



Cold
Spring
Harbor
Laboratory

Advanced Sequencing Technologies & Applications

<http://meetings.cshl.edu/courses.html>



Cold
Spring
Harbor
Laboratory

Introduction to IGV The Integrative Genomics Viewer

Malachi Griffith, Obi Griffith, Jason Walker, Alex Wagner
Advanced Sequencing Technologies & Applications
November 6- 19, 2017



Visualization Tools in Genomics

- there are **over 40 different genome browsers**, which to use?
- depends on
 - task at hand
 - kind and size of data
 - data privacy

HT-seq Genome Browsers



Integrative
Genome
Viewer



UCSC
Genome Browser
Cancer Genome Browser



Trackster
(part of Galaxy)

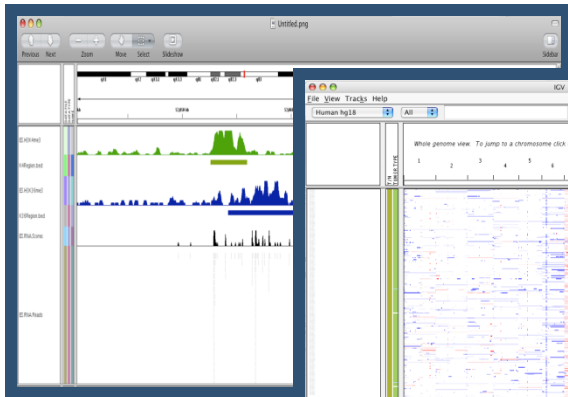


Savant
Genome
Browser

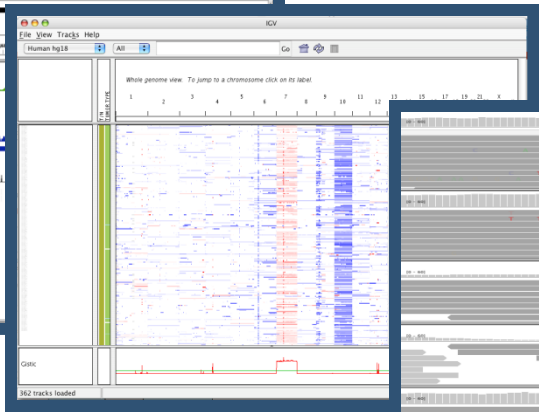
- task at hand : visualizing HT-seq reads, especially good for inspecting variants
- kind and size of data : large BAM files, stored locally or remotely
- data privacy : run on the desktop, can keep all data private
- UCSC Genome Browser has been retro-fitted to display BAM files
- Trackster is a genome browser that can perform visual analytics on small windows of the genome, deploy full analysis with Galaxy

Integrative Genomics Viewer (IGV)

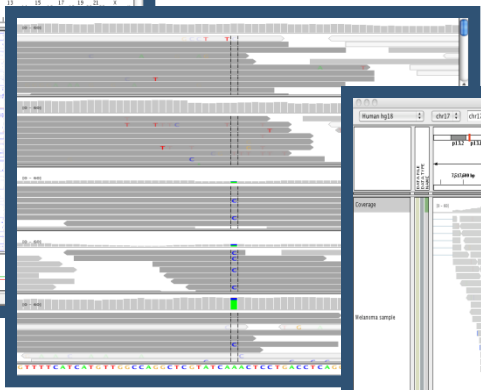
Desktop application for the interactive visual exploration of integrated genomic datasets



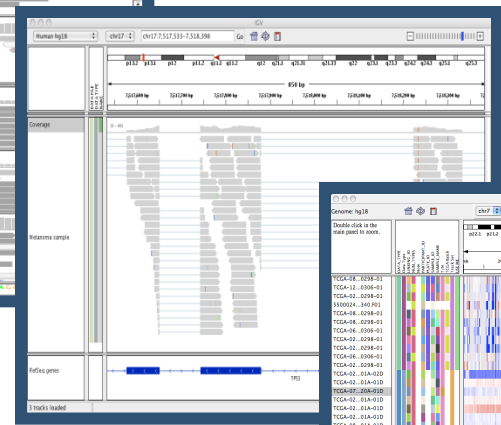
Epigenomics



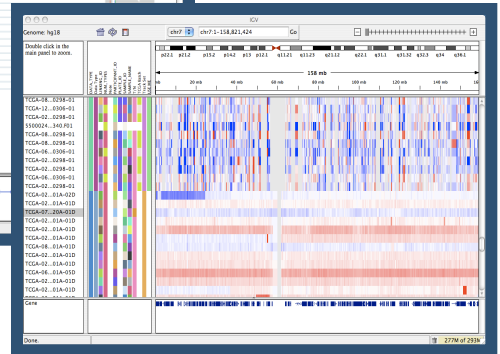
Microarrays



NGS alignments



RNA-Seq



mRNA, CNV, Seq

<http://www.broadinstitute.org/igv>

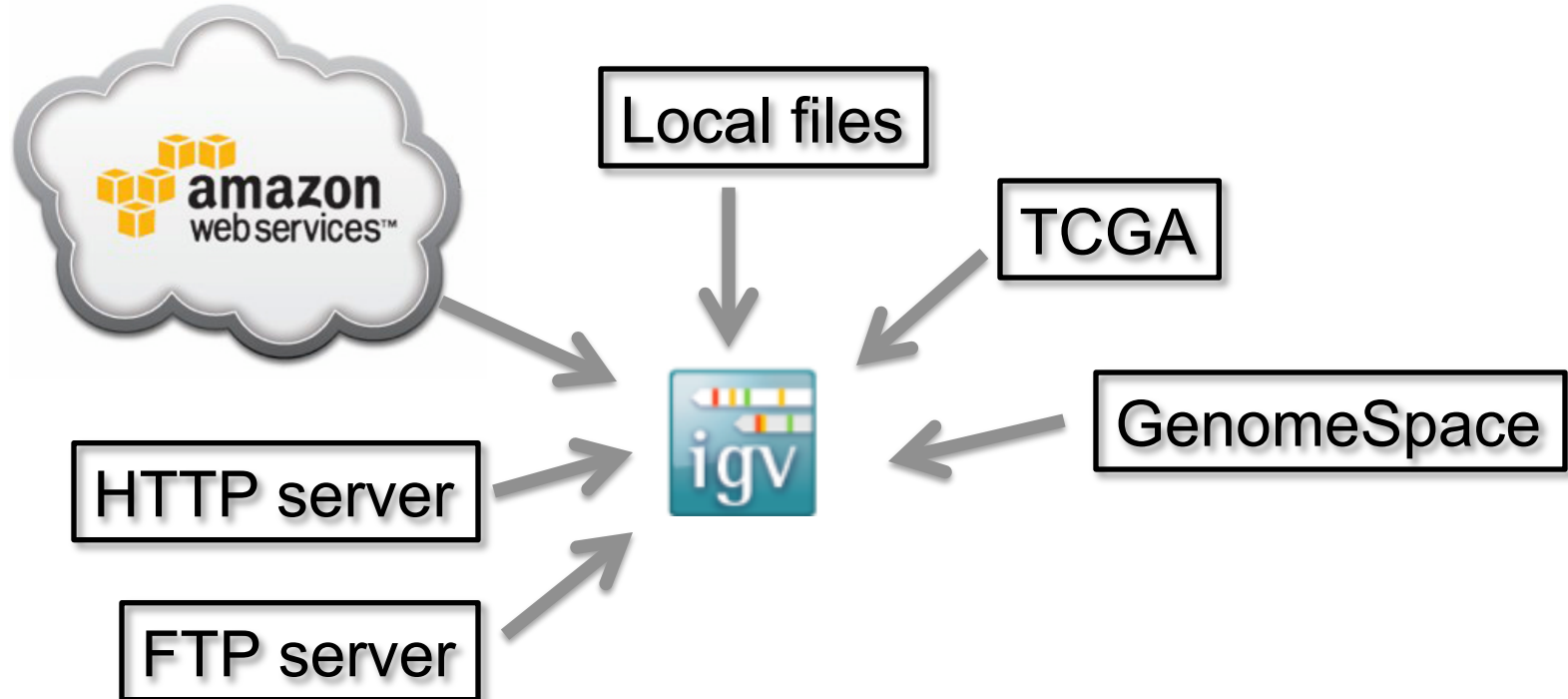
>85,000 registrations (2014)

Features

With IGV you can...

- Explore large genomic datasets with an intuitive, easy-to-use interface.
- Integrate multiple data types with clinical and other sample information.
- View data from multiple sources:
 - local, remote, and “cloud-based”.
- Automation of specific tasks using command-line interface

IGV data sources

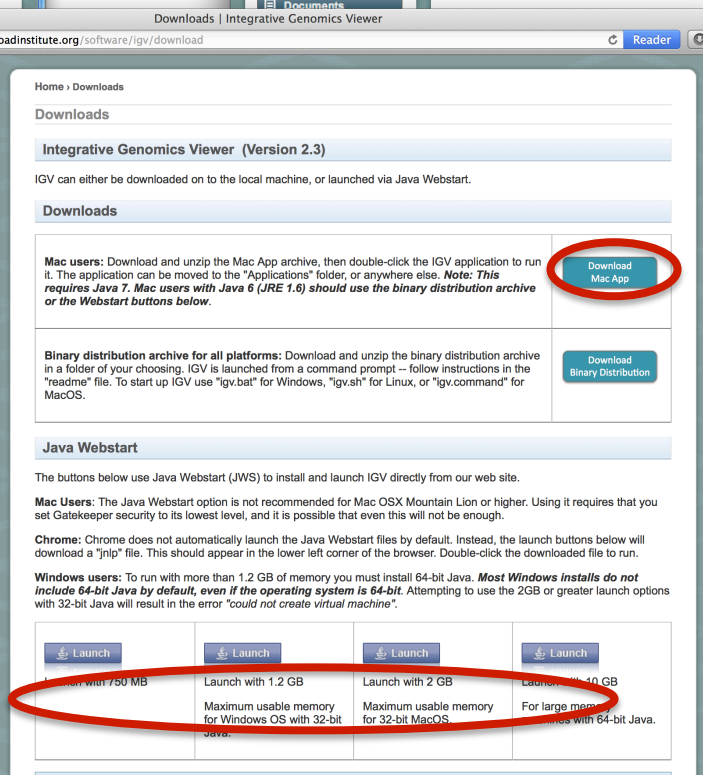
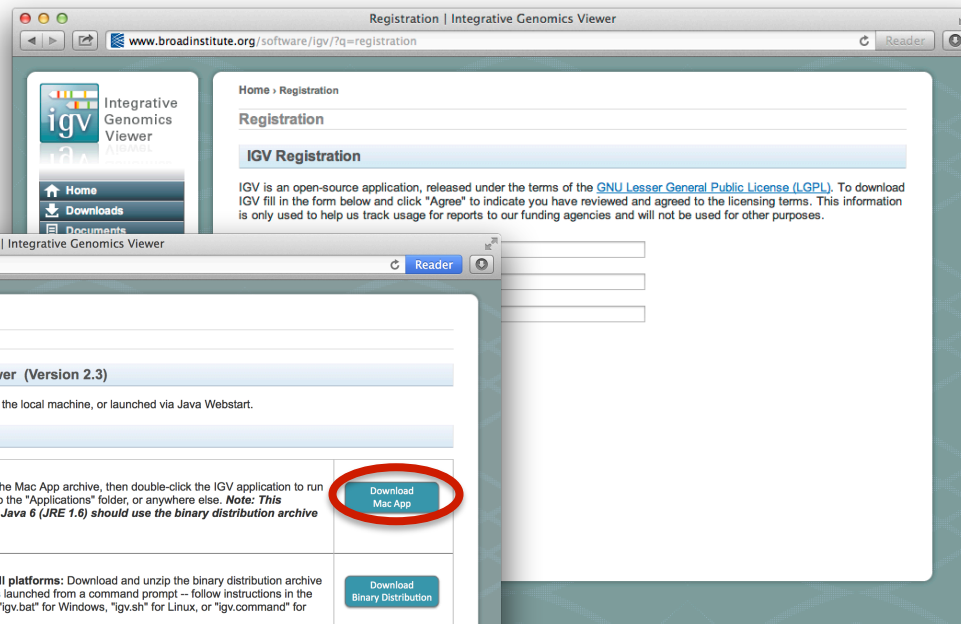
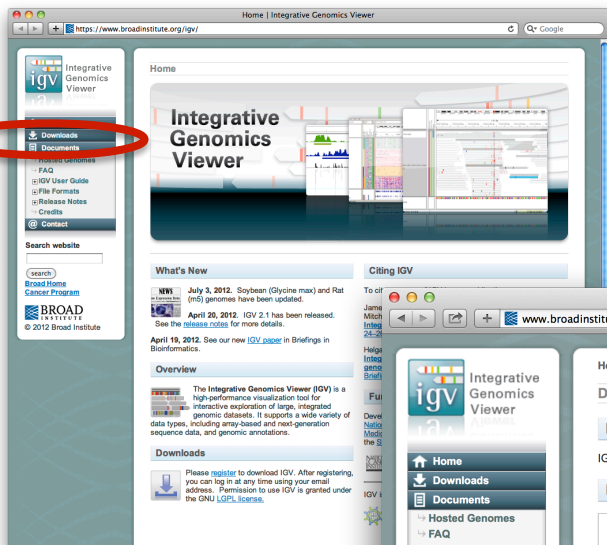


- View **local** files without uploading.
- View **remote** files without downloading the whole dataset.

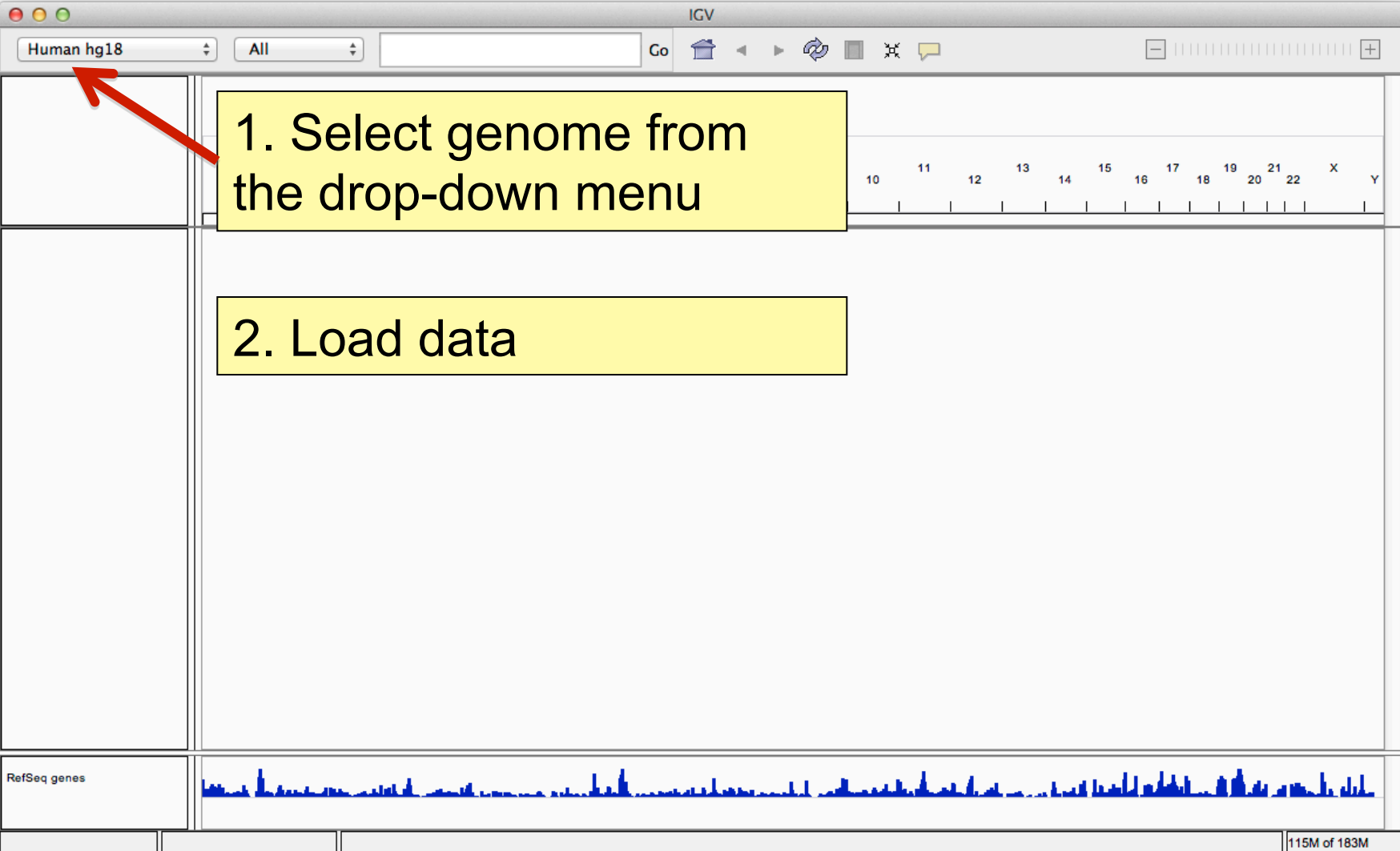
Using IGV: the basics

- Launch IGV
- Select a reference genome
- Load data
- Navigate through the data
 - WGS data
 - SNVs
 - structural variations

Launch IGV



Launch IGV



The screenshot shows the IGV web interface. At the top, there is a header bar with the text "IGV". Below the header, there is a navigation bar containing a dropdown menu with "Human hg18" selected, a "Go" button, and several navigation icons. A red arrow points to the "Human hg18" dropdown menu. Below the navigation bar, there is a large white area with a yellow box containing the text "1. Select genome from the drop-down menu". Below this, there is another yellow box containing the text "2. Load data". At the bottom of the interface, there is a track labeled "RefSeq genes" showing a blue bar chart. In the bottom right corner, there is a status bar showing "115M of 183M".

1. Select genome from the drop-down menu

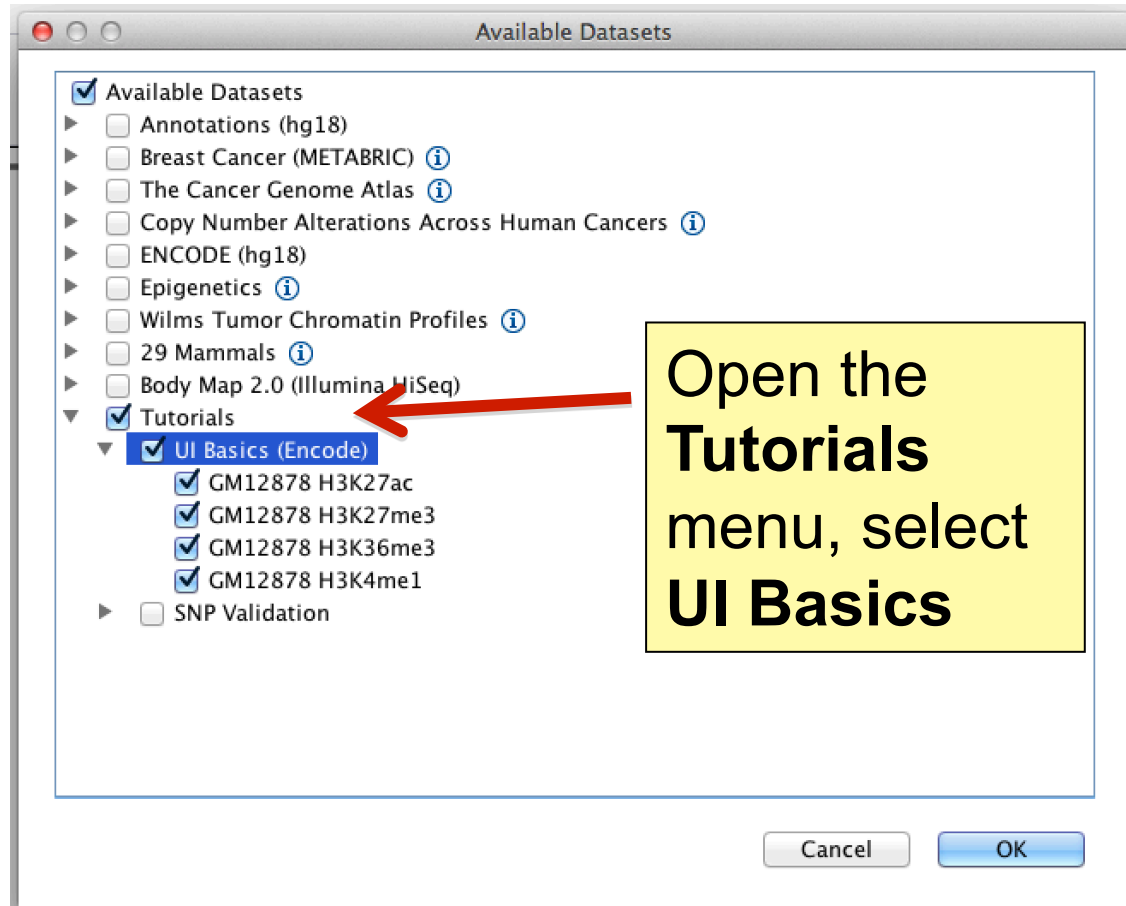
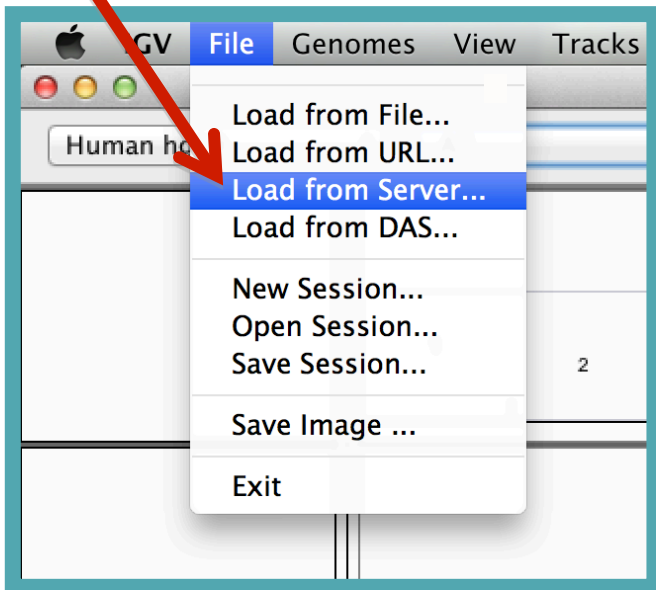
2. Load data

RefSeq genes

115M of 183M

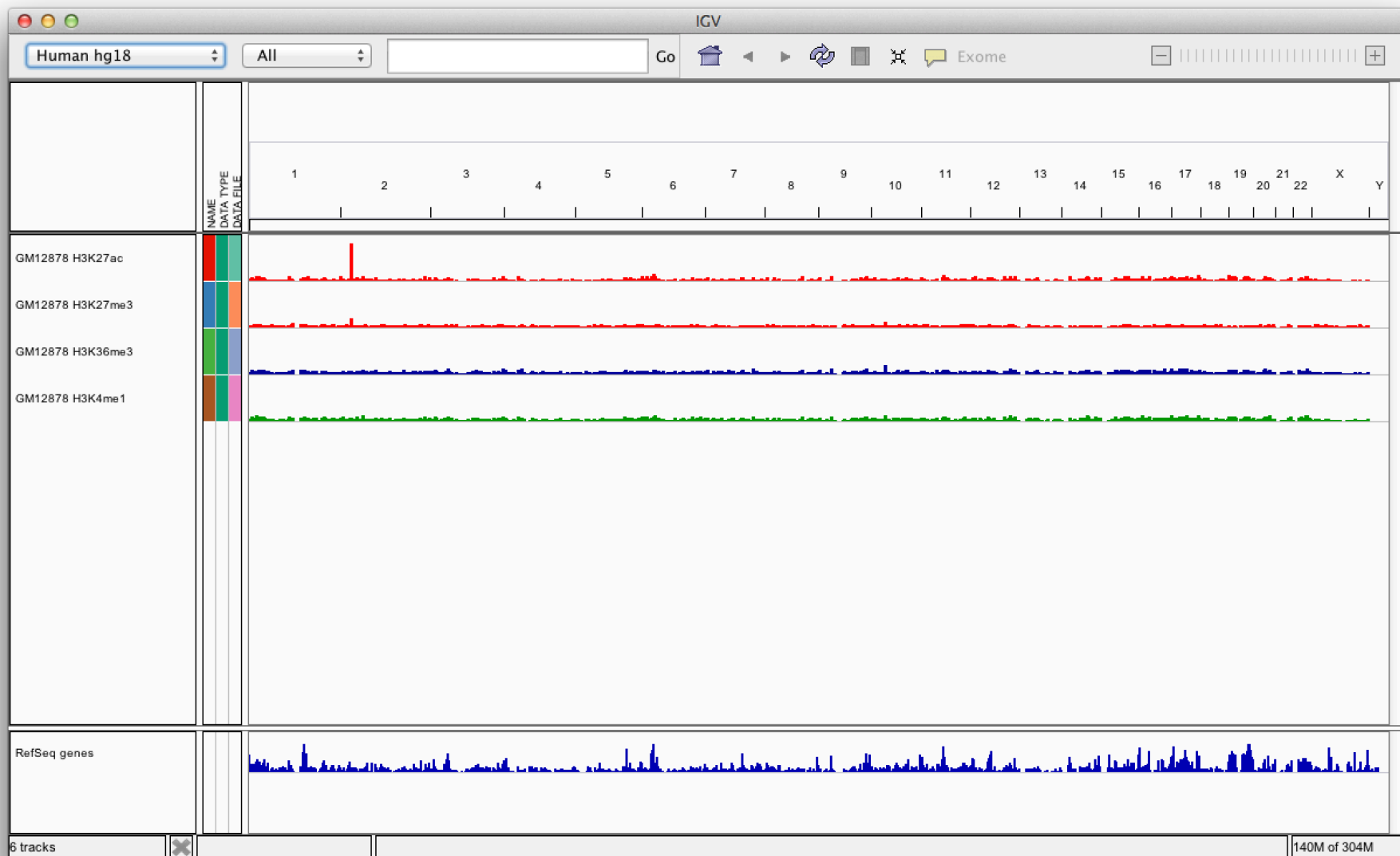
Load data

Select **File > Load from Server...**

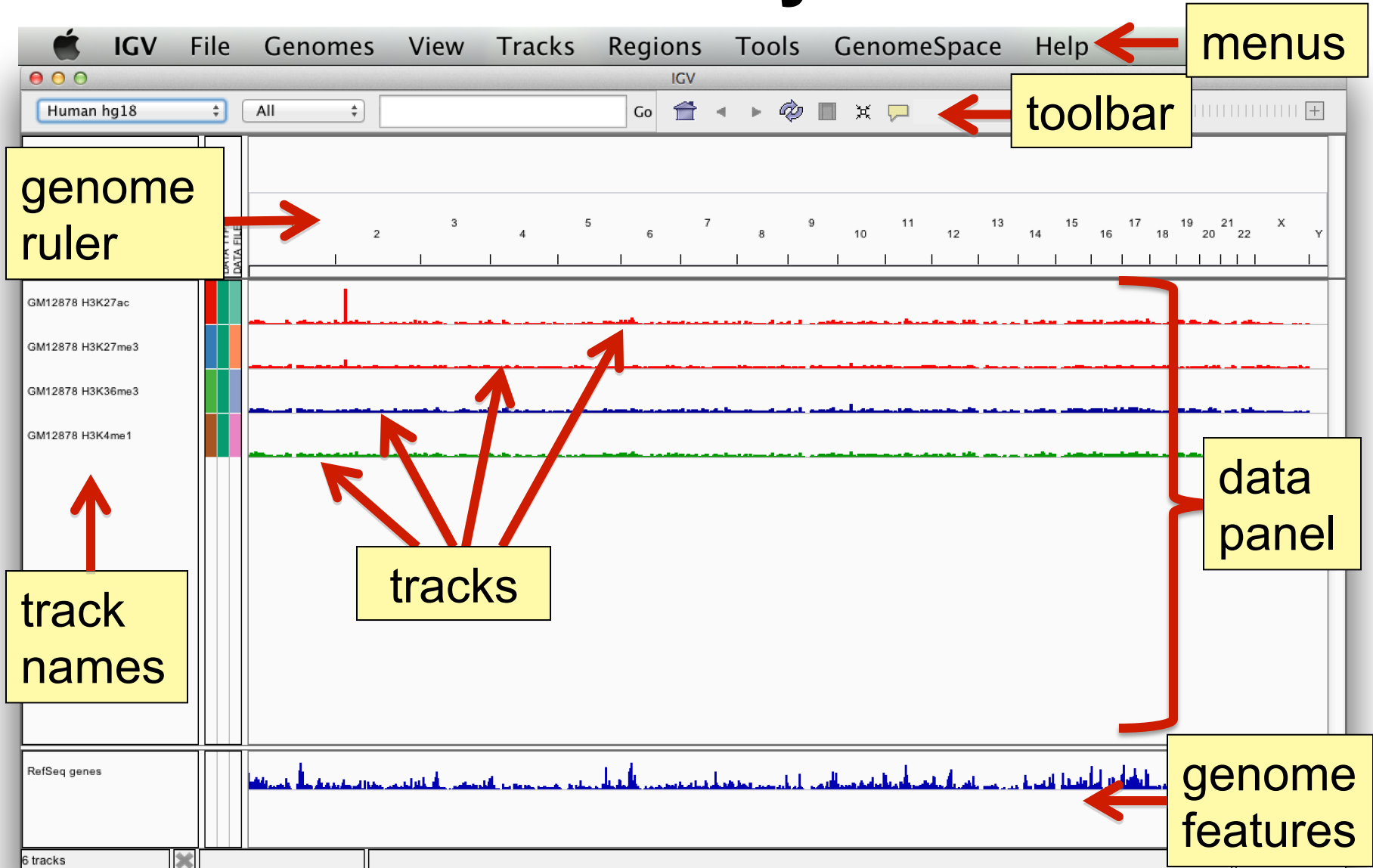


Open the **Tutorials** menu, select **UI Basics**

Screen layout



Screen layout

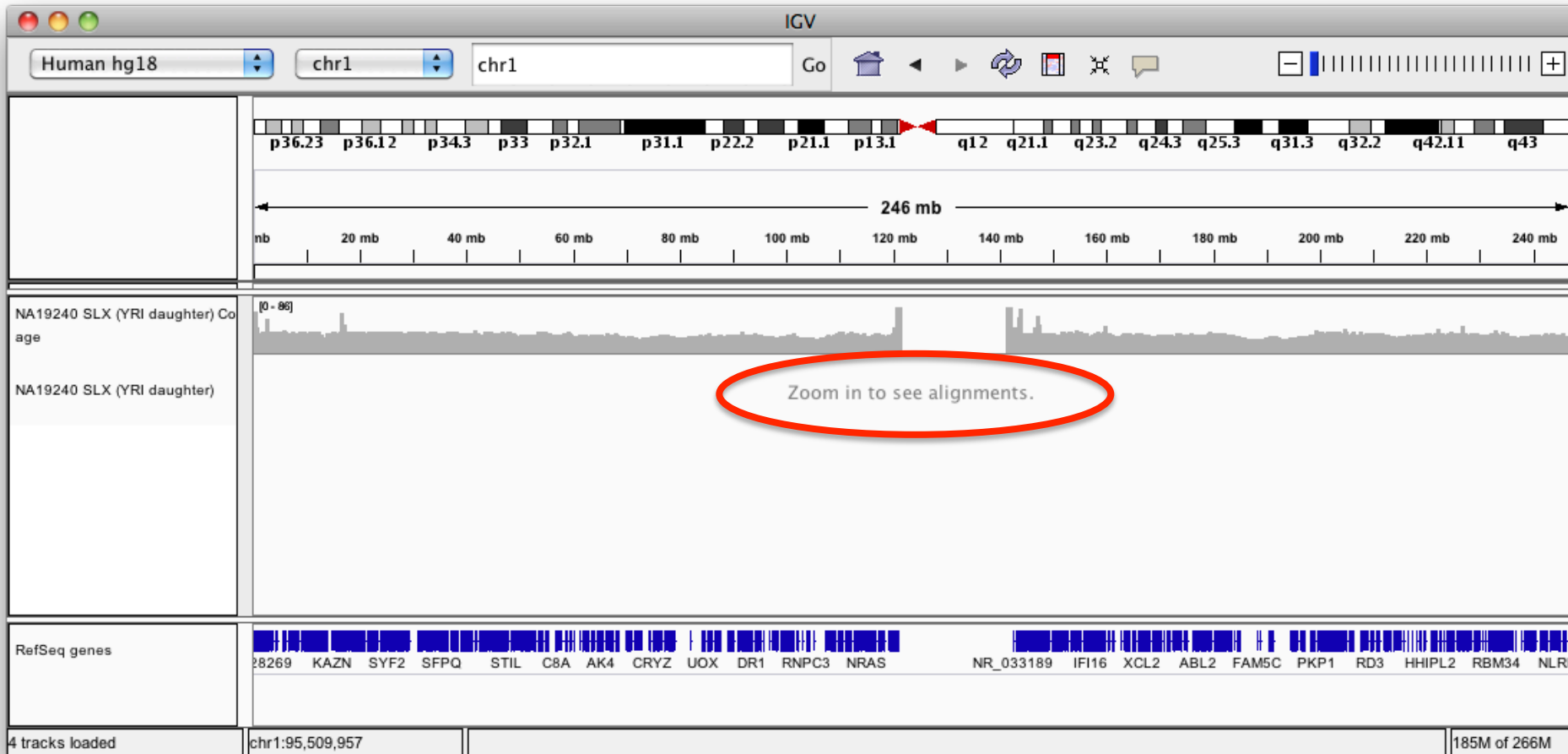


File formats and track types

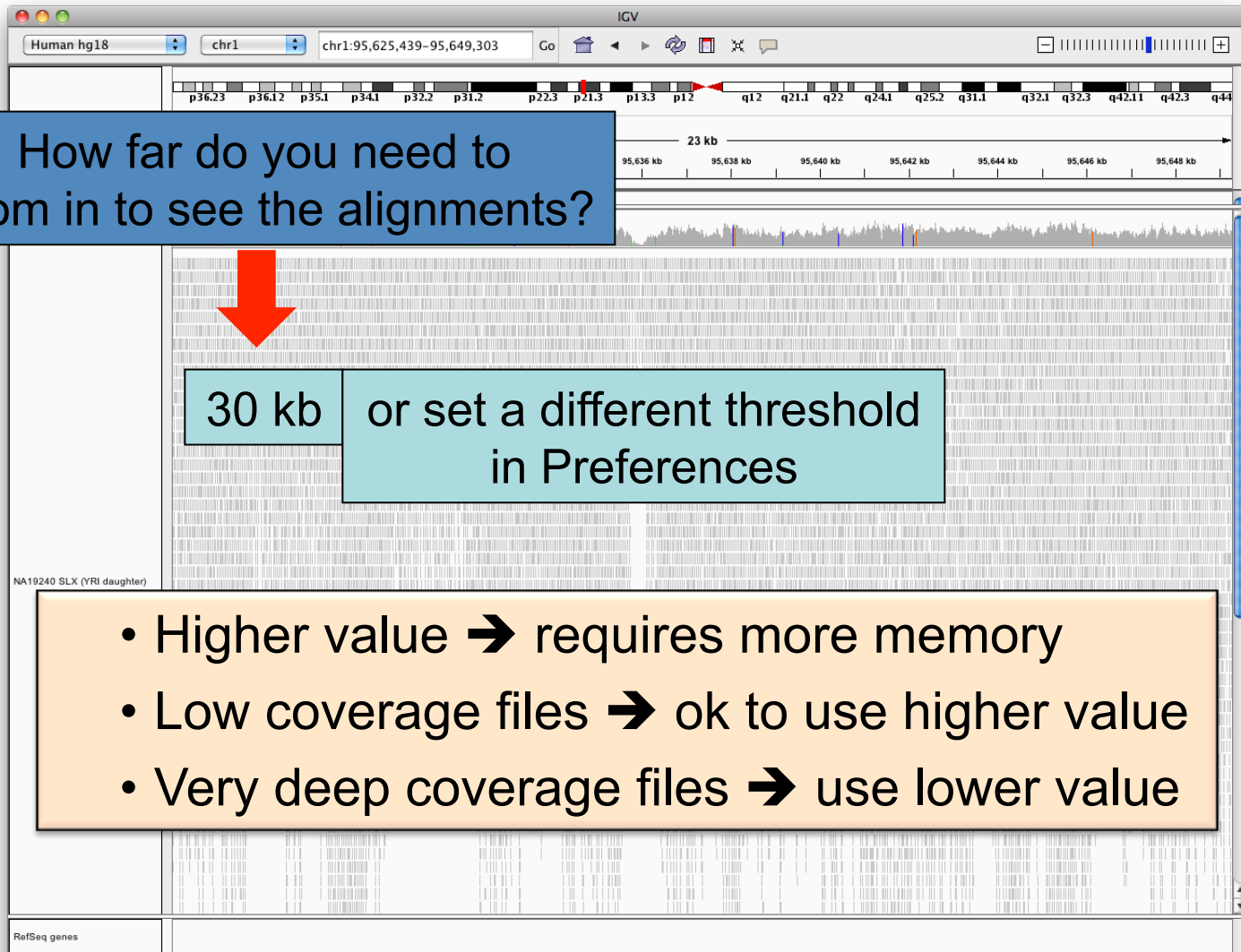
- The **file format** defines the track type.
- The **track type** determines the display options
 - [BAM](#)
 - [BED](#)
 - [BedGraph](#)
 - [bigBed](#)
 - [bigWig](#)
 - [Birdsuite Files](#)
 - [CBS](#)
 - [CN](#)
 - [Cufflinks Files](#)
 - [Custom File Formats](#)
 - [Cytoband](#)
 - [FASTA](#)
 - [GCT](#)
 - [genePred](#)
 - [GFF](#)
 - [GISTIC](#)
 - [Goby](#)
 - [GWAS](#)
 - [IGV](#)
 - [LOH](#)
 - [MAF](#)
 - [Merged BAM File \(.bam.list\)](#)
 - [MUT](#)
 - [PSL](#)
 - [RES](#)
 - [SAM](#)
 - [Sample Information](#)
 - [SEG](#)
 - [SNP](#)
 - [TAB](#)
 - [TDF](#)
 - [Track Line](#)
 - [Type Line](#)
 - [VCF](#)
 - [WIG](#)
- For current list see: www.broadinstitute.org/igv/FileFormats

Viewing alignments

Whole chromosome view



Viewing alignments – Zoom in



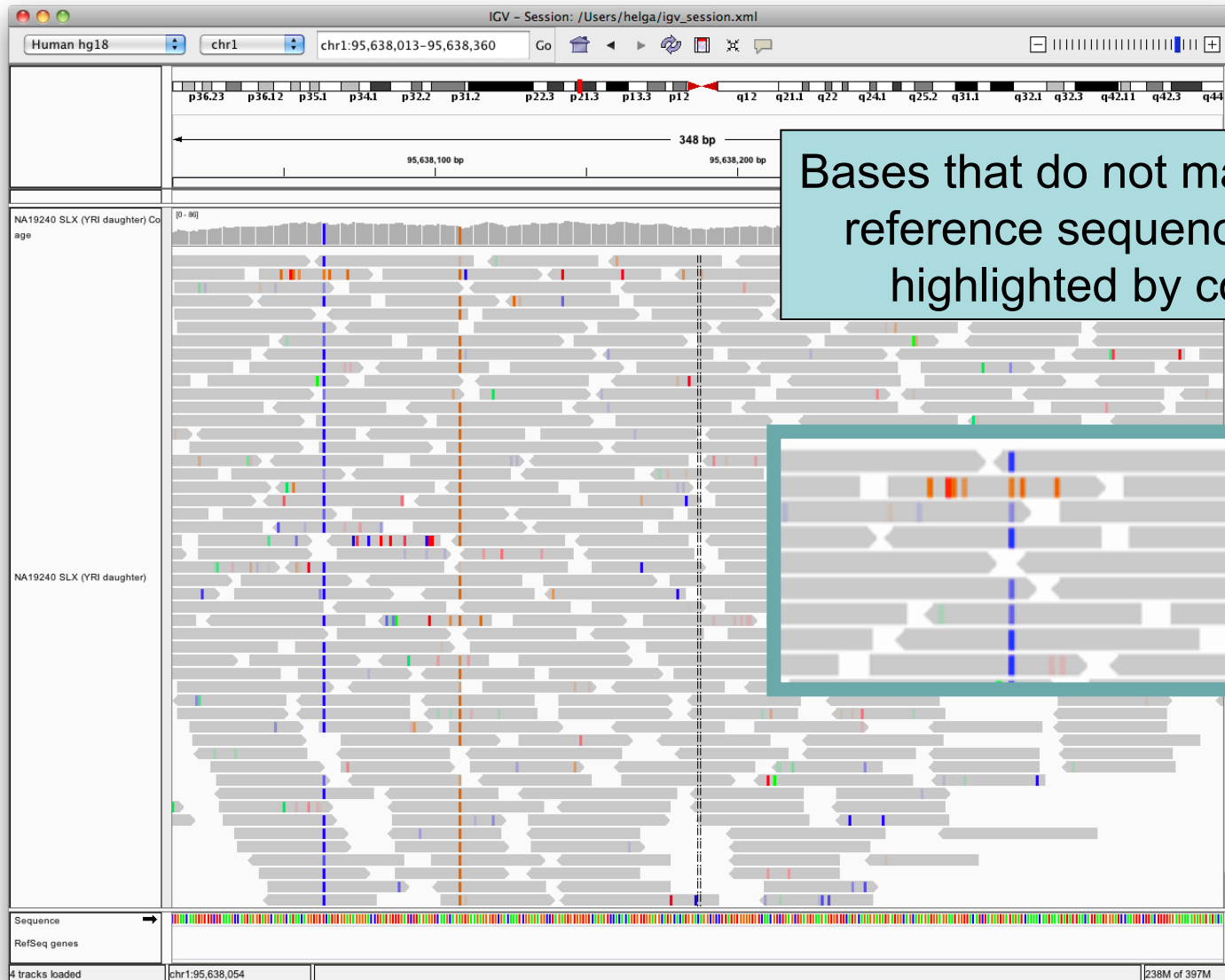
Human hg18 chr1 chr1:95,625,439-95,649,303

How far do you need to zoom in to see the alignments?

30 kb or set a different threshold in Preferences

- Higher value → requires more memory
- Low coverage files → ok to use higher value
- Very deep coverage files → use lower value

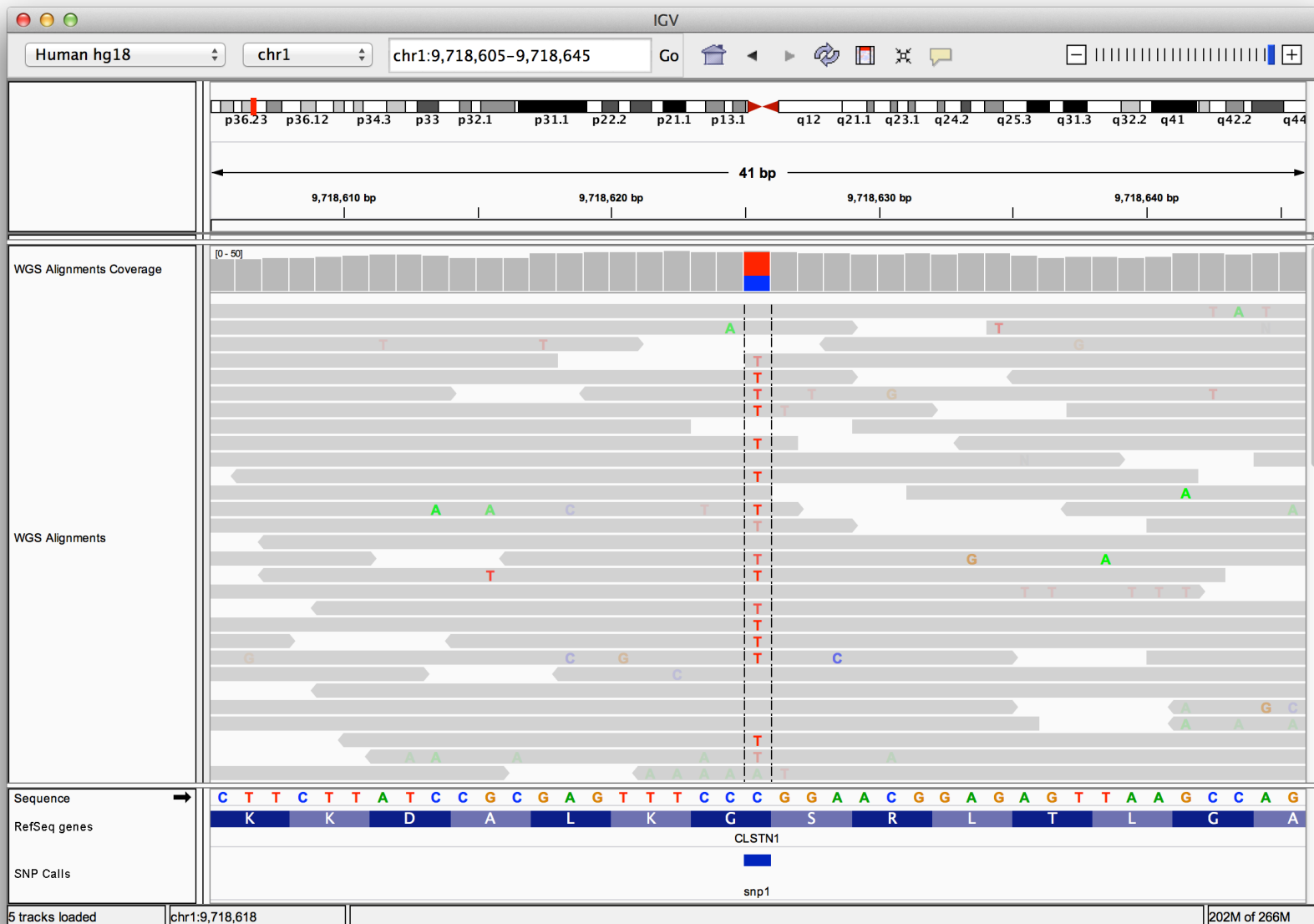
Viewing alignments – Zoom in



SNVs and Structural variations

- Important metrics for evaluating the validity of SNVs:
 - Coverage
 - Amount of support
 - Strand bias / PCR artifacts
 - Mapping qualities
 - Base qualities
- Important metrics for evaluating SVs:
 - Coverage
 - Insert size
 - Read pair orientation

Viewing SNPs and SNVs



Viewing SNPs and SNVs



Viewing Structural Events

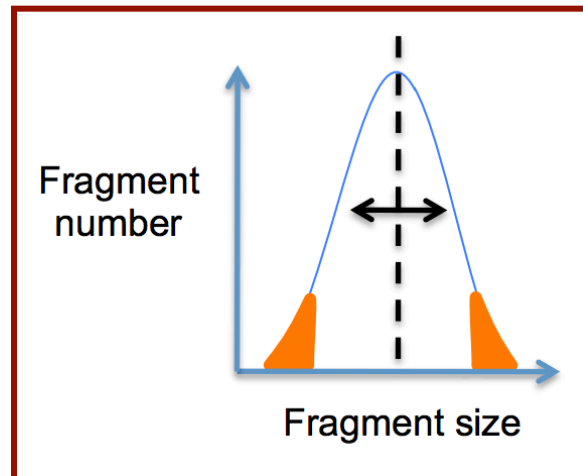
- Paired reads can yield evidence for genomic “structural events”, such as deletions, translocations, and inversions.
- Alignment coloring options help highlight these events based on:
 - Inferred insert size (template length)
 - Pair orientation (relative strand of pair)

Paired-end sequencing

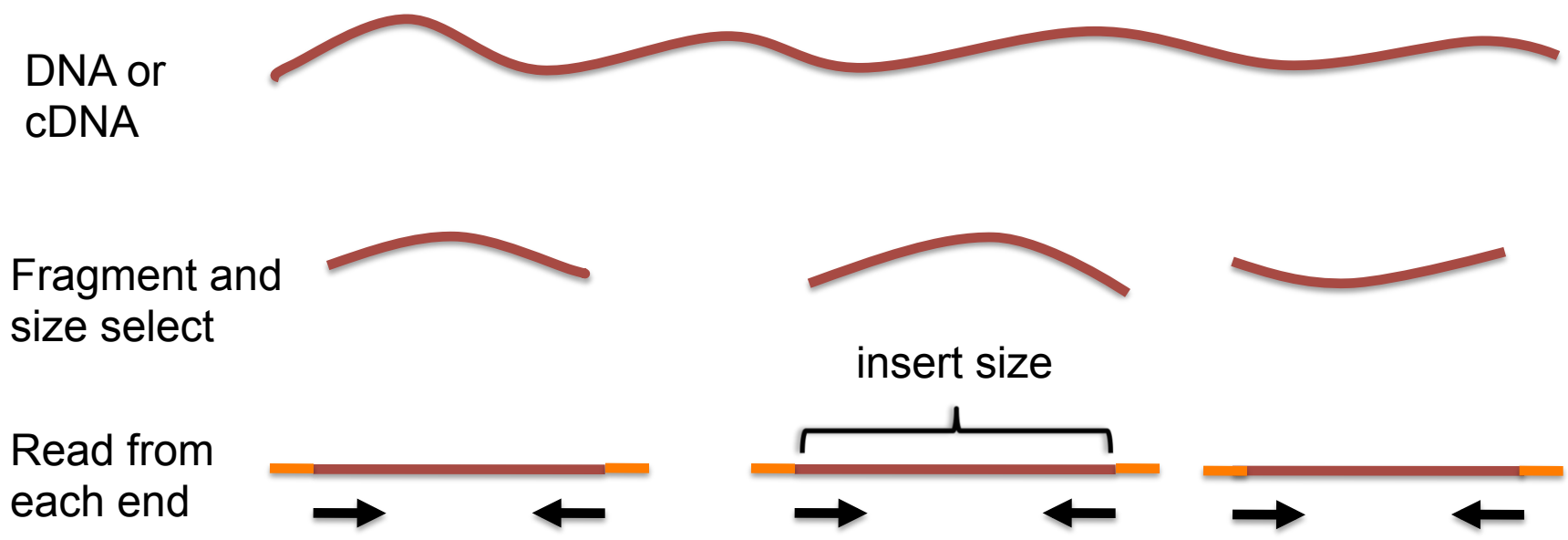
DNA or
cDNA



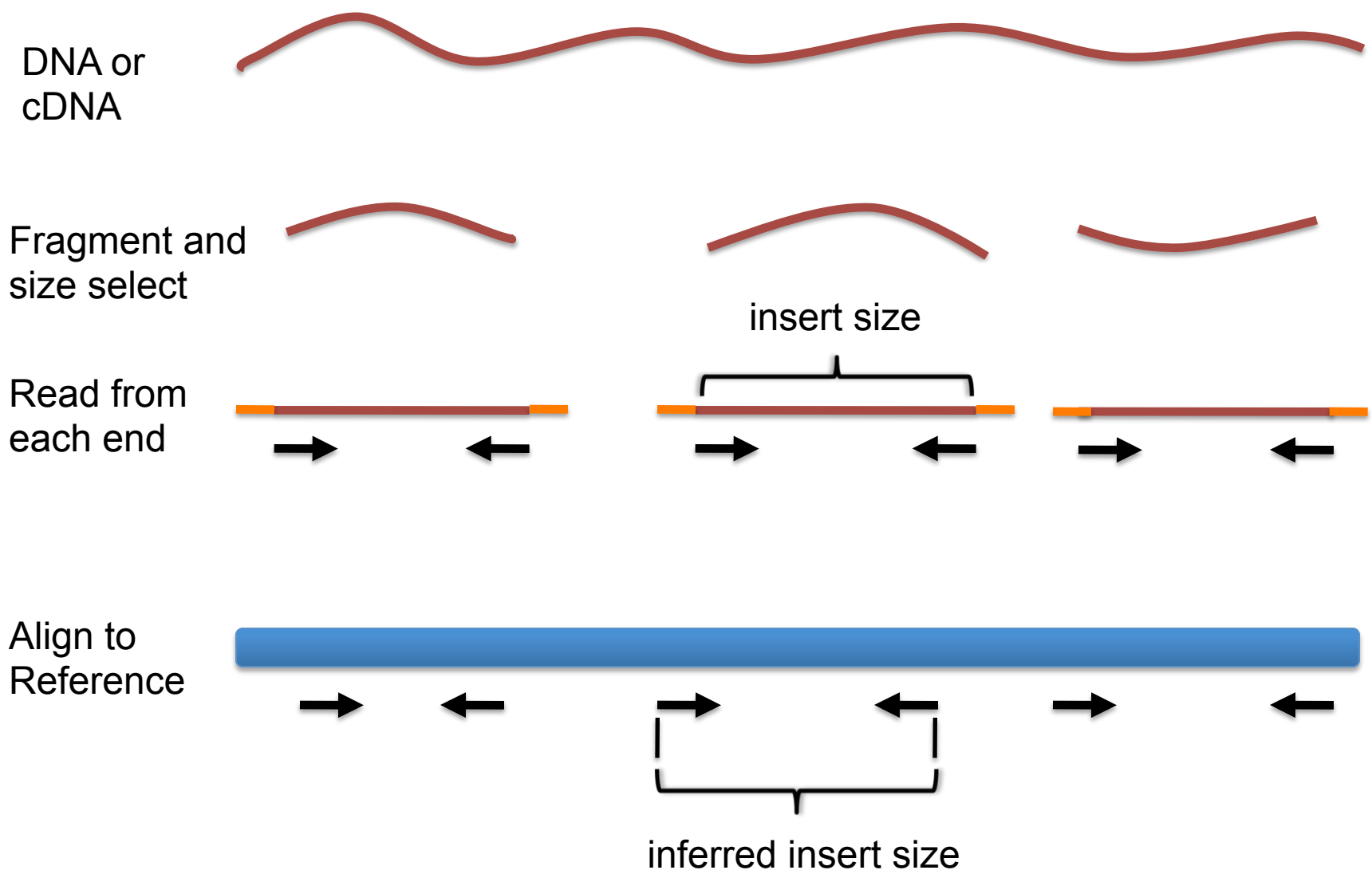
Fragment and
size select



Paired-end sequencing



Paired-end sequencing



Interpreting inferred insert size

The “inferred insert size” can be used to detect structural variants including

- Deletions
- Insertions
- Inter-chromosomal rearrangements: (Undefined insert size)

Deletion

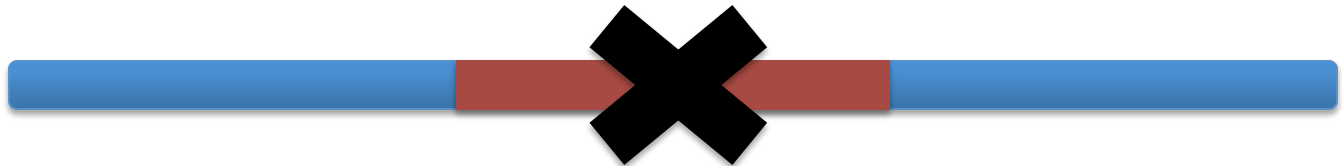
What is the effect of a deletion on inferred insert size?

Deletion

Reference
Genome



Subject



Deletion

Reference
Genome



Subject

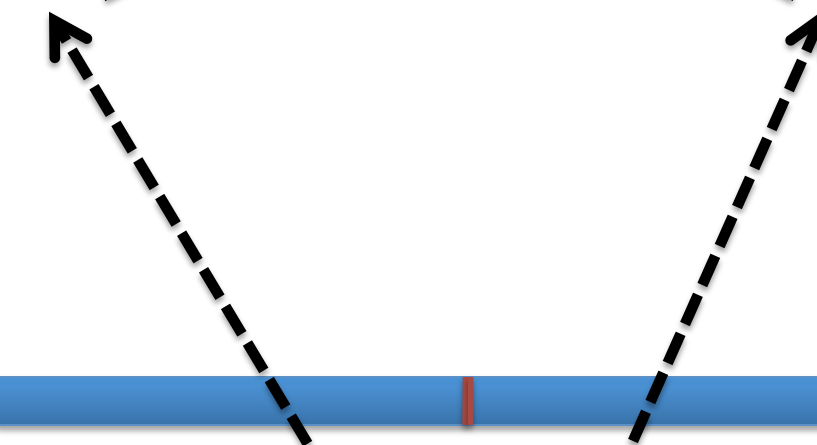


Deletion

Reference
Genome



Subject



Deletion

Inferred insert size is $>$ expected value

Reference
Genome



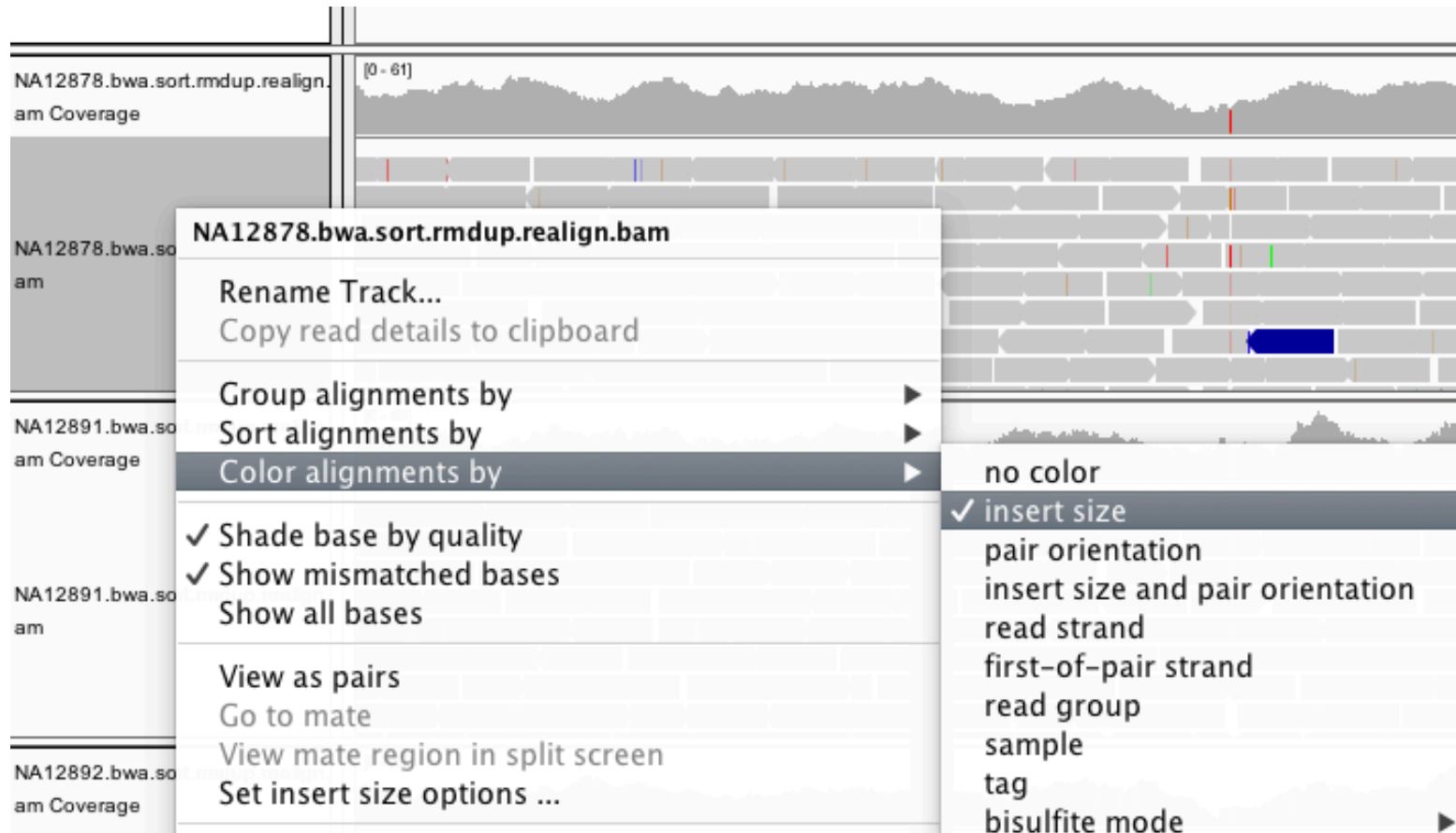
inferred insert size

Subject



expected insert size

Color by insert size

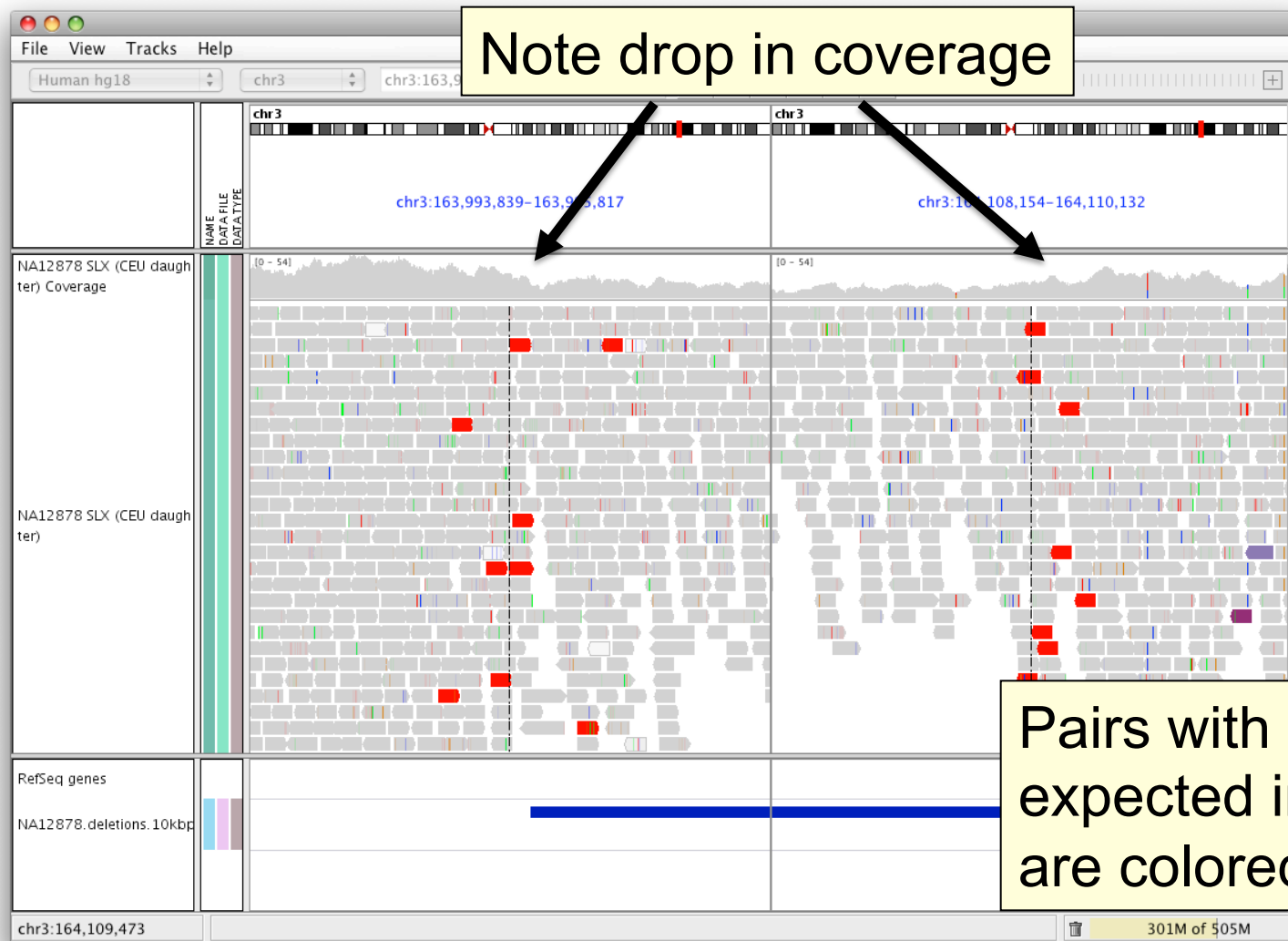


The screenshot shows a genome browser interface with several tracks. The top track is labeled 'NA12878.bwa.sort.rmdup.realign.am Coverage' and shows a coverage plot with a range of [0 - 61]. Below it are alignment tracks for 'NA12878.bwa.sort.rmdup.realign.am', 'NA12891.bwa.sort.rmdup.realign.am Coverage', 'NA12891.bwa.sort.rmdup.realign.am', and 'NA12892.bwa.sort.rmdup.realign.am Coverage'. A context menu is open over the 'NA12878.bwa.sort.rmdup.realign.am' track, listing various actions. The 'Color alignments by' option is selected, and its sub-menu is also open, showing options like 'no color', 'insert size', 'pair orientation', etc. The 'insert size' option is currently selected in the sub-menu.



NA12878.bwa.sort.rmdup.realign.am

- Rename Track...
- Copy read details to clipboard
- Group alignments by ▶
- Sort alignments by ▶
- Color alignments by ▶**
 - no color
 - insert size**
 - pair orientation
 - insert size and pair orientation
 - read strand
 - first-of-pair strand
 - read group
 - sample
 - tag
 - bisulfite mode ▶
- ✓ Shade base by quality
- ✓ Show mismatched bases
- Show all bases
- View as pairs
- Go to mate
- View mate region in split screen
- Set insert size options ...

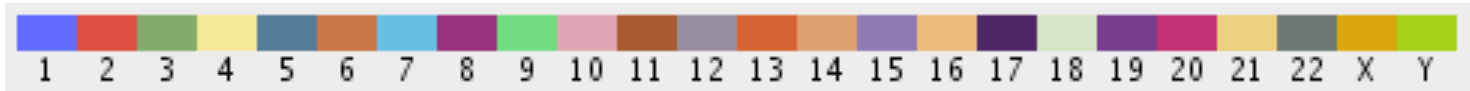
Deletion



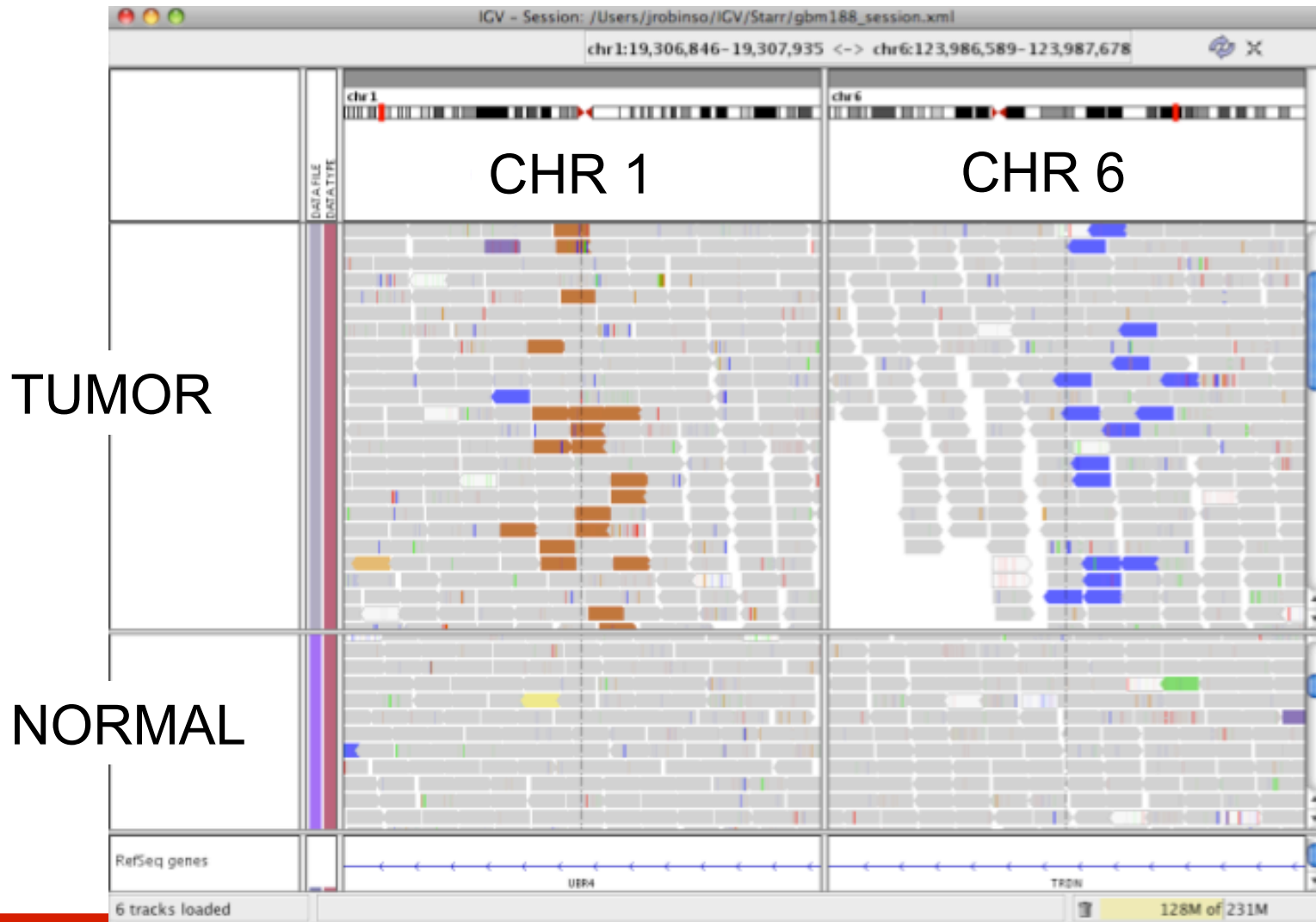
Insert size color scheme

- Smaller than expected insert size: 
- Larger than expected insert size: 
- Pairs on different chromosomes

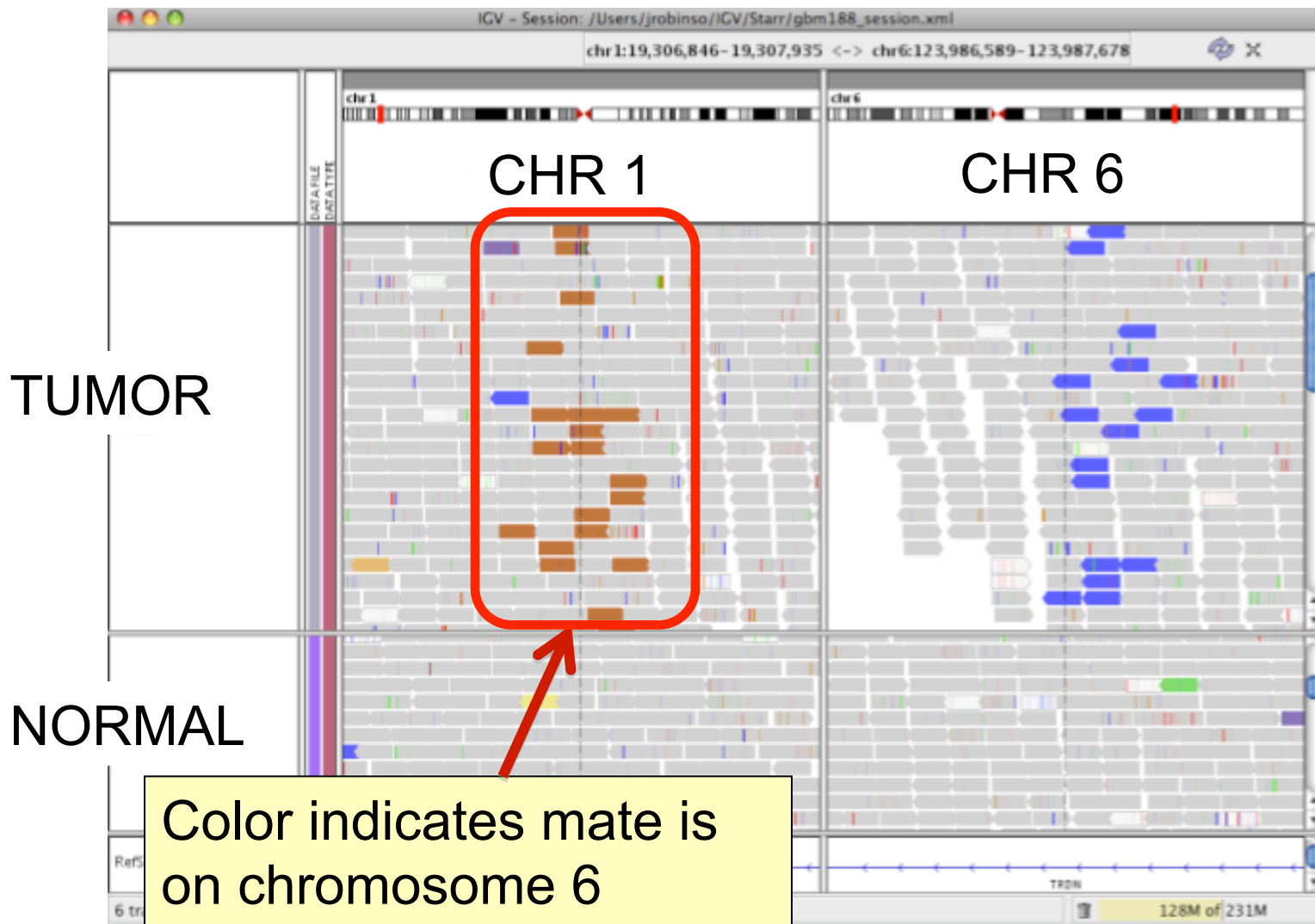
Each end colored by chromosome of its mate



Rearrangement



Rearrangement



Interpreting Read-Pair Orientations

Orientation of paired reads can reveal structural events:

- Inversions
- Duplications
- Translocations
- Complex rearrangements

Orientation is defined in terms of

- read strand, left *vs* right, *and*
- read order, first *vs* second

Inversion

Reference
genome



Inversion

Reference
genome

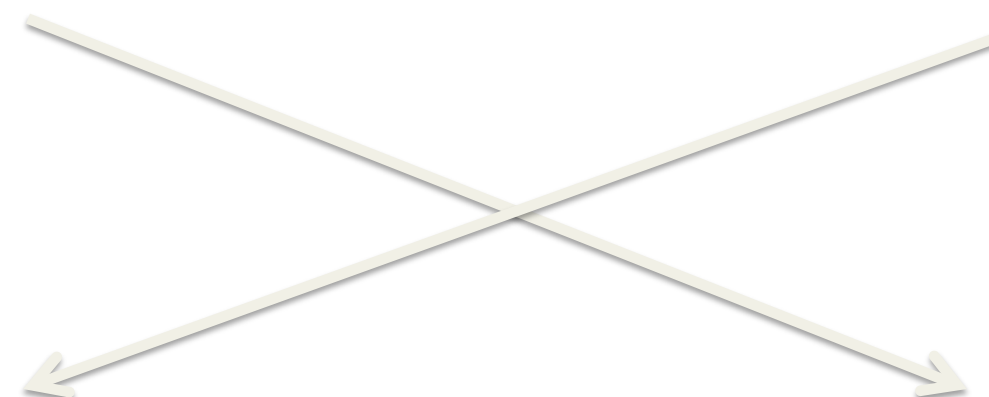


Inversion

Reference
Genome



Subject



Inversion

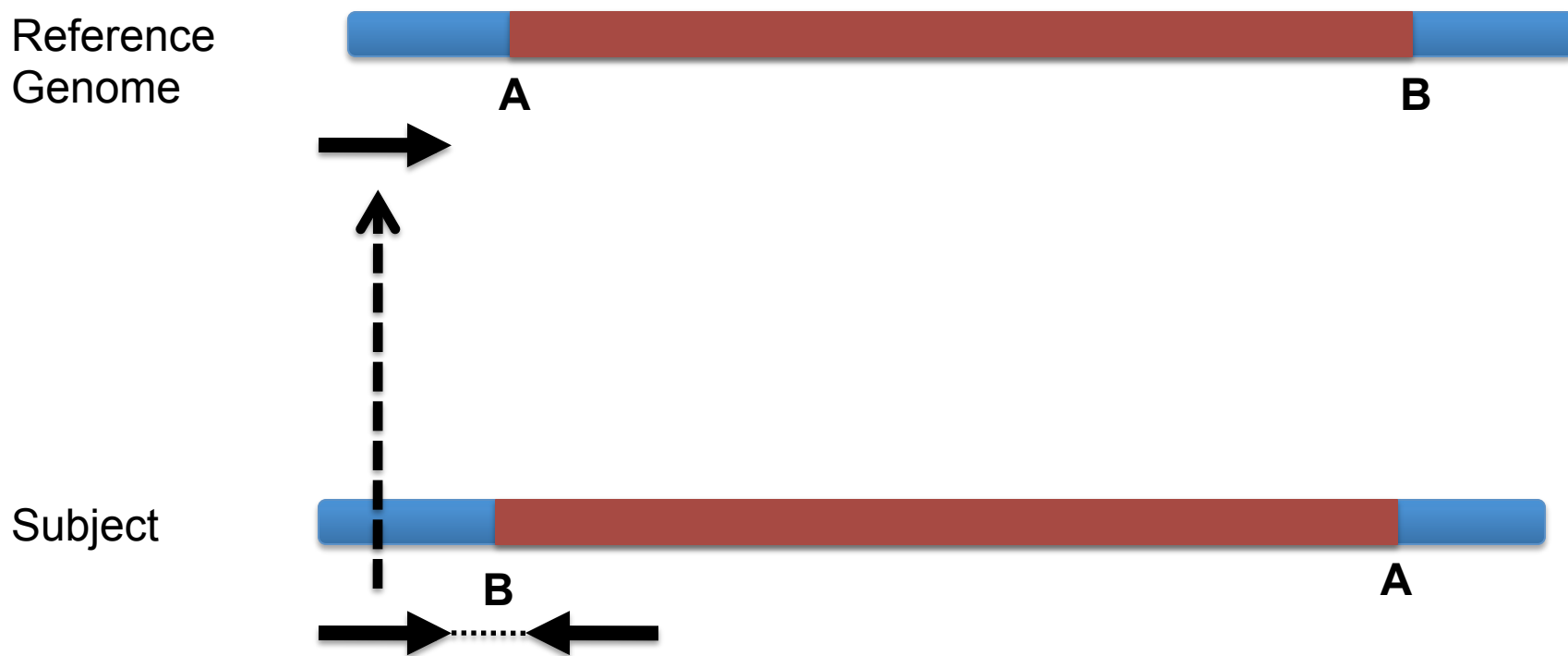
Reference
Genome



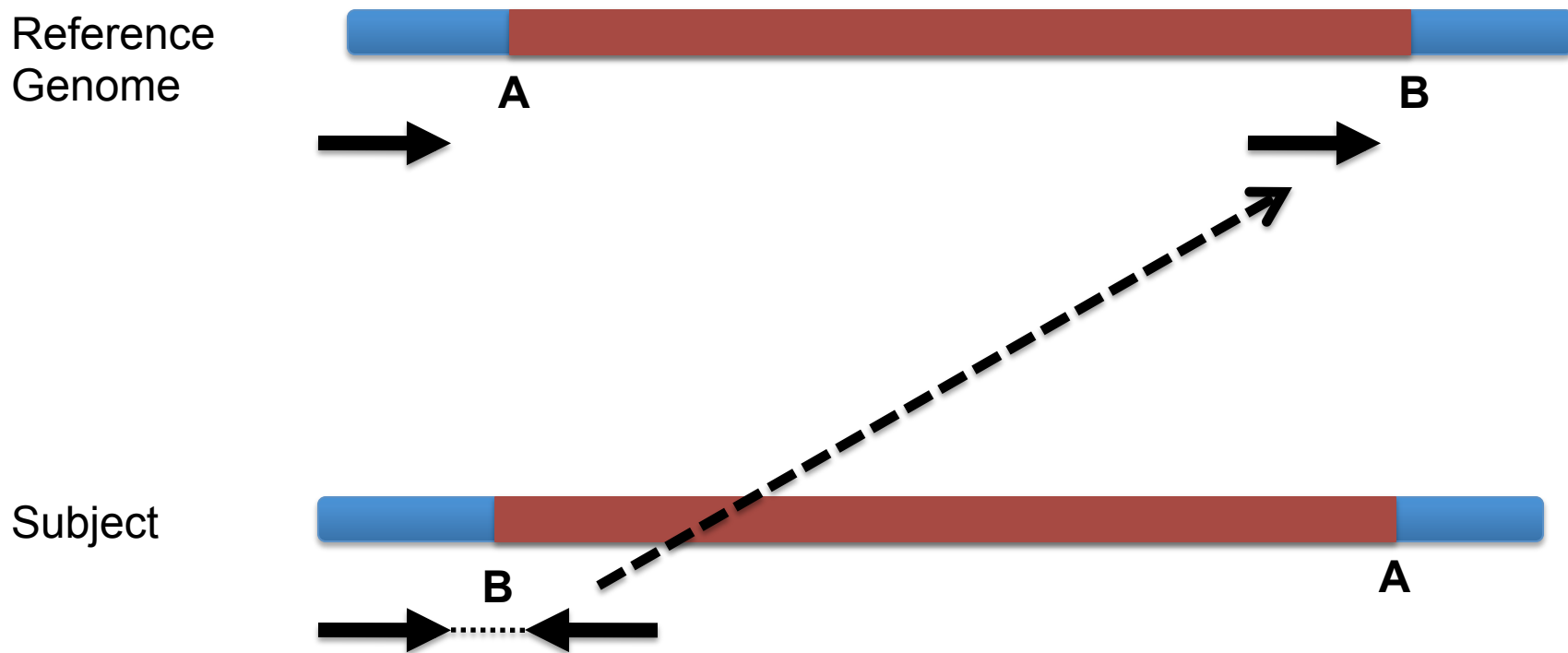
Subject



Inversion

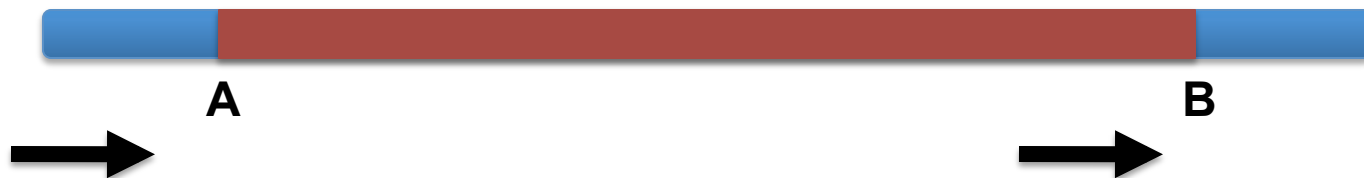


Inversion



Inversion

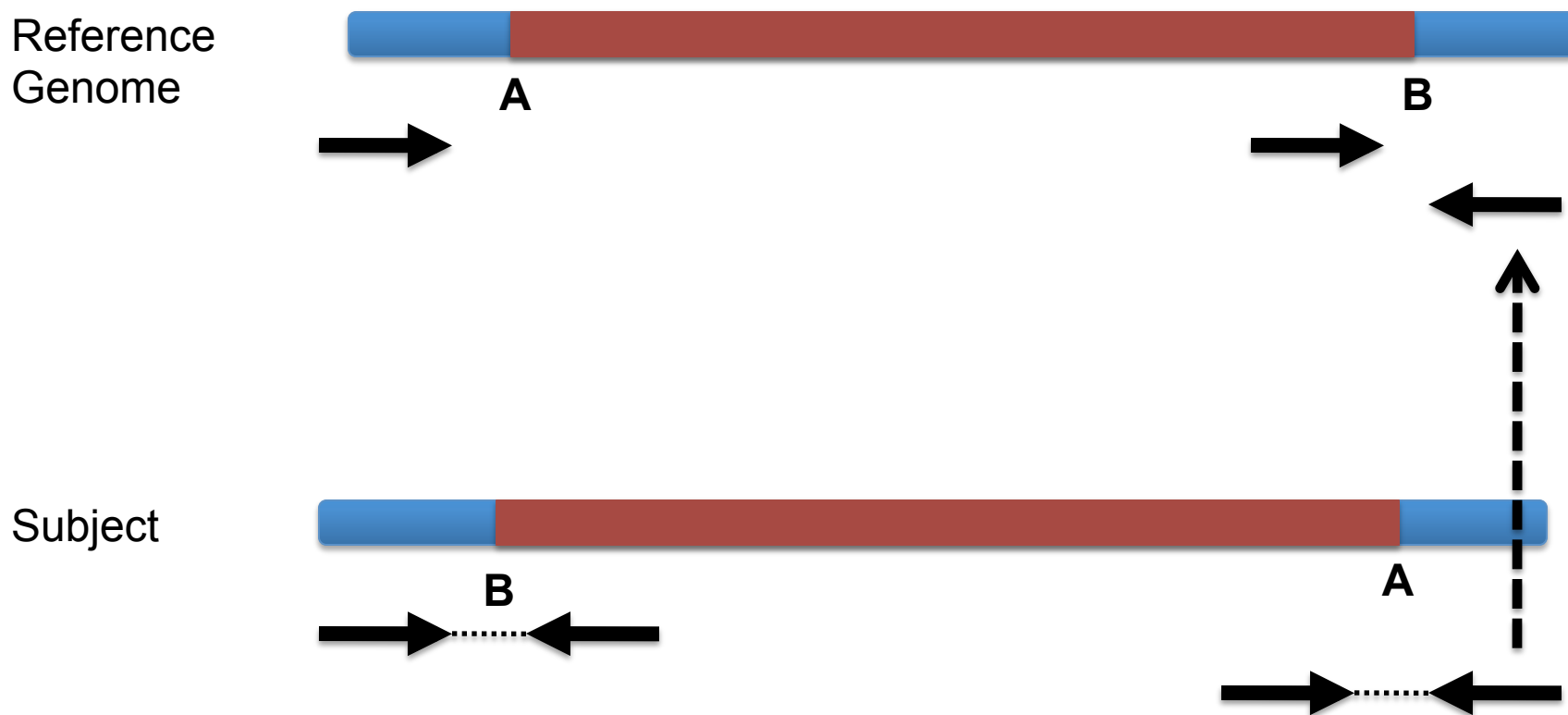
Reference
Genome



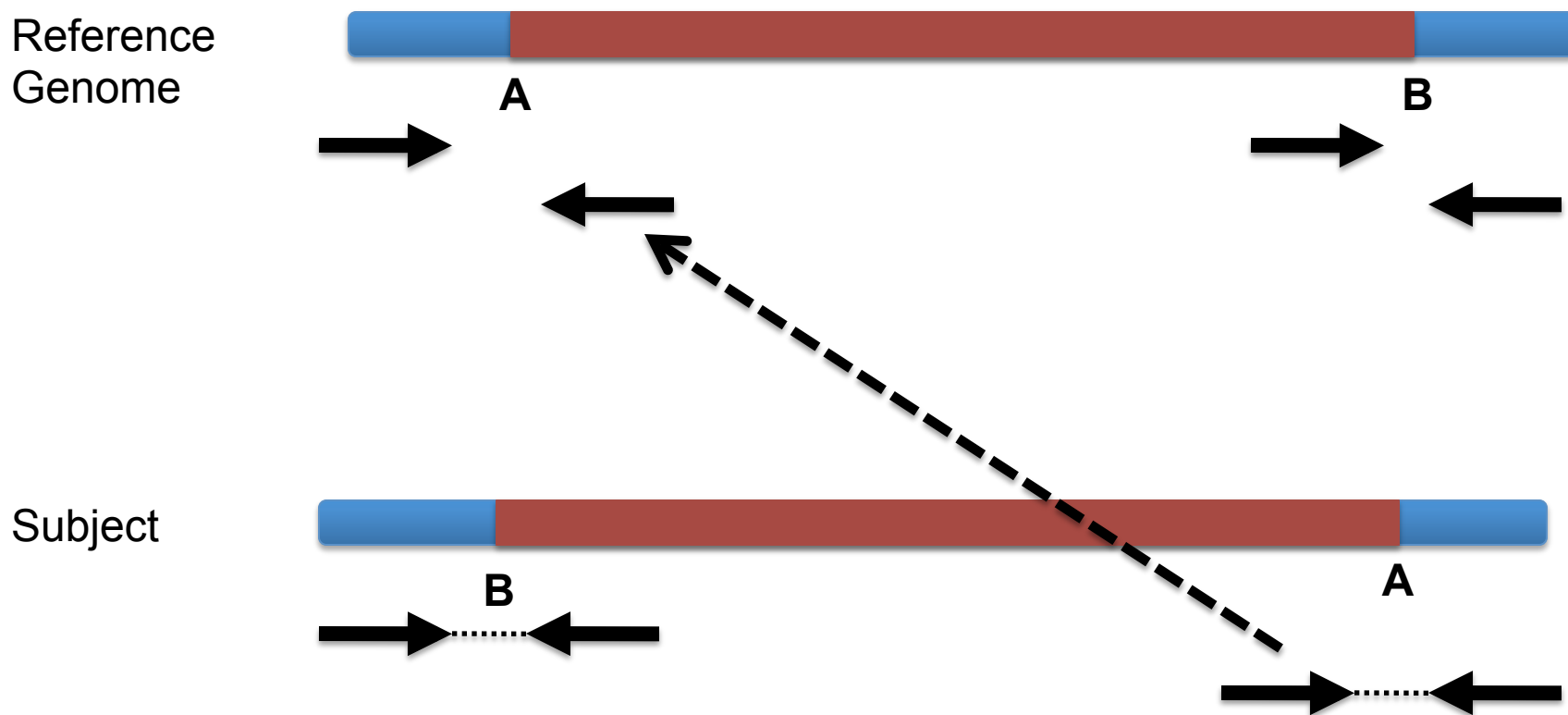
Subject



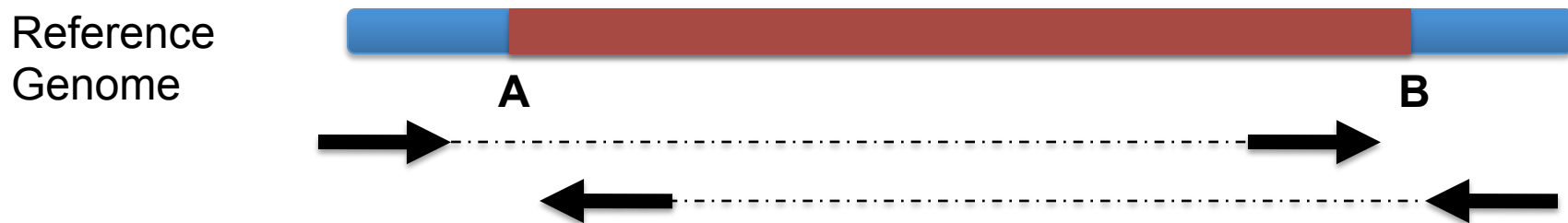
Inversion



Inversion



Inversion



Inversion



Anomaly: expected orientation of pair is inward facing ($\longrightarrow \longleftarrow$)

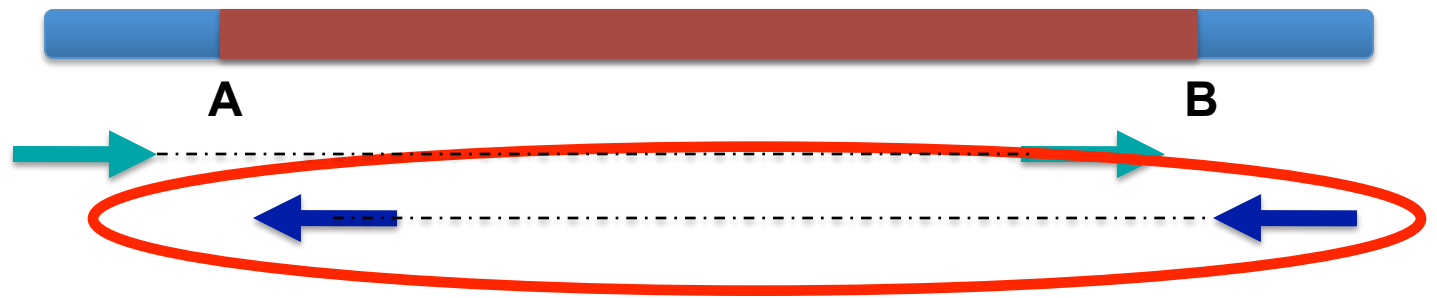
Inversion



“Left” side pair

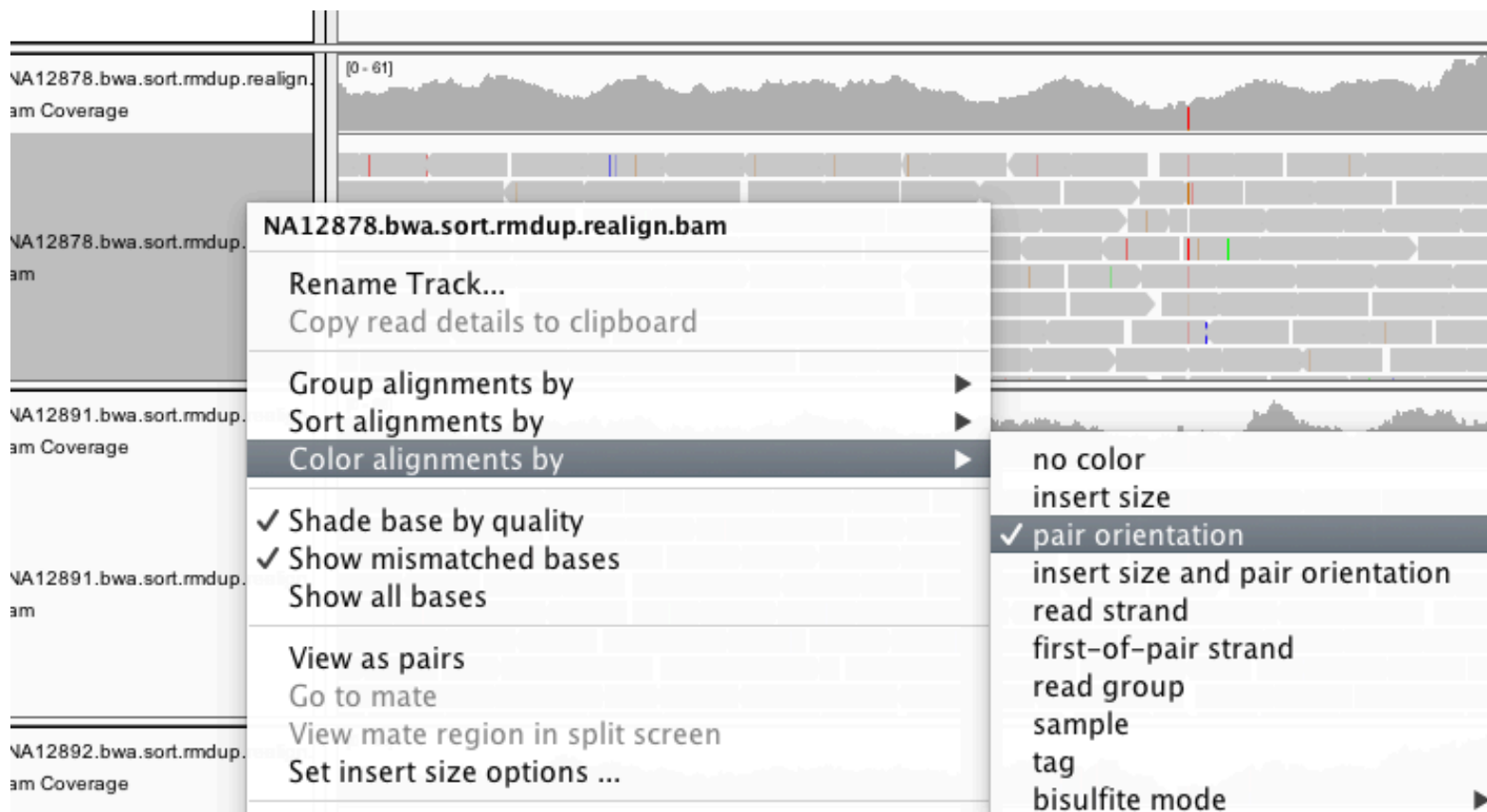
Inversion

Reference
Genome



“Right” side pair

Color by pair orientation

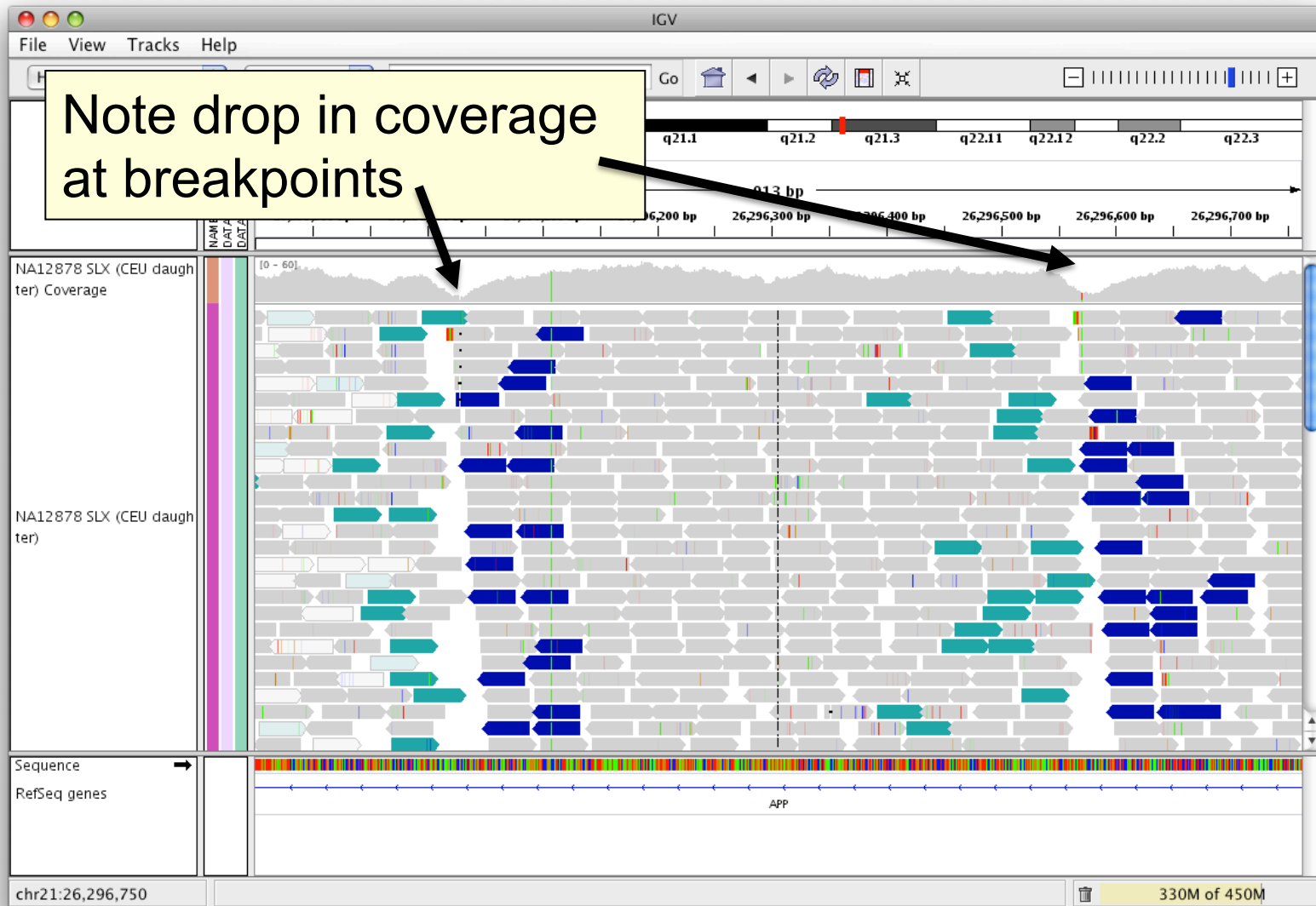


The screenshot shows a genome browser interface with a track titled "NA12878.bwa.sort.rmdup.realign.bam". A context menu is open over the track, listing various options. The "Color alignments by" option is selected, and its sub-menu is visible, showing "pair orientation" as the chosen option. The background shows a coverage plot and a read alignment track with various colored bars representing different pair orientations.

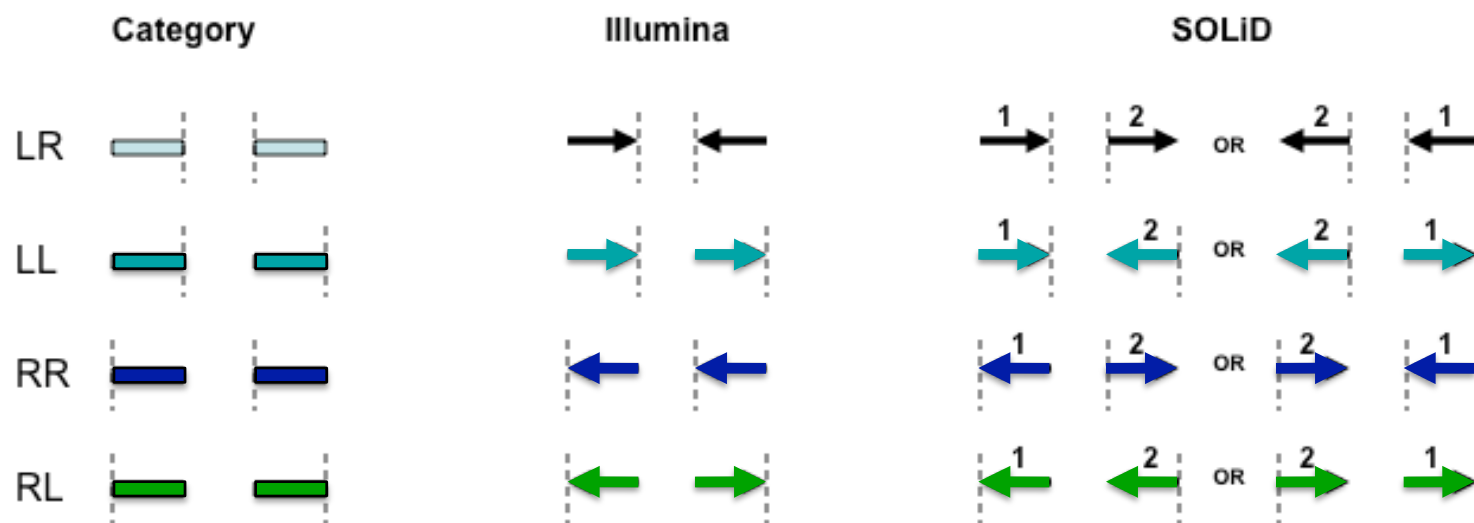
NA12878.bwa.sort.rmdup.realign.bam

- Rename Track...
- Copy read details to clipboard
- Group alignments by ▶
- Sort alignments by ▶
- Color alignments by ▶**
 - no color
 - insert size
 - pair orientation**
 - insert size and pair orientation
 - read strand
 - first-of-pair strand
 - read group
 - sample
 - tag
 - bisulfite mode ▶
- ✓ Shade base by quality
- ✓ Show mismatched bases
- Show all bases
- View as pairs
- Go to mate
- View mate region in split screen
- Set insert size options ...

Inversion



Interpretation of read pair orientations



- LR Normal reads.
The reads are left and right (respectively) of the unsequenced part of the sequenced DNA fragment when aligned back to the reference genome.
- LL,RR Implies inversion in sequenced DNA with respect to reference.
- RL Implies duplication or translocation with respect to reference.

These categories only apply to reads where both mates map to the same chromosome.

Figure courtesy of Bob Handsaker

IGV hands-on tutorial

[https://github.com/griffithlab/
rnaseq_tutorial/wiki/IGV-Tutorial](https://github.com/griffithlab/rnaseq_tutorial/wiki/IGV-Tutorial)

Break