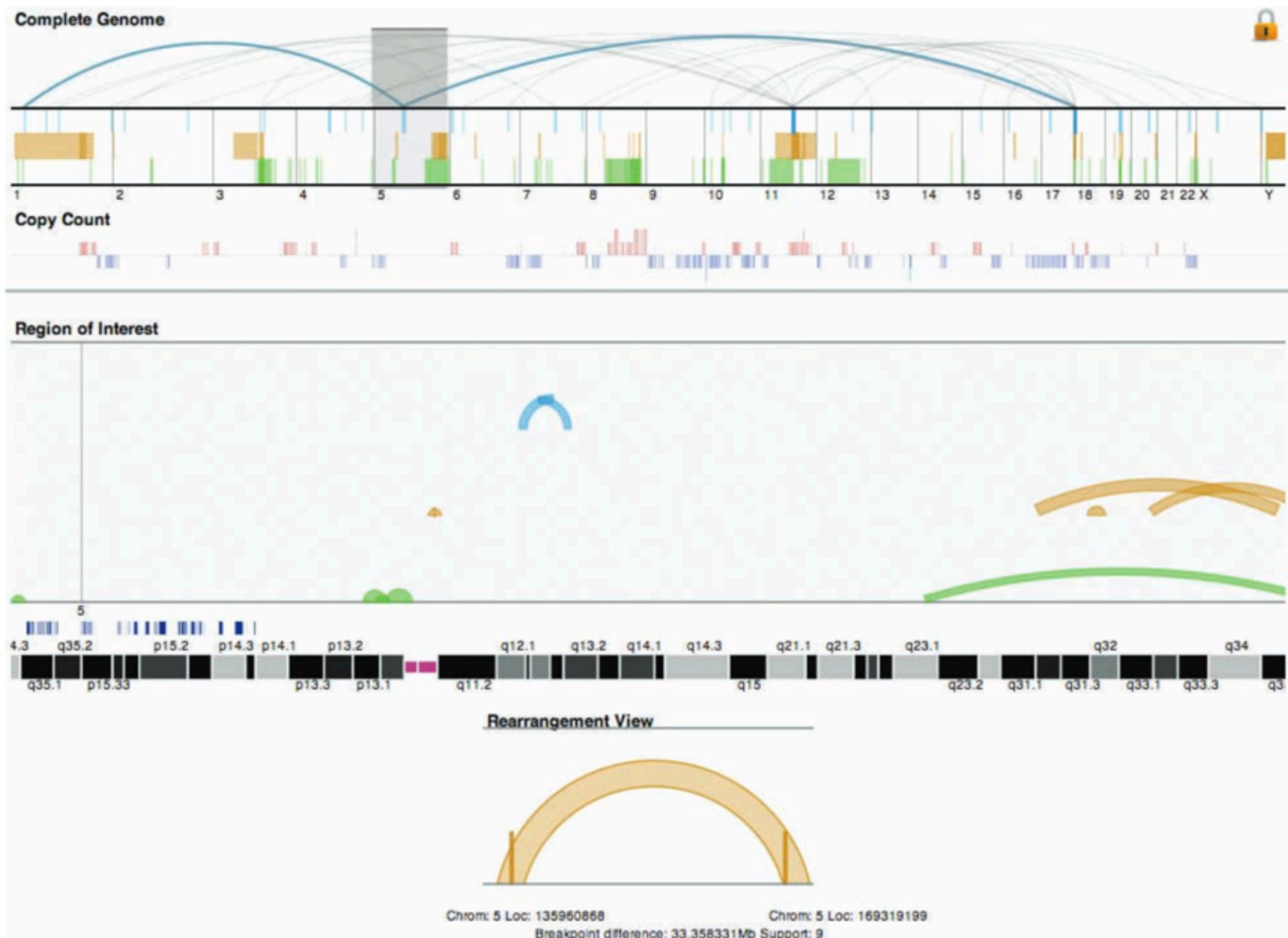
Genome Interactions

Structural Variants

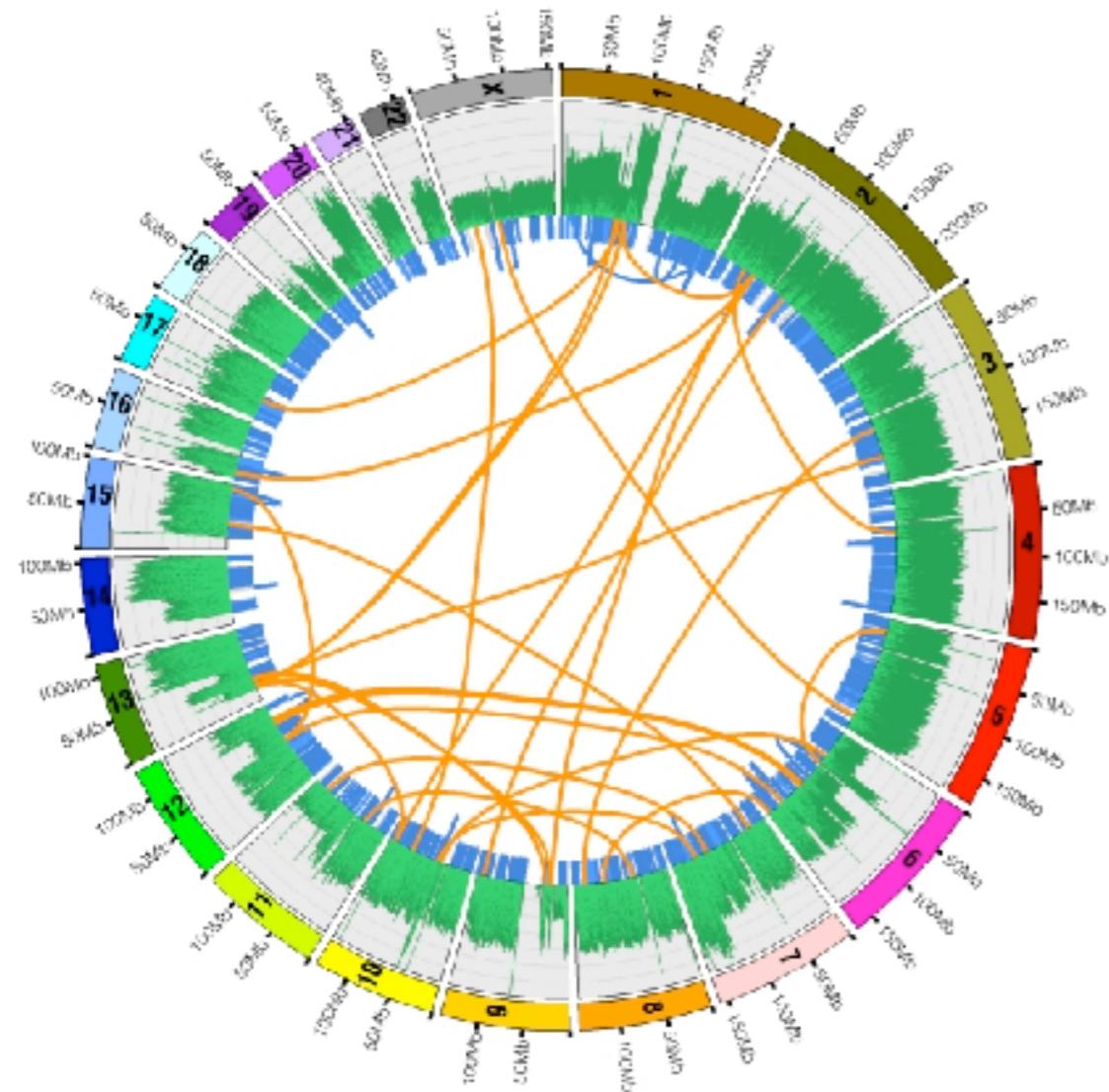
See Gremlin Paper Presentation



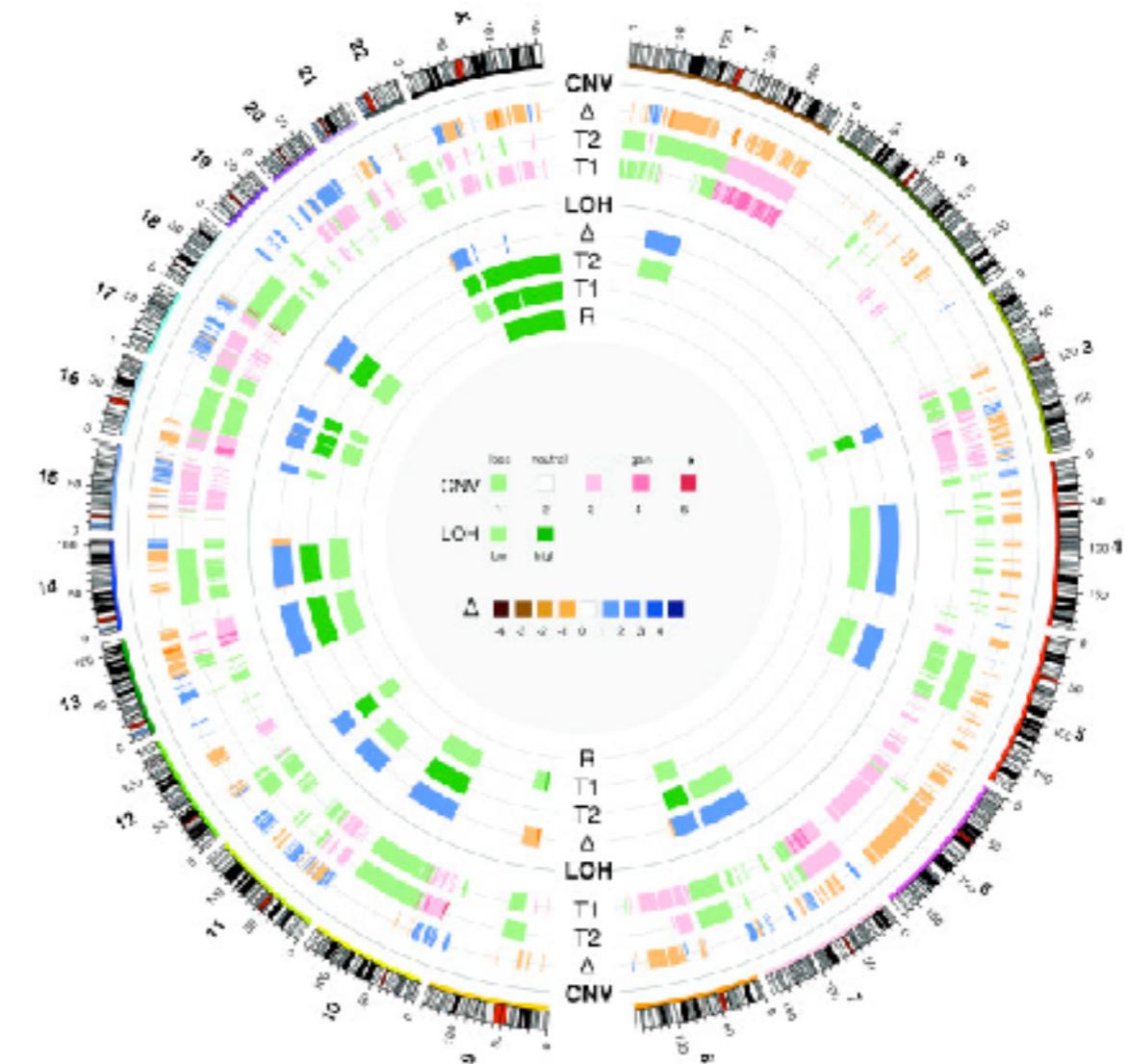
See Gremlin Paper Presentation



Sequences: Circos

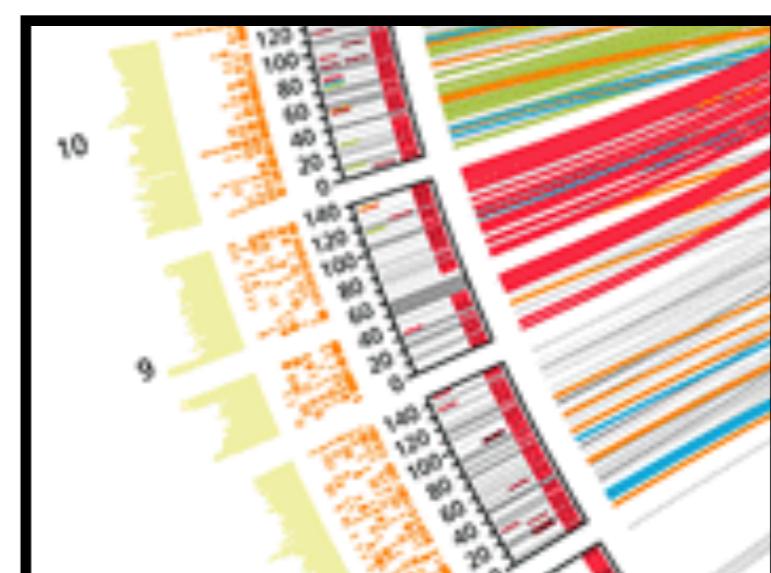
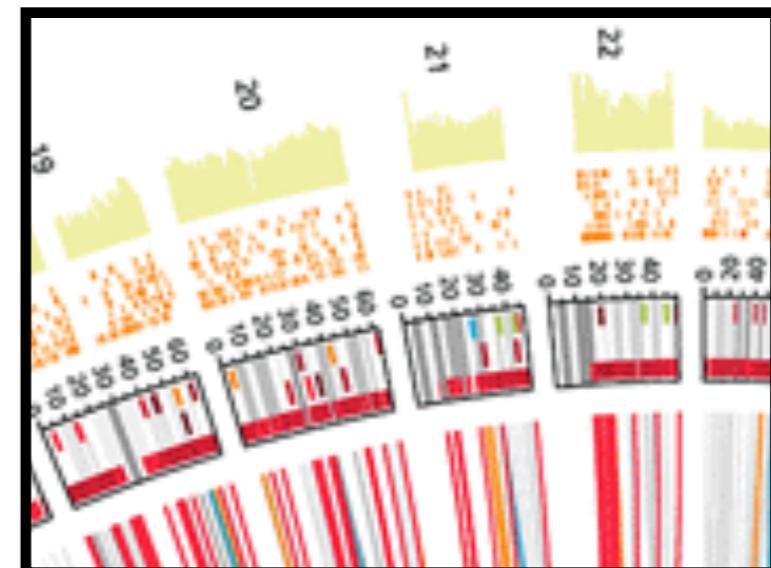


Clark et al. 2009, PLoS Genetics



Jones et al. 2010, Genome Biology

Sequences: Circos



HUMAN GENOME (hg18)



HUMAN GENOME (hg18)

CLINICAL ATTRIBUTES

 Reset View.

SAMPLE

GENDER

 Male Female

TUMOR STAGE

 I IIa IIb IIc
 IIIa IIIb IVa IVb

AGE

 30-44 45-69
 60-74 75-89

HISTOLOGY

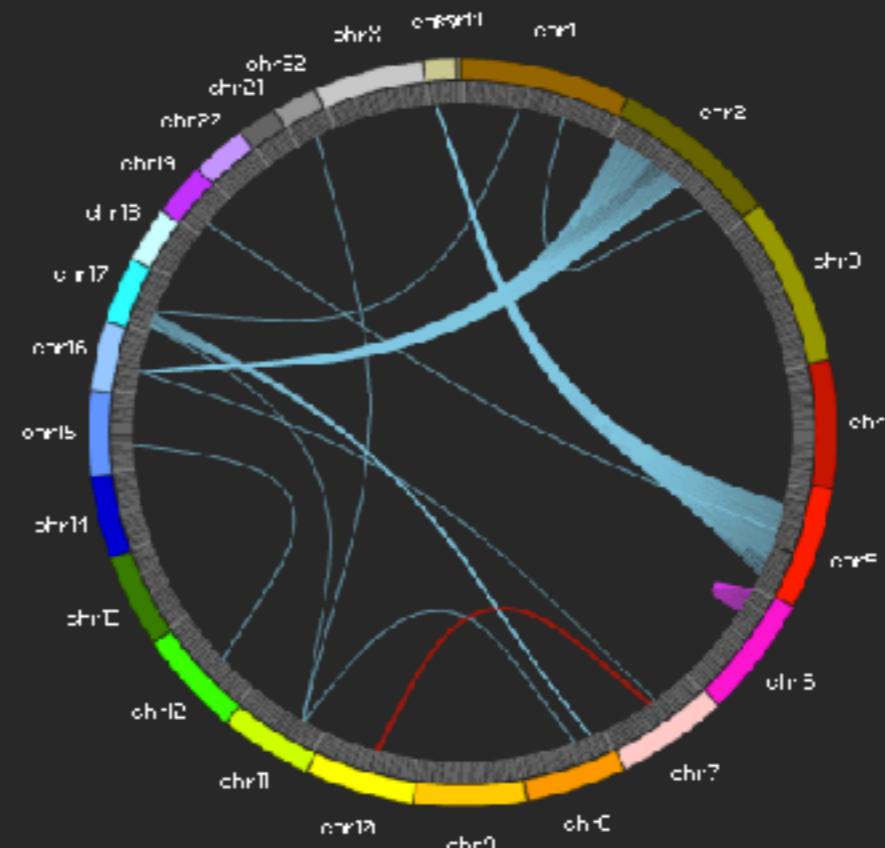
 Color PC

Colon Mucinous AD

Rectal HU

Rectal Mucinous AD

FPF-ITX



ID	TCGAID	TYPE	PAIRS_SUPPT	TUMOR_COUNT	NORMAL_COUNT	COPY_RATIO	DEST_CHROM	DEST_BRK	ORG_CHROM	START	END	SIZE	PVALUE_DEPT
56995	TCGA-AB-2901	ITX	ITX/3ITX	73	68	1.16	chr7	81627174	chr10	80572278	80572622	645	0.181
56997	TCGA-AB-2901	ITX	ITX/3ITX	33	48	0.76	chr17	10079441	chr11	37385505	37385892	388	0.911
57000	TCGA-HG-2901	ITX	ITX/2ITX	11	35	0.35	chr12	55276251	chr6	37281819	37282091	273	1
57001	TCGA-AB-2901	ITX	ITX/3ITX	1638	1824	1.11	chr2	132735667	chr16	33666389	33667166	798	0.001
57000	TCGA-AB-2901	ITX	ITX/2ITX	1015027	1445012	0.77	chr0	30205102	chr17	7100400	20020250	10520700	1
57007	TCGA-AB-2901	ITX	ITX/3ITX	120	140	0.00	chr1	00745410	chr17	20007700	20000320	120	0.74
57009	TCGA-AB-2901	ITX	ITX/2ITX	46	57	0.89	chr2	150707698	chr1	162506924	162507345	422	0.755
57012	TCGA-AB-2901	ITX	ITX/3ITX	49	52	1.04	chr11	43407788	chr22	27395208	27398017	410	0.494
57013	TCGA-AB-2901	ITX	ITX/3ITX	5200287	9849329	1.03	chr15	33871145	chr2	9508546	132729533	123120968	0
57020	TCGA-HG-2901	ITX	2ITX/ITX	6899616	9640101	1.02	chrY	10630290	chr5	92406120	153009355	120503376	0
57018	TCGA-HG-2901	ITX	ITX/4ITX	1058	1148	1.02	chr19	23976736	chr5	7182353	7182777	395	0.359

ILM04 GENOME (Chr10)

CLINICAL ATTRIBUTES

Race: White

SAMPLE

SEX

Male Female

TUMOR STAGE

I IIA IIB IIC
IIA IIIB IIIA IIIIB

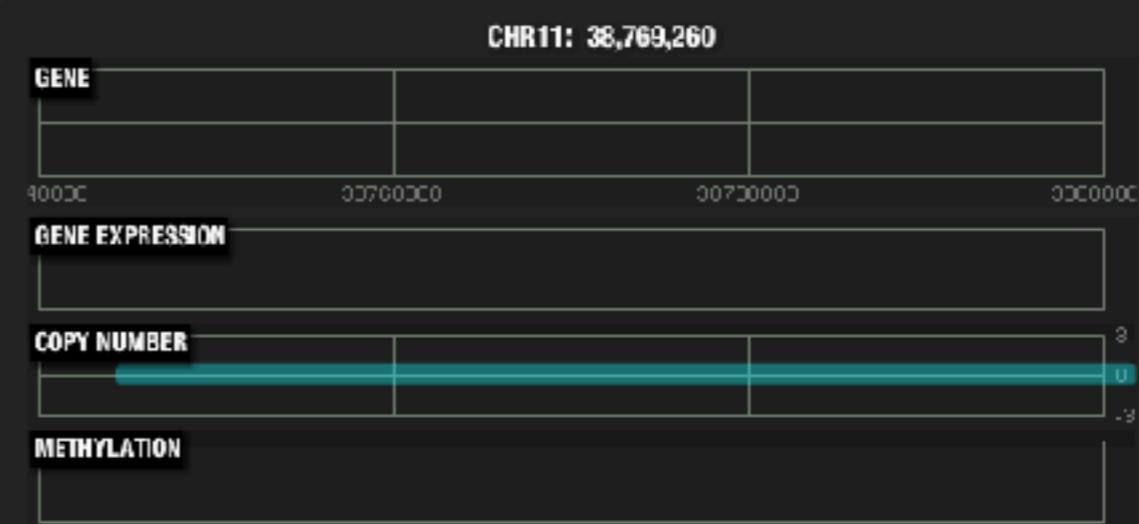
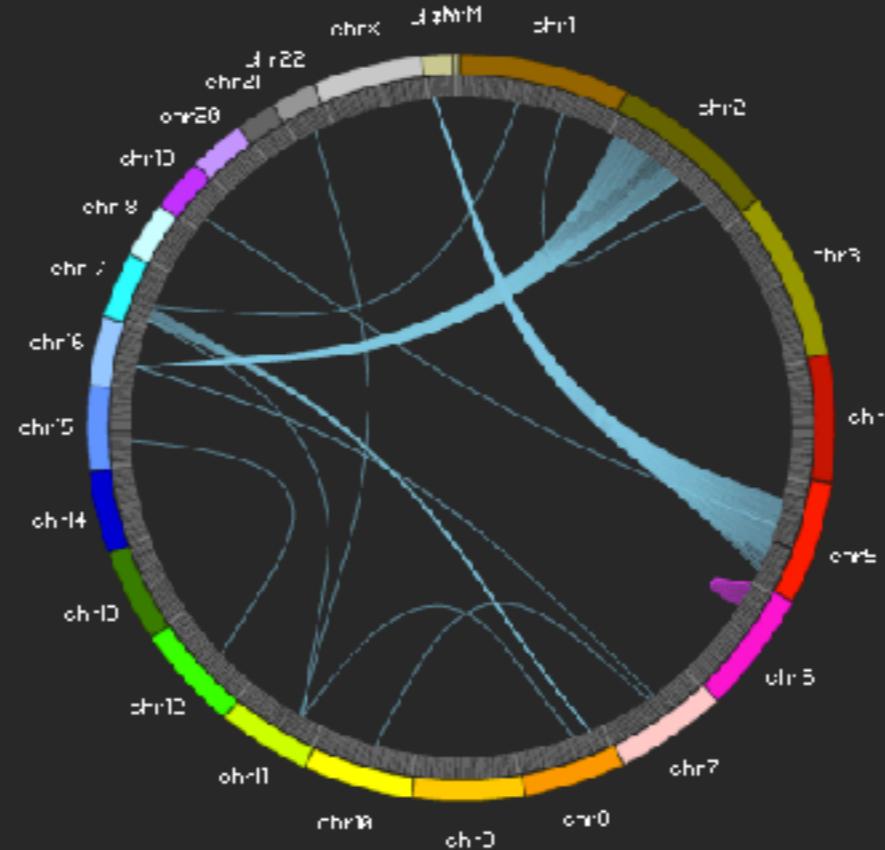
TCC

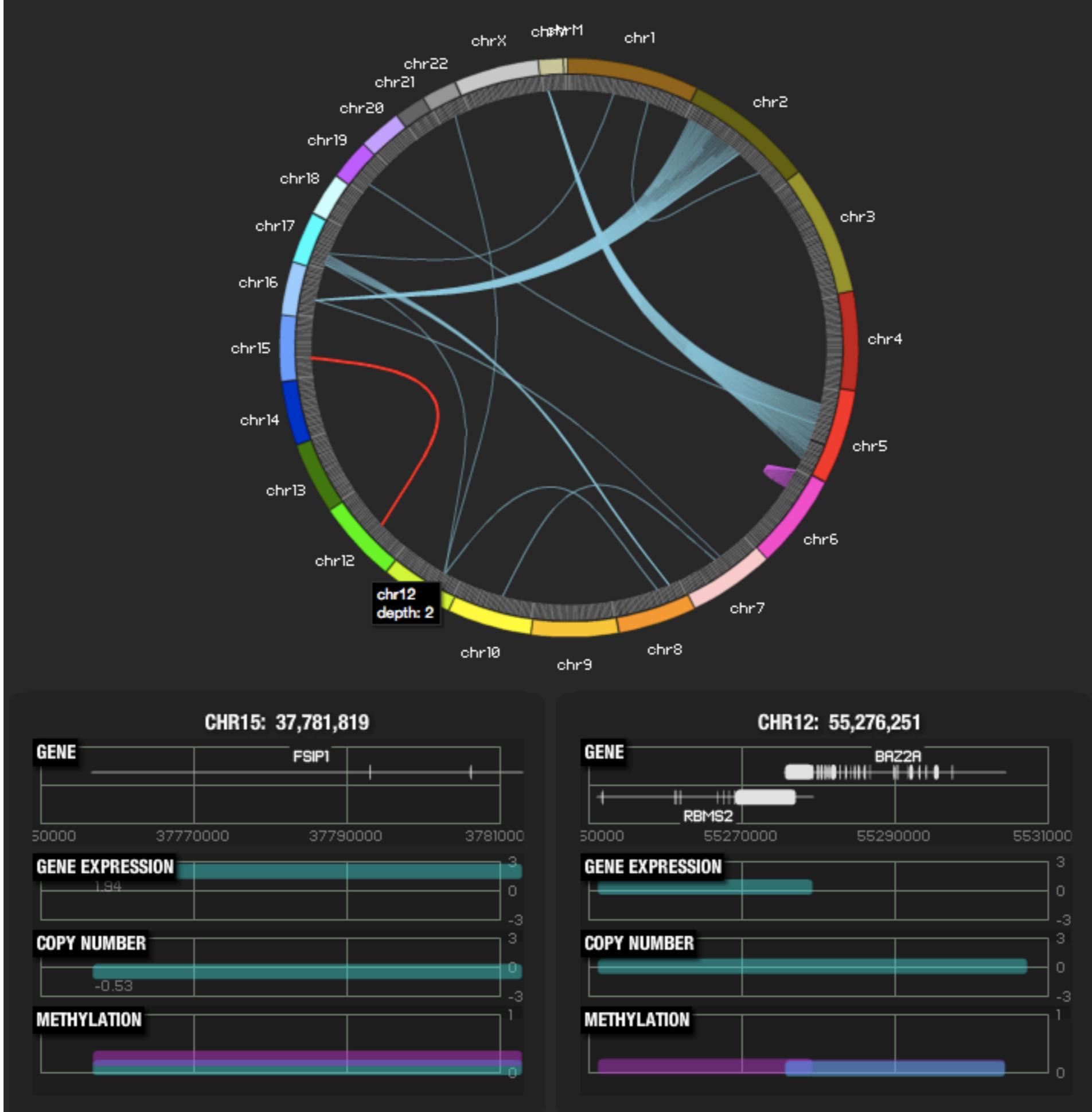
30-49 45-59
60-74 75-80

HISTOLOGY

Cancer PD
Colon Mucinous AD
Rectal AD
Vaginal Mucinous HJ

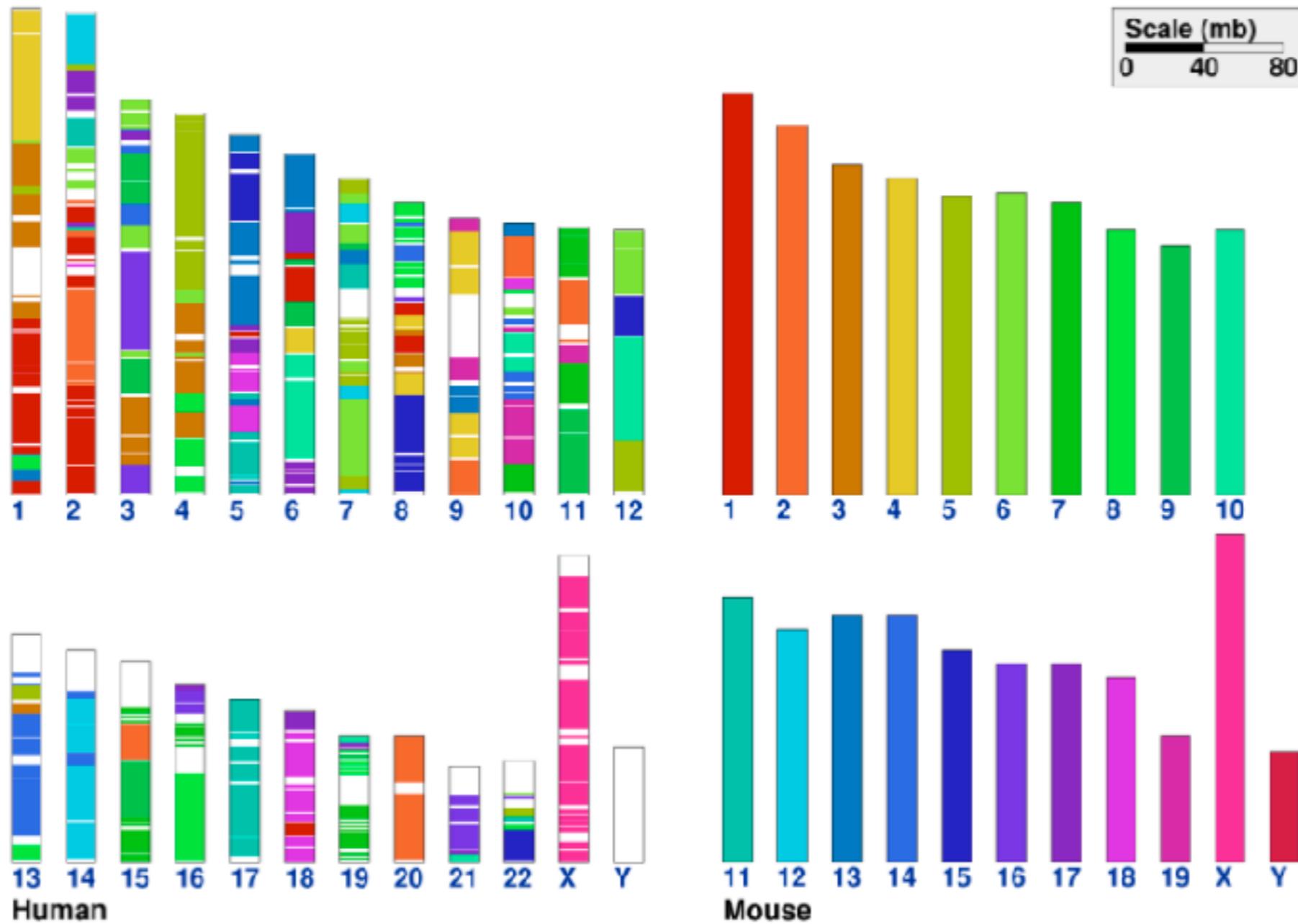
FPS.50



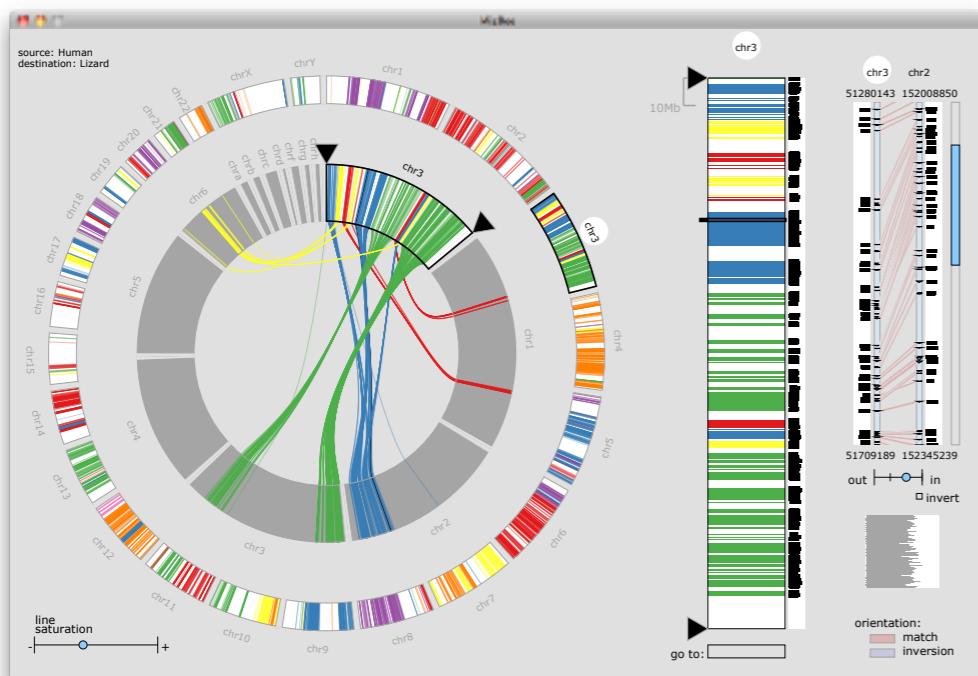


Synteny

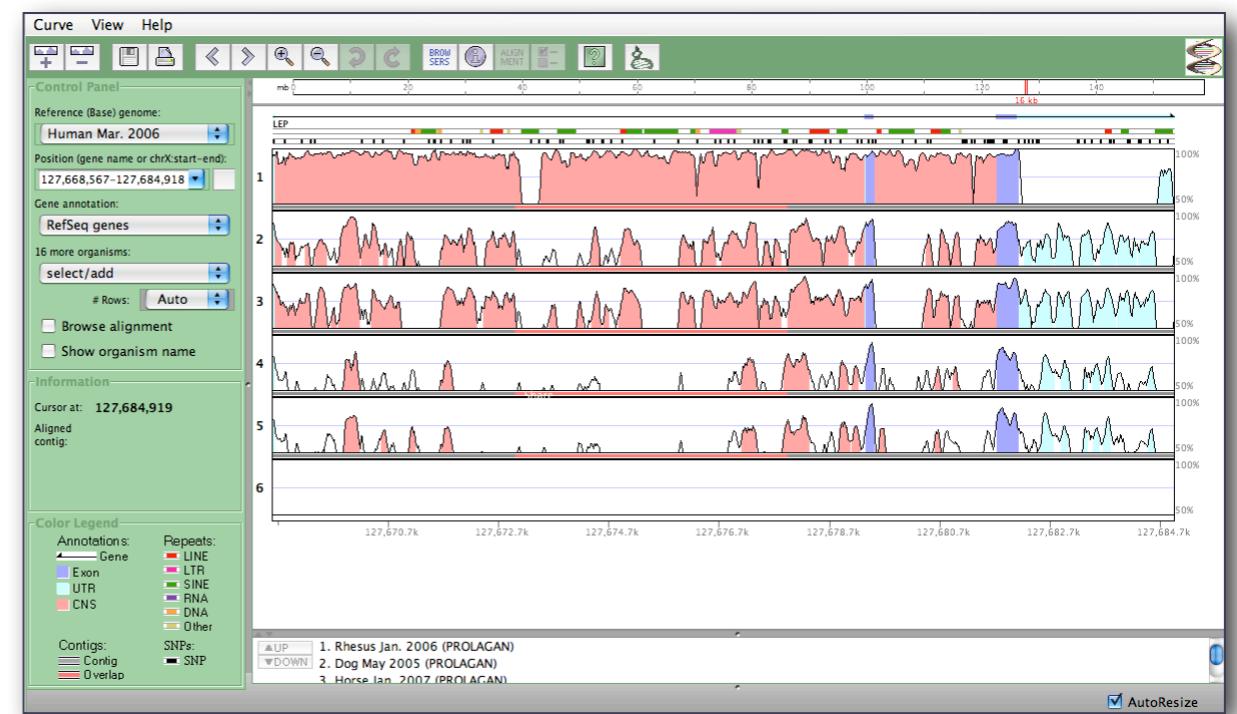
Sequences: Comparative Genomics



Sequences: Comparative Genomics

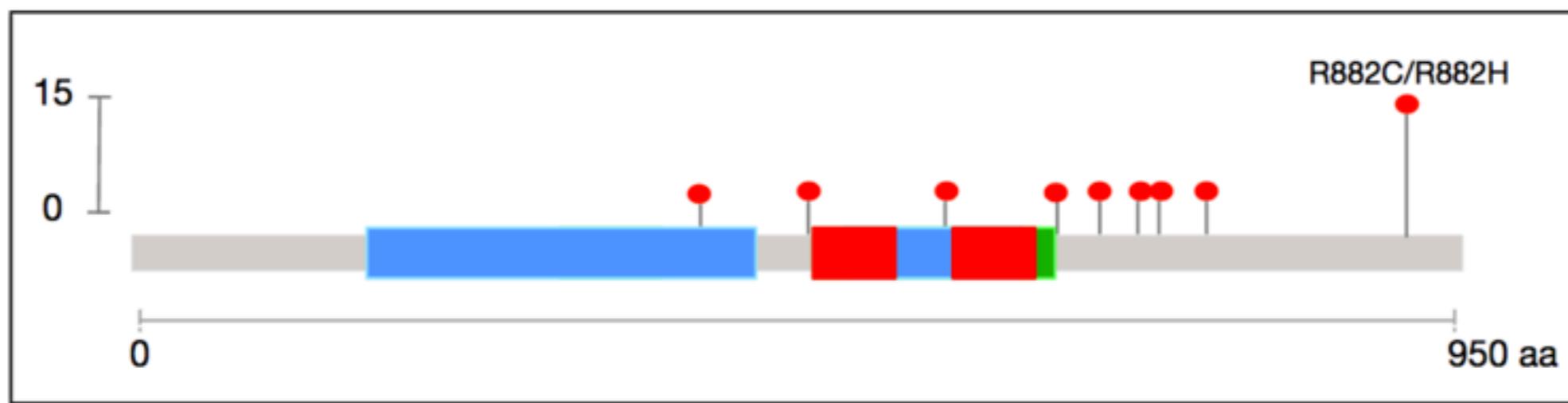


<http://www.mizbee.org>

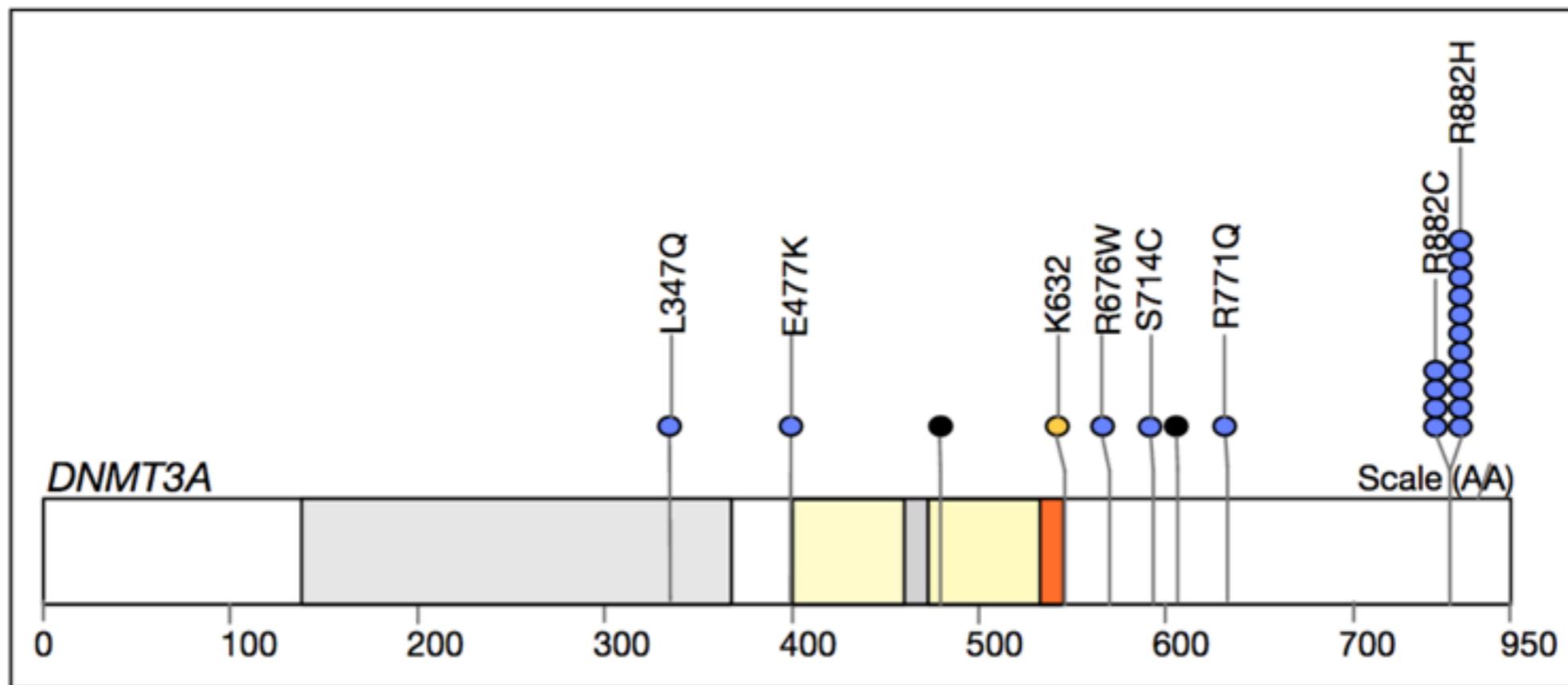


<http://genome.lbl.gov/vista>

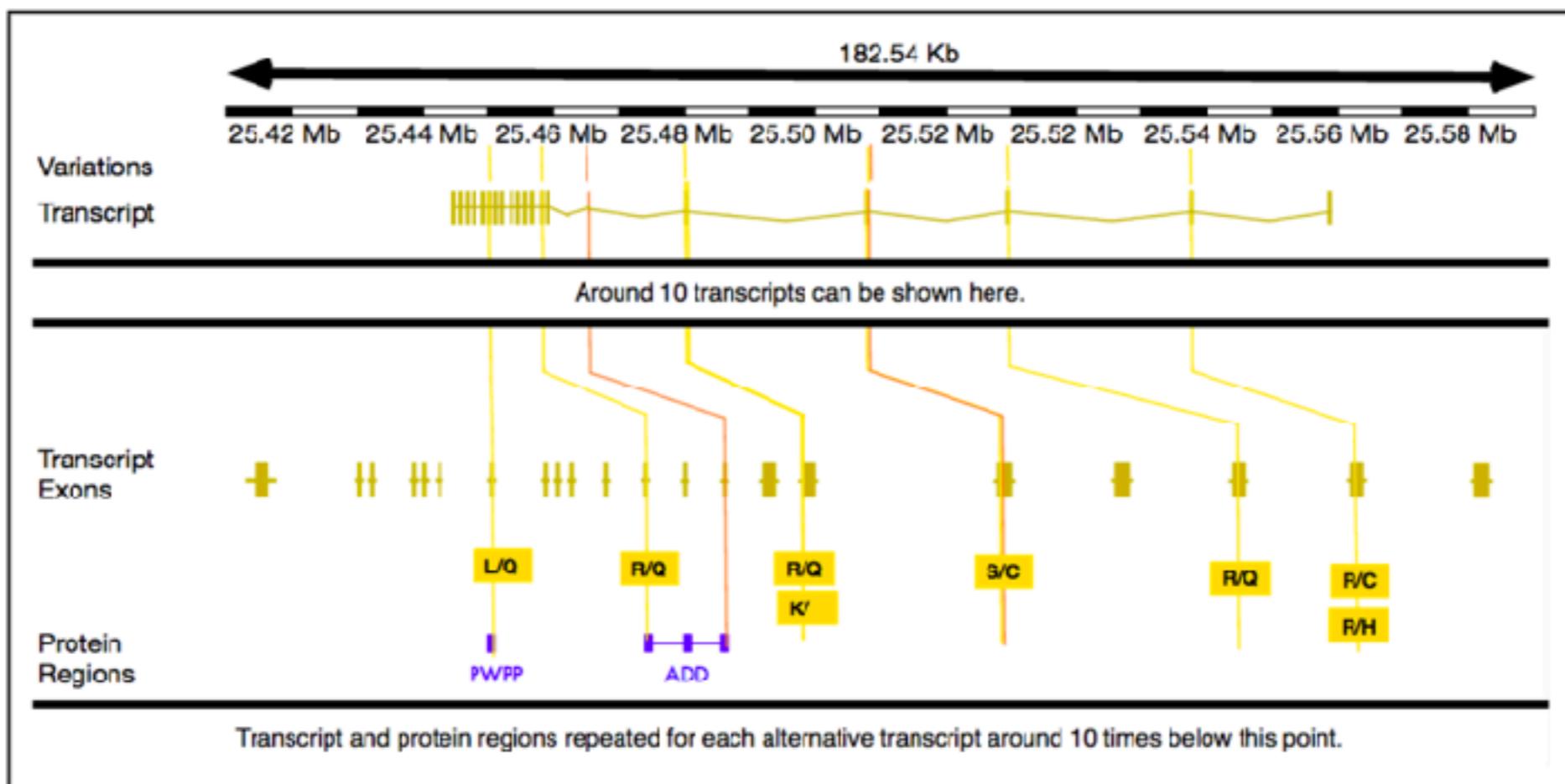
SNPs and Indels



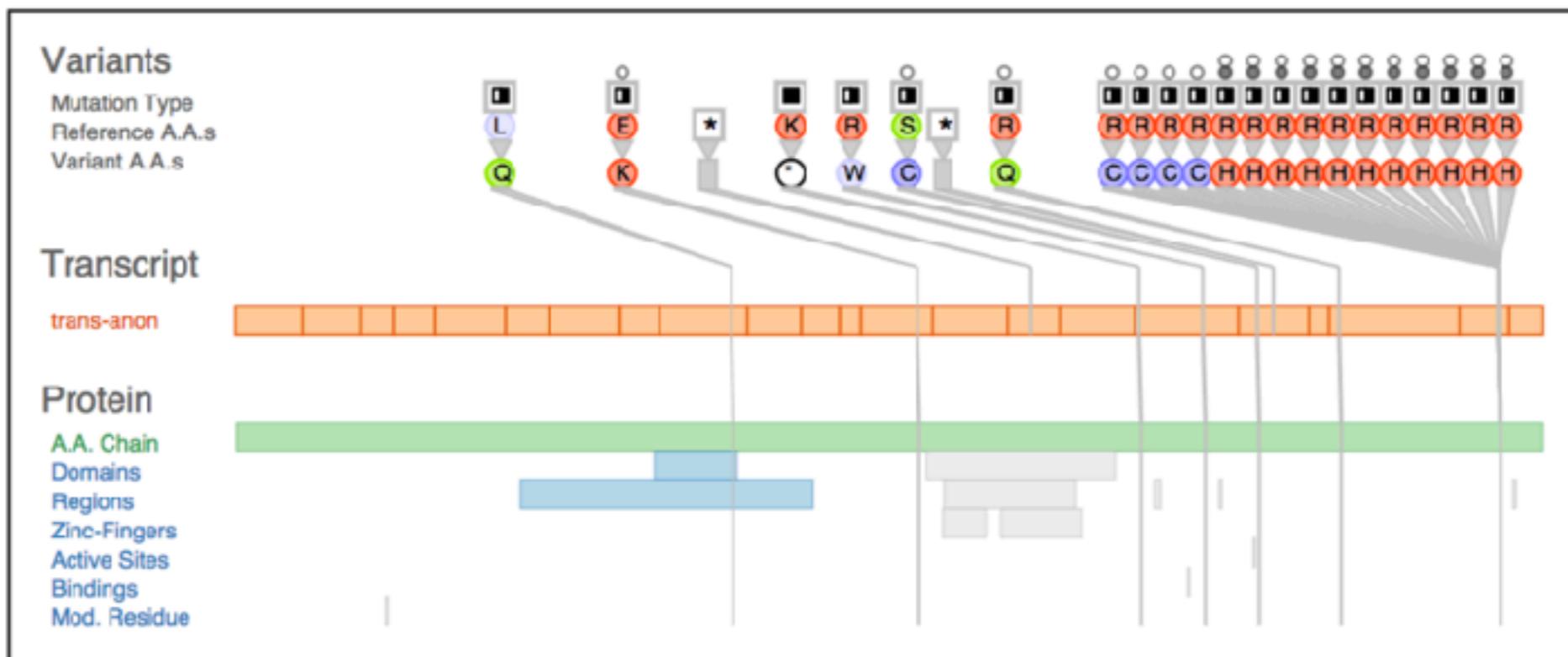
(a)



(b)



(c)

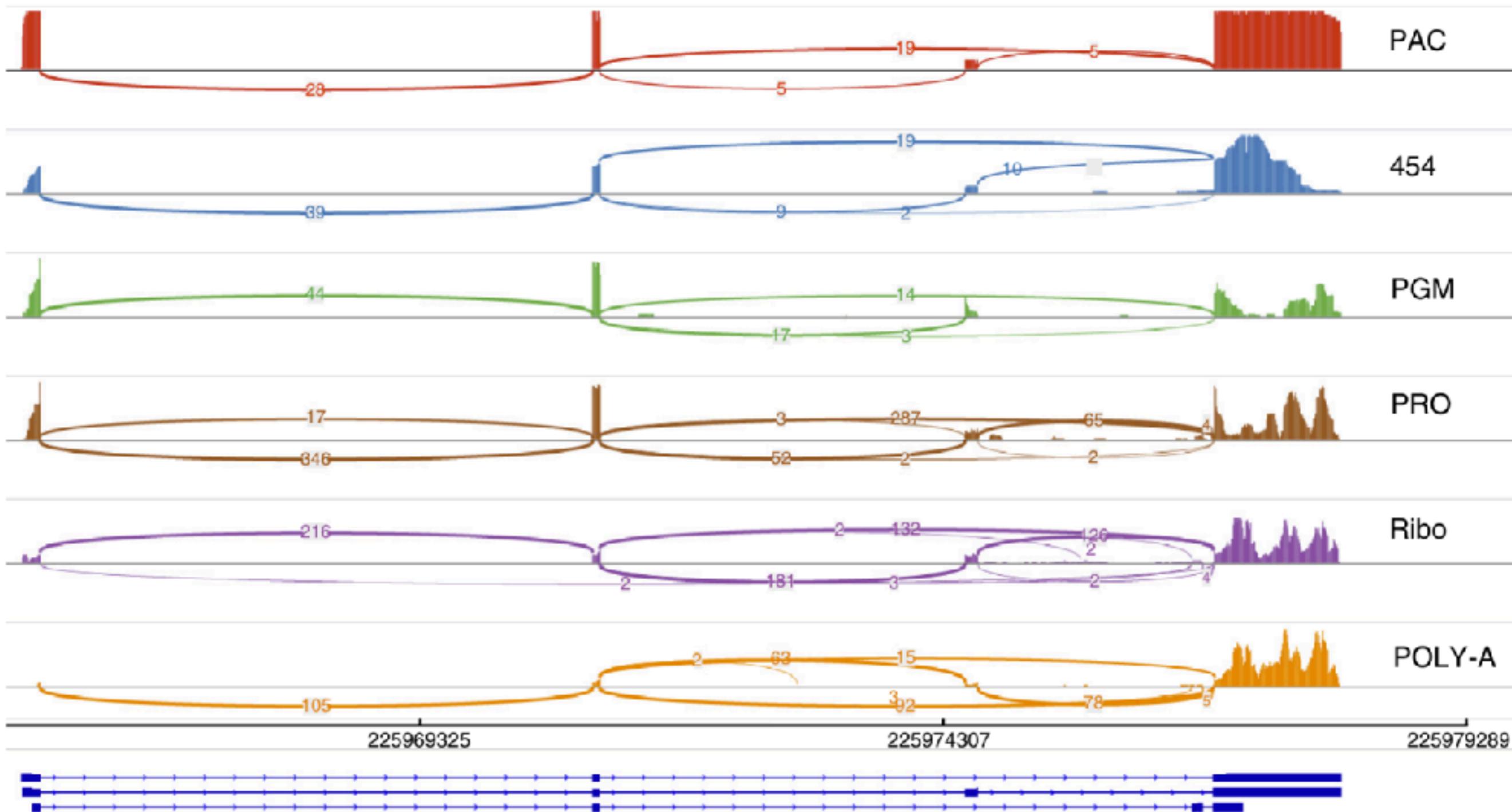


(d)

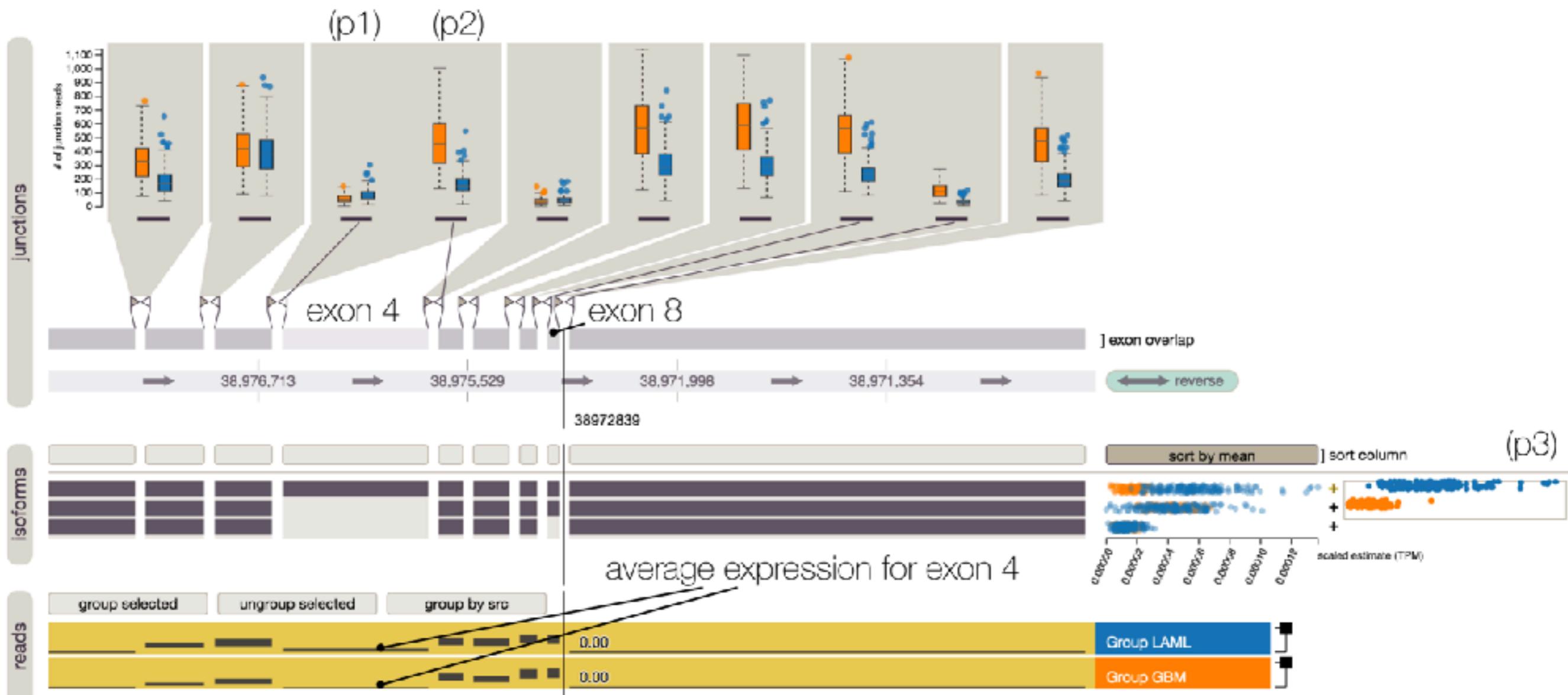
VariantView, Ferstay et al.

Alternative Splicing

Sashimi Plots

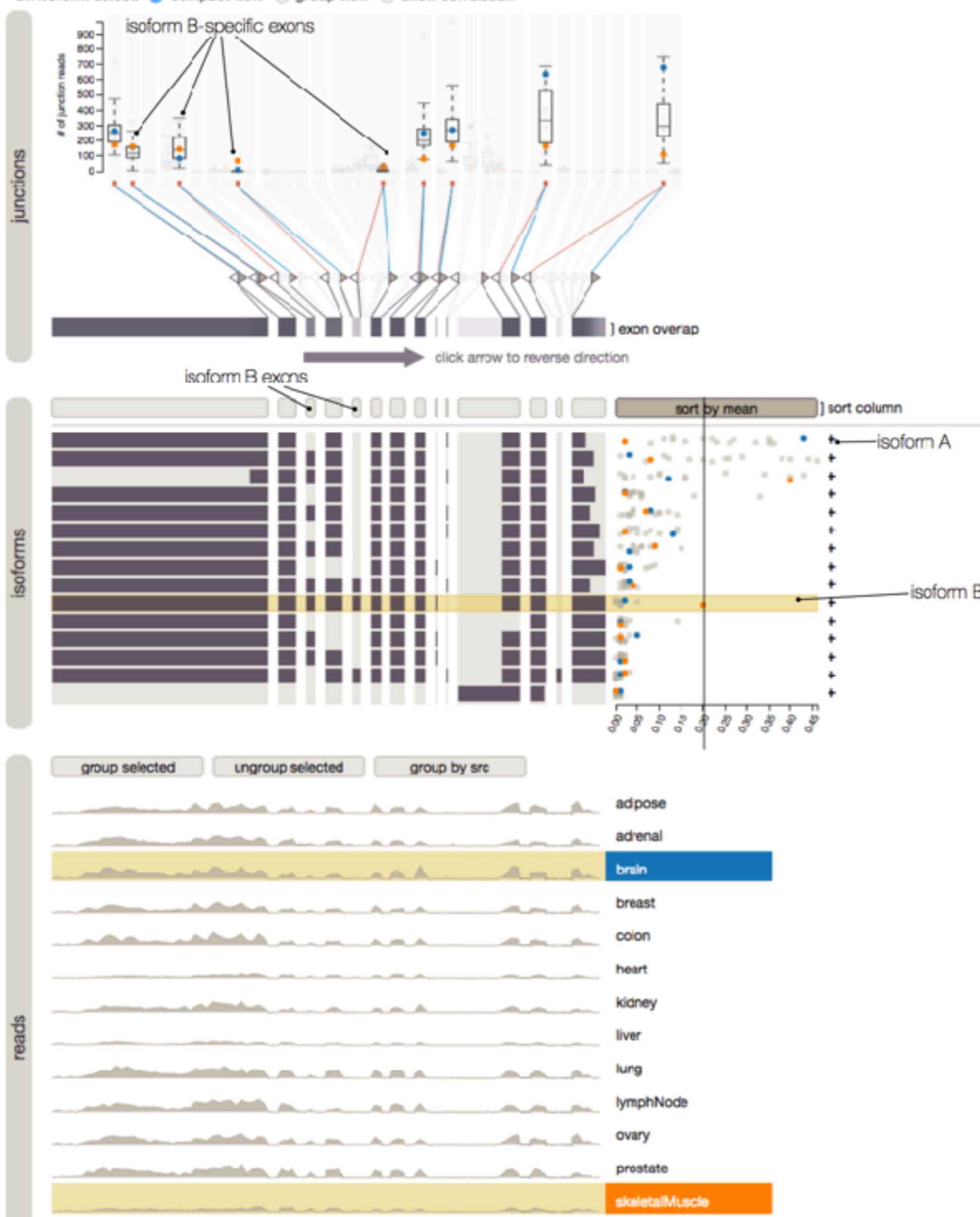


Vials



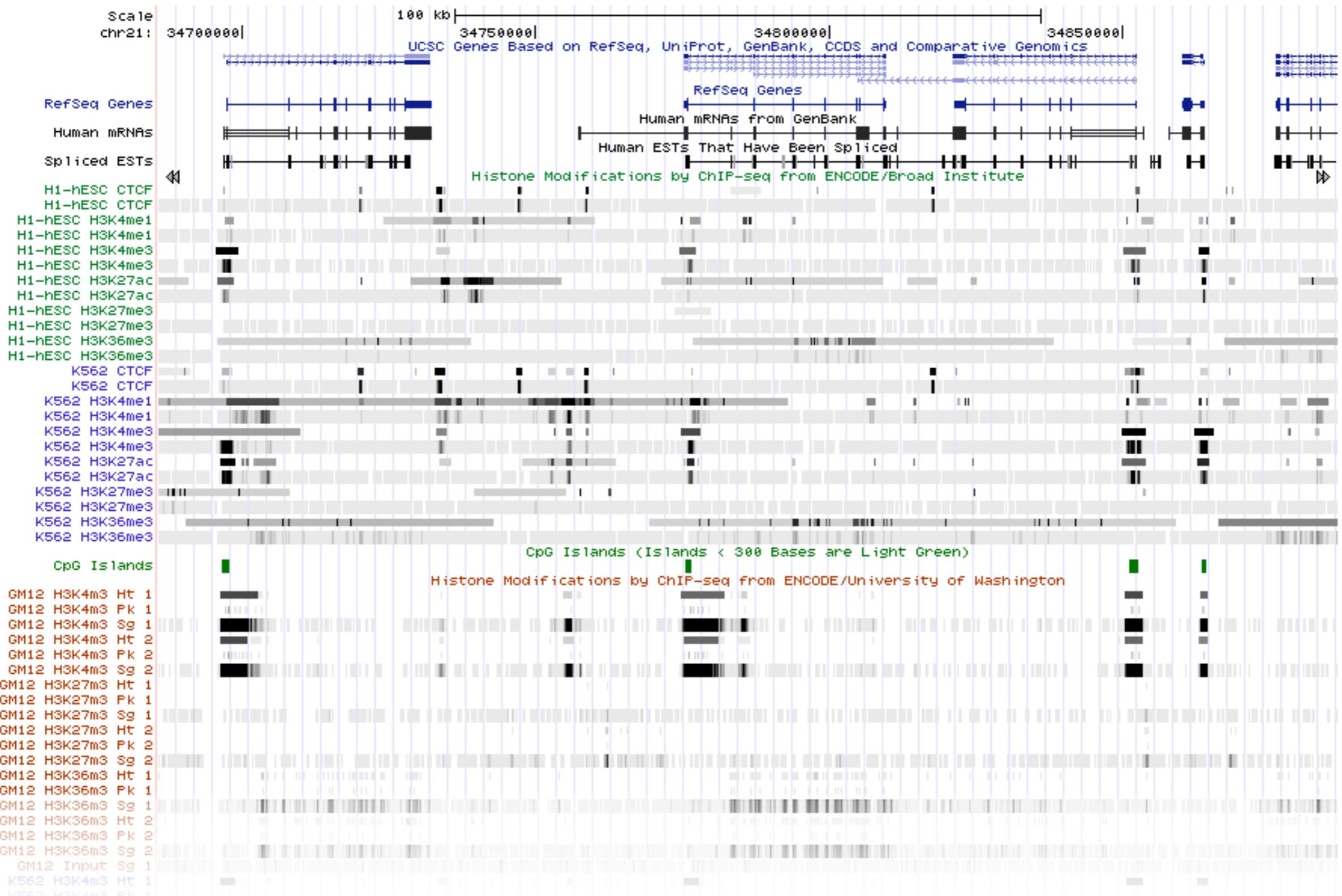
On Isotom-select: compact view group view show correlation

Vials



Genome Sequence

Sequences: UCSC Genome Browser



Sequences: UCSC Genome Browser

“squished”



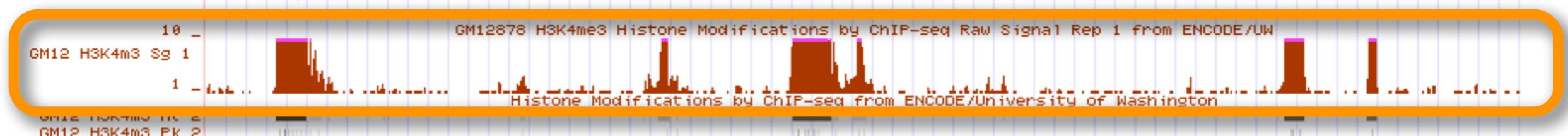
“dense”



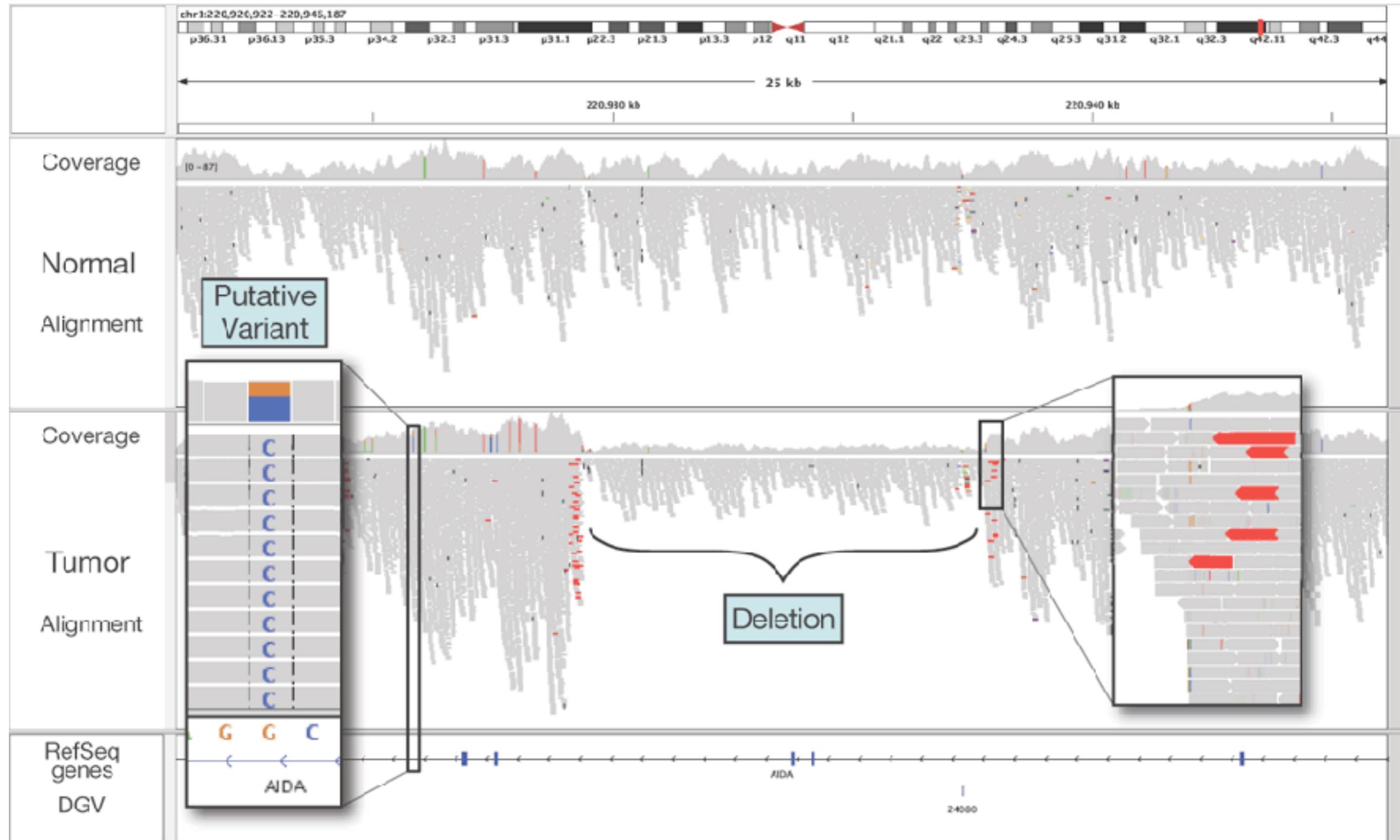
“packed”



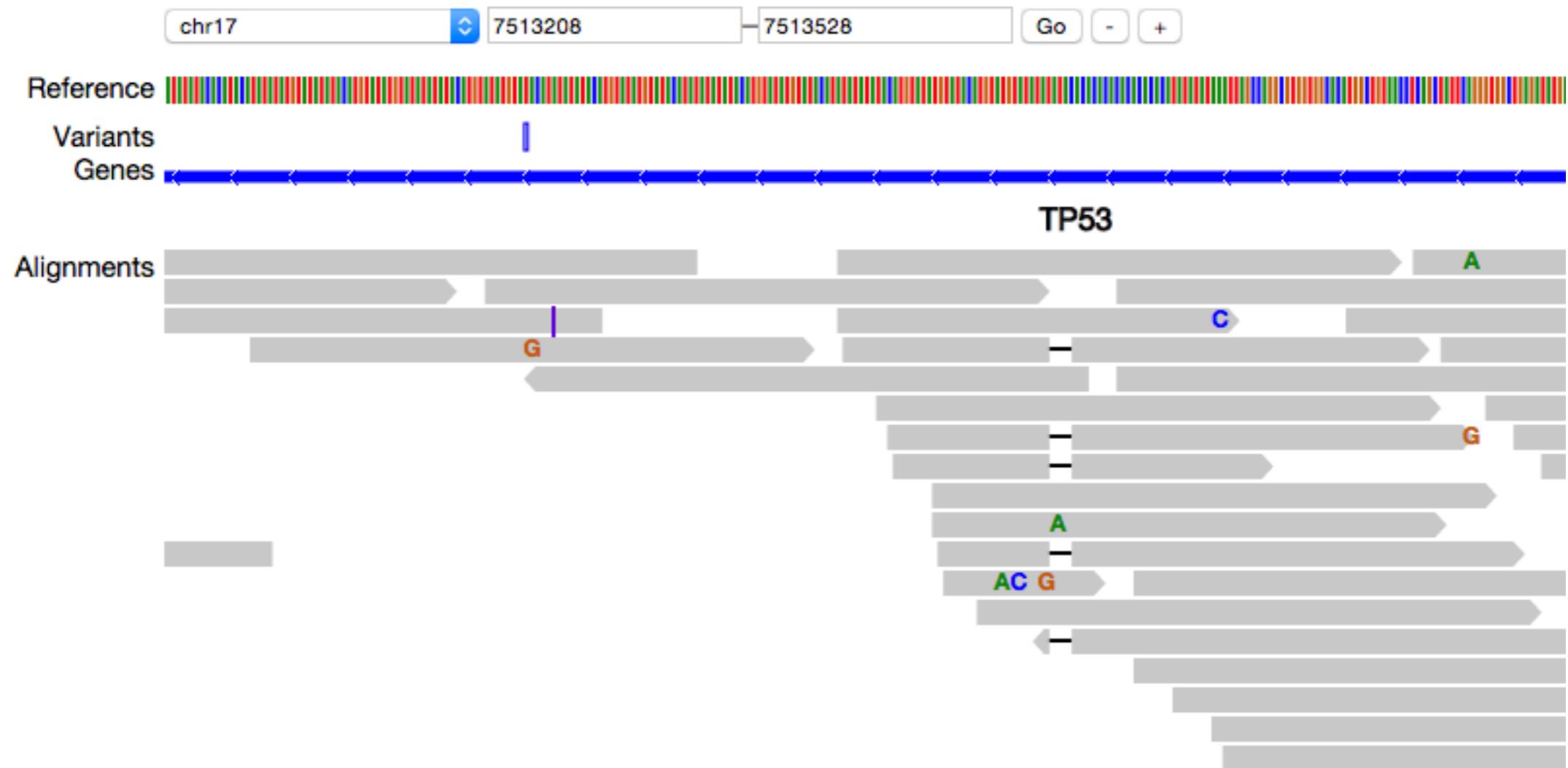
“full”



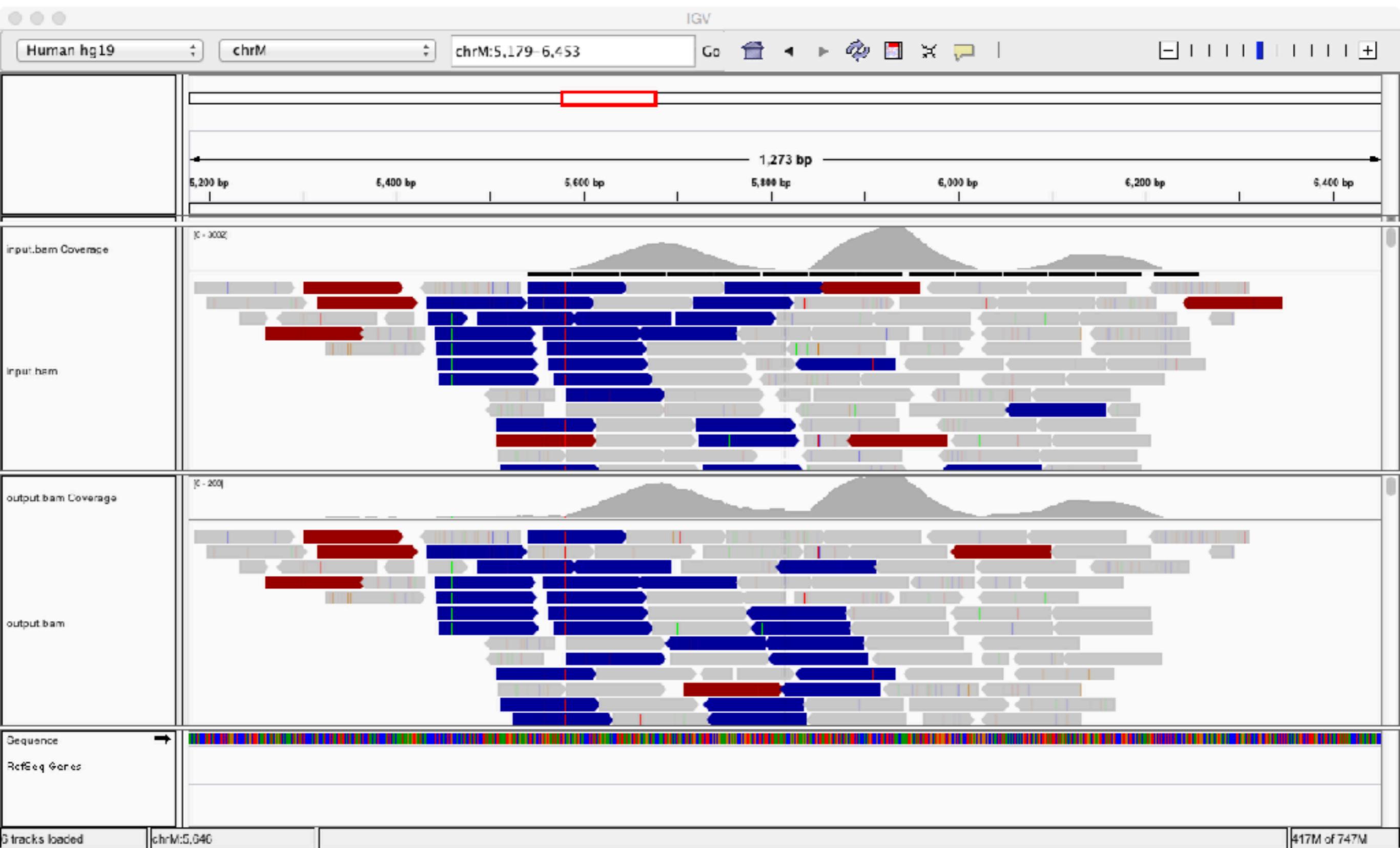
Sequences: Integrative Genomics Viewer (IGV)



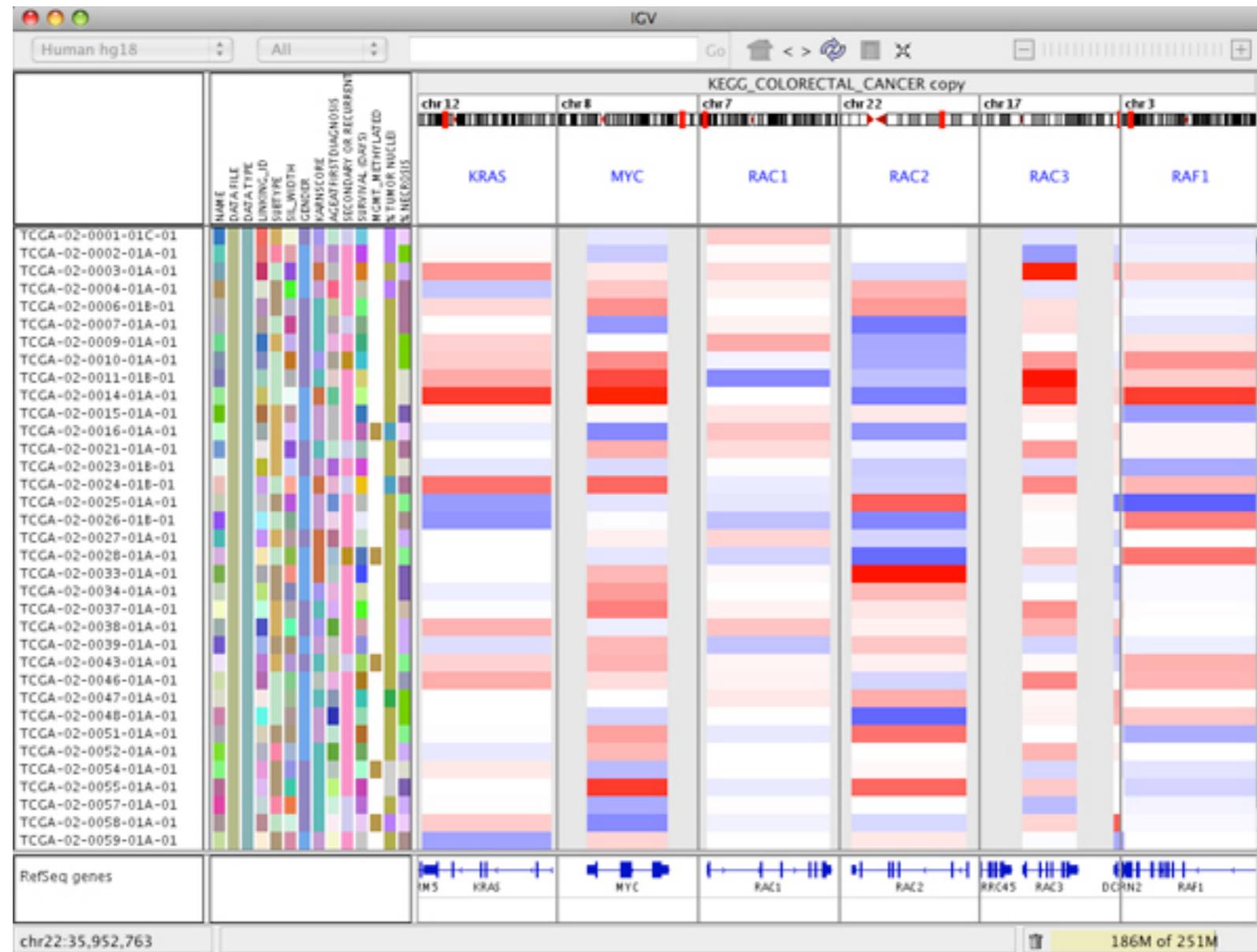
PileUp Visualization



PileUp Visualization

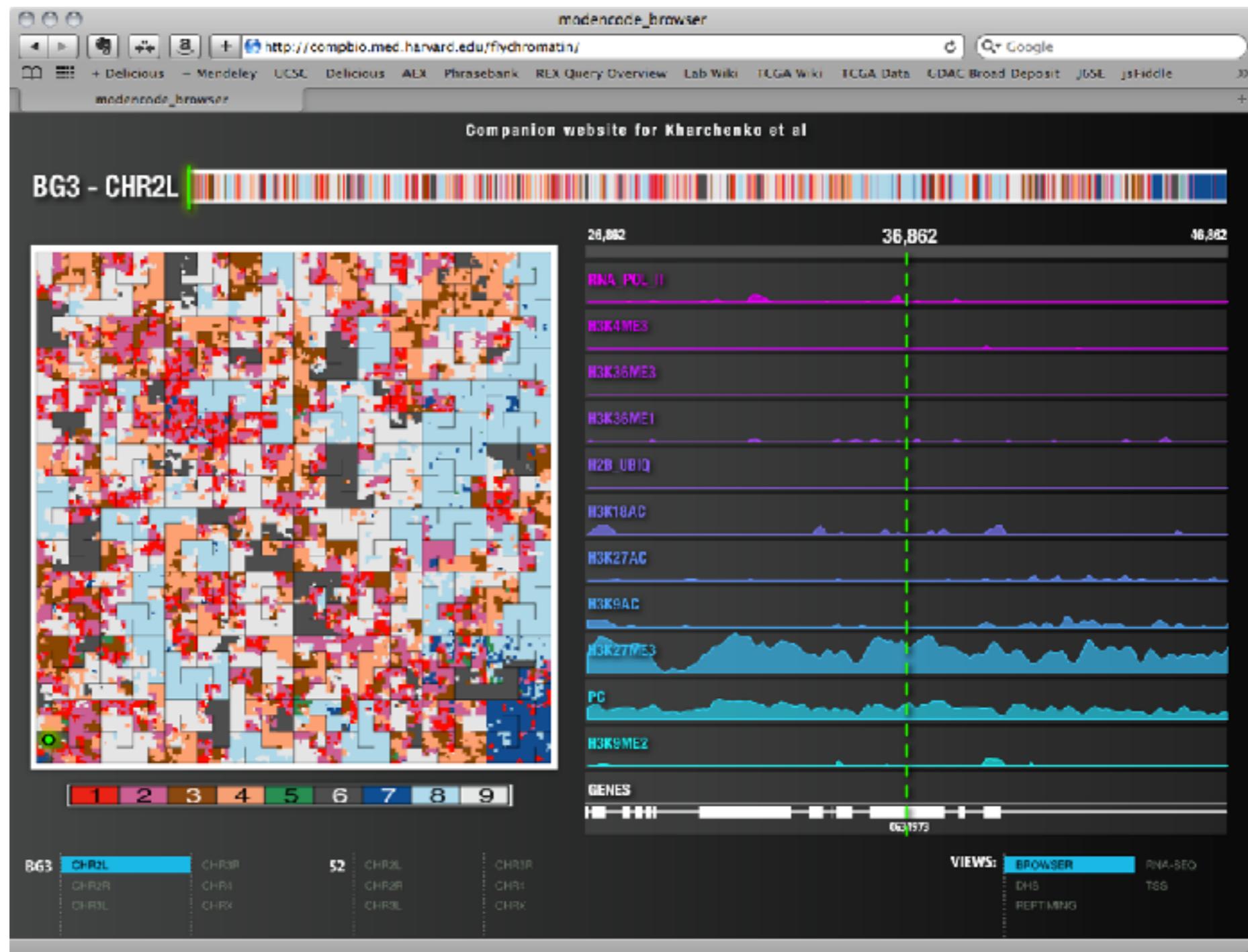


Sequences: Integrative Genomics Viewer (IGV)



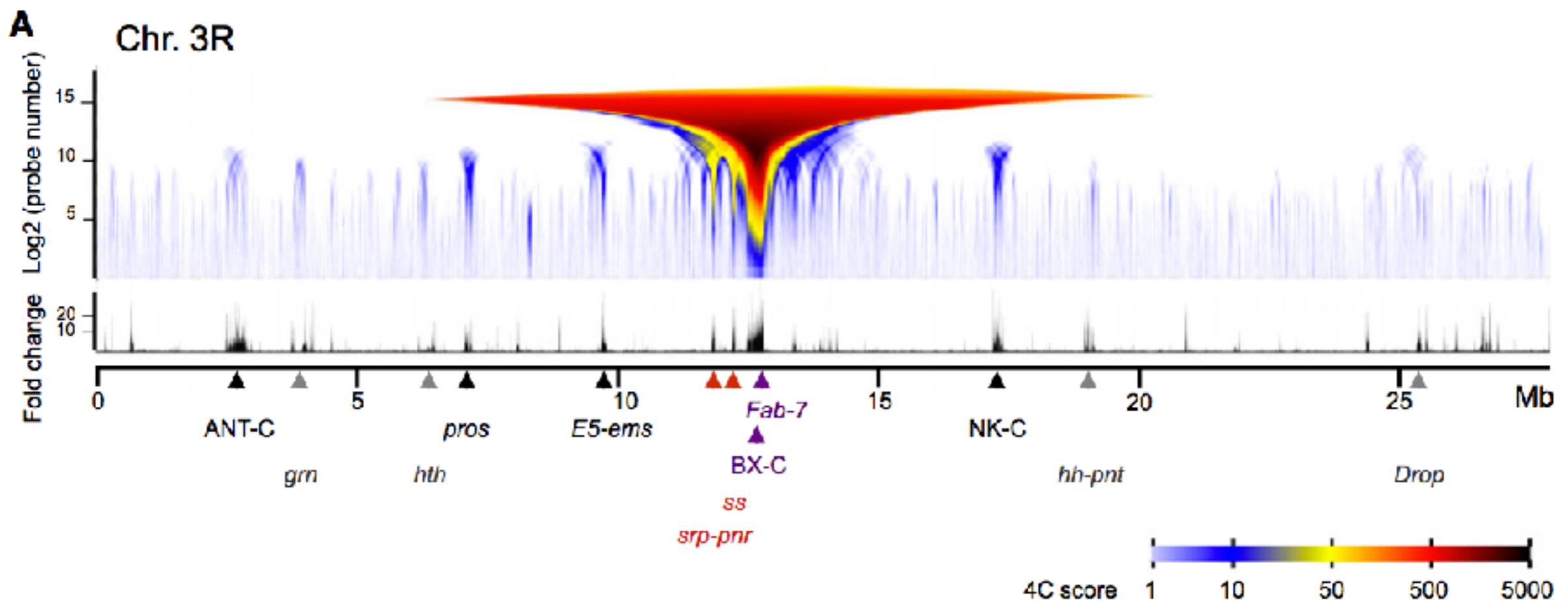
Epigenomic Profiles

Sequences: Epigenomics

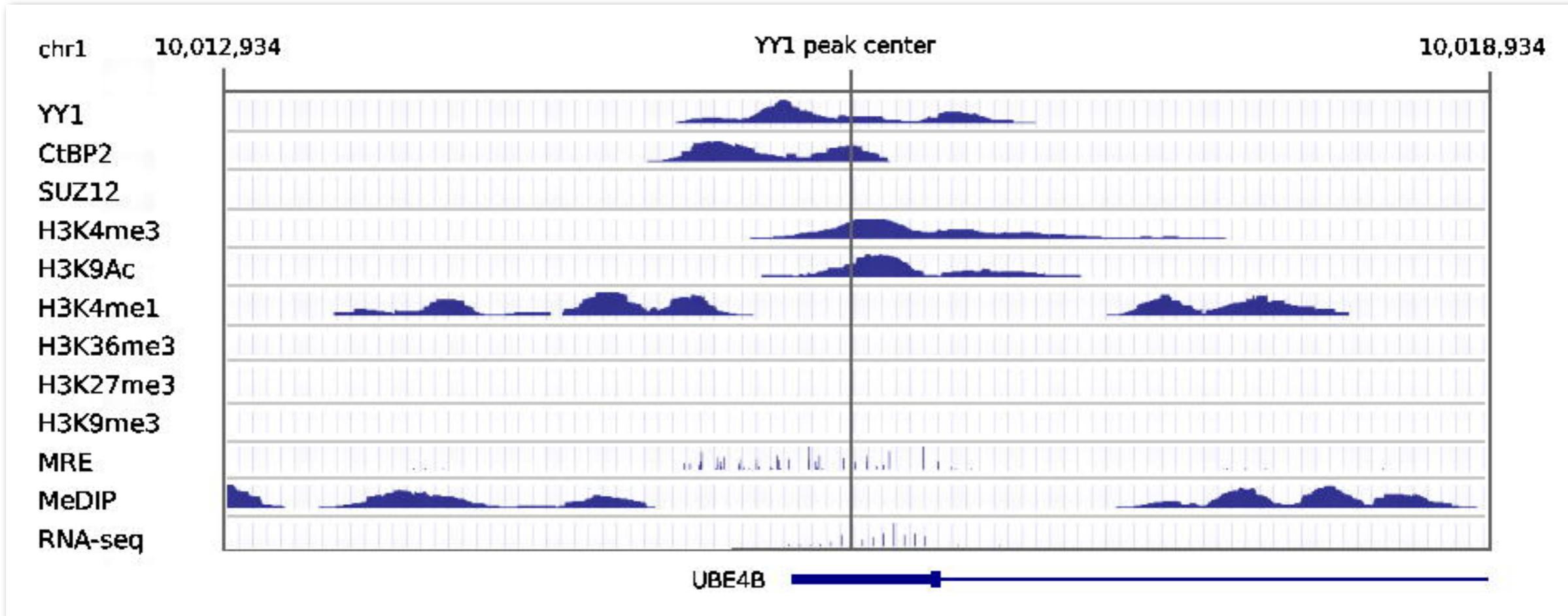


Sequences: Epigenomics

Domainogram



Sequences: Epigenomics



Sequences: Epigenomics

Spark Profile Summaries

