



Cold
Spring
Harbor
Laboratory

Introduction to IGV The Integrative Genomics Viewer

Kelsy Cotto, Felicia Gomez,
Obi Griffith, Malachi Griffith, Huiming Xia
Advanced Sequencing Technologies & Applications
November 4- 17, 2019



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)

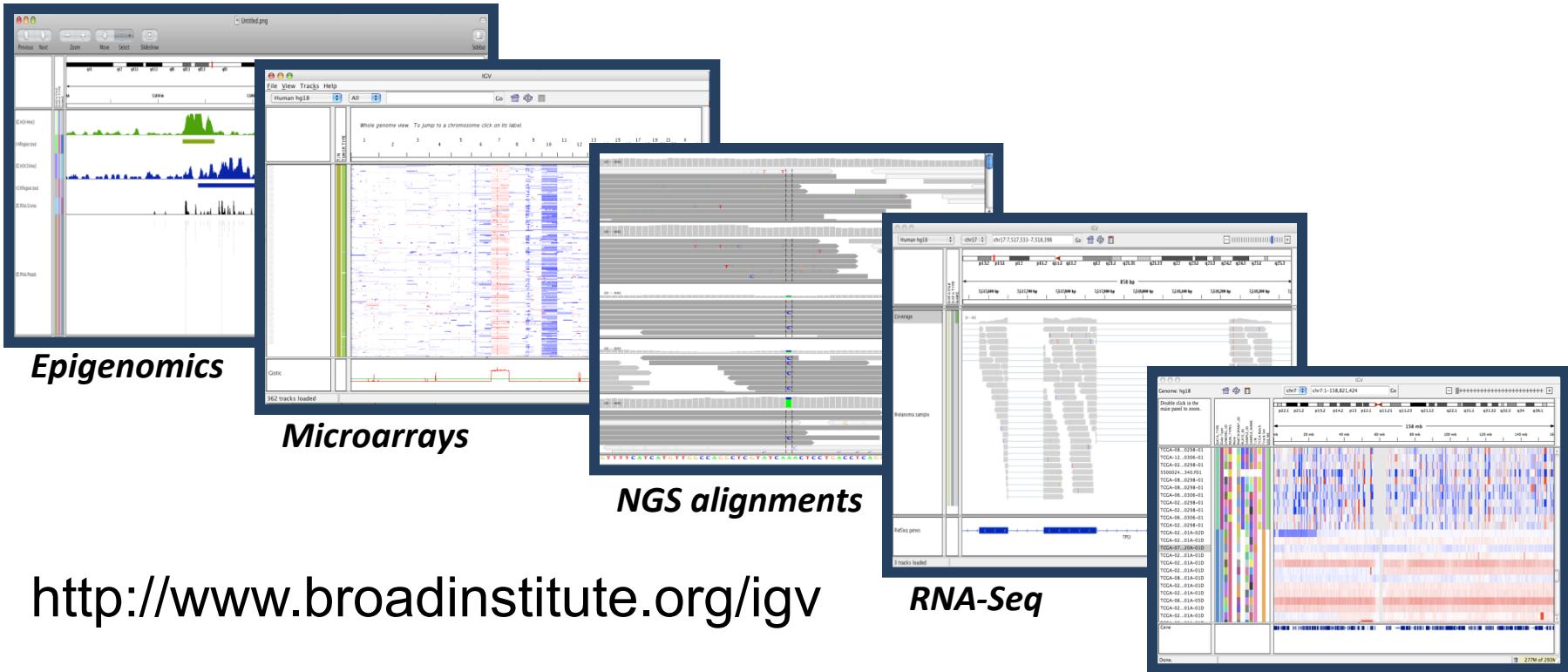


Savant
Genome
Browser

- 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>

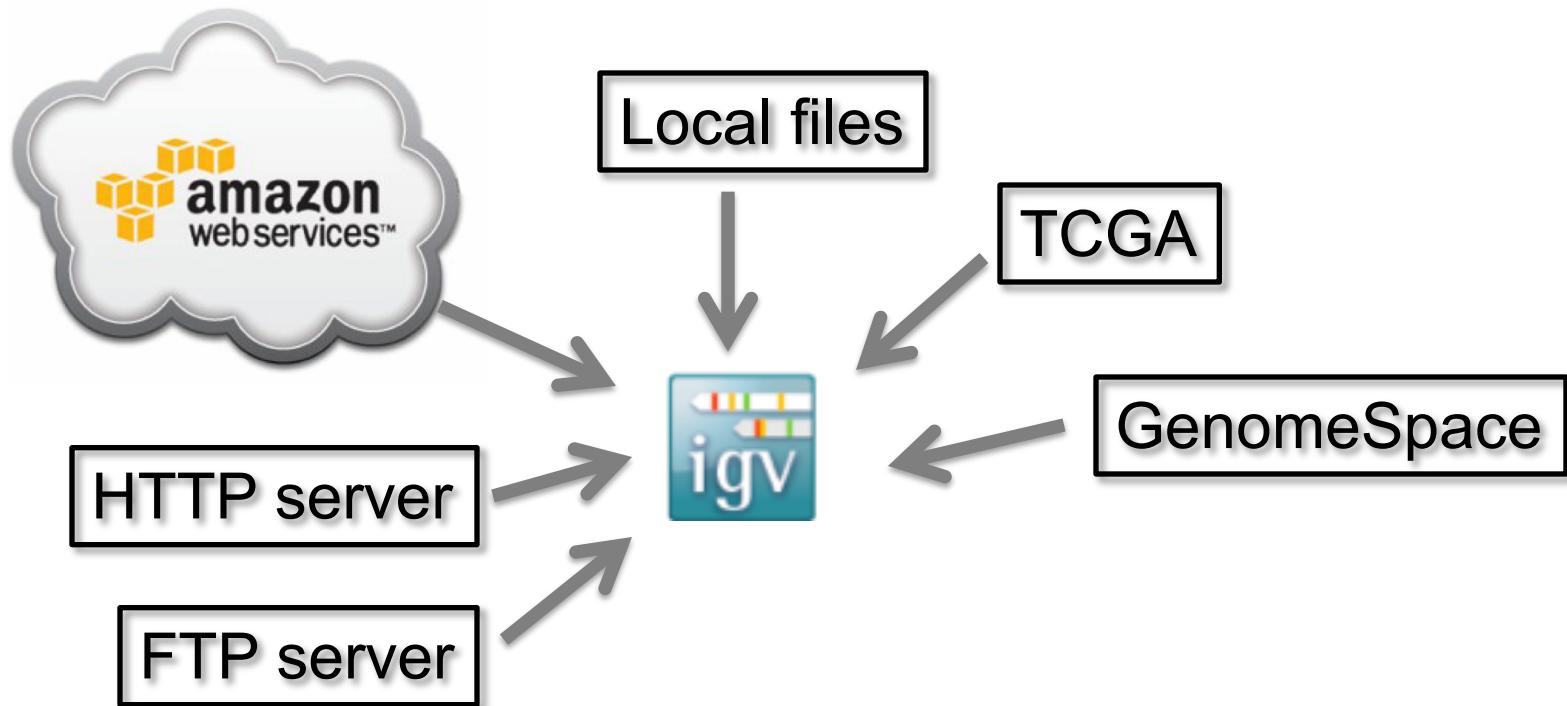
>85,000 registrations (2014)

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

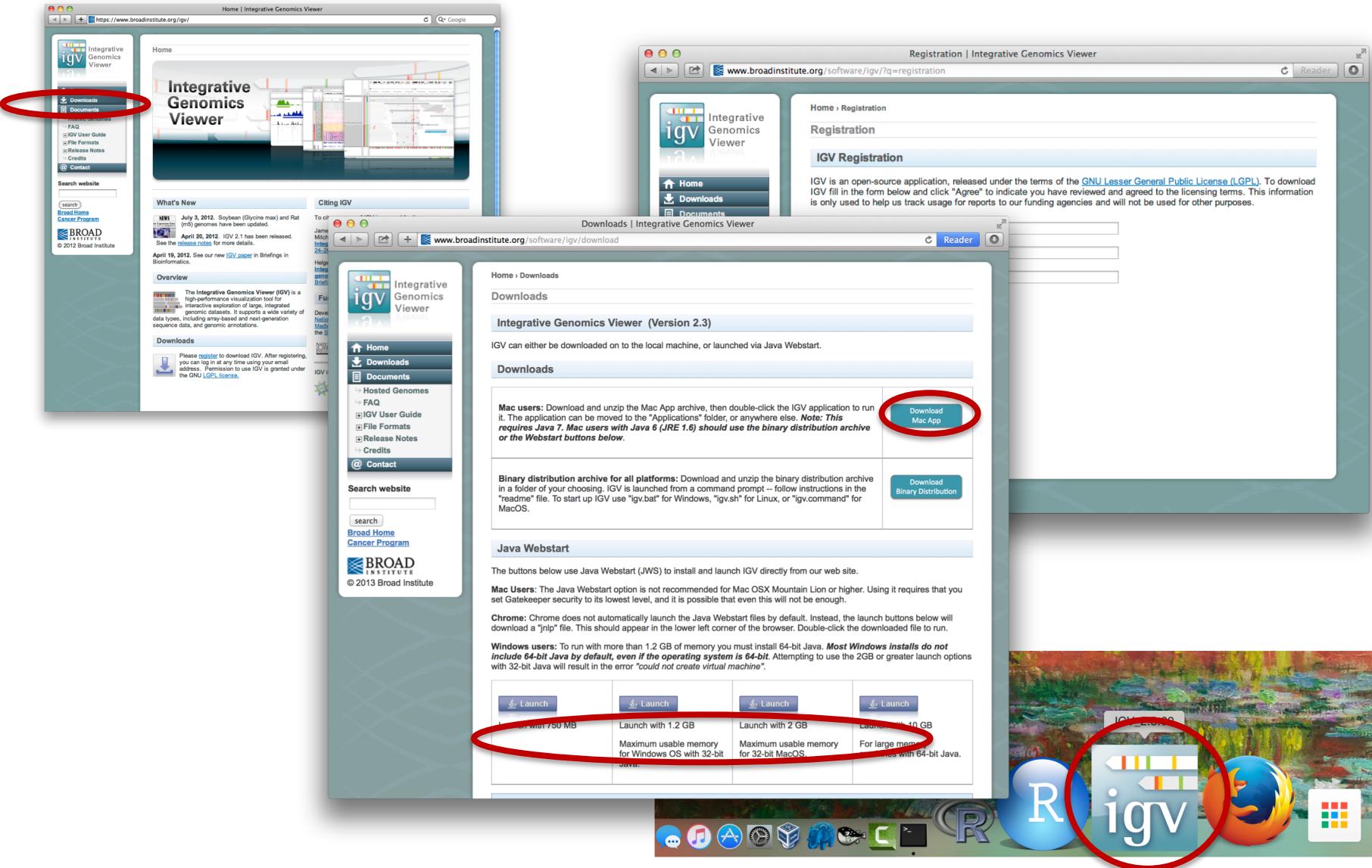


- View **local** files without uploading.
- View **remote** files without downloading the whole dataset.

Using IGV: the basics

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

Launch IGV



IGV Downloads

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for genomic data. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Downloads

Please register to download IGV. After registering, you can log in at any time using your email address and password to use IGV as granted under the GNU GPL license.

Mac users: Download and unzip the Mac App archive, then double-click the IGV application to run it. The application can be moved to the "Applications" folder, or anywhere else. **Note: This requires Java 7.** **Mac users with Java 6 (JRE 1.6) should use the binary distribution archive or the Webstart buttons below.**

Binary distribution archive for all platforms: Download and unzip the binary distribution archive in a folder of your choosing. IGV is launched from a command prompt -- follow instructions in the "readme" file. To start up IGV use "igv.bat" for Windows, "igv.sh" for Linux, or "igv.command" for MacOS.

Java Webstart

The buttons below use Java Webstart (JWS) to install and launch IGV directly from our web site.

Mac Users: The Java Webstart option is not recommended for Mac OSX Mountain Lion or higher. Using it requires that you set Gatekeeper security to its lowest level, and it is possible that even this will not be enough.

Chrome: Chrome does not automatically launch the Java Webstart files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

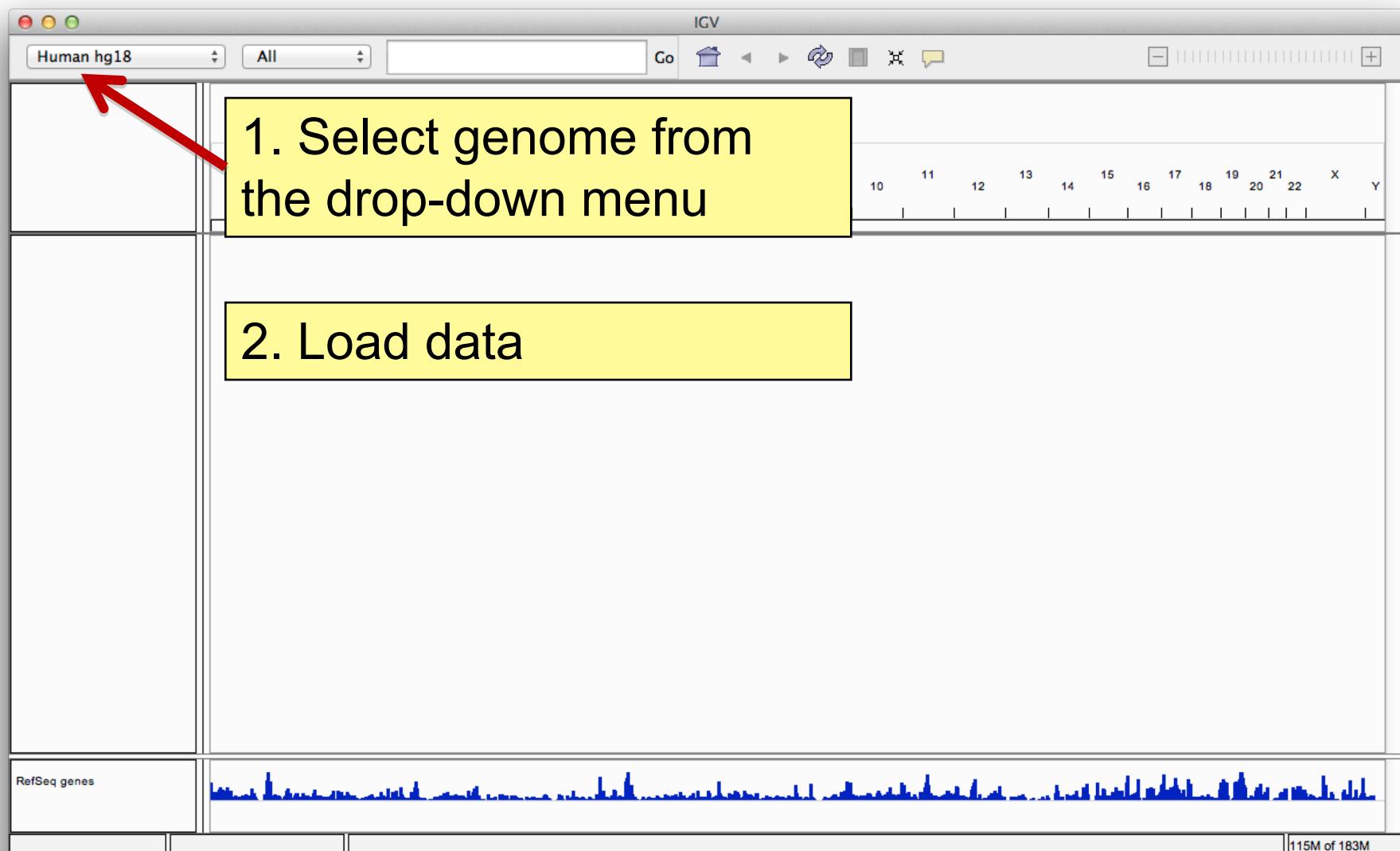
Windows users: To run with more than 1.2 GB of memory you must install 64-bit Java. **Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit.** Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error "could not create virtual machine".

Launch	Launch	Launch	Launch
Launch with 750 MB Maximum usable memory for Windows OS with 32-bit Java.	Launch with 1.2 GB Maximum usable memory for 32-bit Mac OS.	Launch with 2 GB Maximum usable memory for 32-bit Mac OS.	Launch with 10 GB For large memory needs with 64-bit Java.

IGV Application Icons

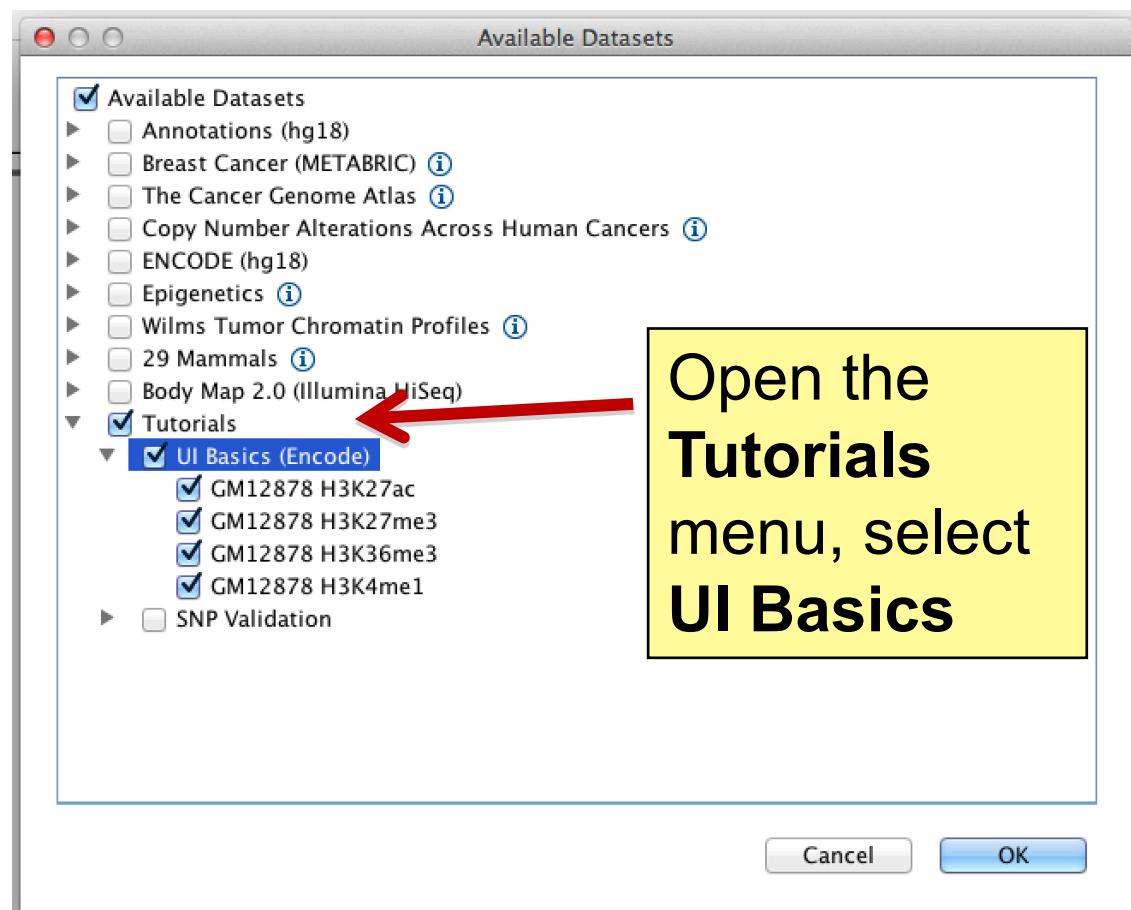
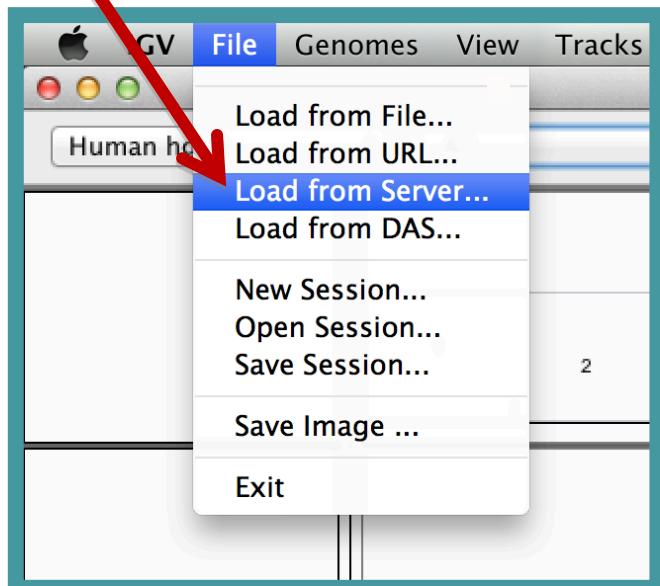
IGV application icons are shown on a desktop background, including the IGV icon with a red circle, the R logo, the Firefox logo, and the Google Chrome logo.

Launch IGV

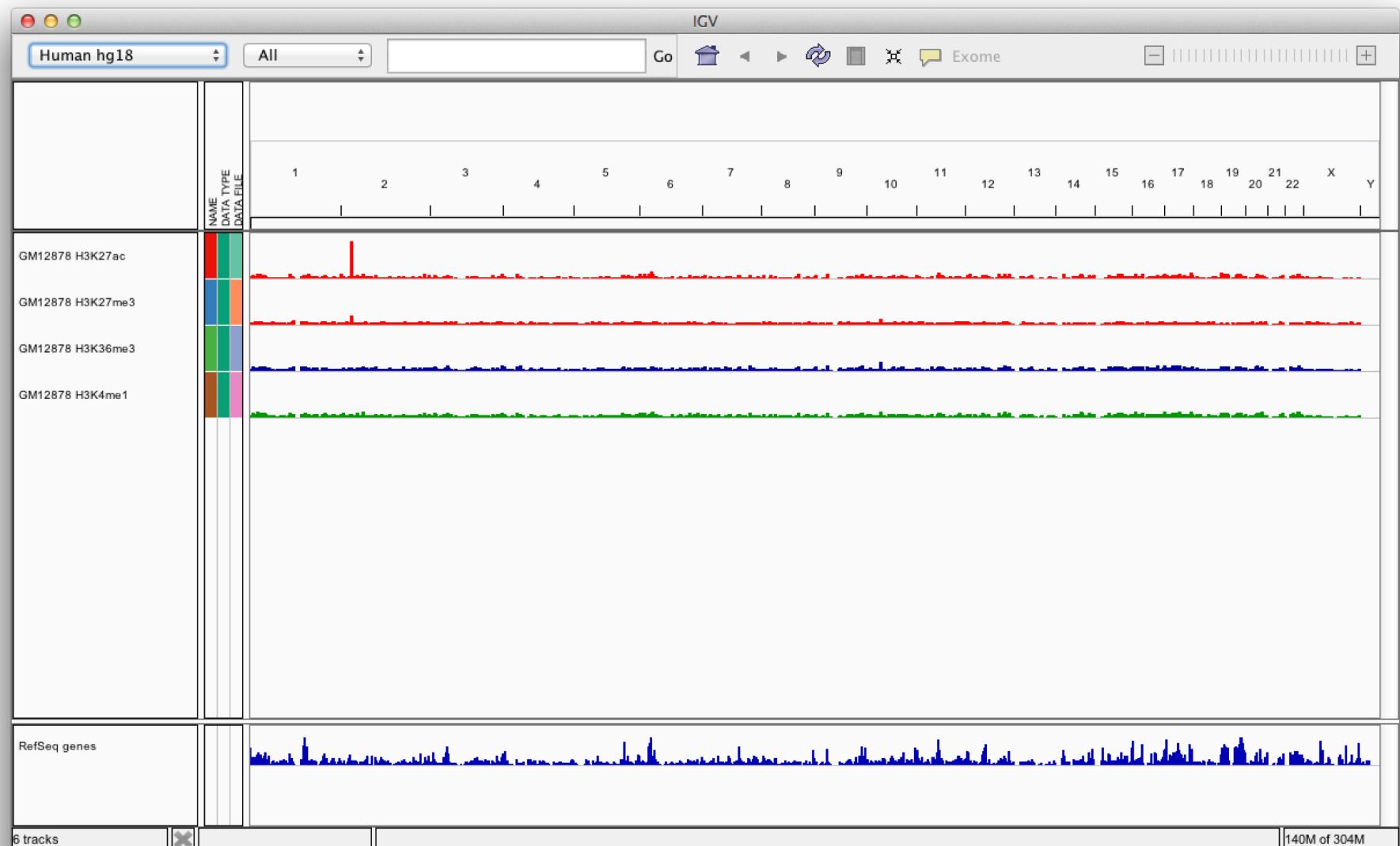


Load data

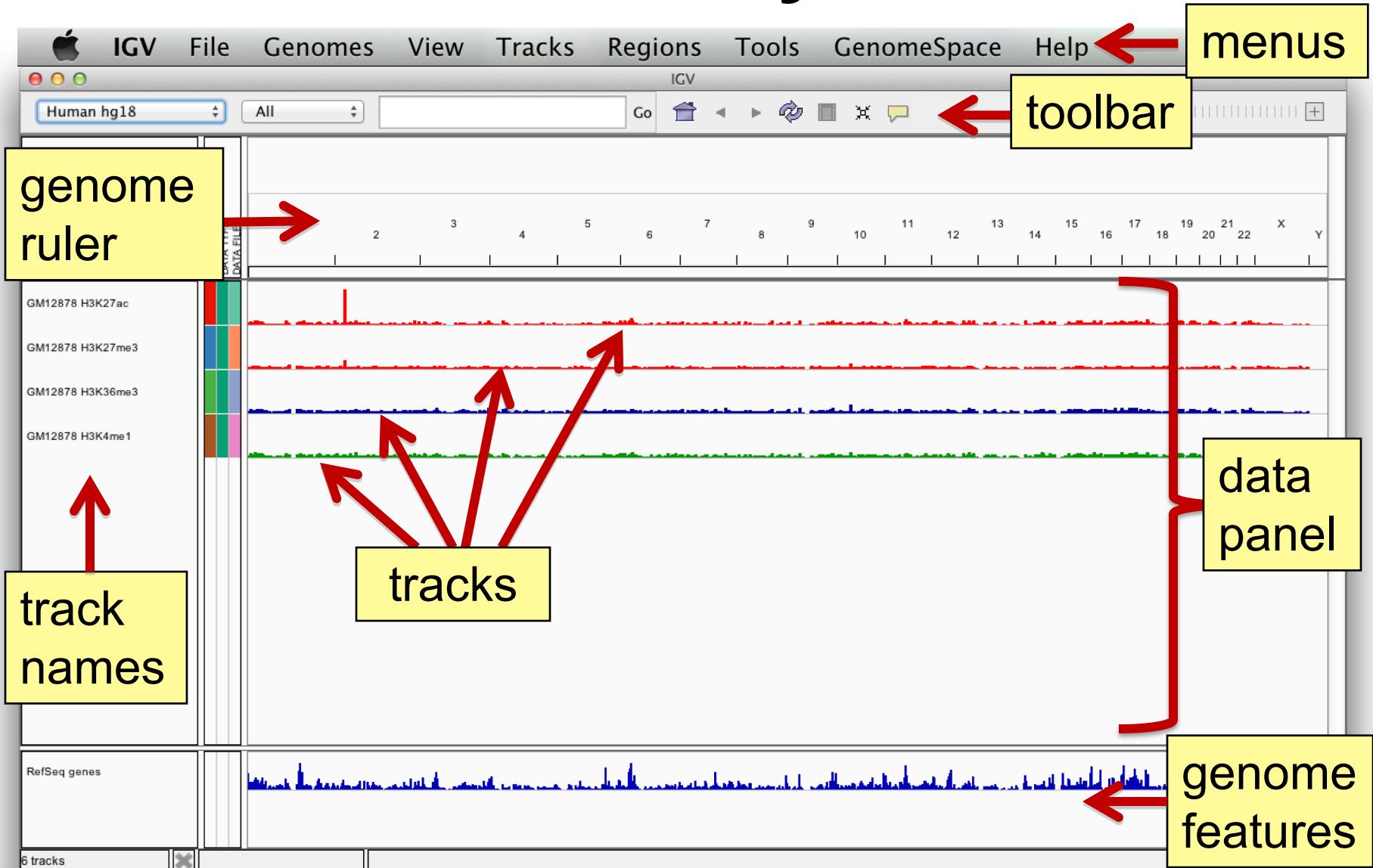
Select File > Load from Server...



Screen layout



Screen layout



File formats and track types

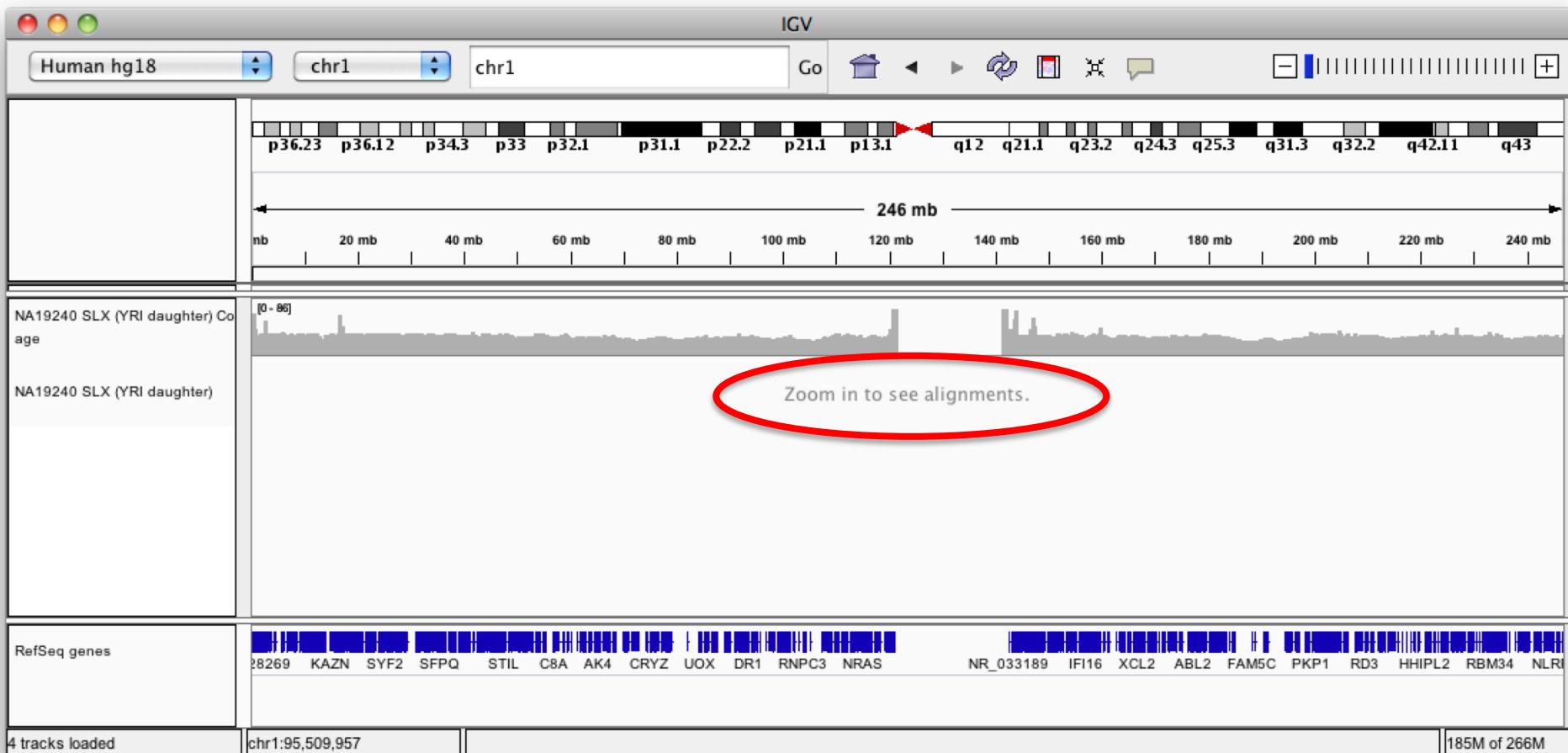
- The **file format** defines the track type.
- The **track type** determines the display options

- [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](#)

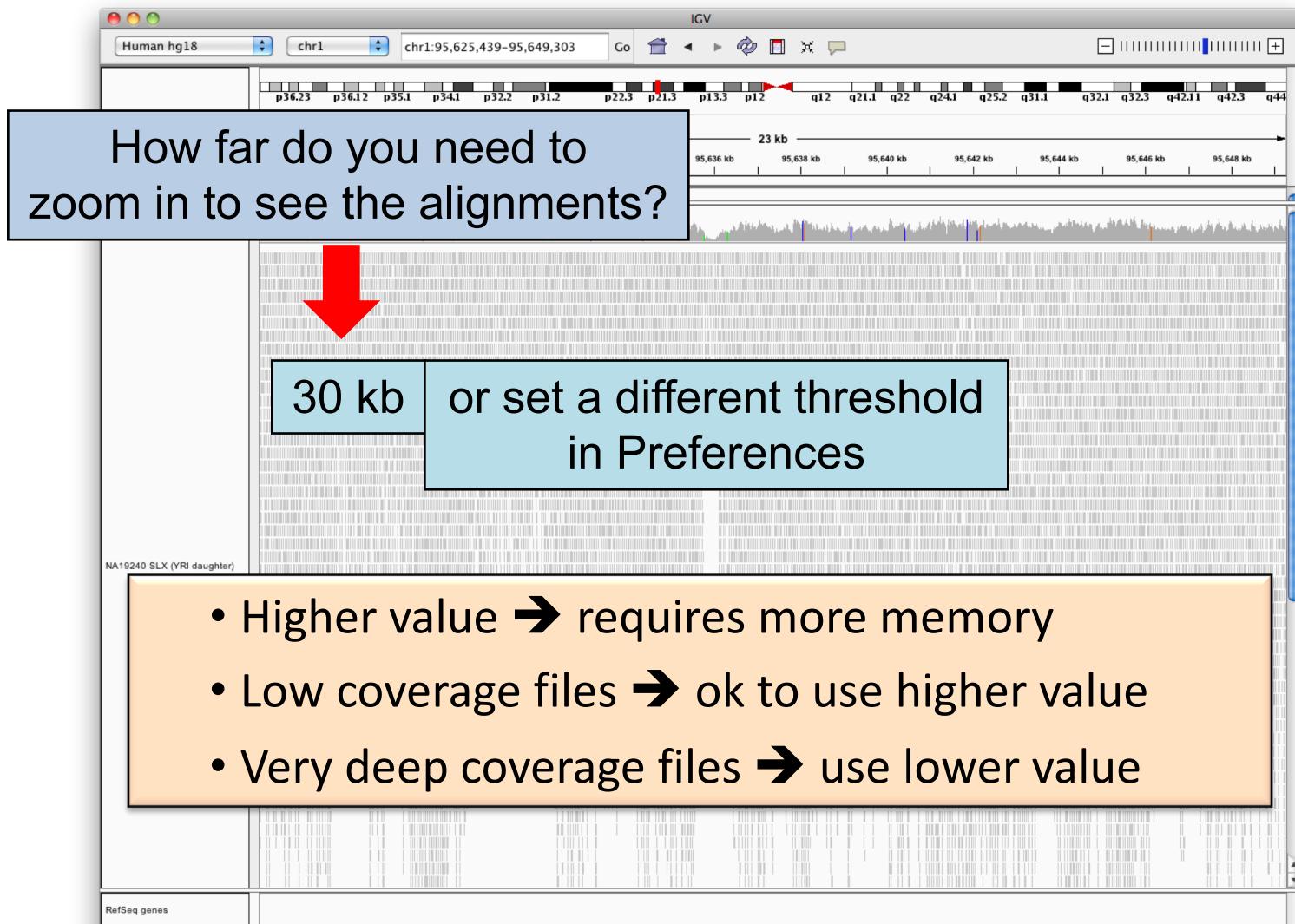
- For current list see: www.broadinstitute.org/igv/FileFormats

Viewing alignments

Whole chromosome view



Viewing alignments – Zoom in



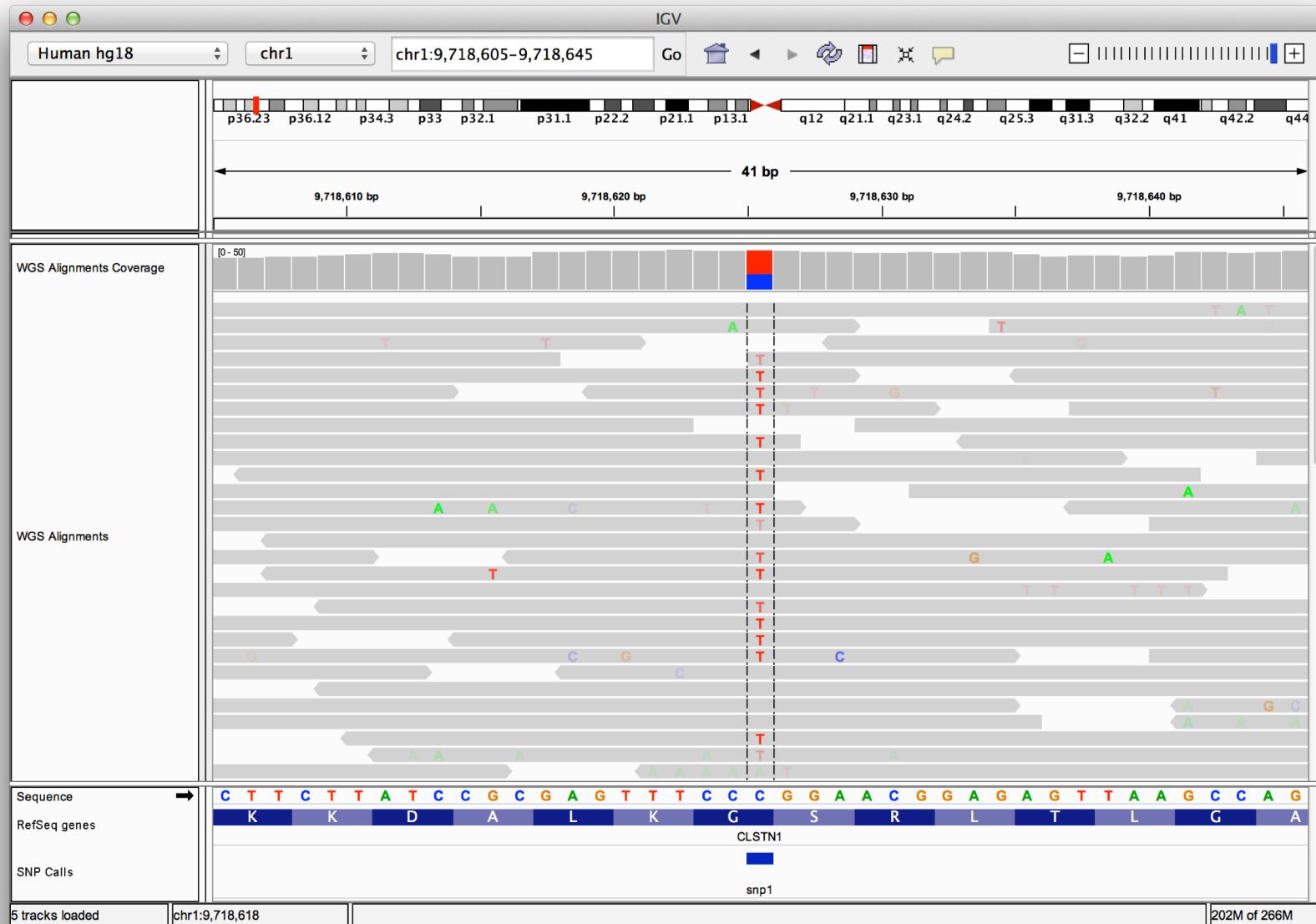
Viewing alignments – Zoom in



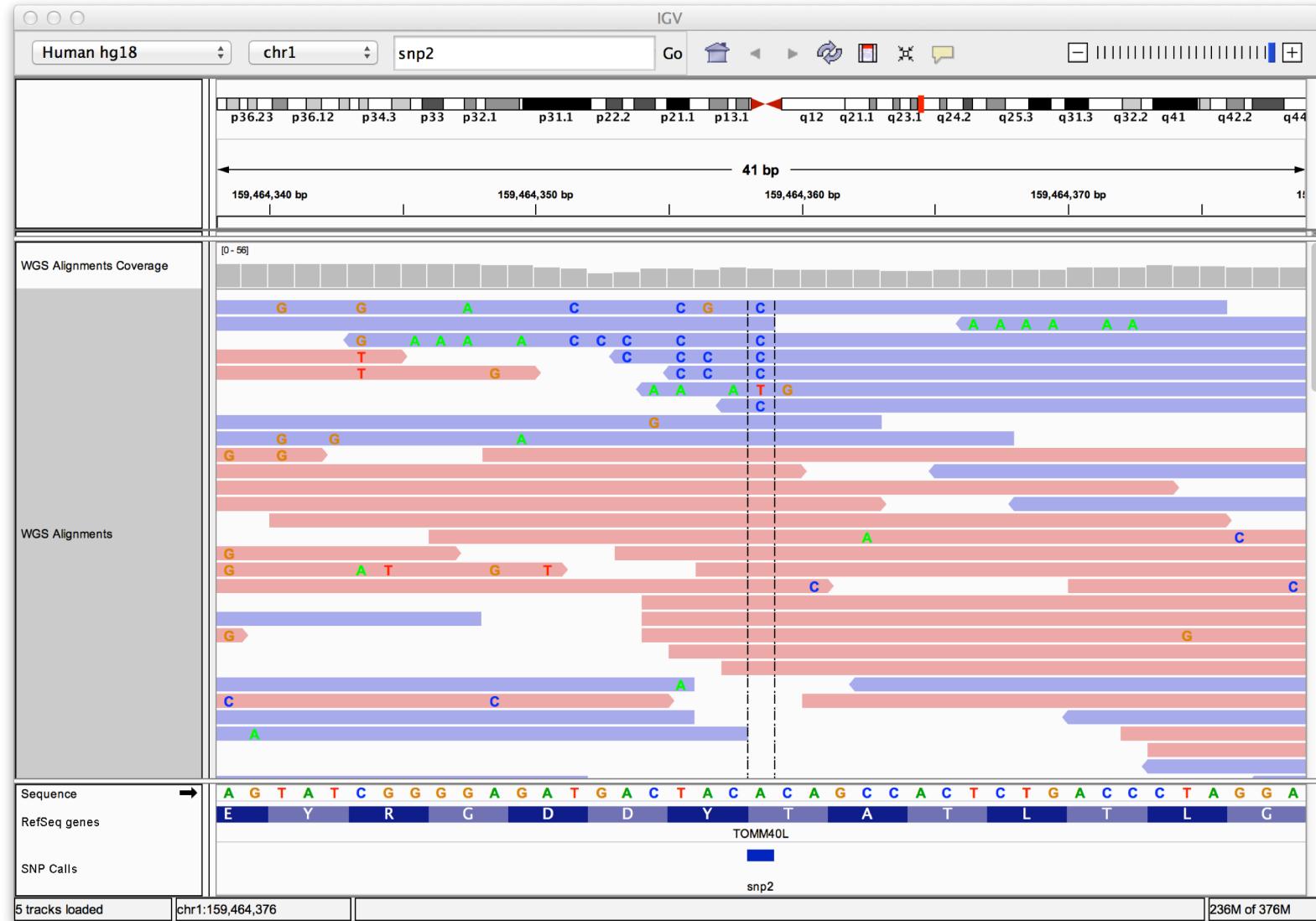
SNVs and Structural variations

- 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



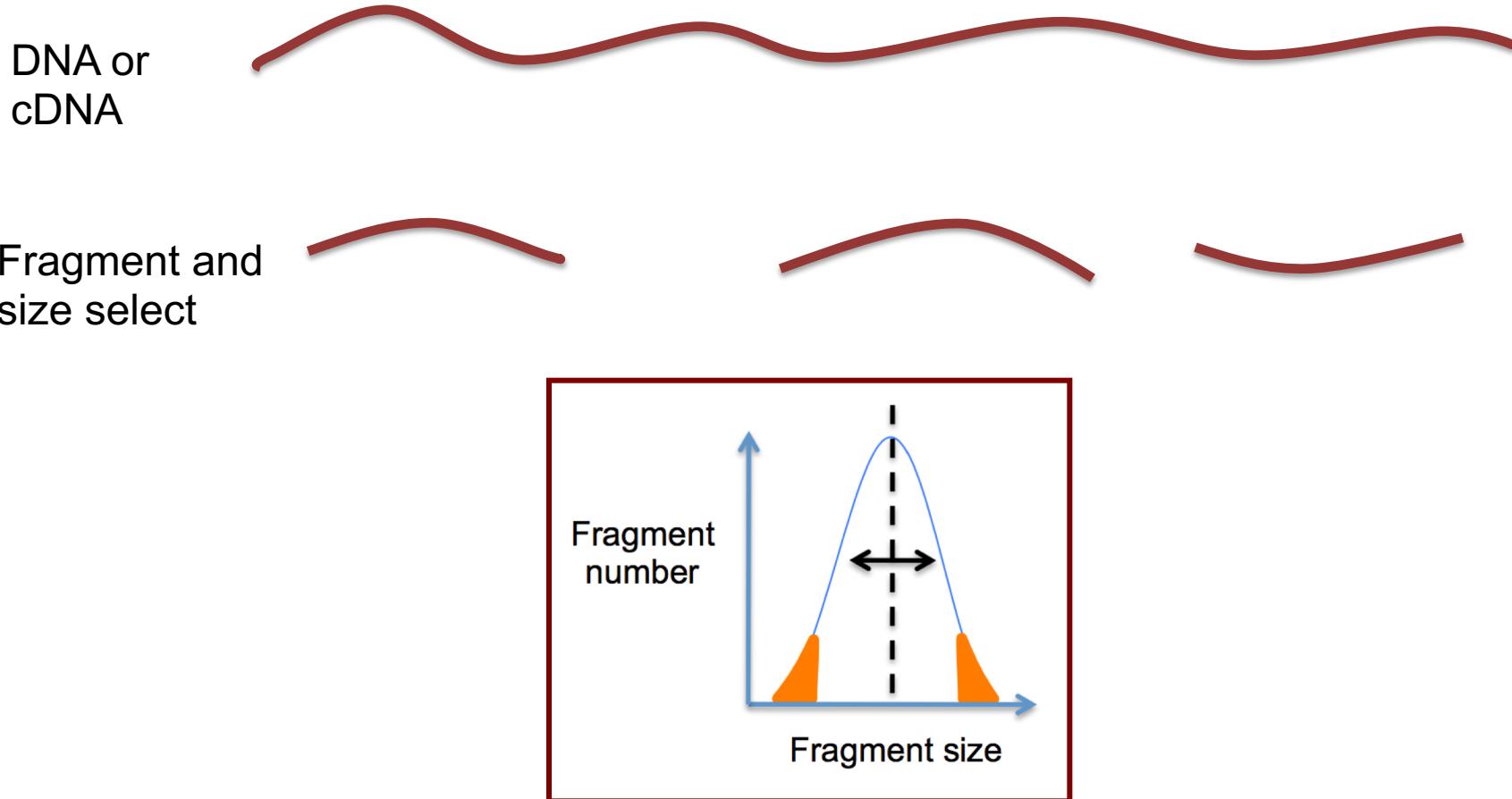
Viewing SNPs and SNVs



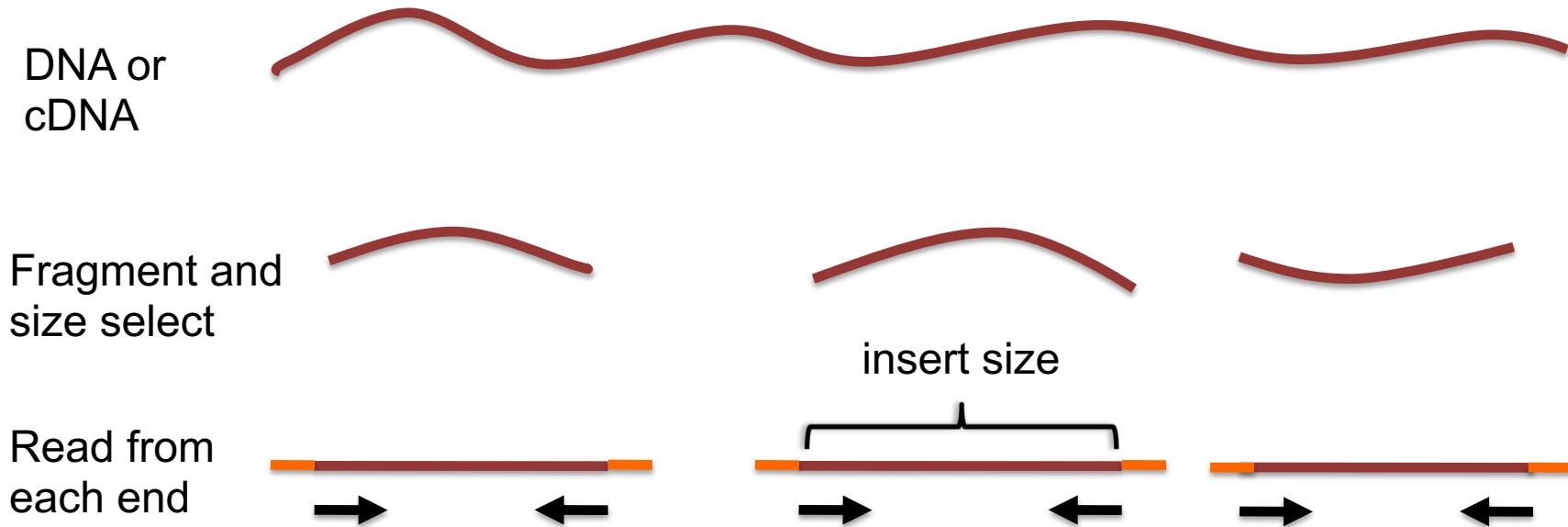
Viewing Structural Events

- 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)

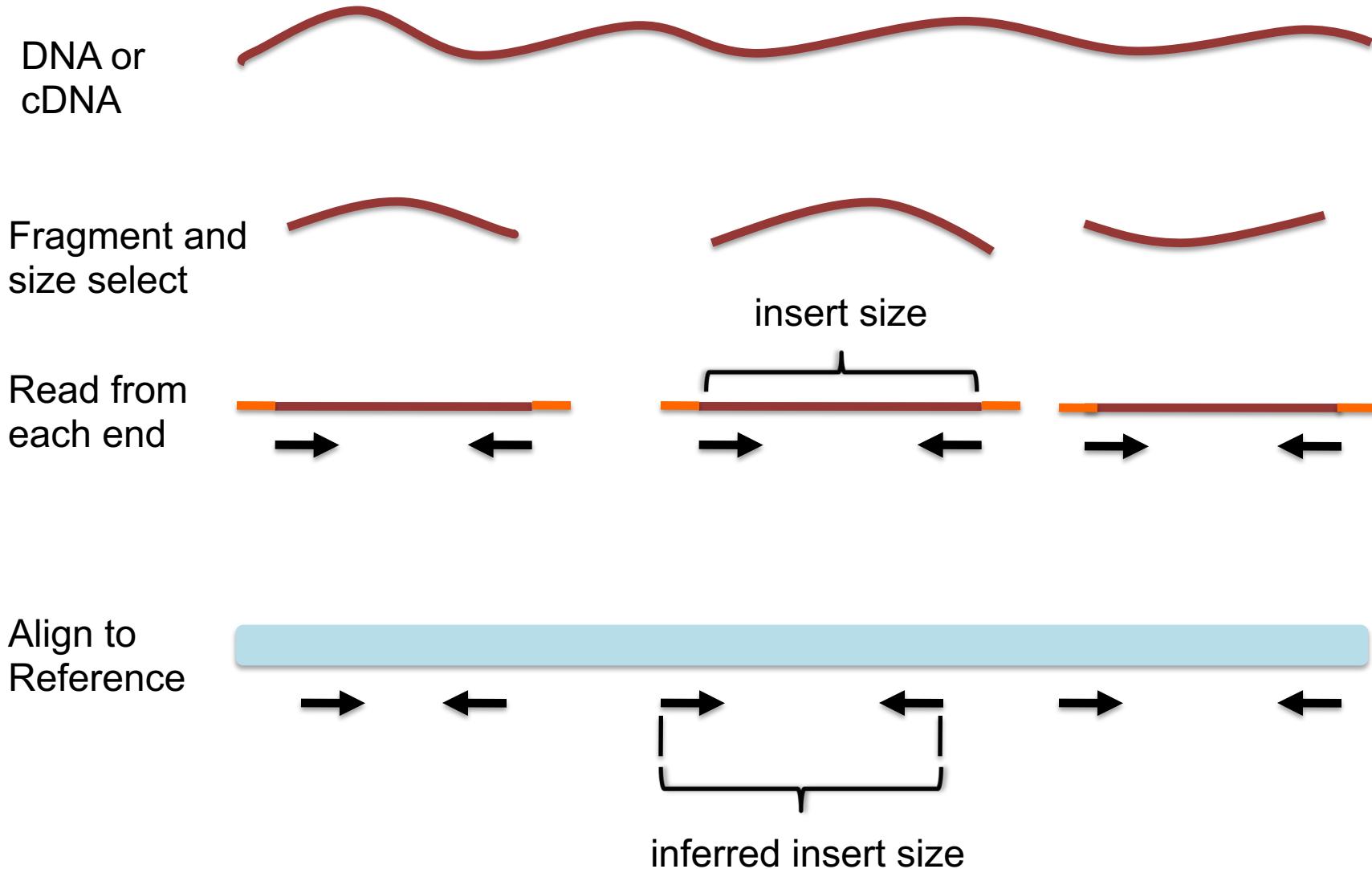
Paired-end sequencing



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

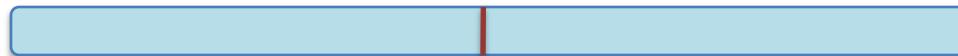


Deletion

Reference
Genome



Subject



Deletion

Reference
Genome



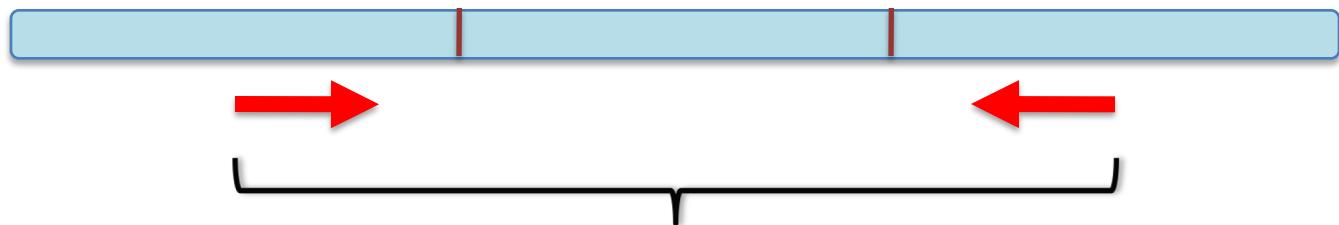
Subject



Deletion

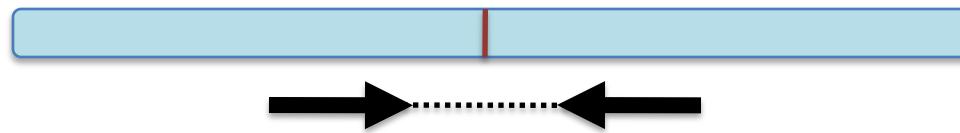
Inferred insert size is > expected value

Reference
Genome



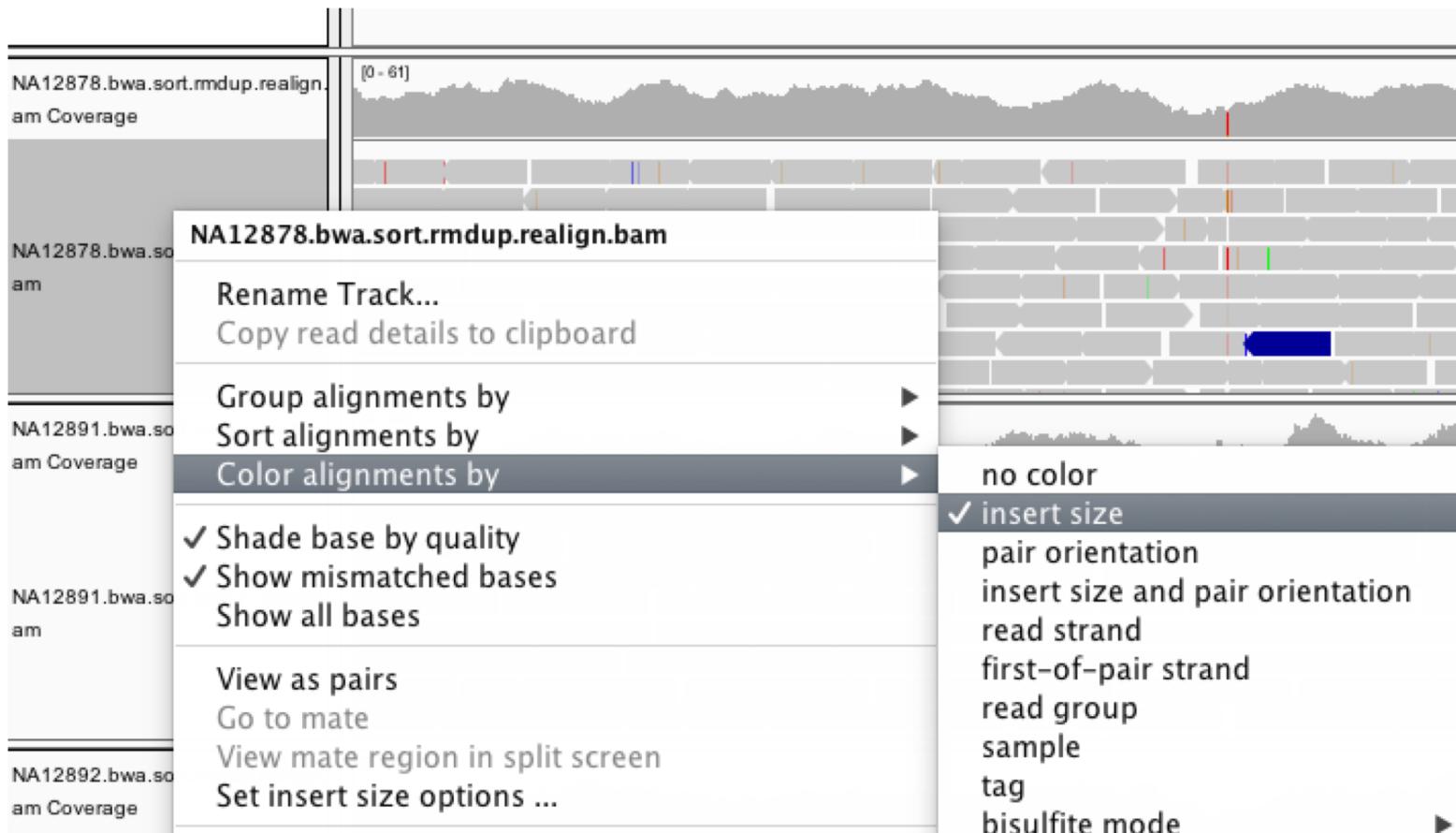
inferred insert size

Subject

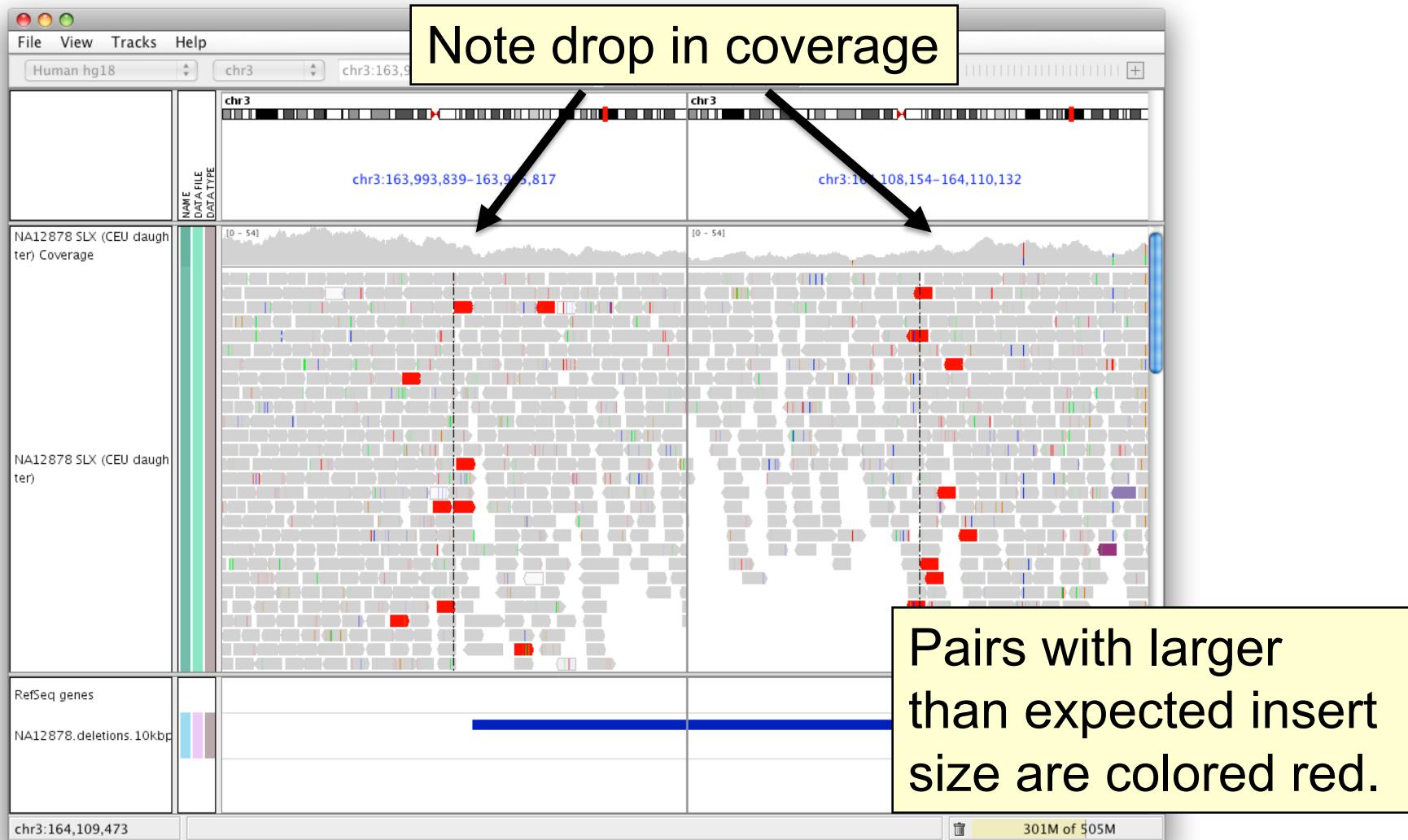


expected insert size

Color by insert size



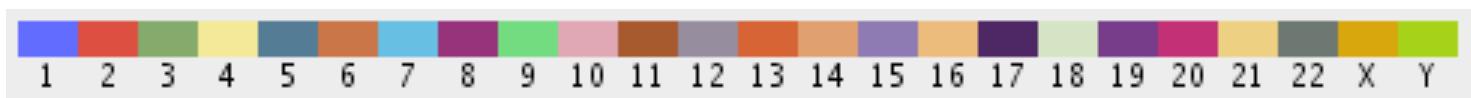
Deletion



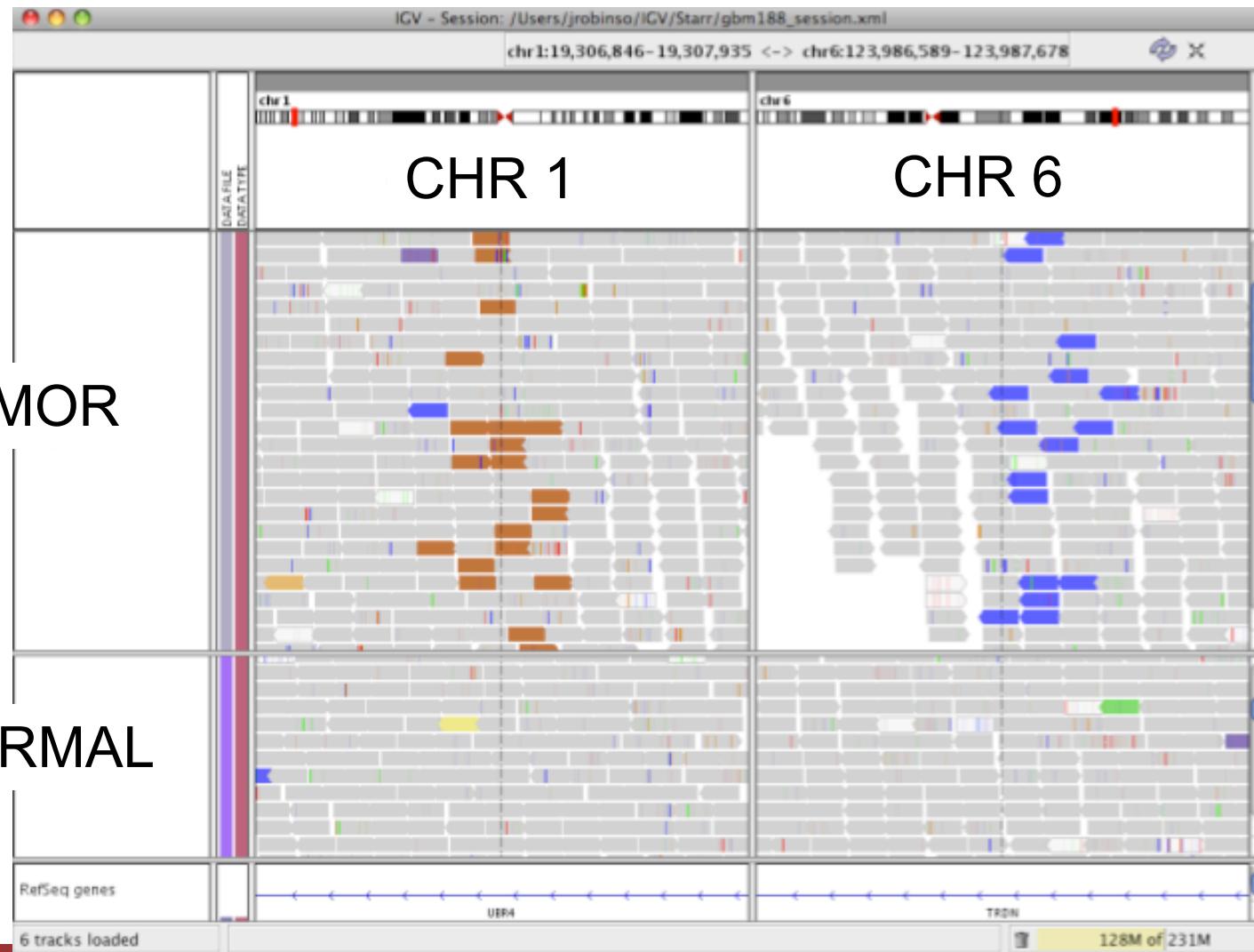
Insert size color scheme

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

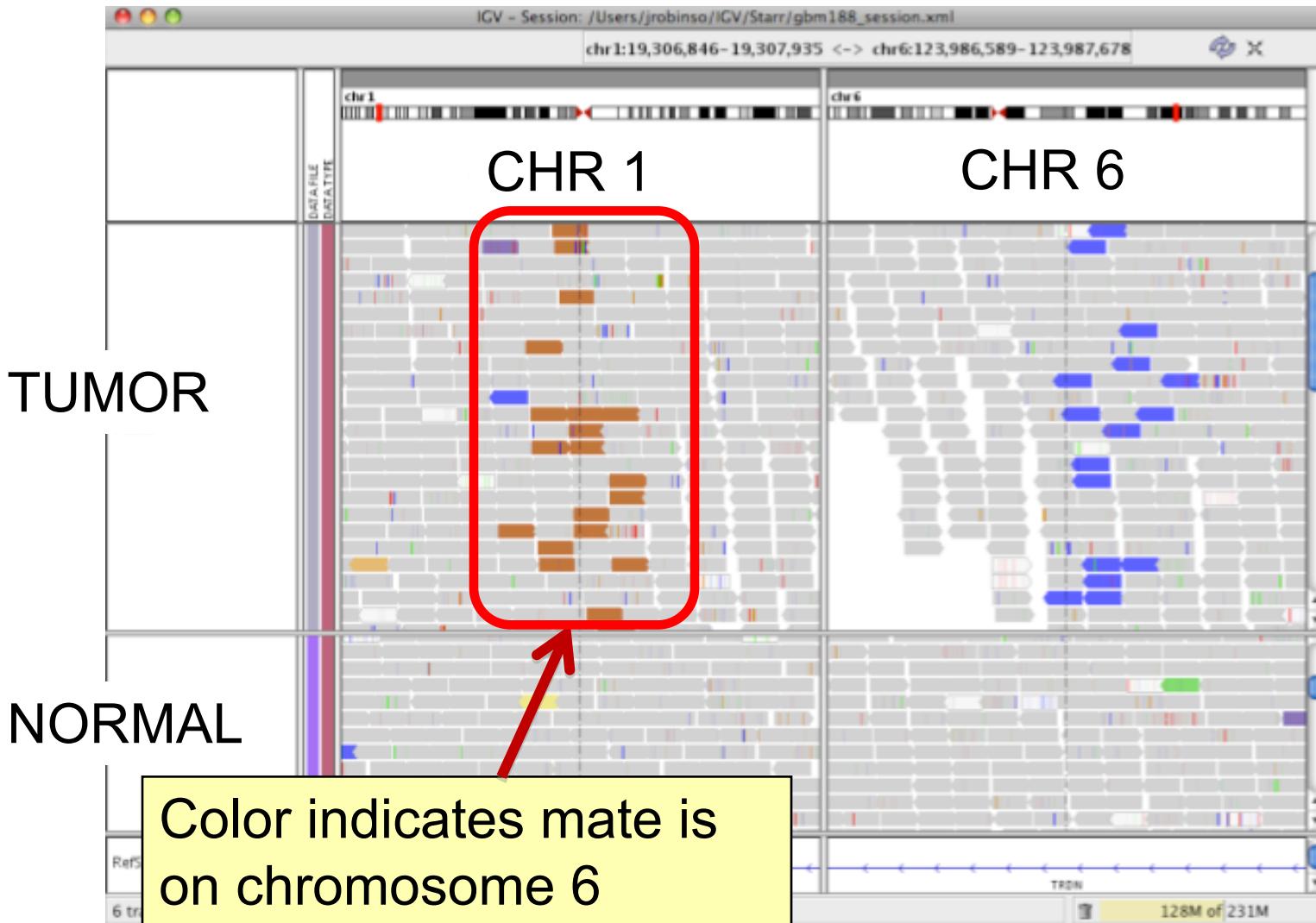
Each end colored by chromosome of its mate



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



A

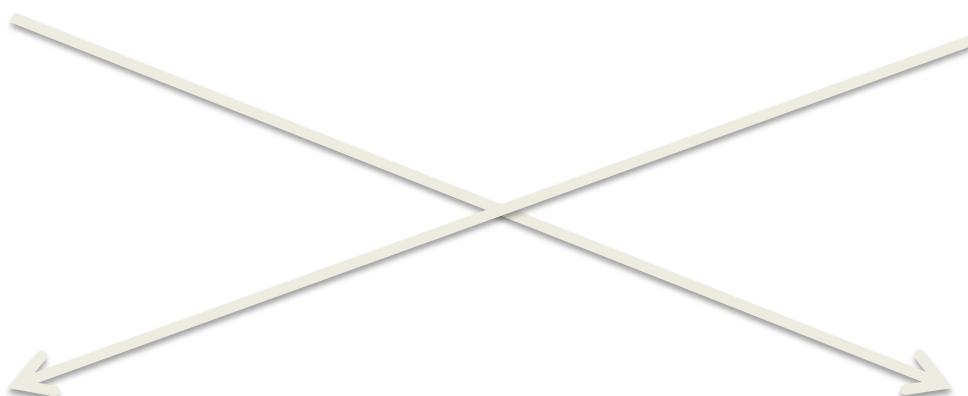
B

Subject



B

A



Inversion

Reference
Genome

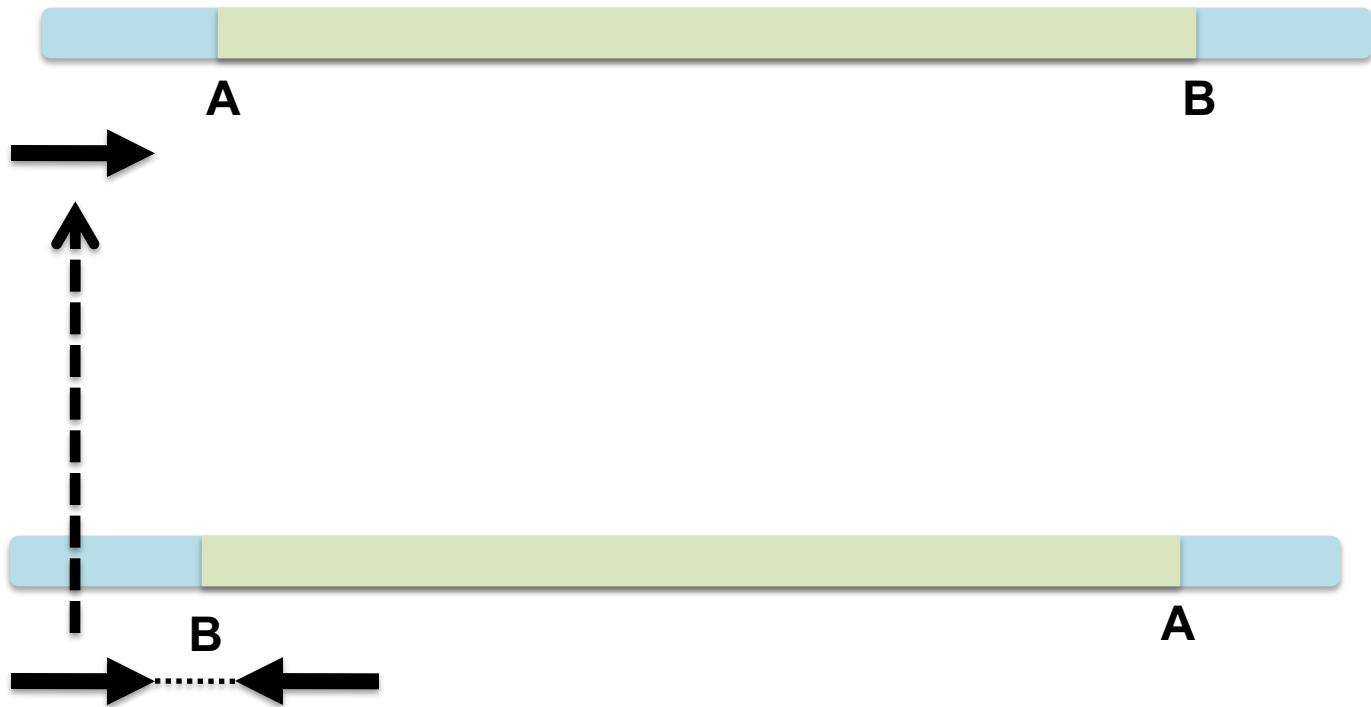


Subject



Inversion

Reference
Genome



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

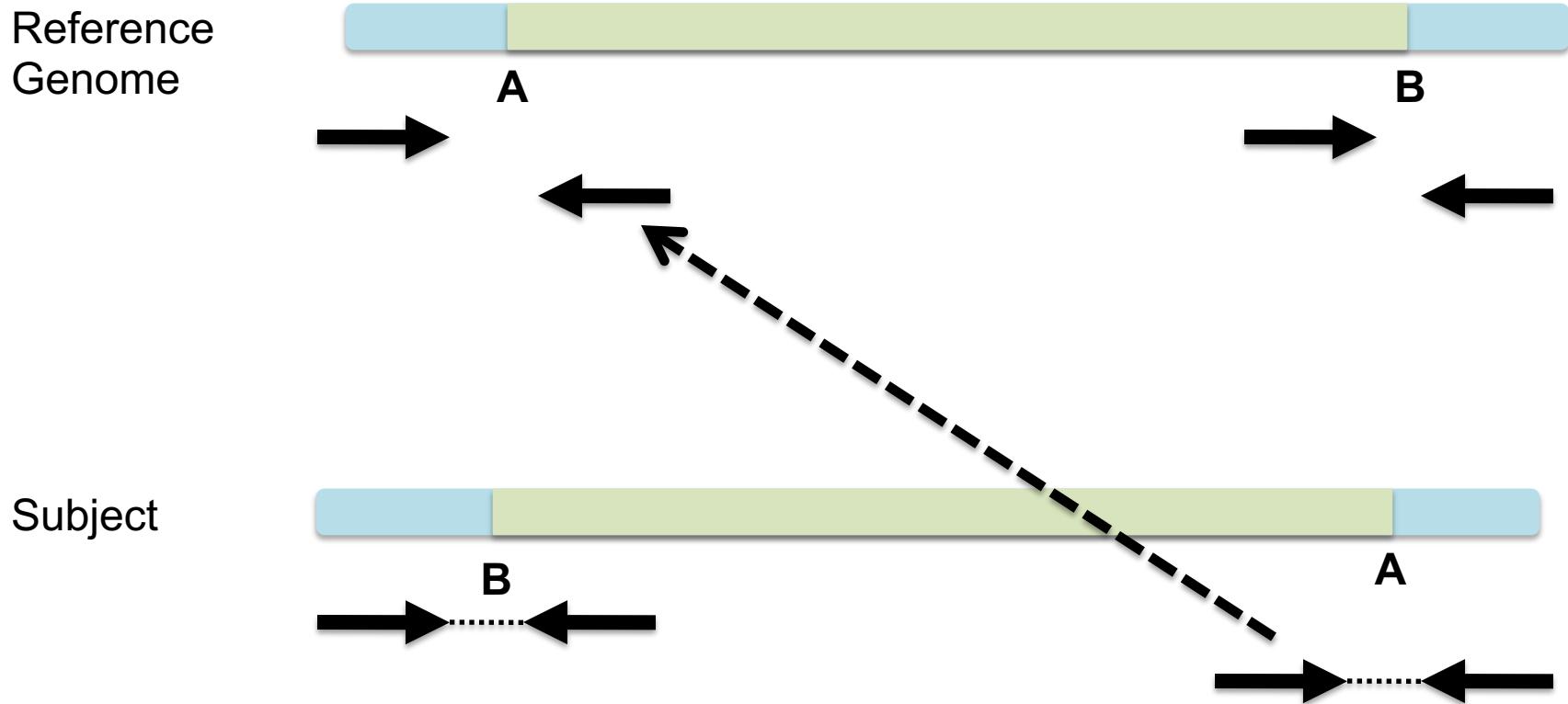
Reference
Genome



Subject

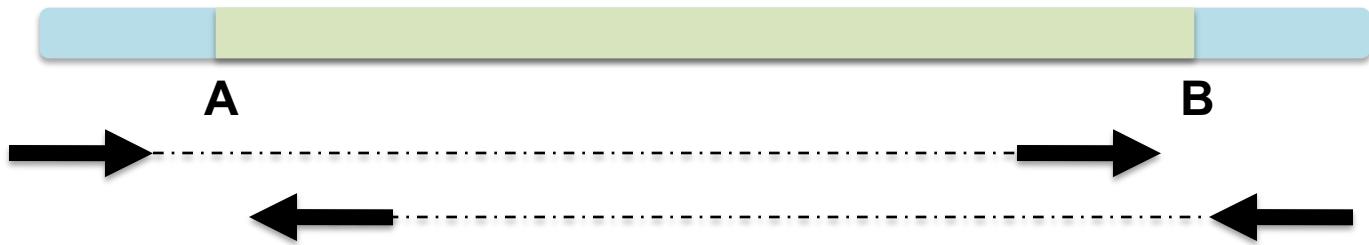


Inversion



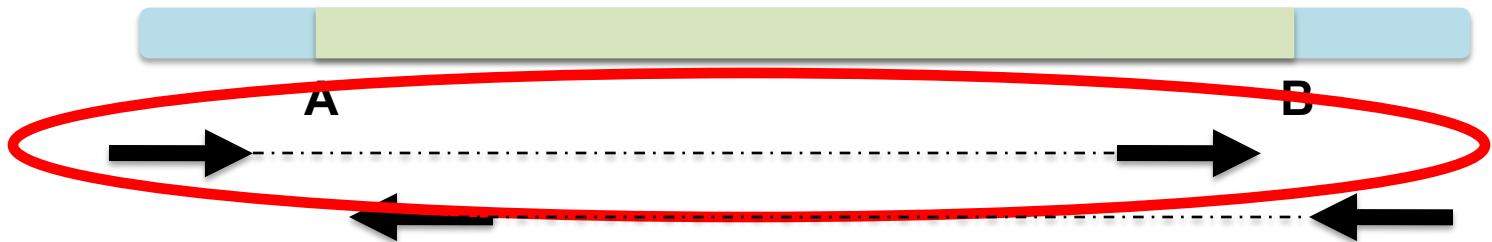
Inversion

Reference
Genome



Inversion

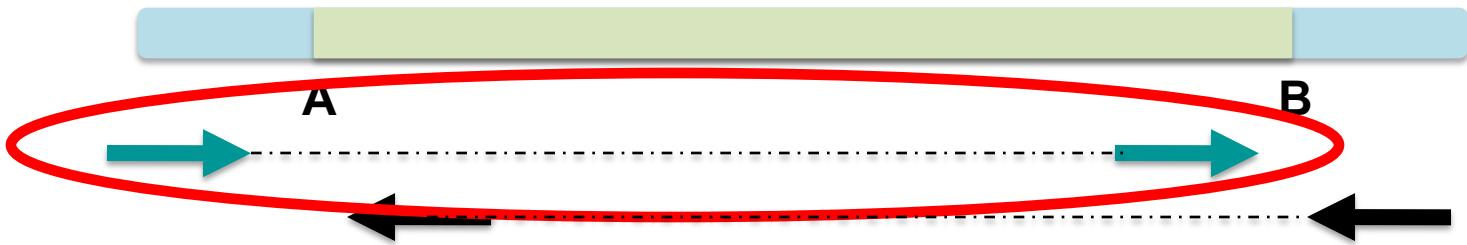
Reference
Genome



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

Inversion

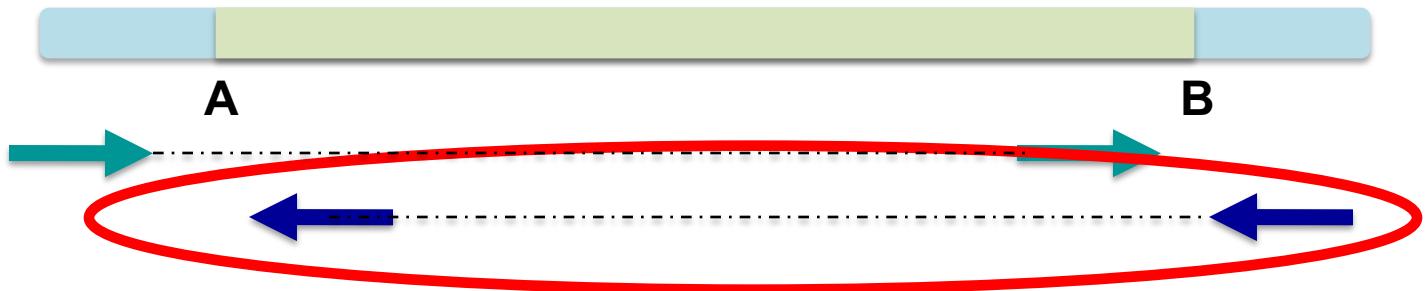
Reference
Genome



“Left” side pair

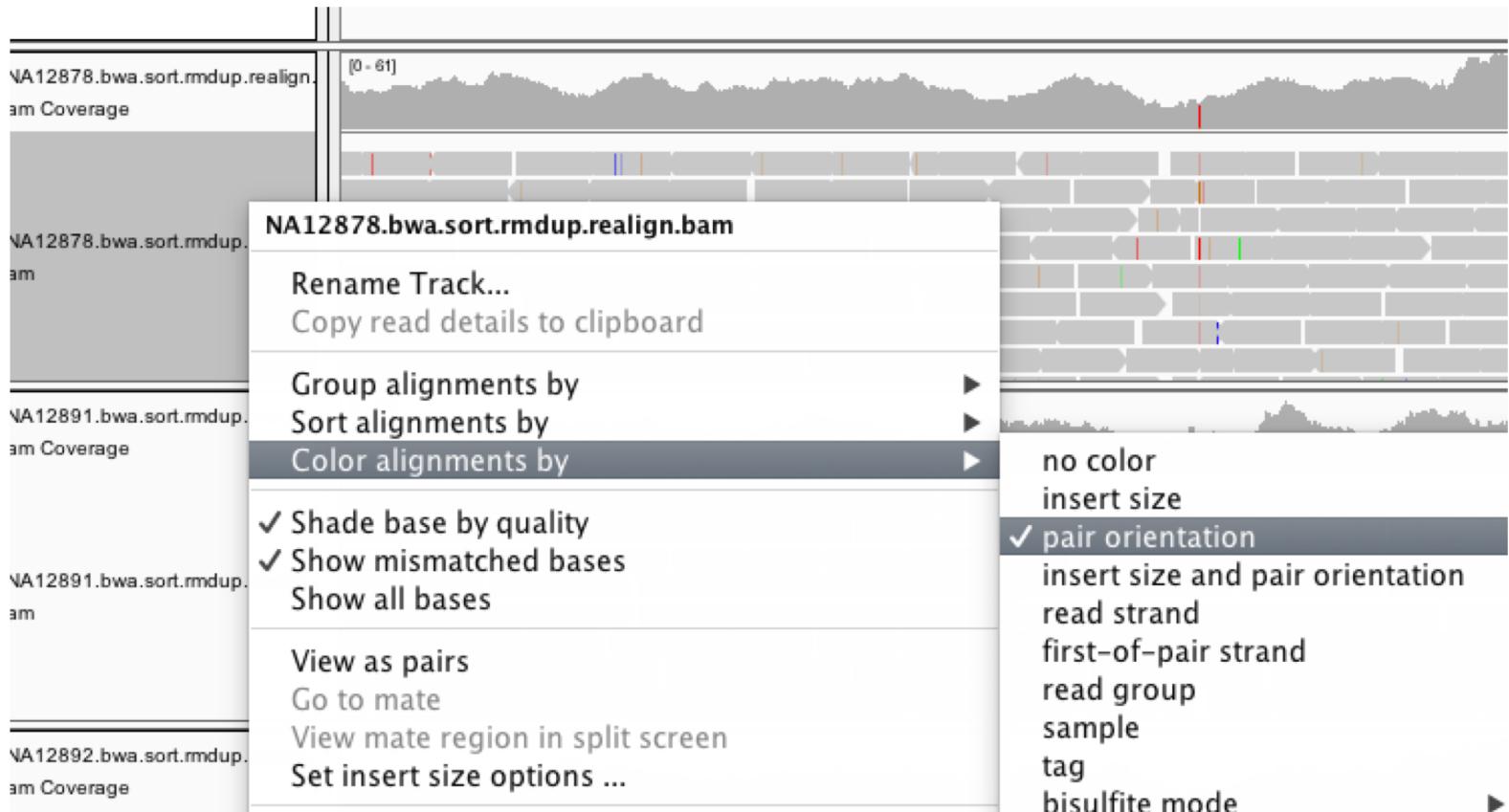
Inversion

Reference
Genome

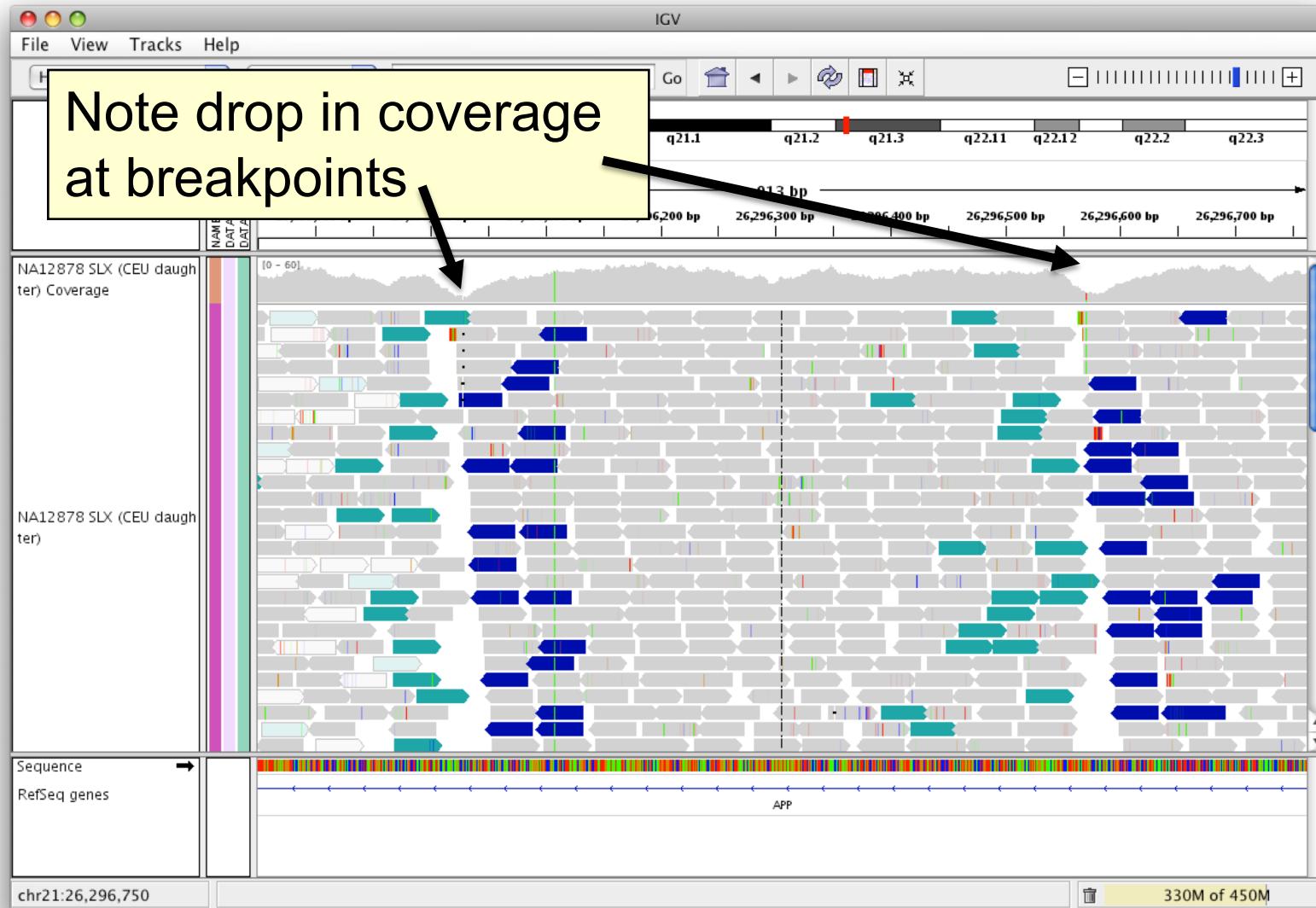


“Right” side pair

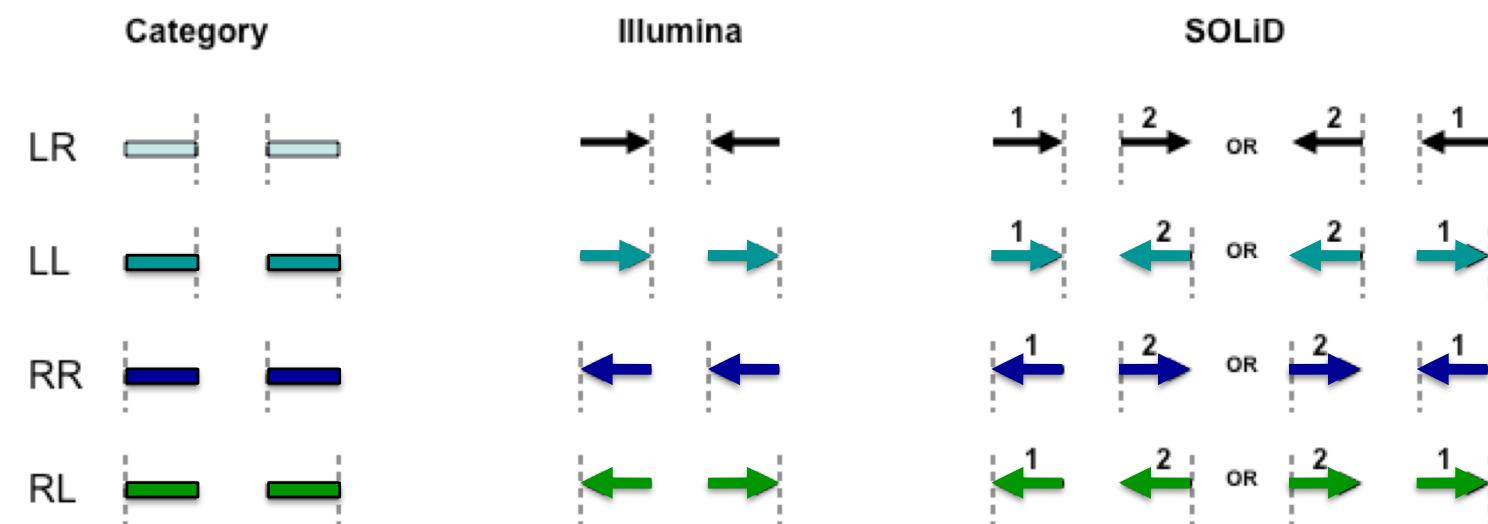
Color by pair orientation



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.

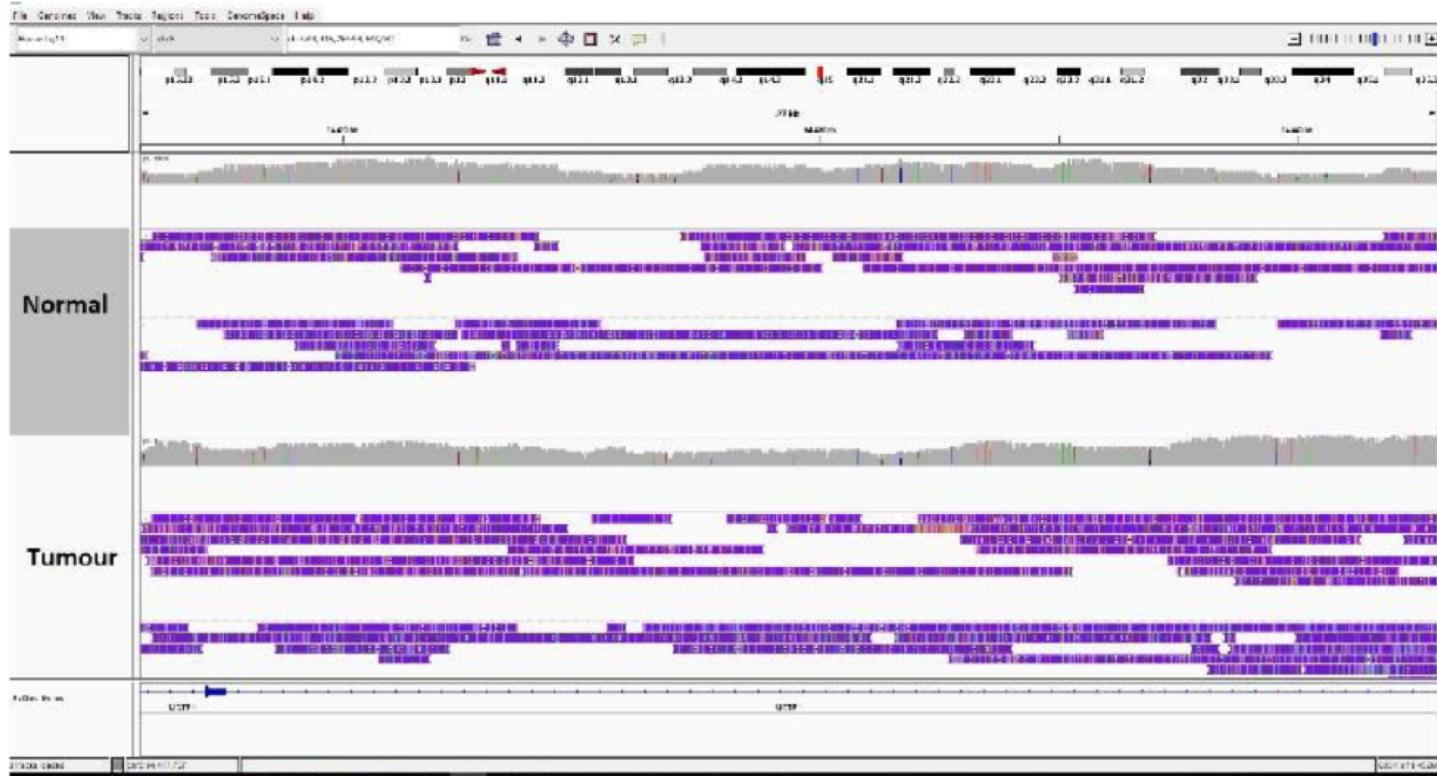
LL,RR Implies inversion in sequenced DNA with respect to reference.

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

Long read considerations



- Commonly see lots of small indels and single base errors that are simply noise
- Can be removed to be able to view the data more cleanly

Long read considerations

Setting an indel threshold hides noise from small indels

File Genomes View Tracks Regions Tools GenomeSpace Help

Human hg19

Human hg19

Preferences... Color Legends ... Set user specific preferences

Show Name Panel
Show Attribute Display
Select Attributes to Show...
Show Header Panel
Reorder Panels...
Go to

General Tracks Variants Charts Alignments Probes Proxy Advanced Cram

Track Display Options

On initial load show: Alignment Track Coverage Track Splice Junction Track

Alignment Track Options

Visibility range threshold (kb): 1000 Range at which alignments become visible

Downsample reads Max read count: 100 per window size (bases): 50

Shade mismatched bases by quality: 5 to 20

Mapping quality threshold: 0

Label indels > 1 bases

Flag clipping > 0 bases

Hide indels < 20 bases

Filter duplicate reads

Filter vendor failed reads

Filter secondary alignments

Show center line

Flag unmapped pairs

Show soft-clipped bases

Quick consensus mode

Filter supplementary alignments

Hidden SAM ta... SA,MD,XA,RG

Coverage Track Options

Coverage allele-fraction threshold: 0.2 Quality weight allele fraction

Splice Junction Track Options

Show flanking regions Min flanking width: 0 Min junction coverage: 1

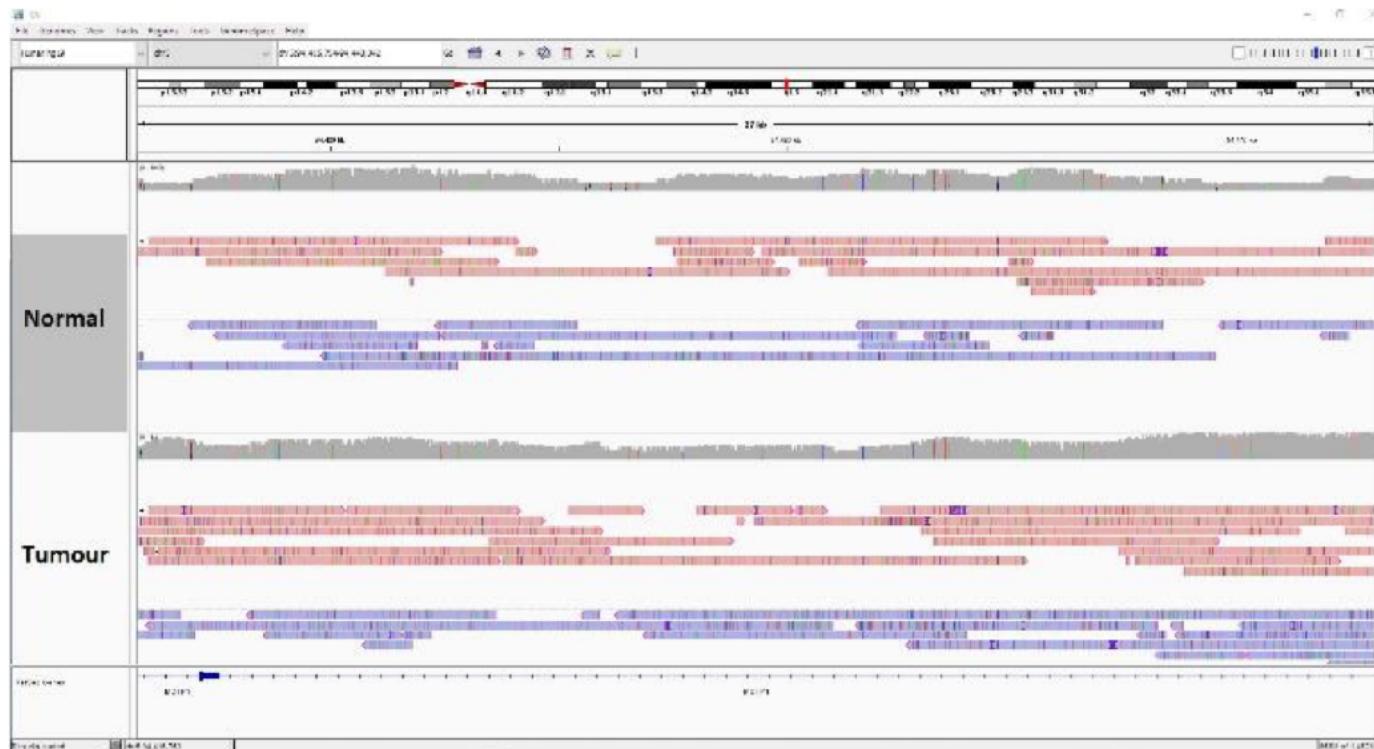
Insert Size Options

Defaults Minimum (bp): 50 Compute Minimum (percentile): 0.5

Maximum (bp): 1000 Maximum (percentile): 99.5

OK Cancel

Long read considerations



- Reads are not all purple dashes
- Next step would be to call a consensus at each position

Long read considerations

The screenshot shows a software interface with a menu bar (File, Genomes, View, Tracks, Regions, Tools, GenomeSpace, Help) and a genome selection dropdown (Human hg19). The 'View' menu is open, showing options like Preferences..., Color Legends..., Show Name Panel, Show Attribute Display, Show Header Panel, Reorder Panels..., and Go to. The main window displays 'Track Display Options' with several tabs: General, Tracks, Variants, Charts, Alignments, Probes, Proxy, Advanced, and Cram. The 'Alignments' tab is selected. Under 'Alignment Track Options', there are checkboxes for Alignment Track, Coverage Track, and Splice Junction Track. Other settings include Visibility range threshold (kb), Downsample reads Max read count per window size (bases), Shade mismatched bases by quality, Mapping quality threshold, Label indels > [1] bases, Flag clipping > [0] bases, Hide indels < [20] bases, Filter duplicate reads, Filter vendor failed reads, Filter secondary alignments, Show center line, Hidden SAM ta... SA,MD,XA,RG, Flag unmapped pairs, Show soft-clipped bases, Quick consensus mode (which is checked), and Filter supplementary alignments. An arrow points from the 'Quick consensus mode' checkbox to a yellow callout box. The callout box contains the text: "Option for generating consensus sequences". The 'Coverage Track Options' section includes a Coverage allele-fraction threshold input field (0.2) and a Quality weight allele fraction checkbox. The 'Splice Junction Track Options' section includes a Show flanking regions checkbox, a Min flanking width input field (0), and a Min junction coverage input field (1). The 'Insert Size Options' section includes a Defaults Minimum (bp) input field (50), a Compute Minimum (percentile) checkbox (checked), a Maximum (bp) input field (1000), and a Maximum (percentile) input field (99.5). At the bottom right are OK and Cancel buttons.

Long read considerations



- Much easier to parse through the genomic data
- Large insertions and deletions are also labelled now

Manual Review Standard Operating Procedure (SOP) paper

© American College of Medical Genetics and Genomics

ARTICLE | Genetics
inMedicine

Open

Standard operating procedure for somatic variant refinement of sequencing data with paired tumor and normal samples

Erica K. Barnell, BS¹, Peter Ronning, BS¹, Katie M. Campbell, BS¹, Kilannin Krysiak, PhD^{1,2}, Benjamin J. Ainscough, PhD^{1,3}, Lana M. Sheta¹, Shahil P. Pema¹, Alina D. Schmidt, BS¹, Megan Richters, BS¹, Kelsy C. Cotto, BS¹, Arpad M. Danos, PhD¹, Cody Ramirez, BS¹, Zachary L. Skidmore, MEng¹, Nicholas C. Spies, BS¹, Jasreet Hundal, MS¹, Malik S. Sediqzad¹, Jason Kunisaki, BS¹, Felicia Gomez, PhD¹, Lee Trani, BS¹, Matthew Matlock, BS¹, Alex H. Wagner, PhD¹, S. Joshua Swamidass, MD/PhD^{4,5}, Malachi Griffith, PhD^{1,2,3,6} and Obi L. Griffith, PhD^{1,2,3,6}

Purpose: Following automated variant calling, manual review of aligned read sequences is required to identify a high-quality list of somatic variants. Despite widespread use in analyzing sequence data, methods to standardize manual review have not been described, resulting in high inter- and intralab variability.

Methods: This manual review standard operating procedure (SOP) consists of methods to annotate variants with four different calls and 19 tags. The calls indicate a reviewer's confidence in each variant and the tags indicate commonly observed sequencing patterns and artifacts that inform the manual review call. Four individuals were asked to classify variants prior to, and after, reading the SOP and accuracy was assessed by comparing reviewer calls with orthogonal validation sequencing.

Results: After reading the SOP, average accuracy in somatic variant identification increased by 16.7% (p value = 0.0298) and average interreviewer agreement increased by 12.7% (p value < 0.001). Manual review conducted after reading the SOP did not significantly increase reviewer time.

Conclusion: This SOP supports and enhances manual somatic variant detection by improving reviewer accuracy while reducing the interreviewer variability for variant calling and annotation.

Genetics in Medicine (2018) <https://doi.org/10.1038/s41436-018-0278-z>

Keywords: somatic variant refinement; manual review