Sequence data visualization and IGV

Chris Miller

BFX-workshop week 05

Adapted from:

Malachi Griffith, Obi Griffith, Fouad Yousif

High-Throughput Biology: From Sequence to Networks

https://github.com/griffithlab/rnaseq_tutorial_wiki/blob/master/LectureFiles/cbw-cshl/2017/IGV_Tutorial_Brief.pptx

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

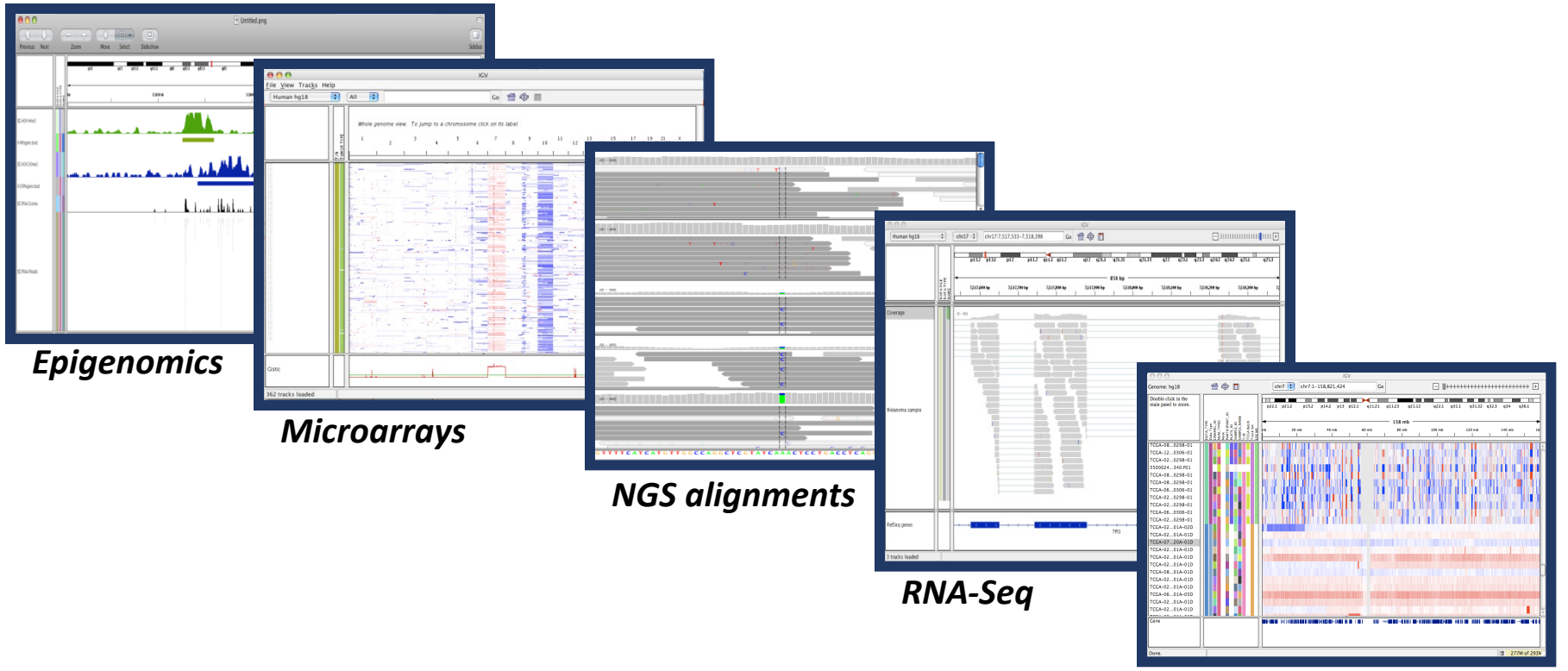
Trackster (part of Galaxy)

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



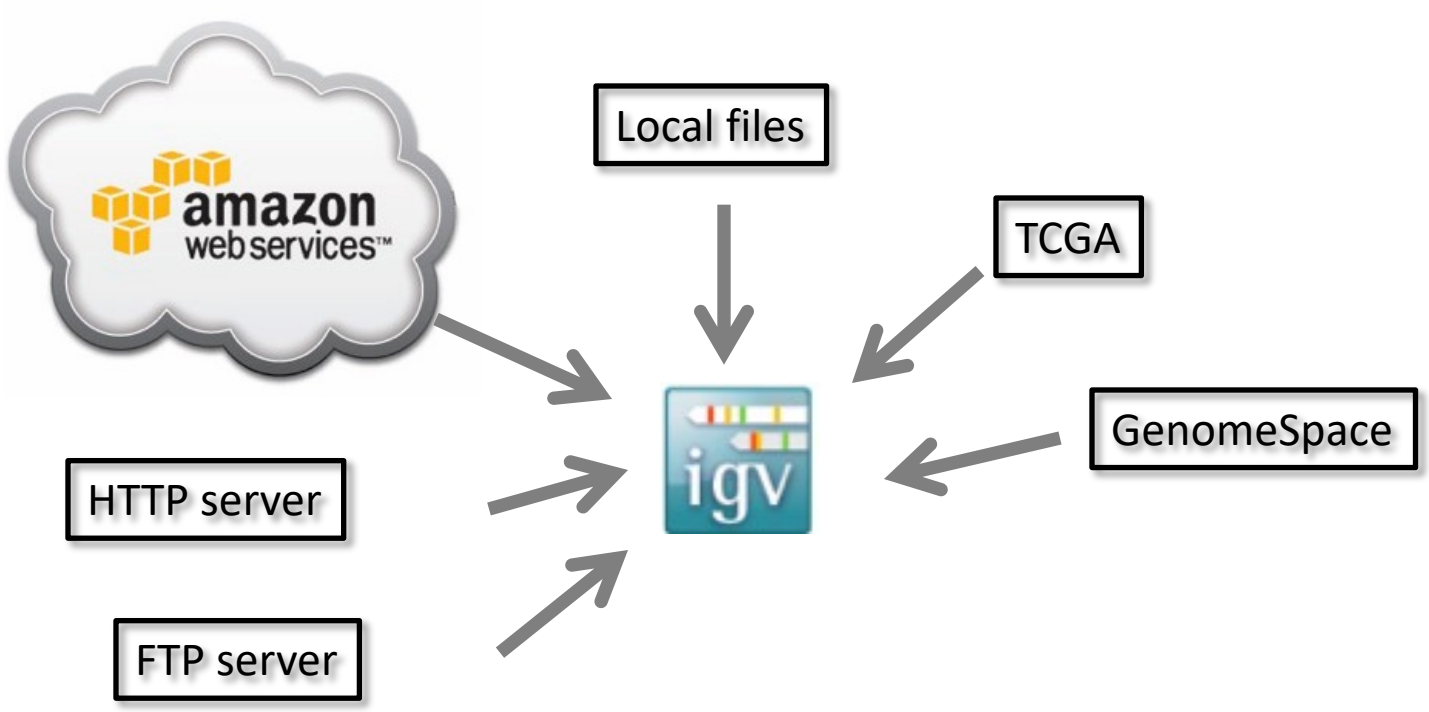
http://www.broadinstitute.org/igv

Features of IGC

With IGV you can…

- intuitive, easy-to-use interface
- Scales well to large data
- Integrate multiple data types
- View data from multiple locations:
  - local, remote, and “cloud-based”.
- Some automation of tasks using command-line interface

IGV data sources

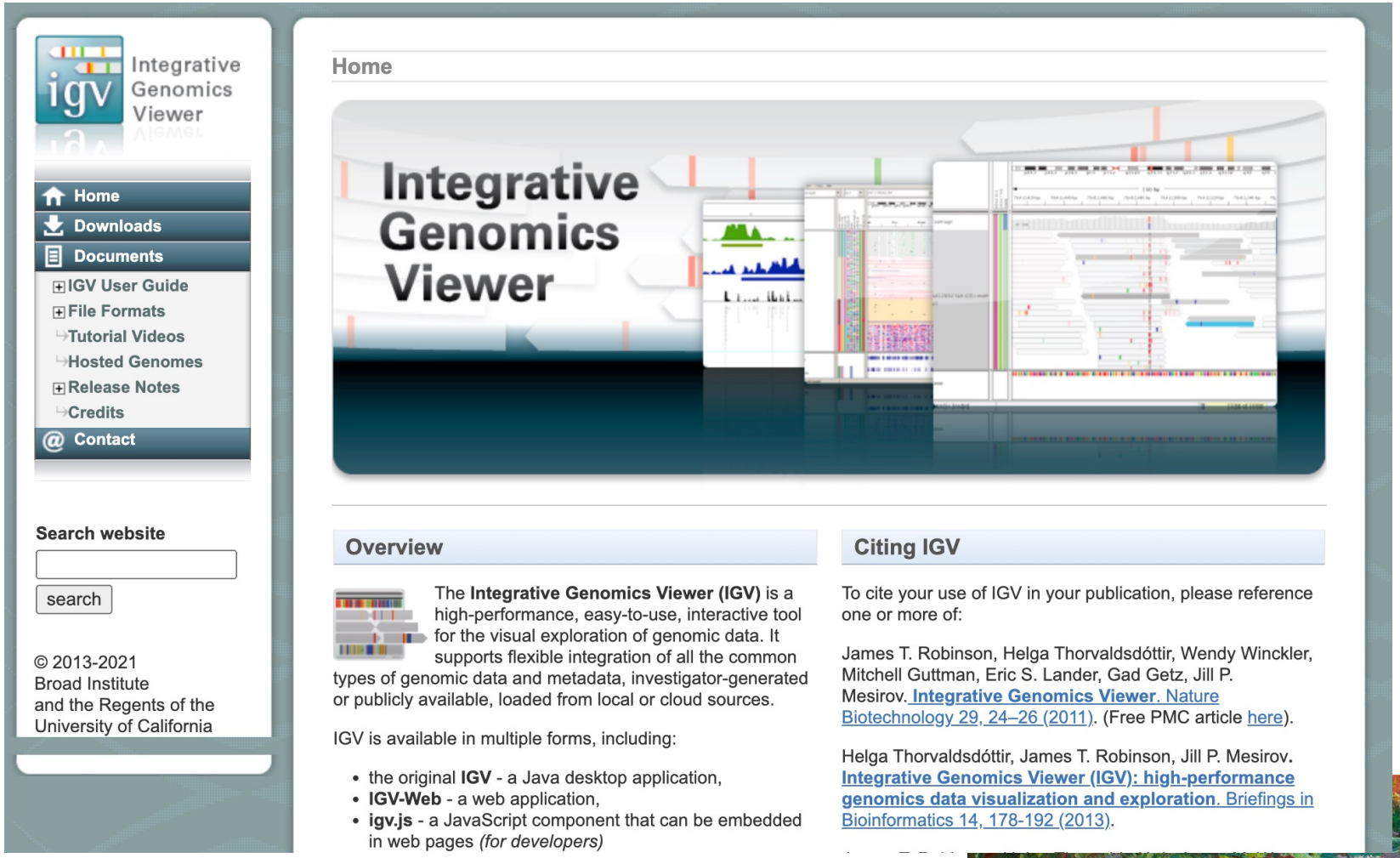


- View **local** files without uploading.
- View **remote** files without downloading the whole dataset.

Using IGV: the basics

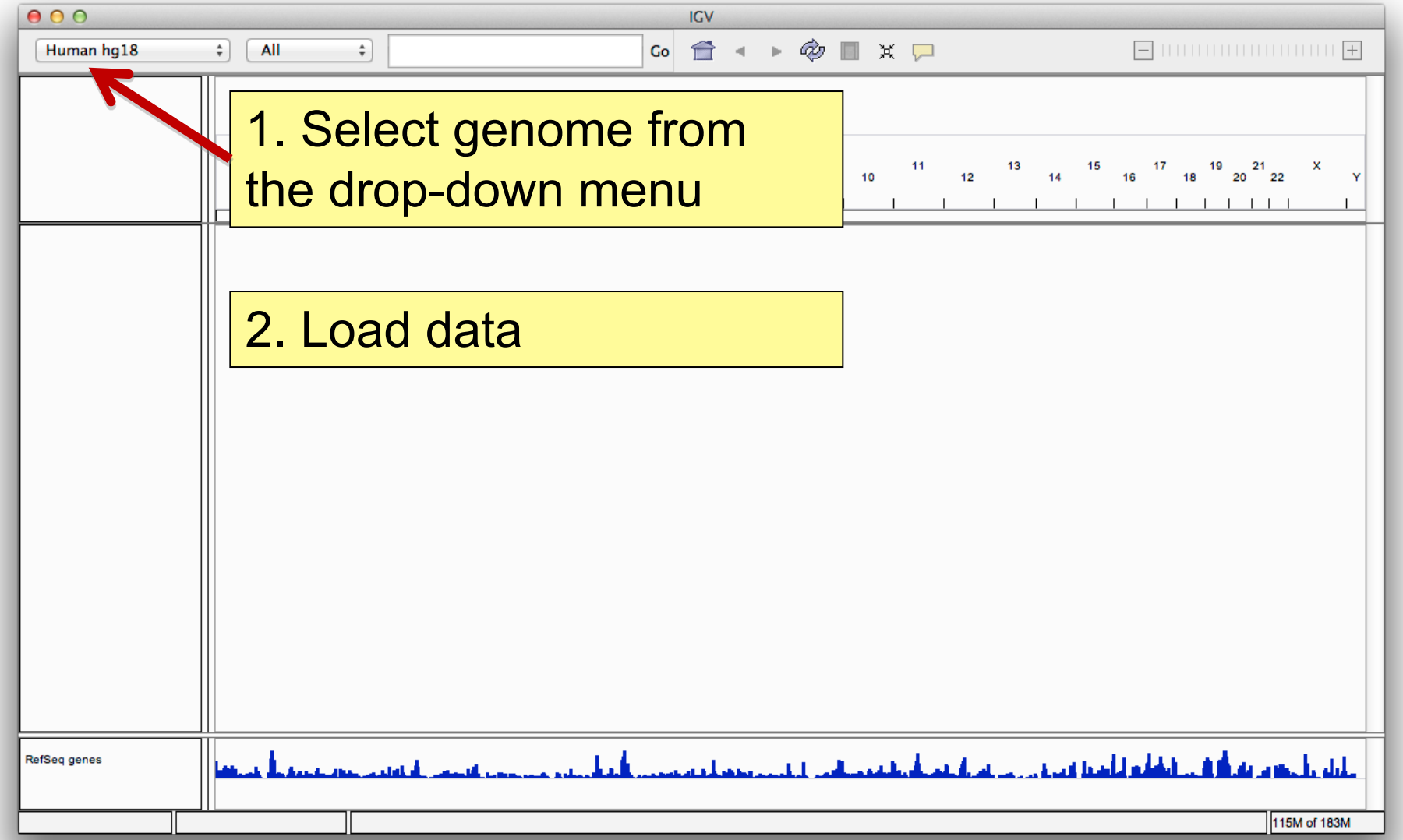
- Launch IGV
- Select a reference genome
- Load data
- Navigate through the data
  - WGS data
    - SNVs
    - structural variations

Launch IGV

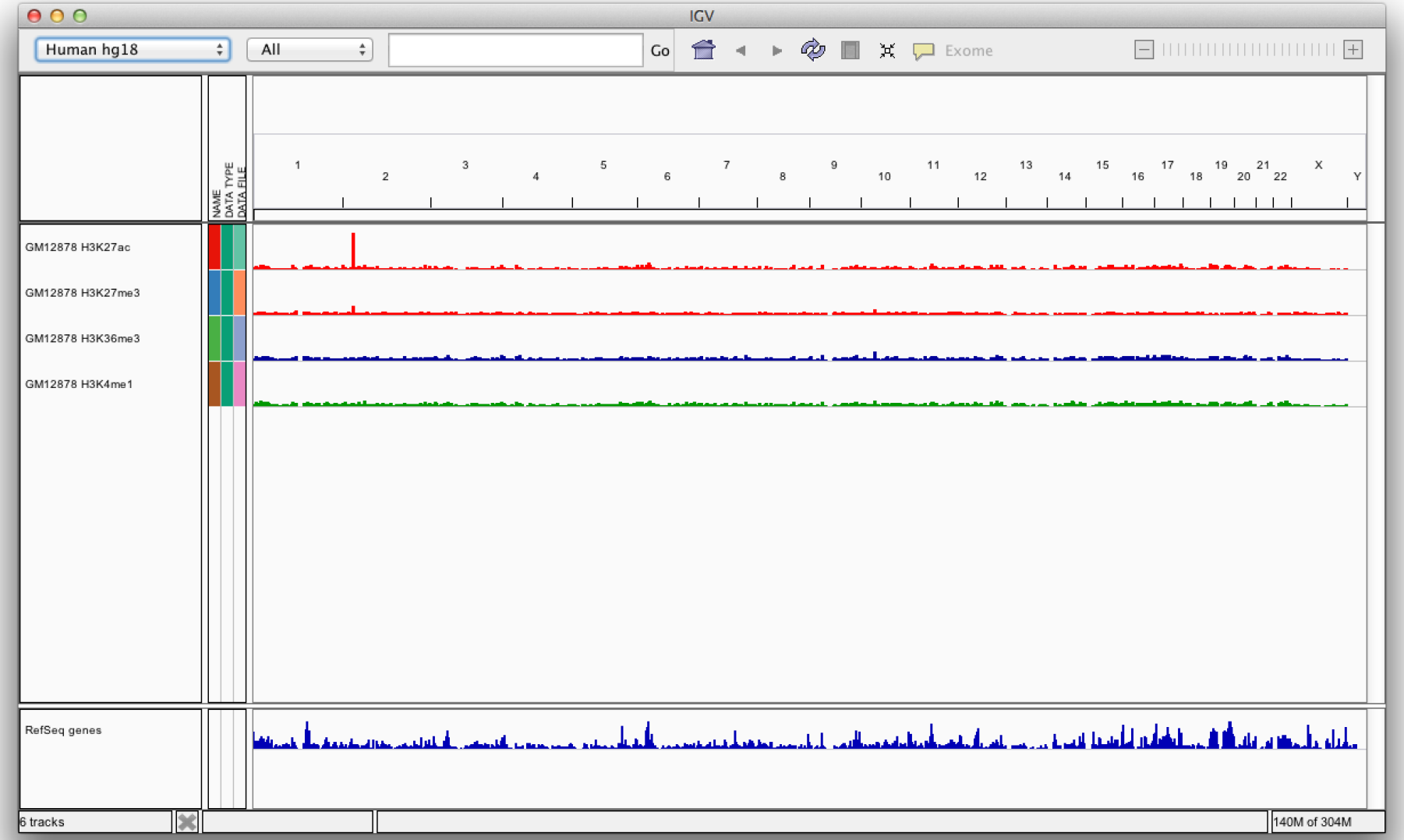

Integrative Genomics Viewer
Home
Downloads
Documents
IGV User Guide
File Formats
Tutorial Videos
Hosted Genomes
Release Notes
Credits
Contact
Search website
search
© 2013-2021
Broad Institute
and the Regents of the
University of California
Home
Integrative Genomics Viewer
Overview
The Integrative Genomics Viewer (IGV) is a high-performance, easy-to-use, interactive tool for the visual exploration of genomic data. It supports flexible integration of all the common types of genomic data and metadata, investigator-generated or publicly available, loaded from local or cloud sources.
IGV is available in multiple forms, including:
the original IGV - a Java desktop application,
IGV-Web - a web application,
igv.js - a JavaScript component that can be embedded in web pages (for developers)
Citing IGV
To cite your use of IGV in your publication, please reference one or more of:
James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011). (Free PMC article here).
Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178-192 (2013).


igv

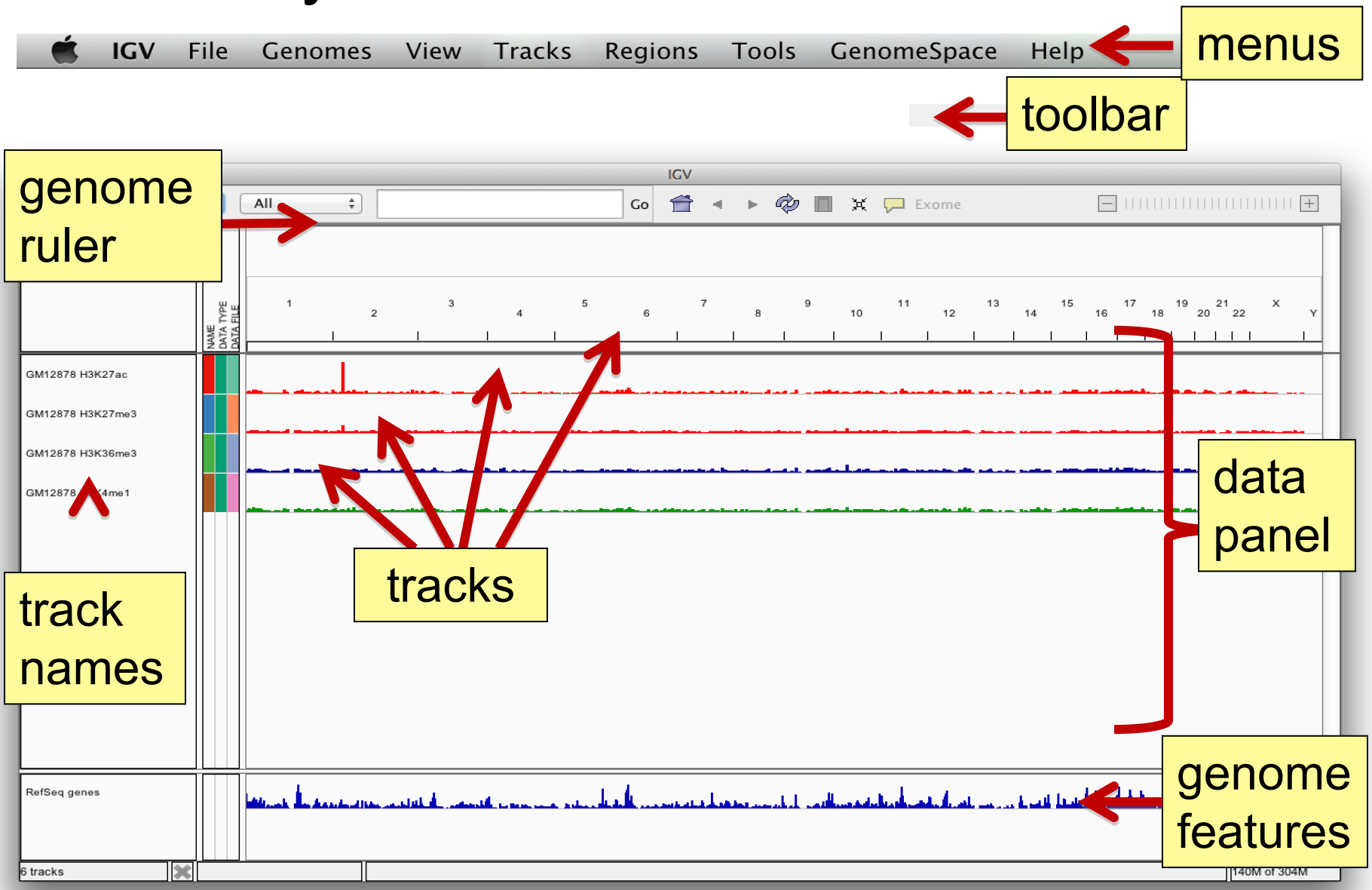
Launch IGV


IGV
Human hg18
All
Go
1. Select genome from the drop-down menu
2. Load data
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
RefSeq genes
115M of 183M

Screen layout



Screen layout



File formats and track types

- The **file format** defines the track type.
- The **track type** determines the display options

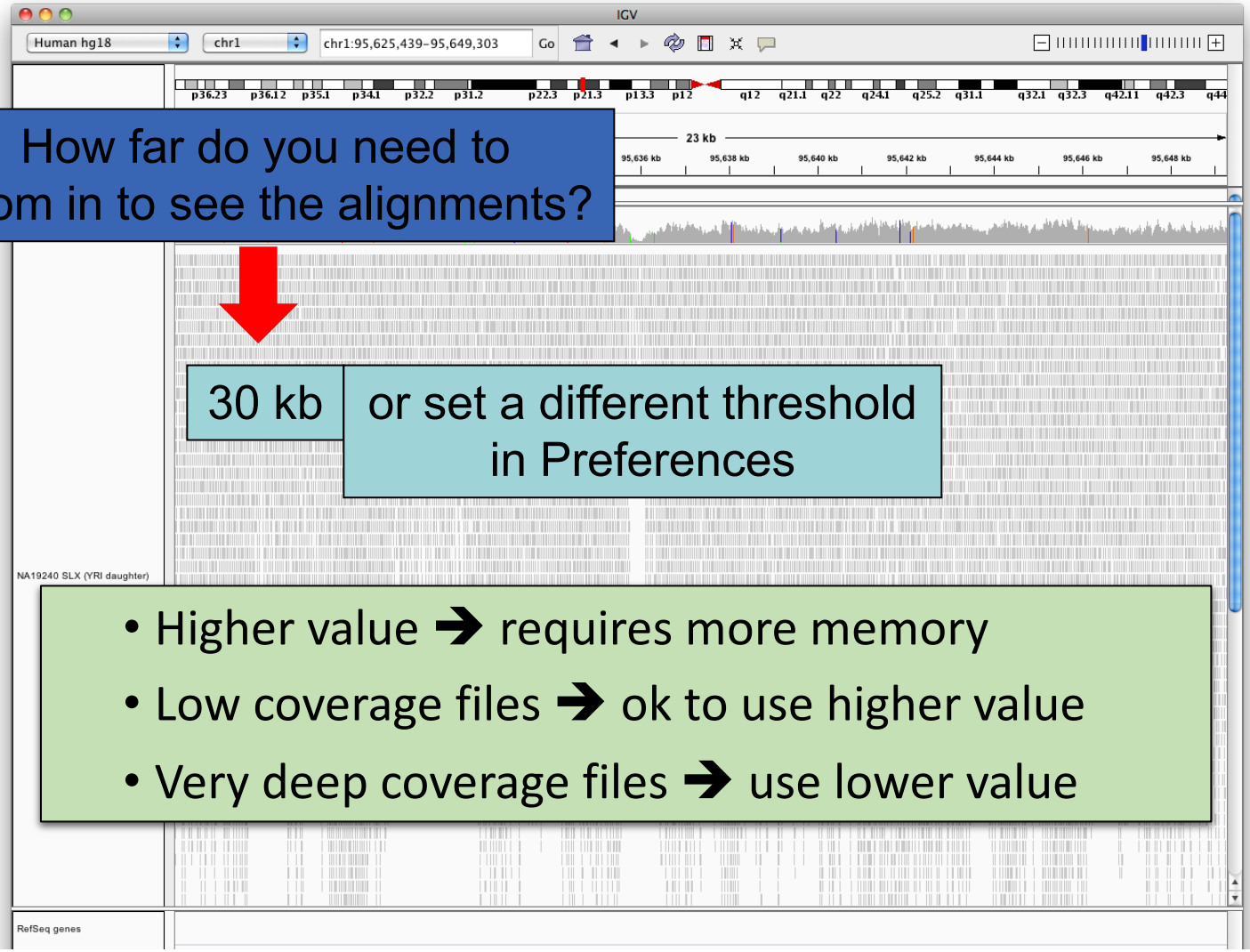
- BAM
- BED
- BedGraph
- bigBed
- bigWig
- Birdsuite Files
- CBS
- CN
- Cufflinks Files
- Custom File Formats
- Cytoband
- FASTA
- GCT
- genePred
- GFF
- GISTIC
- Goby
- GWAS
- IGV
- LOH
- MAF
- Merged BAM File (.bam.list)
- MUT
- PSL
- RES
- SAM
- Sample Information
- SEG
- SNP
- TAB
- TDF
- Track Line
- Type Line
- VCF
- WIG

- For current list see: www.broadinstitute.org/igv/FileFormats

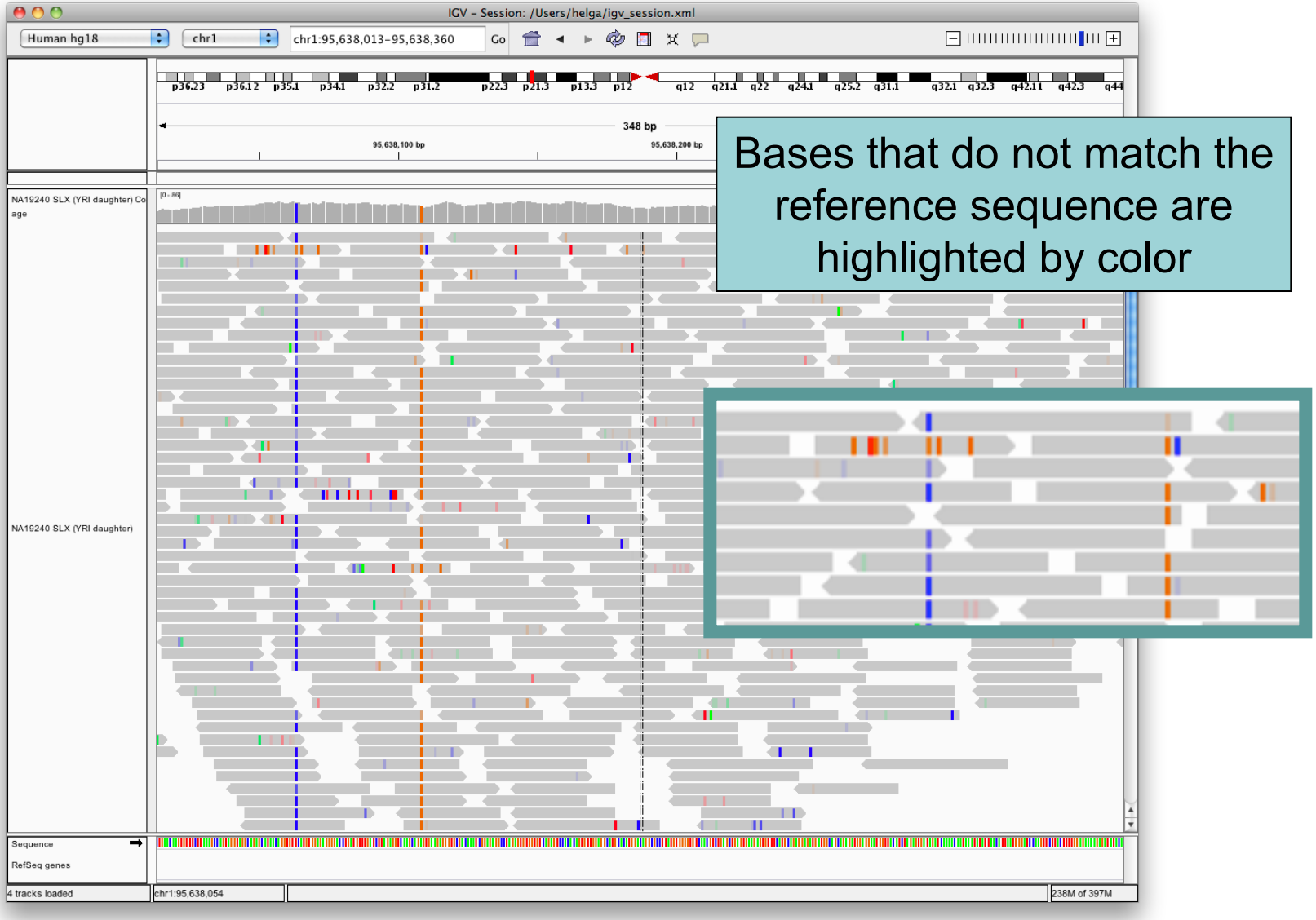
Viewing alignments

Whole chromosome view

Viewing alignments – Zoom in



Viewing alignments – Zoom in

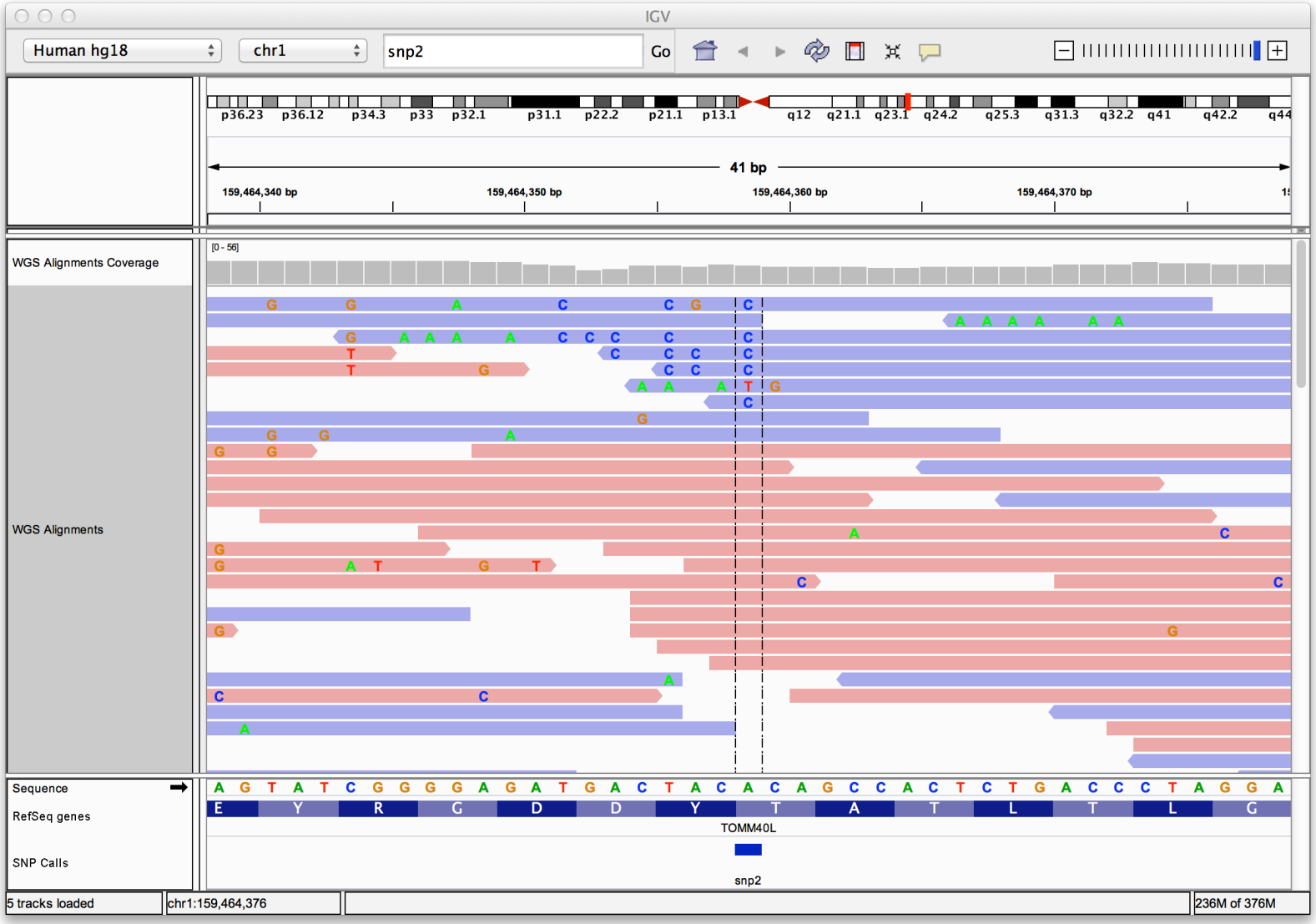


SNVs and Structural variations

- 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



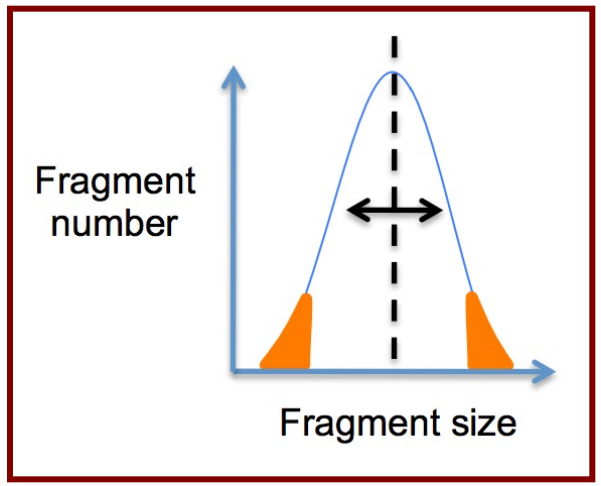
Viewing Structural Events

- 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)

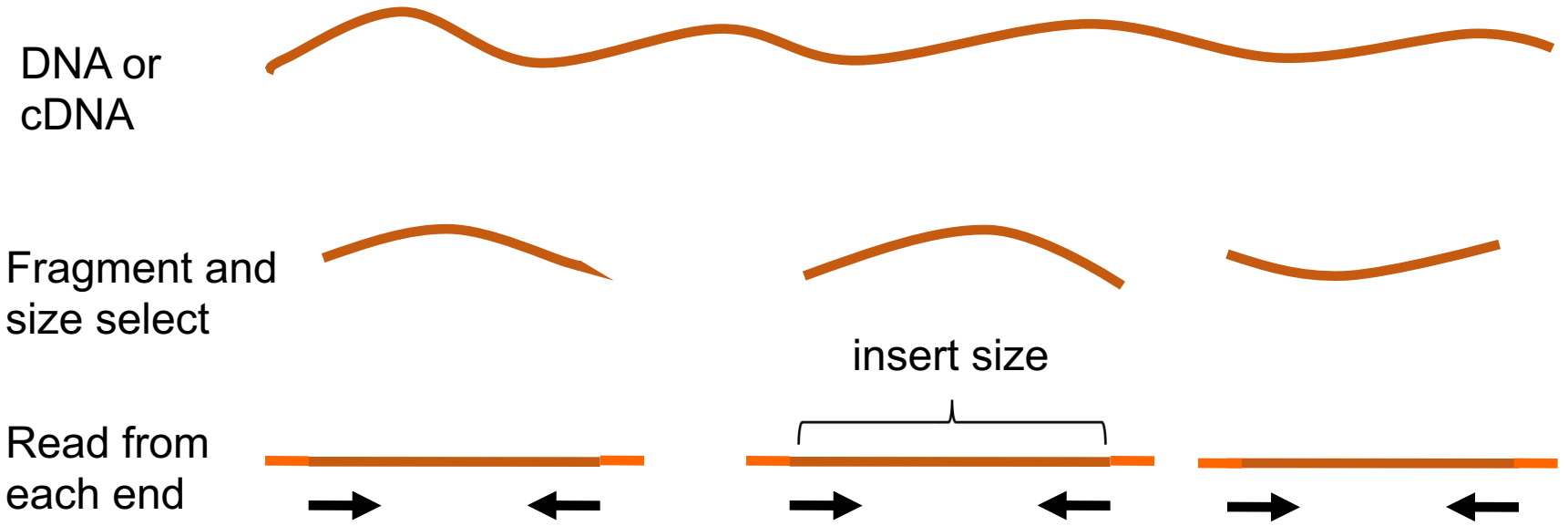
Paired-end sequencing

DNA or
cDNA

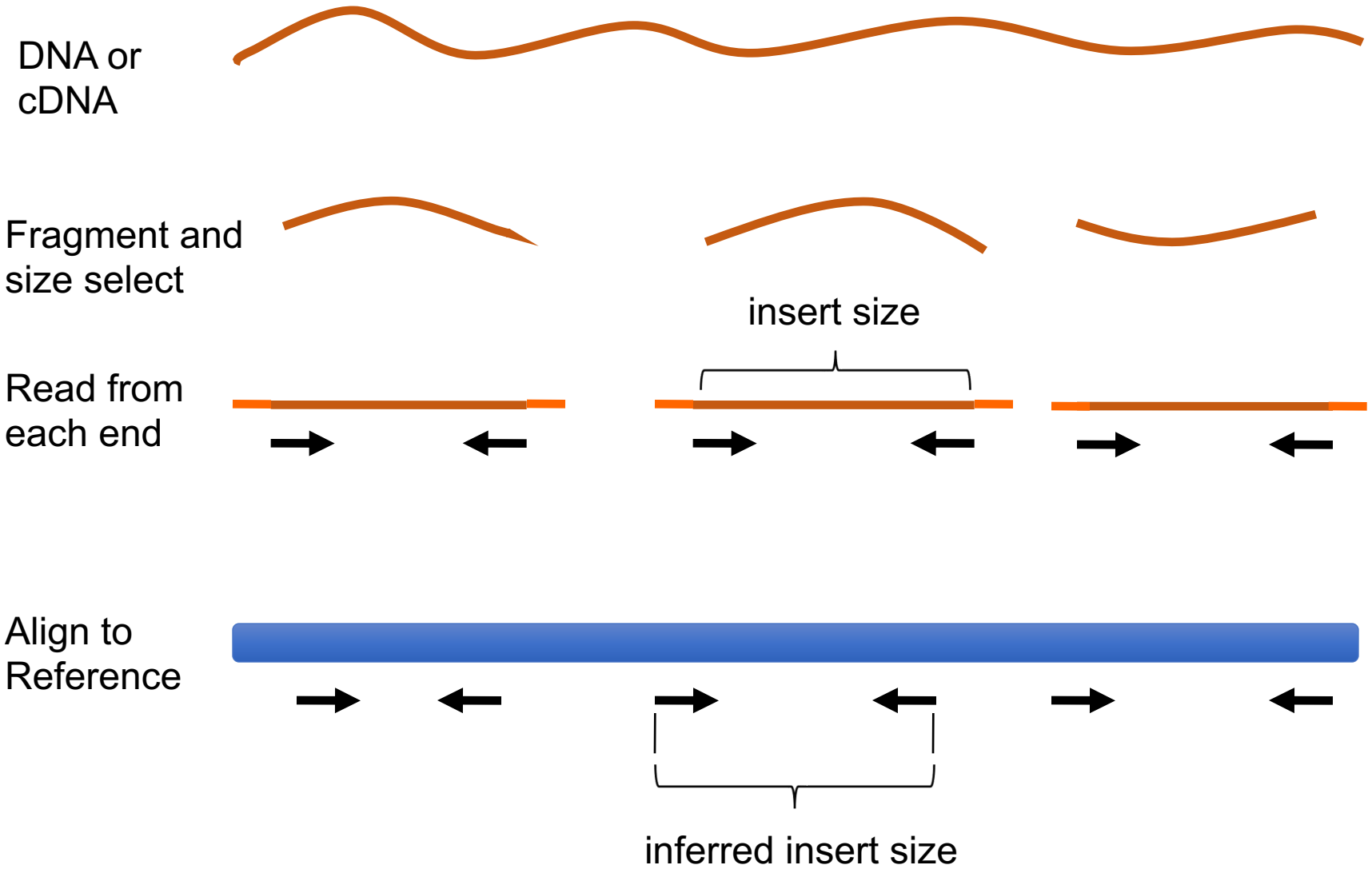
Fragment and
size select



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)

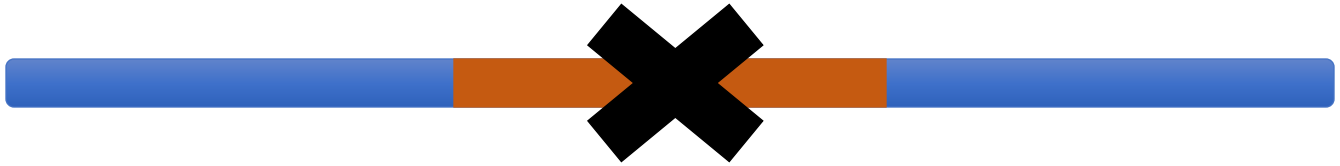
Deletion

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

Deletion

Reference Genome

Subject

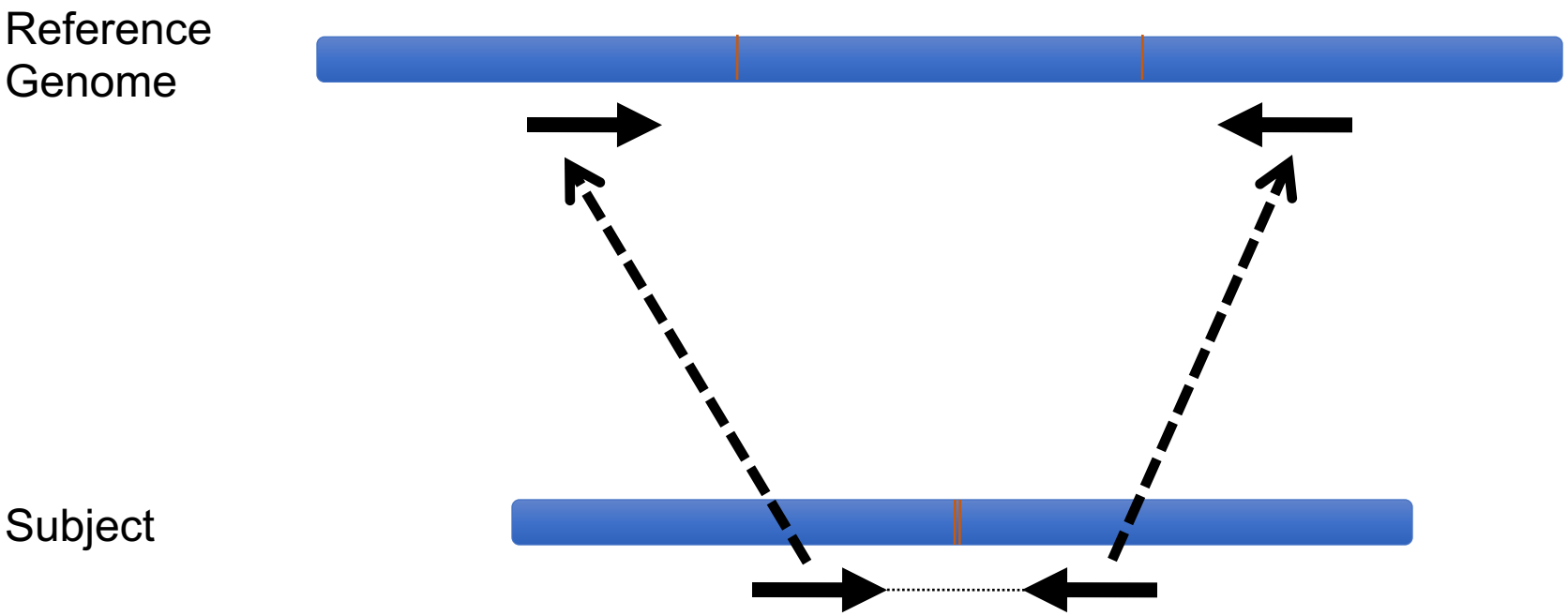


Deletion

Reference Genome

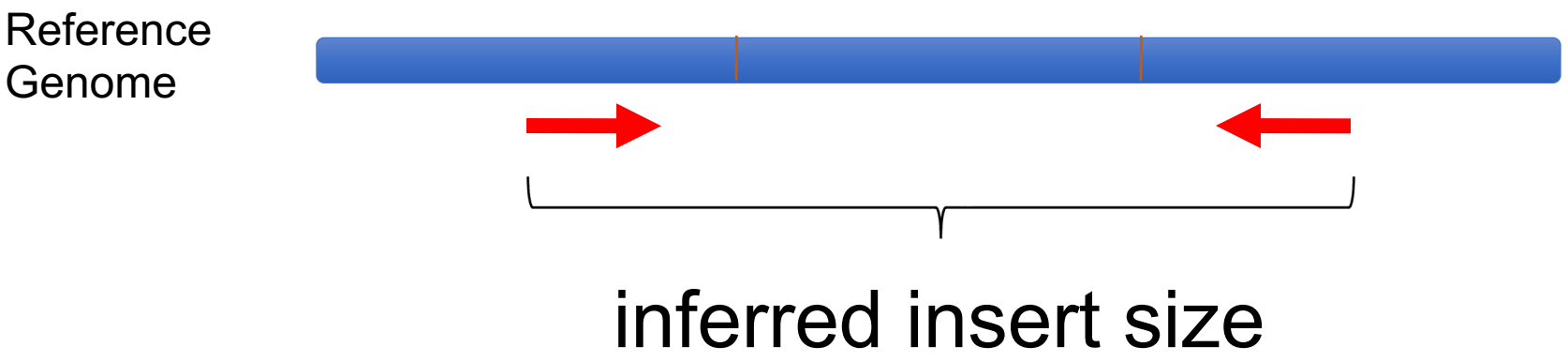
Subject

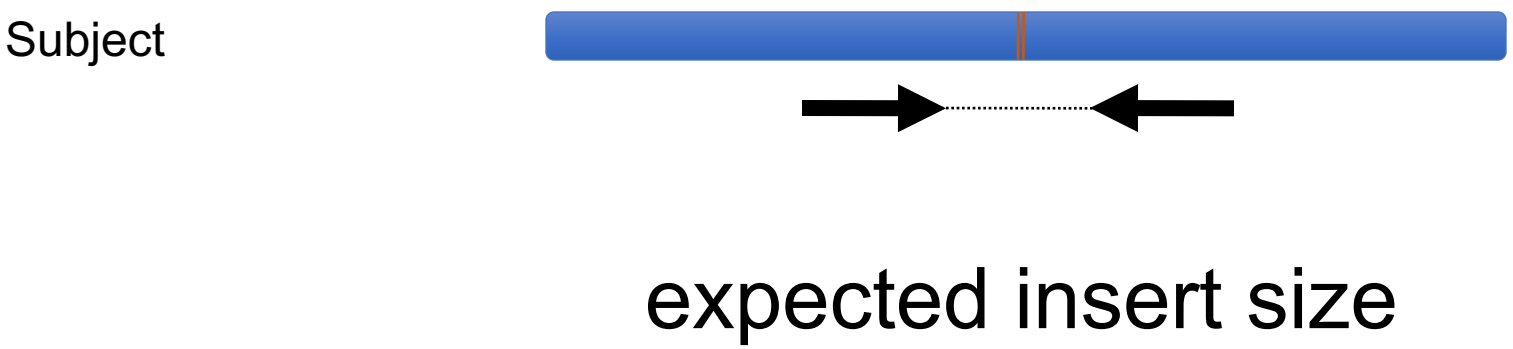
Deletion



Deletion

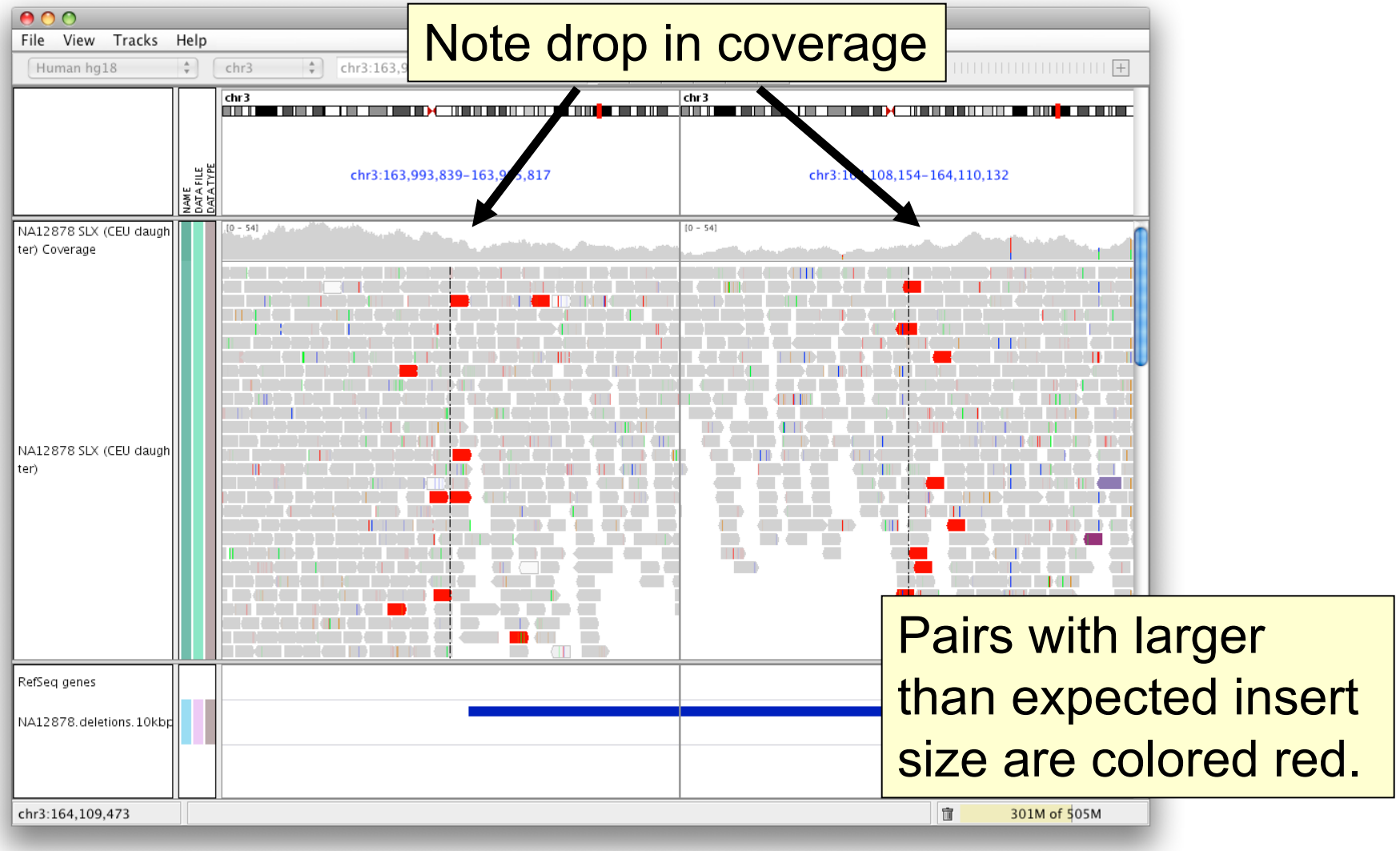
Inferred insert size is > expected value





Color by insert size

Deletion


Note drop in coverage
File View Tracks Help
Human hg18
chr3
chr3:163,9
chr3
chr3
chr3:163,993,839-163,9
chr3:1
108,154-164,110,132
NAME
DATA FILE
DATA TYPE
NA12878 SLX (CEU daugh
ter) Coverage
[0 - 54]
[0 - 54]
NA12878 SLX (CEU daugh
ter)
RefSeq genes
NA12878.deletions.10kbp
chr3:164,109,473
301M of 505M
Pairs with larger than expected insert size are colored red.

Insert size color scheme

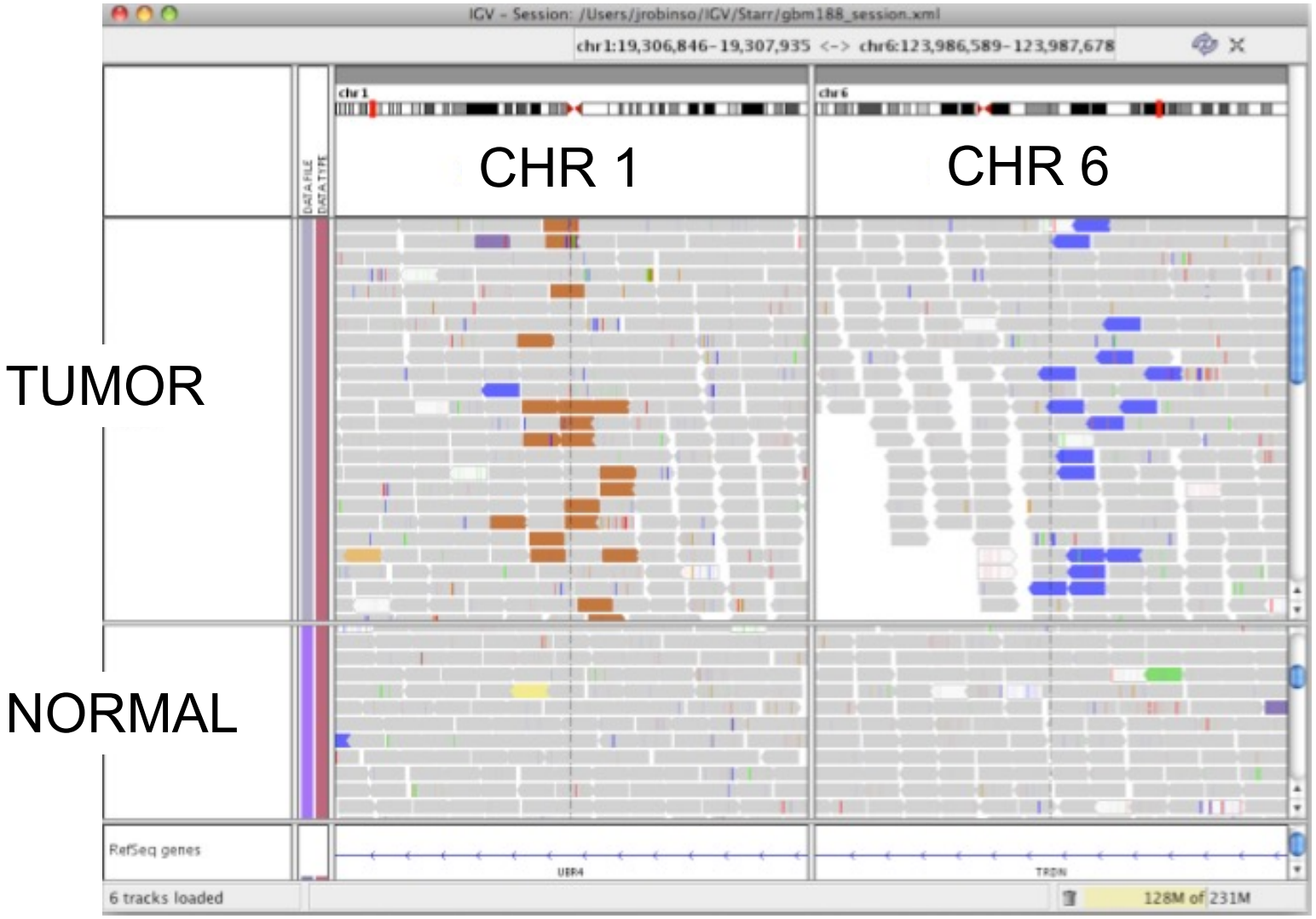
- Smaller than expected insert size:

- Larger than expected insert size:

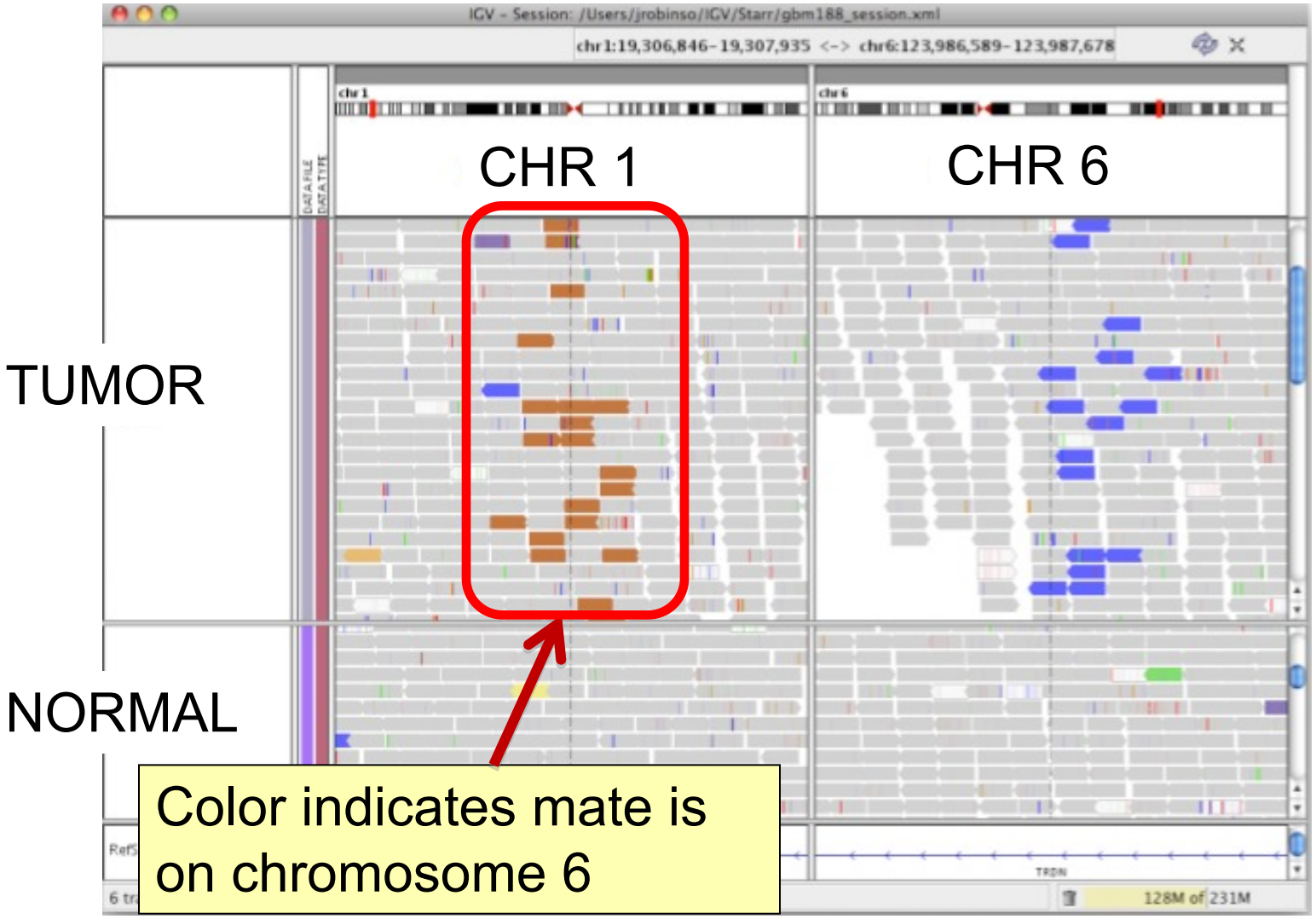
- Pairs on different chromosomes

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

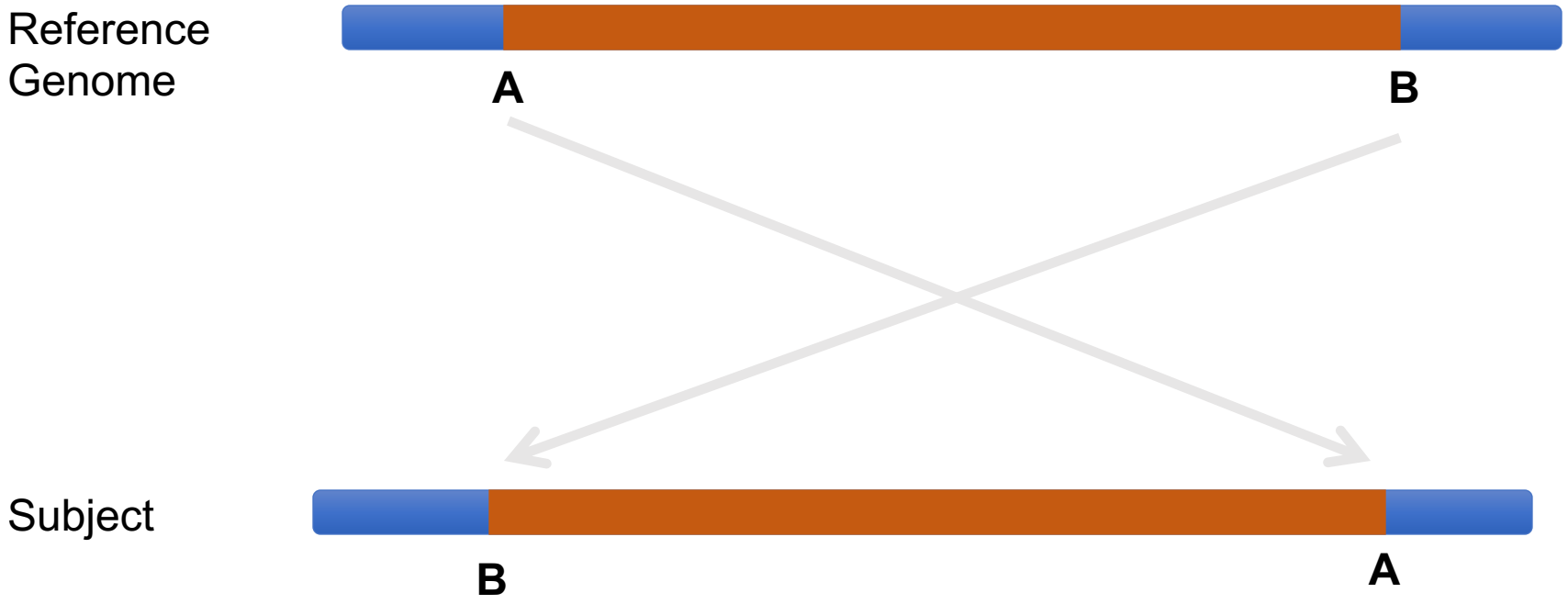
Inversion

Reference genome

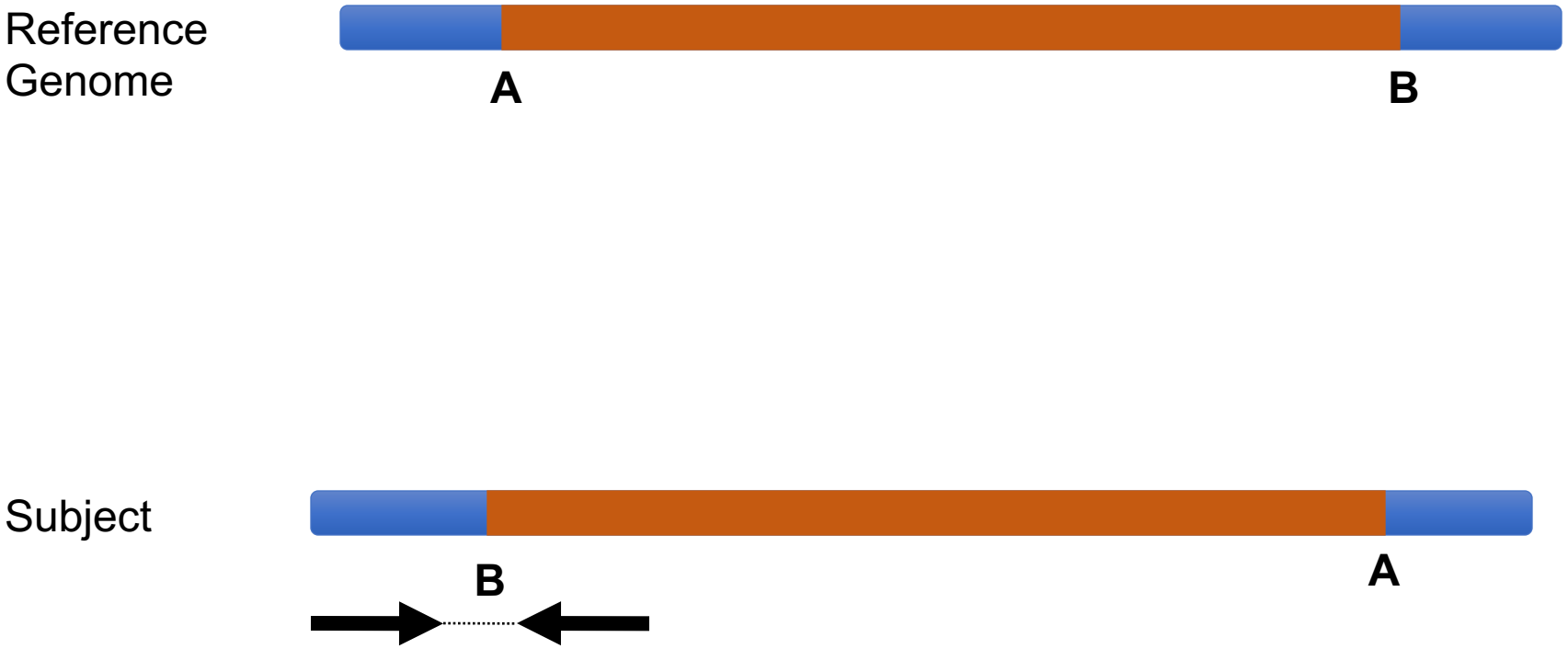
Inversion

Reference genome

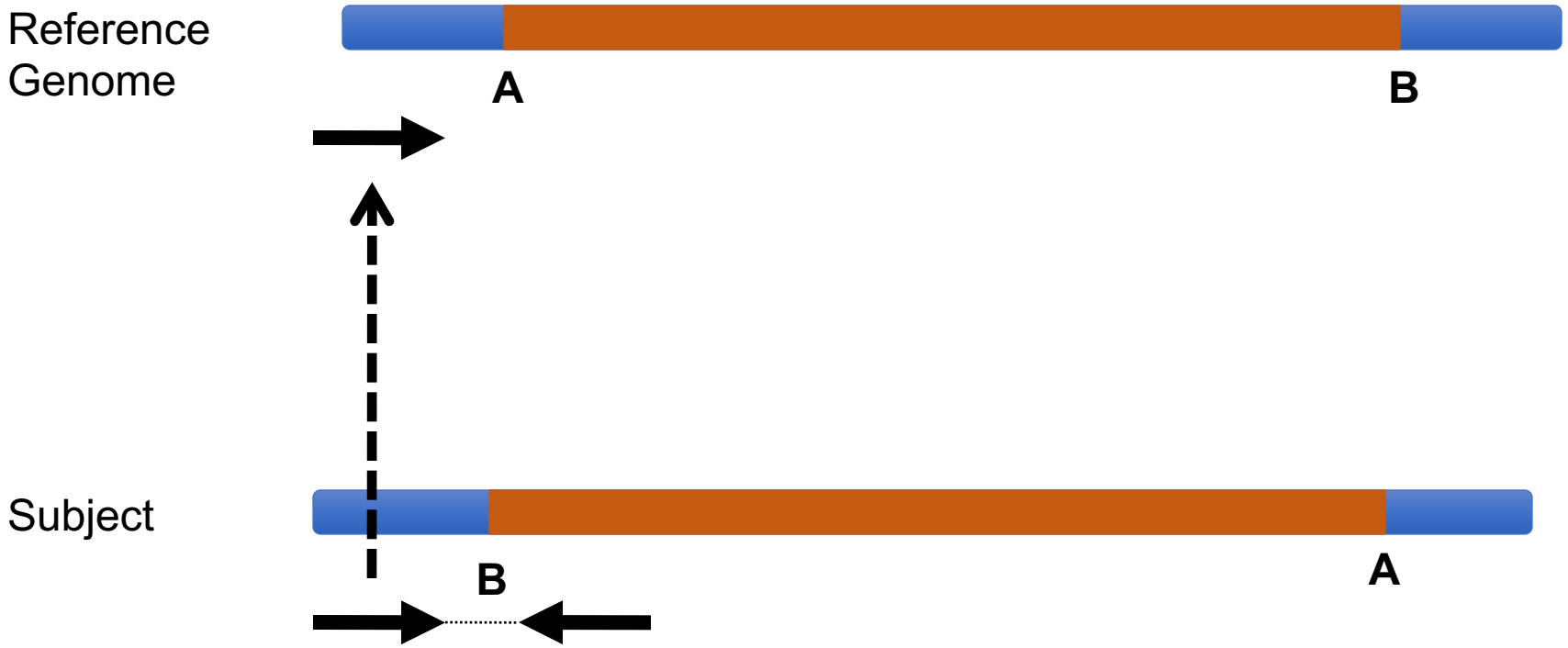
Inversion



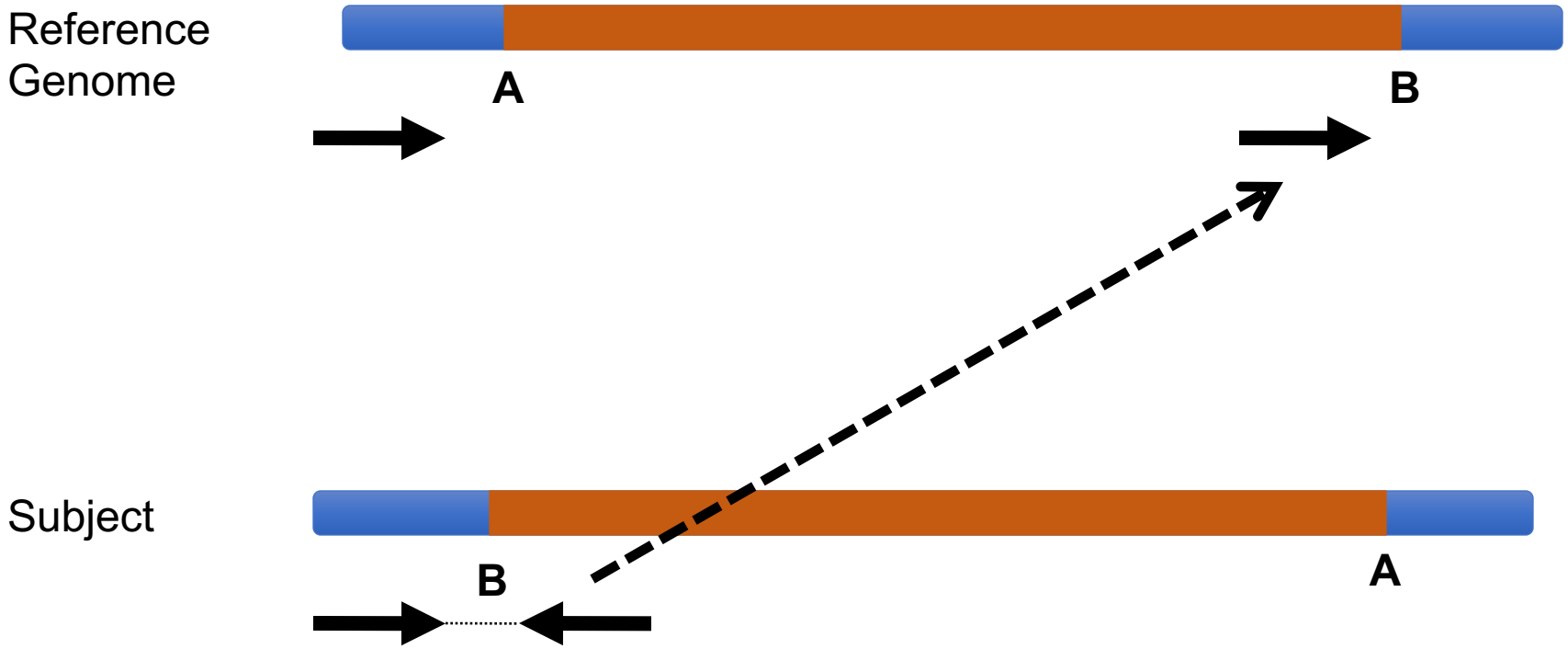
Inversion



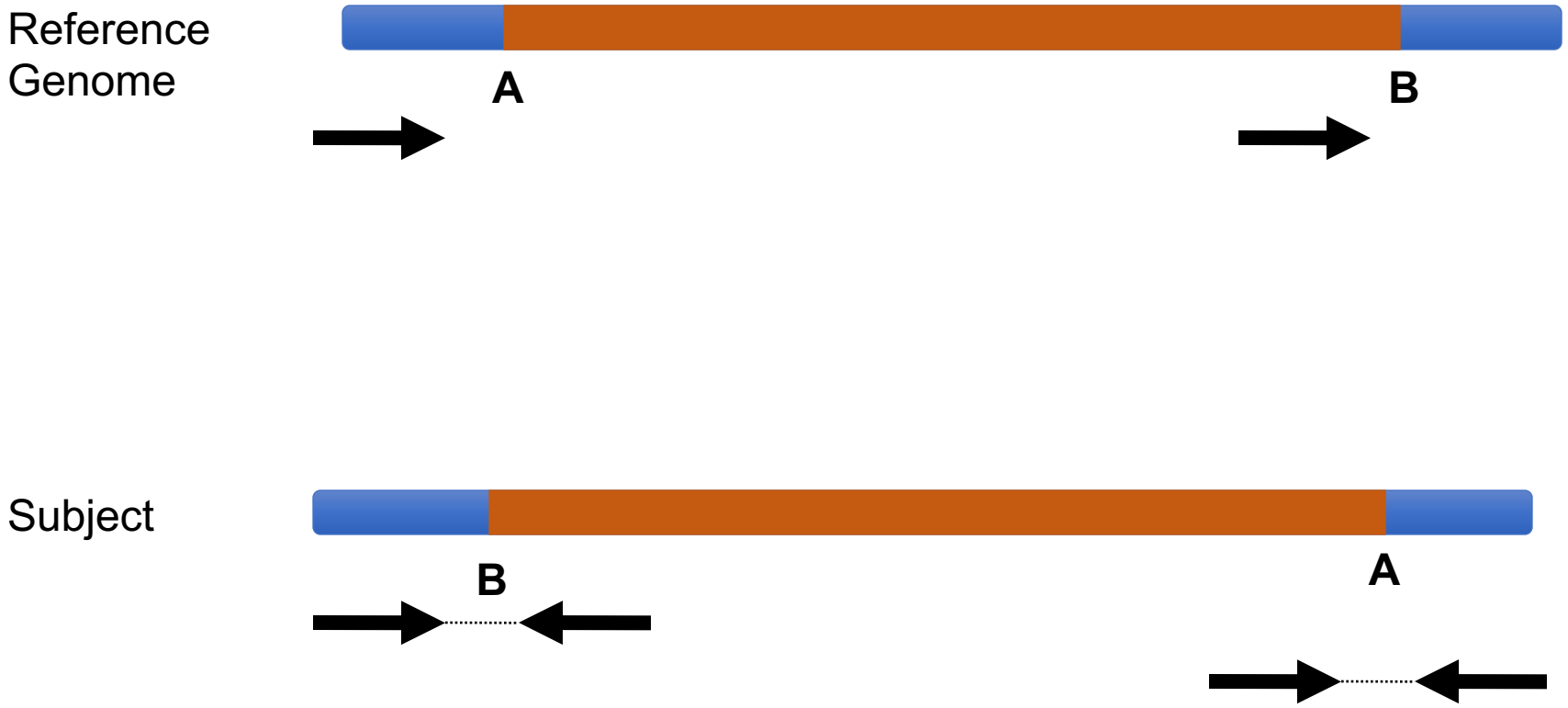
Inversion



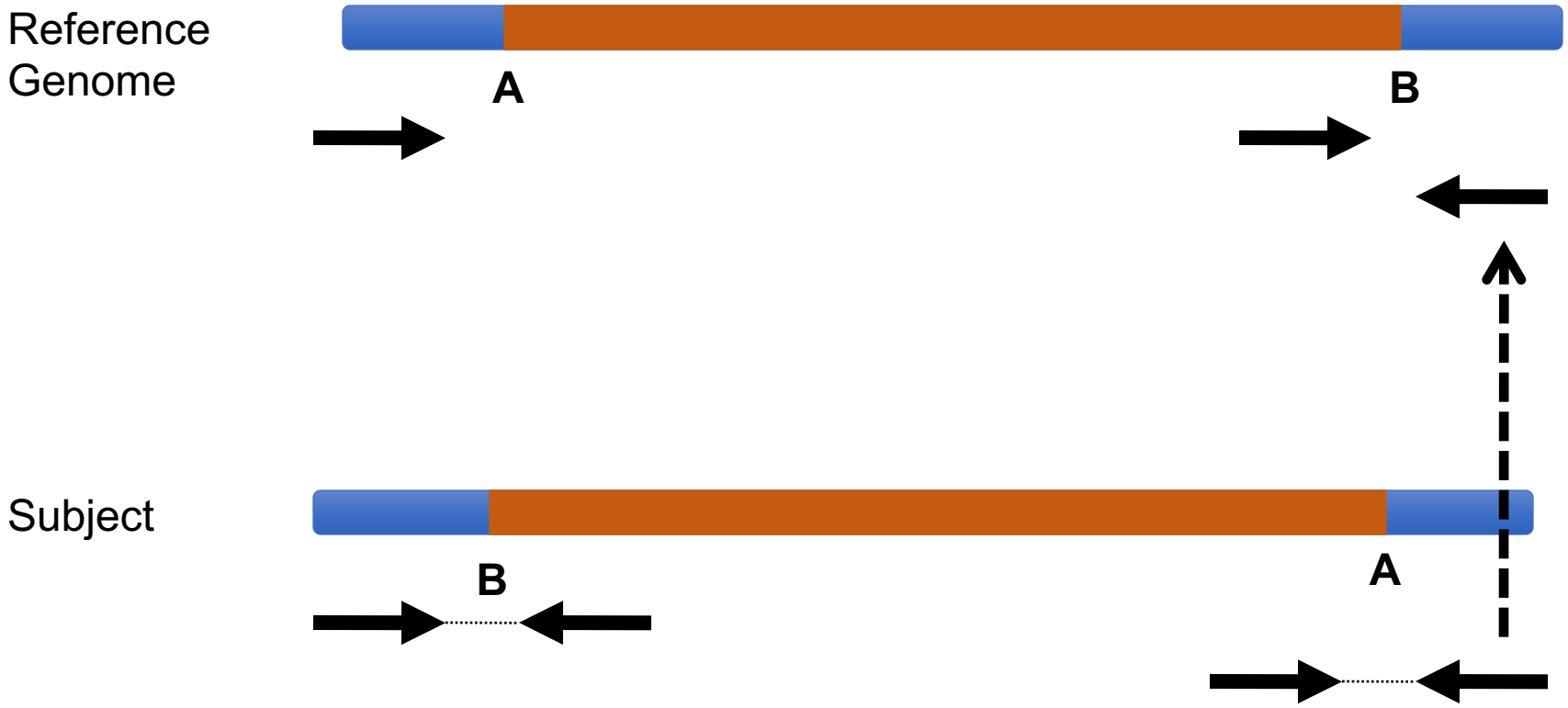
Inversion



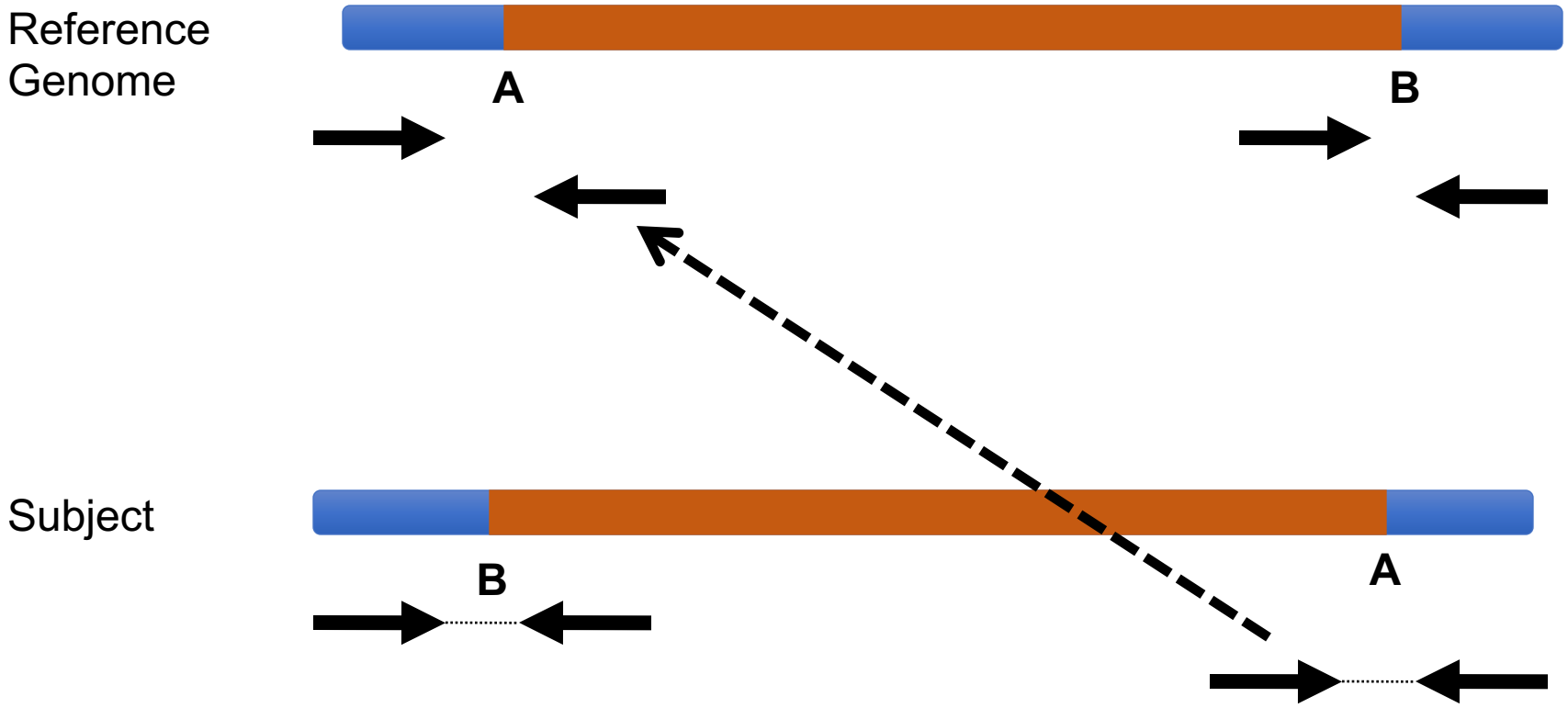
Inversion



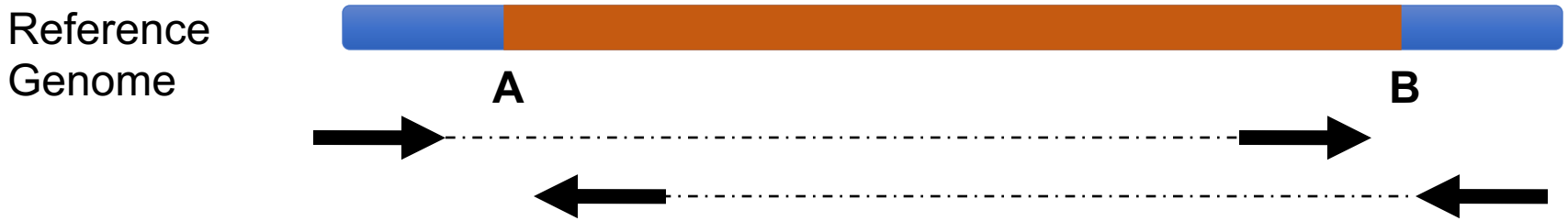
Inversion



Inversion



Inversion



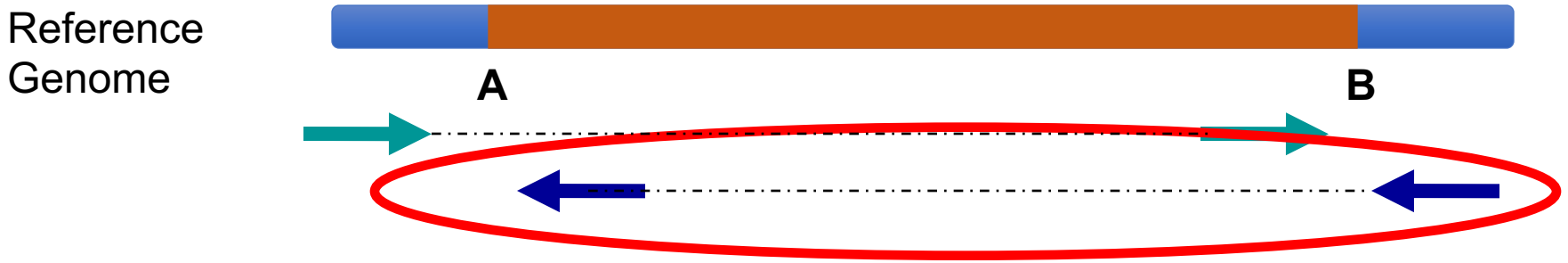
Inversion

Anomaly: expected orientation of pair is inward facing (→ ←)

Inversion

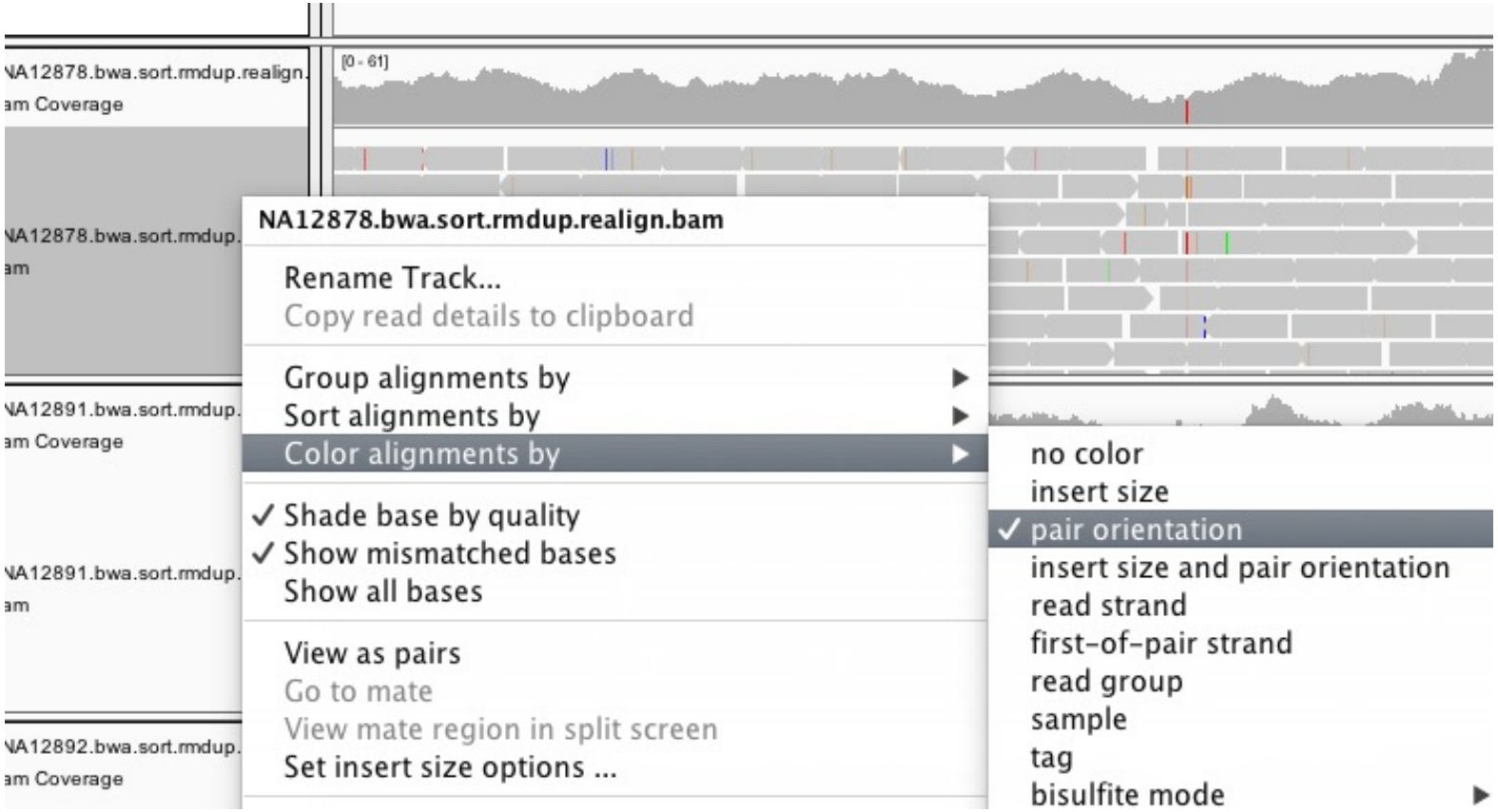
“Left” side pair

Inversion

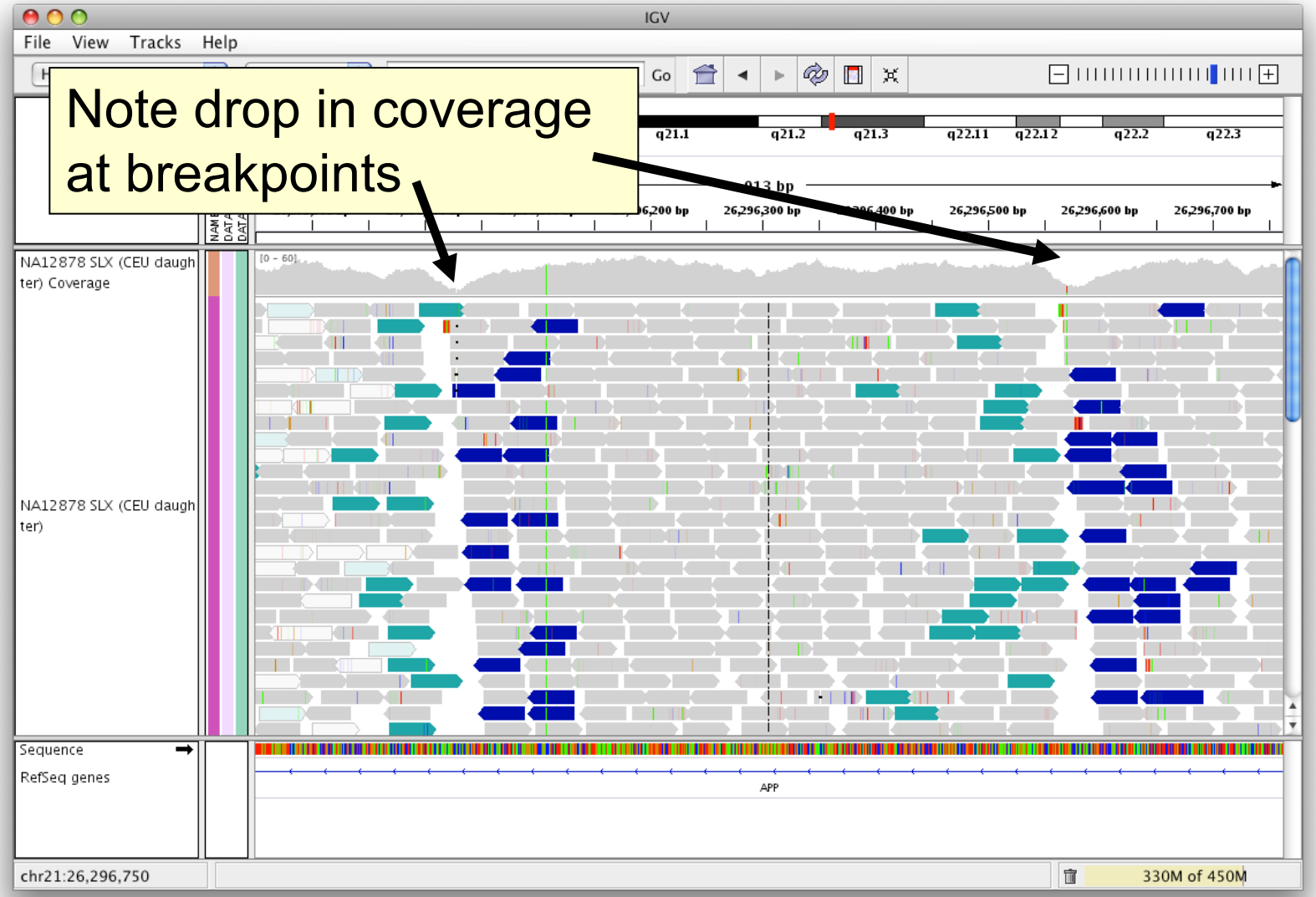


“Right” side pair

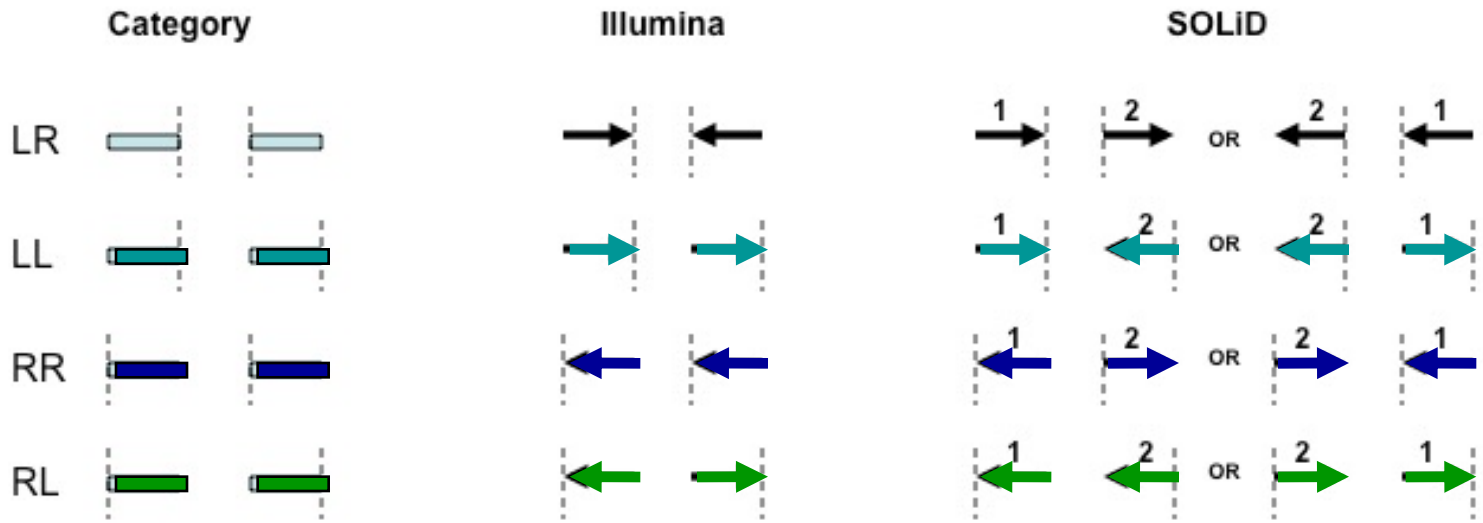
Color by pair orientation



Inversion


IGV
File View Tracks Help
Go
Note drop in coverage at breakpoints
q21.1 q21.2 q21.3 q22.11 q22.12 q22.2 q22.3
26,296,200 bp
26,296,300 bp
26,296,400 bp
26,296,500 bp
26,296,600 bp
26,296,700 bp
NA12878 SLX (CEU daughter) Coverage
[0 - 60]
NA12878 SLX (CEU daughter)
Sequence
RefSeq genes
APP
chr21:26,296,750
330M of 450M

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

Assignment

- https://pmbio.org/module-03-align/0003/03/01/IntroToIGV/