

Introduction to NGS Visualization with the Integrative Genomics Viewer (IGV)

**Programming for Biology 2014
Cold Spring Harbor
Jim Robinson**

Agenda

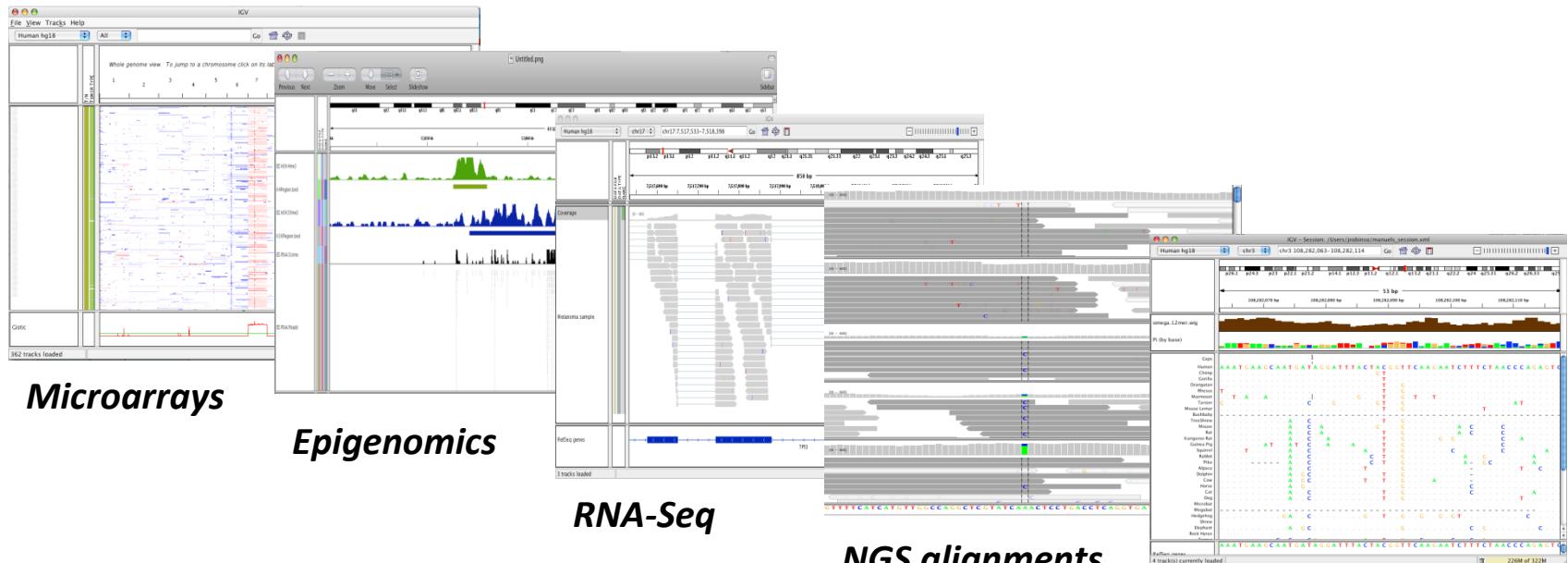


- Introduction
- Using IGV: The Basics
- Data Tracks and File Formats
- NGS Alignments
 - SNPs
 - Structural Events
 - RNA-seq
- igvtools
- Exercises

What is IGV



A desktop application for integrated visualization of multiple data types and annotations in the context of the genome.



Comparative genomics

Installing IGV



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

A screenshot of a web browser displaying the 'Home | Integrative Genomics Viewer' page at https://www.broadinstitute.org/igv/. A red arrow points to the 'Downloads' link in the left sidebar menu.

The page includes the following sections:

- Home**: Main content area featuring a large image of the IGV interface and a 'What's New' section.
- Downloads**: Sidebar menu item highlighted by a red arrow.
- Documents**: Sidebar menu item.
- Hosted Genomes**: Sidebar menu item.
- FAQ**: Sidebar menu item.
- IGV User Guide**: Sidebar menu item.
- File Formats**: Sidebar menu item.
- Release Notes**: Sidebar menu item.
- Credits**: Sidebar menu item.
- Contact**: Sidebar menu item.
- Search website**: Search bar.
- BROAD INSTITUTE**: Logo and copyright information.

What's New

- July 3, 2012.** Soybean (*Glycine max*) and Rat (rn5) genomes have been updated.
- April 20, 2012.** IGV 2.1 has been released. See the [release notes](#) for more details.
- April 19, 2012.** See our new [IGV paper](#) in *Briefings in Bioinformatics*.

Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. 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. Permission to use IGV is granted under the GNU [LGPL license](#).

Citing IGV

To cite your use of IGV in your publication:

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), or
Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. *Briefings in Bioinformatics* 2012.

Funding

Development of IGV is made possible by funding from the [National Cancer Institute](#), the [National Institute of General Medical Sciences](#) of the [National Institutes of Health](#), and the [Starr Cancer Consortium](#).

IGV is participating in the [GenomeSpace](#) initiative.

Installing IGV



The screenshot shows the IGV website's main menu on the left and the 'Log In' page on the right. The main menu includes links for Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, IGV for iPad, Credits, Contact, and a search bar. The 'Log In' page has a message about registration and a red box highlights the 'email address:' field where 'igv-team@broadinstitue.org' is typed. A red arrow points from the text in the yellow box to this field.

For email use
igv-team@broadinstitue.org

Launch IGV



A screenshot of a web browser window showing the 'Downloads' section of the IGV website. The URL is 'www.broadinstitute.org/igv/download'. The page includes a sidebar with links like Home, Downloads, Documents, and Contact. A main content area shows instructions for installing IGV, including sections for 'Mac Application', 'Java Web Start', and 'Binary Distribution'. In the 'Mac Application' section, there is a 'Download Mac App' button highlighted with a red arrow pointing from the yellow callout box.

Download the Mac App bundle and double-click to unzip it.

Using IGV: The Basics

Using IGV: the basics



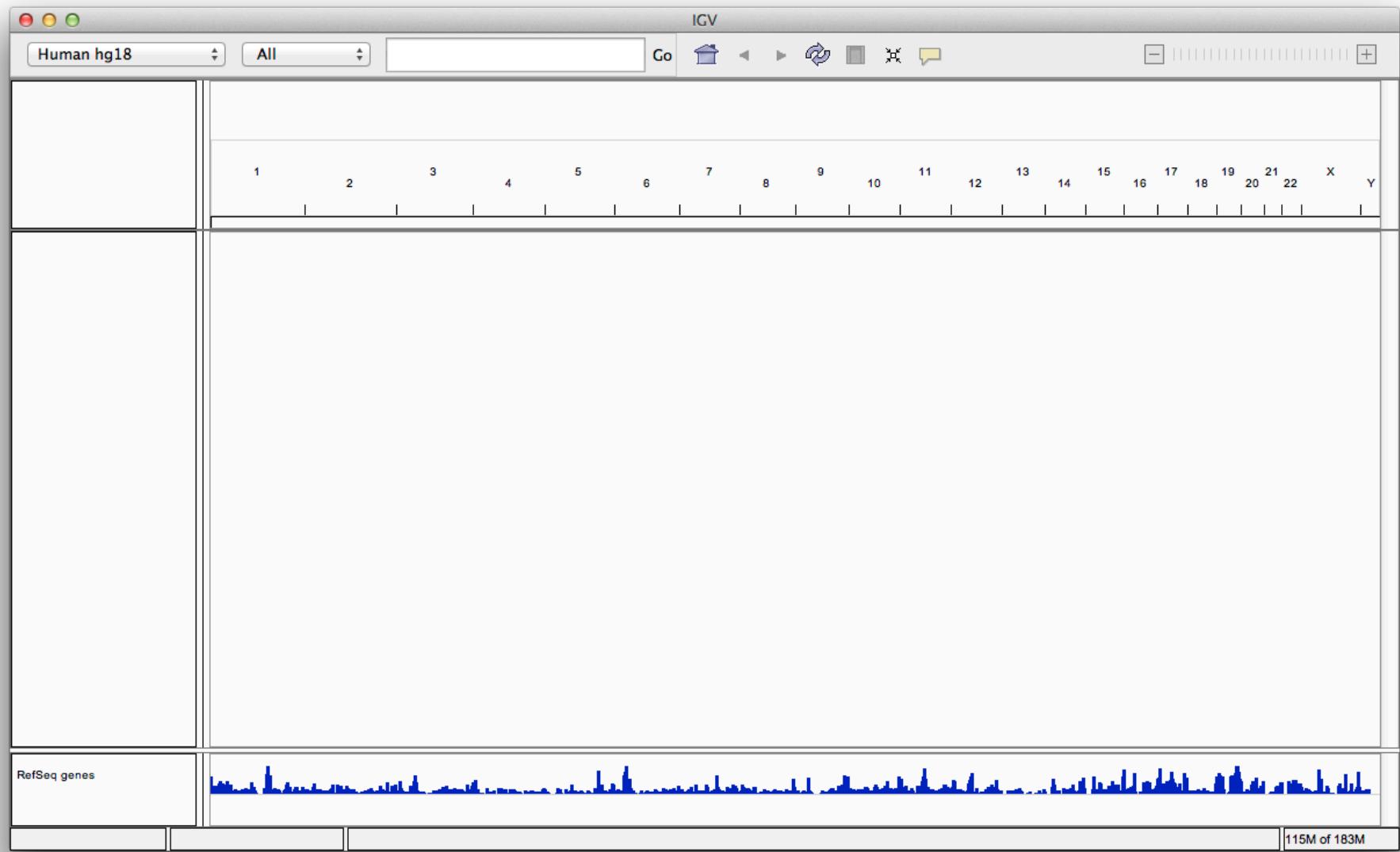
Hands-on exercise

- Launch IGV
- Select a reference genome
- Load data
- Navigate through the data

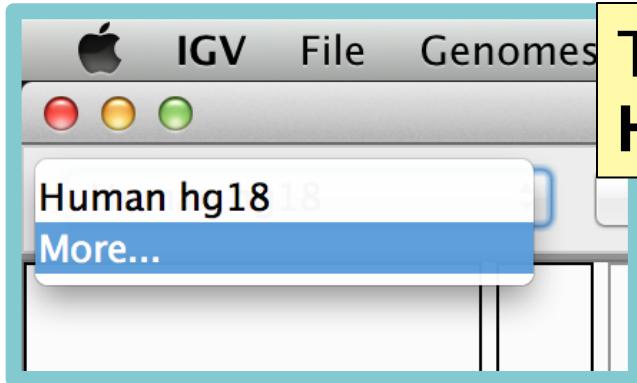
Select the reference genome



A screenshot of the IGV software interface. At the top left, there is a dropdown menu labeled "Human hg18". To its right is another button labeled "All". A yellow callout box with a black border and rounded corners is positioned over the "Human hg18" dropdown, containing the text "Select genome from the drop-down menu". The main window shows a genomic track for chromosome 14, with other chromosomes (15-22, X, Y) visible at the top. Below the track, a "RefSeq genes" track displays blue vertical bars representing gene locations. In the bottom right corner of the main window, the text "115M of 183M" is visible. The entire interface has a light gray background with a dark gray header bar.



Select the reference genome

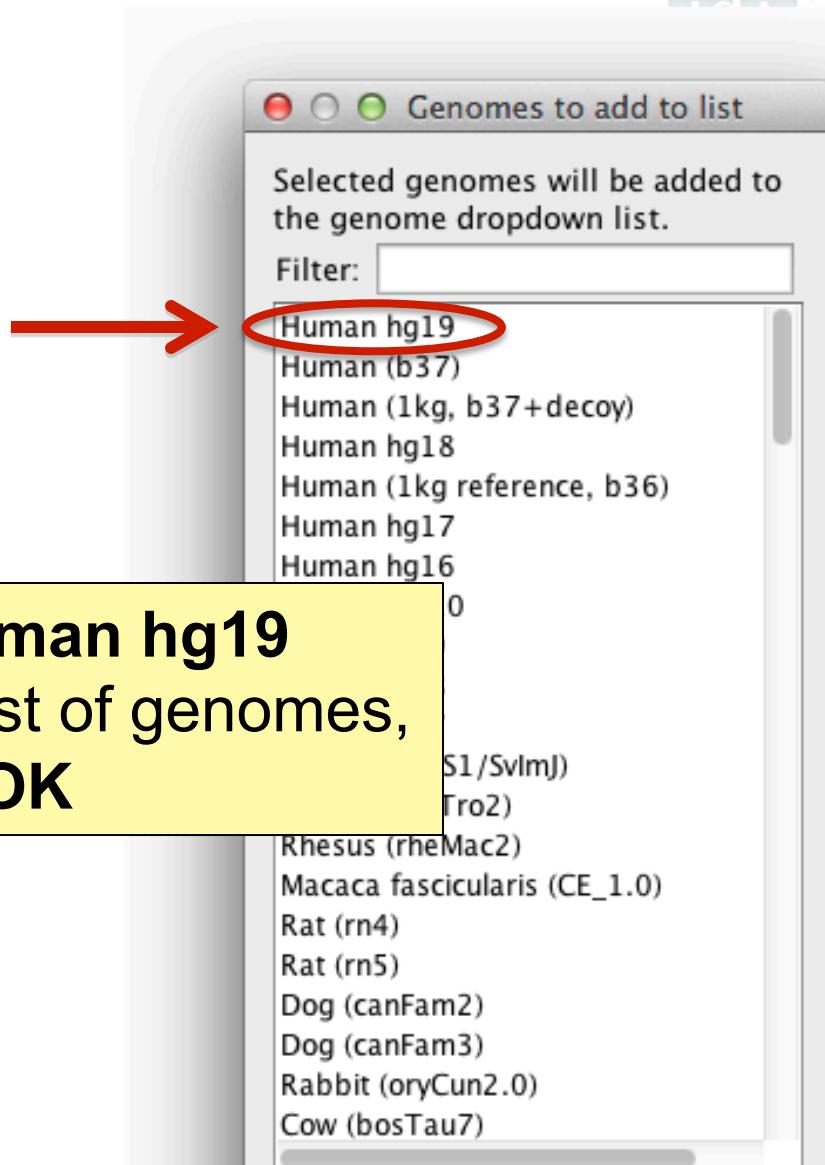
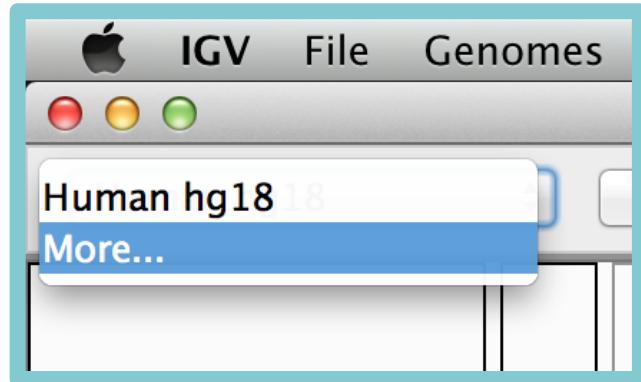


Today, we will use both
Human hg18 and hg19



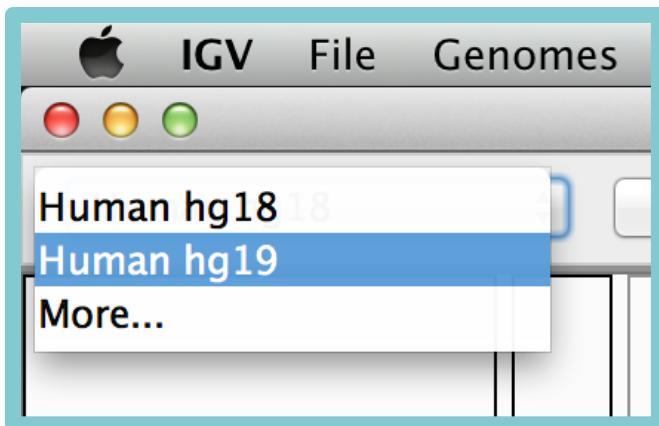
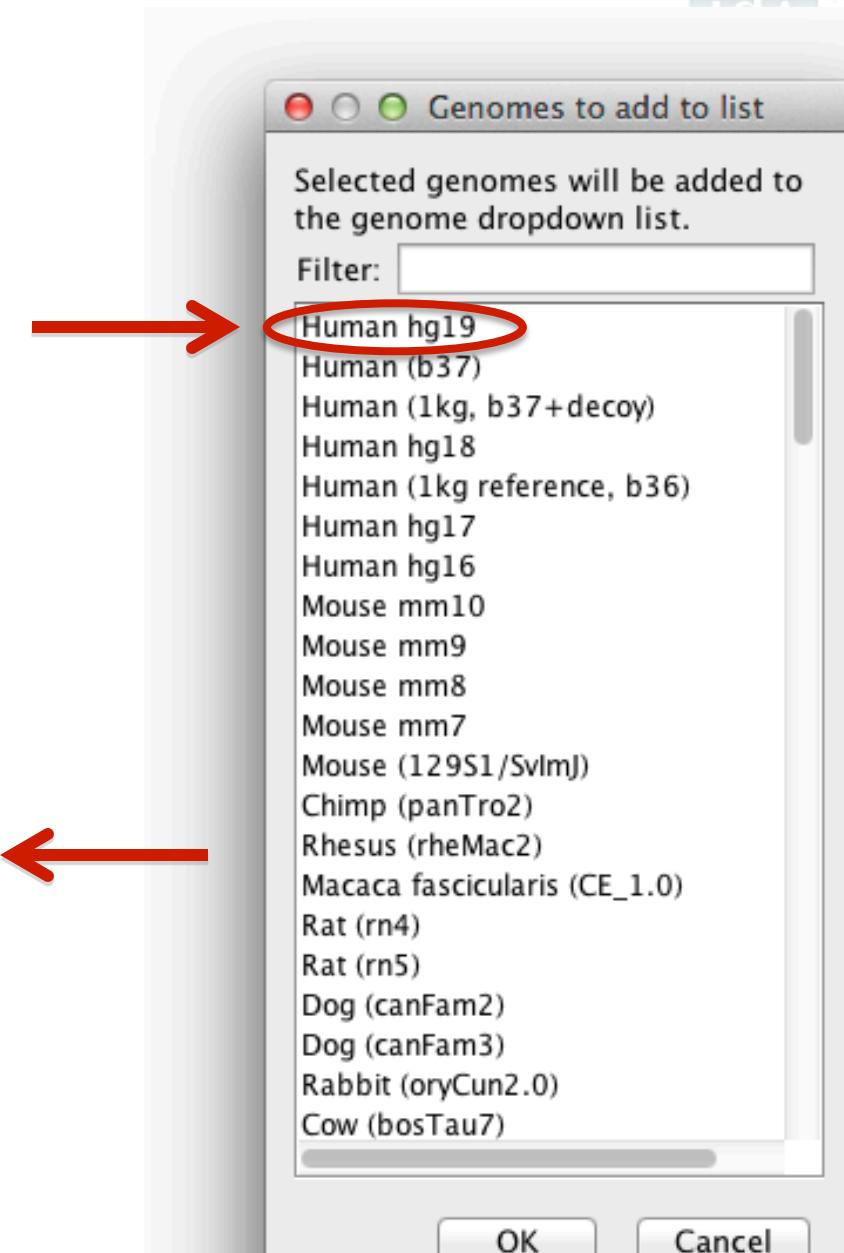
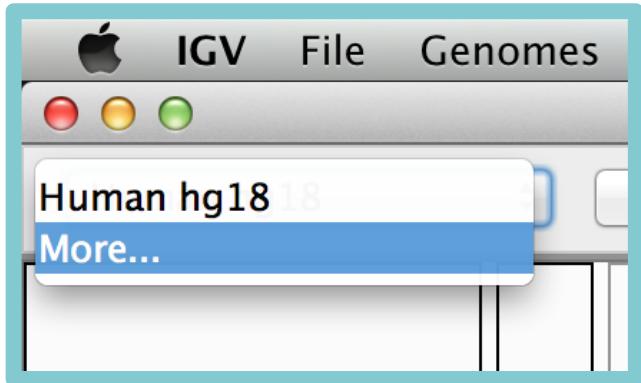
If **Human hg19** is not in the menu,
then click on **More...**

Select the reference genome



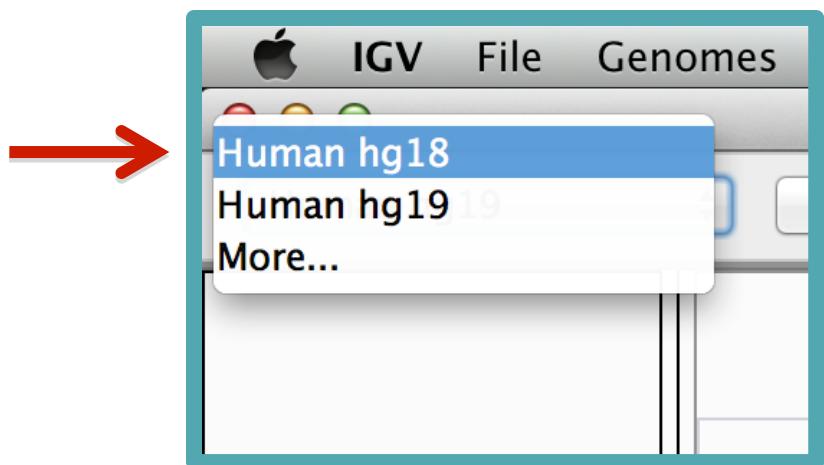
Select **Human hg19**
from the list of genomes,
and click **OK**

Select the reference genome



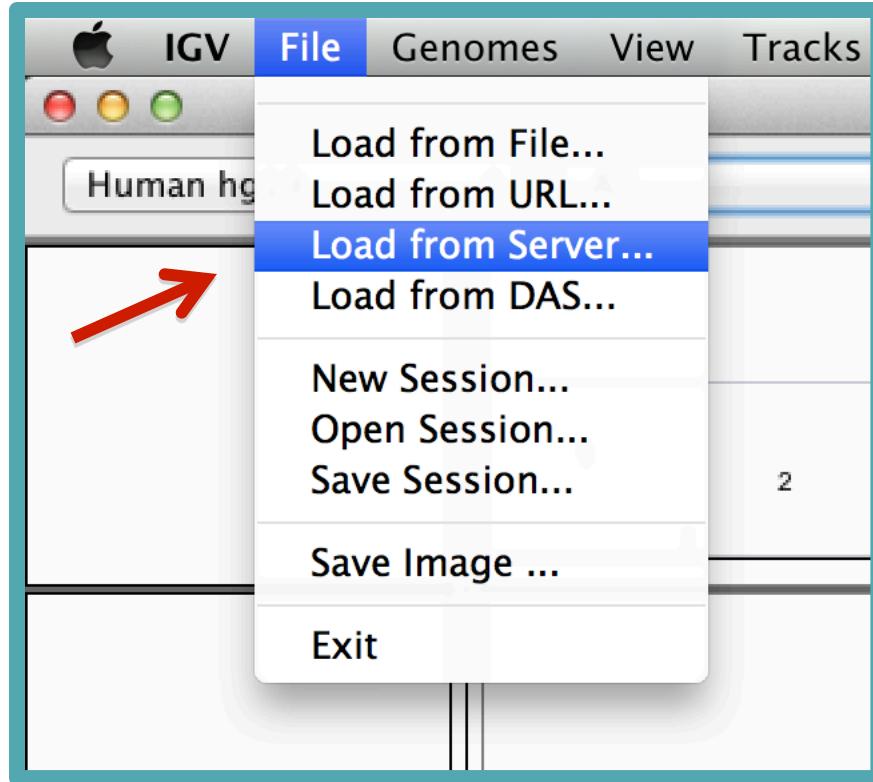
Select the reference genome

Select Human hg18

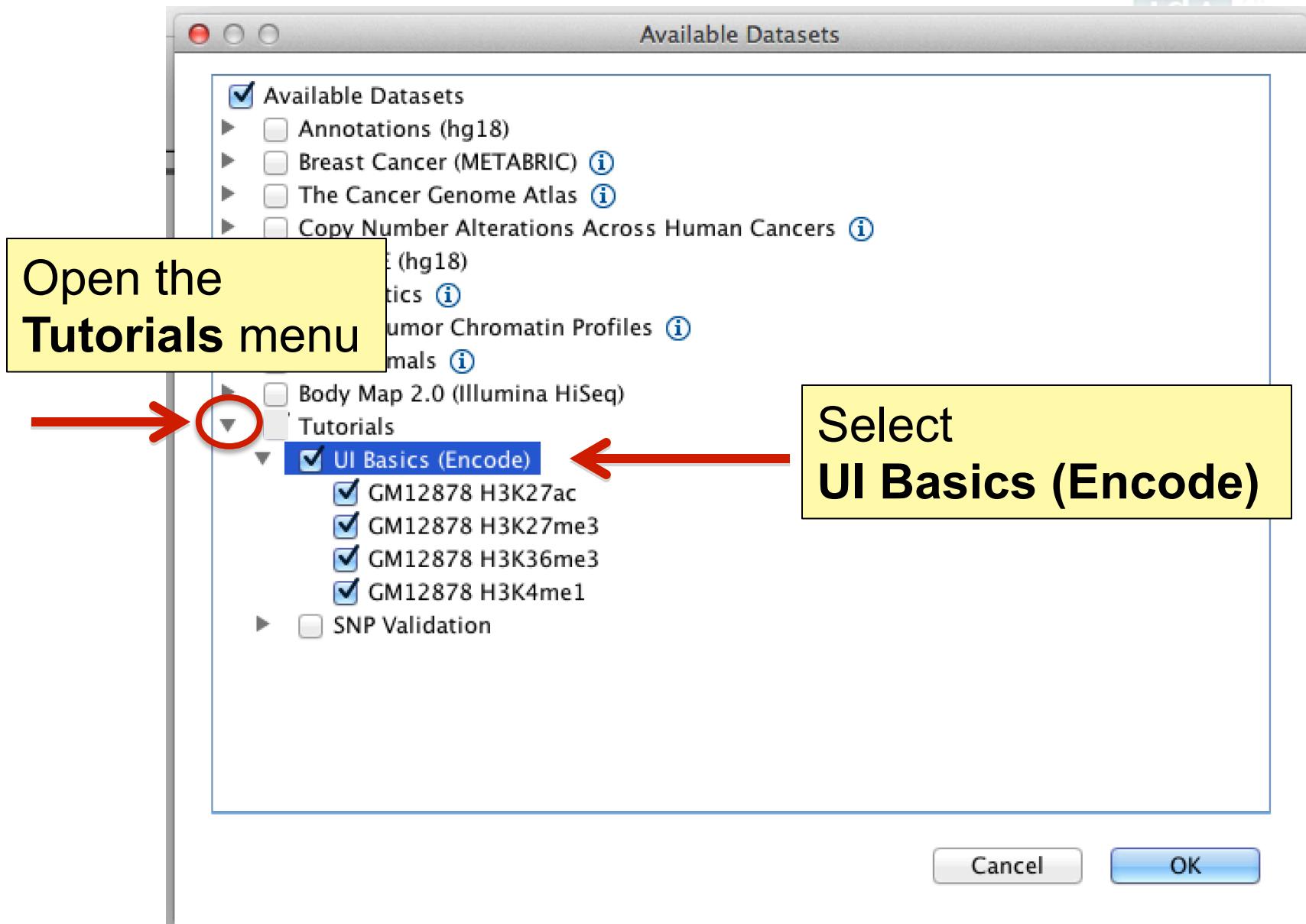


Load data

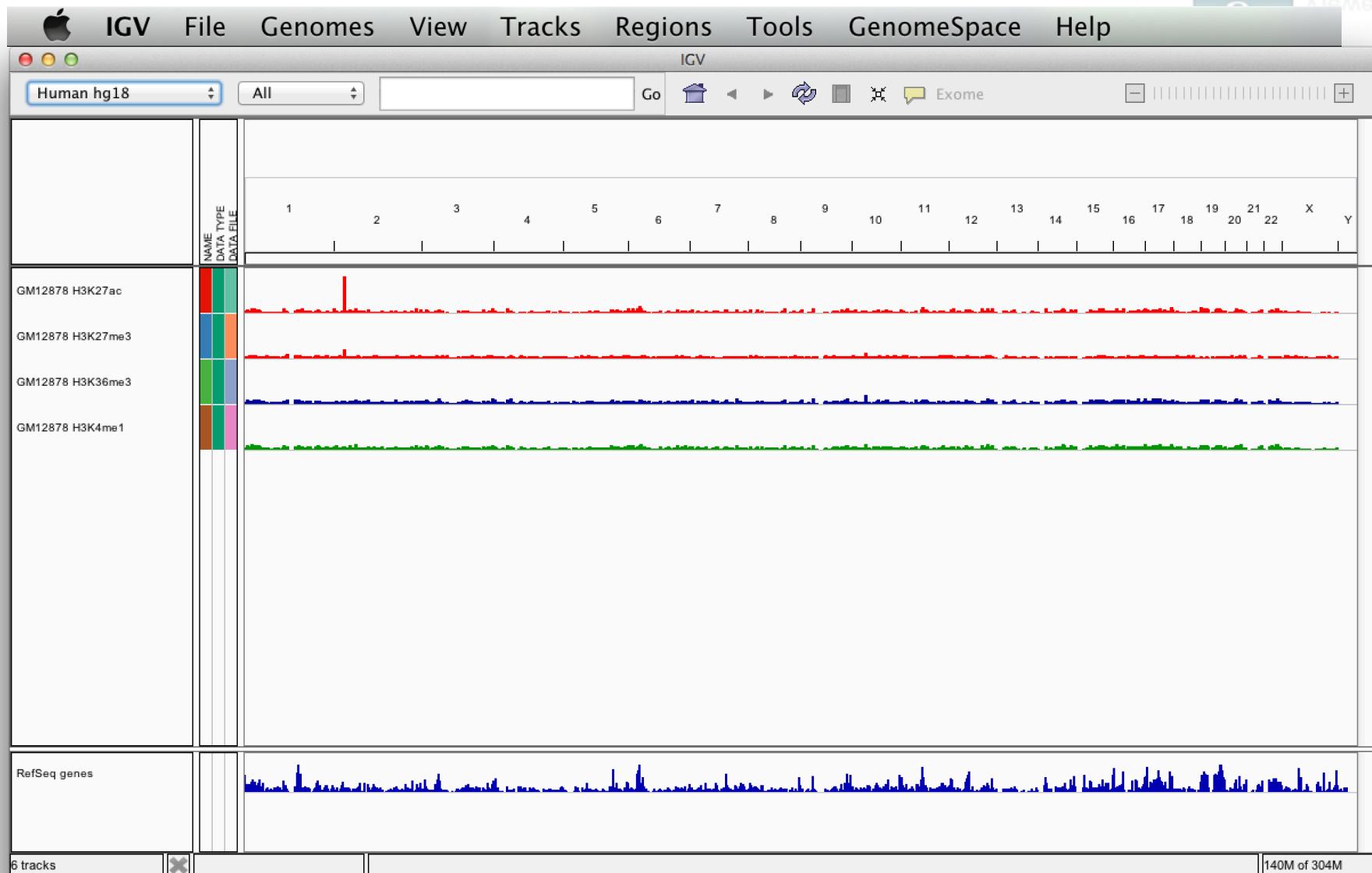
Select File > Load from Server...



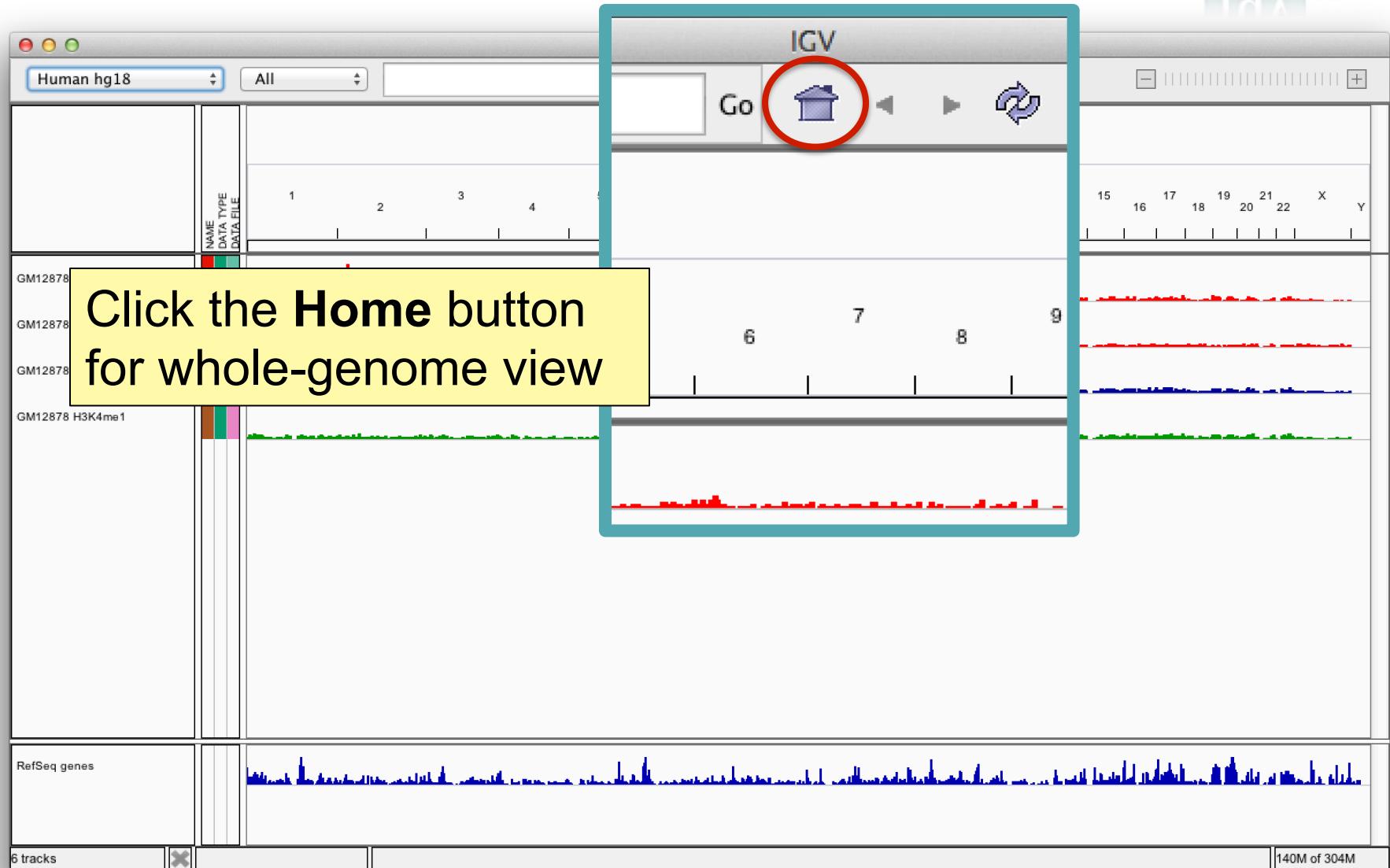
Load data



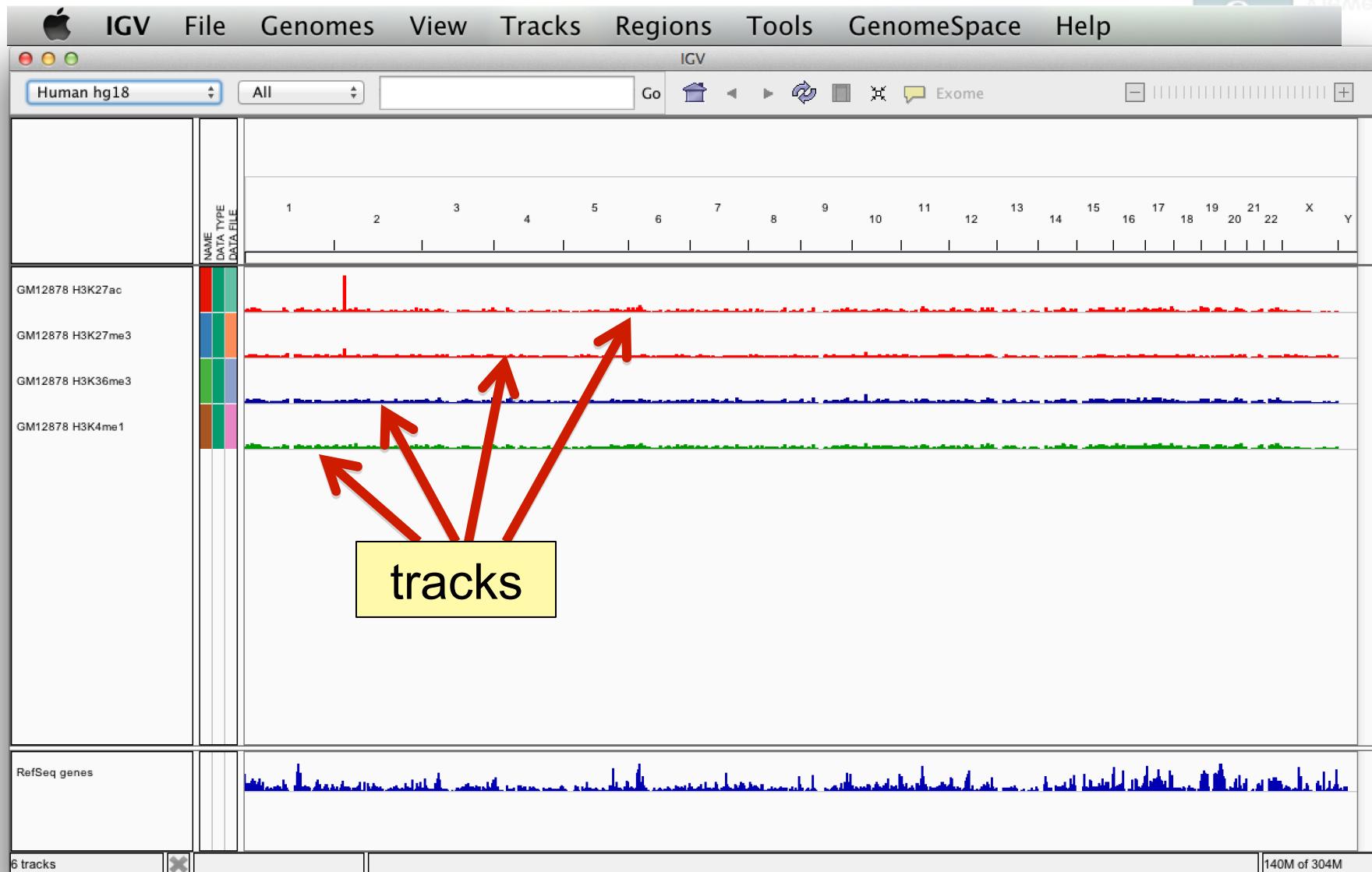
Screen layout



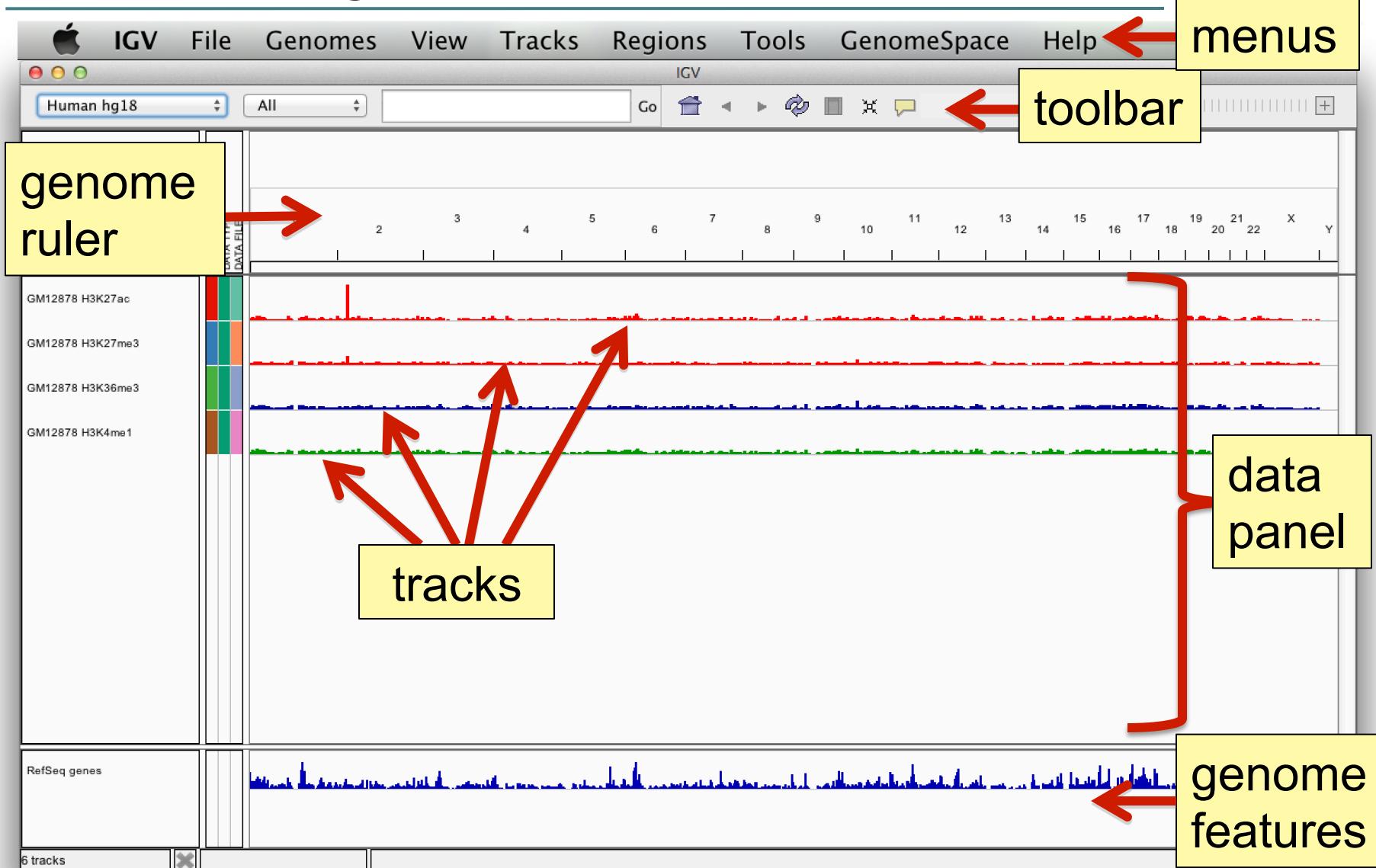
Screen layout



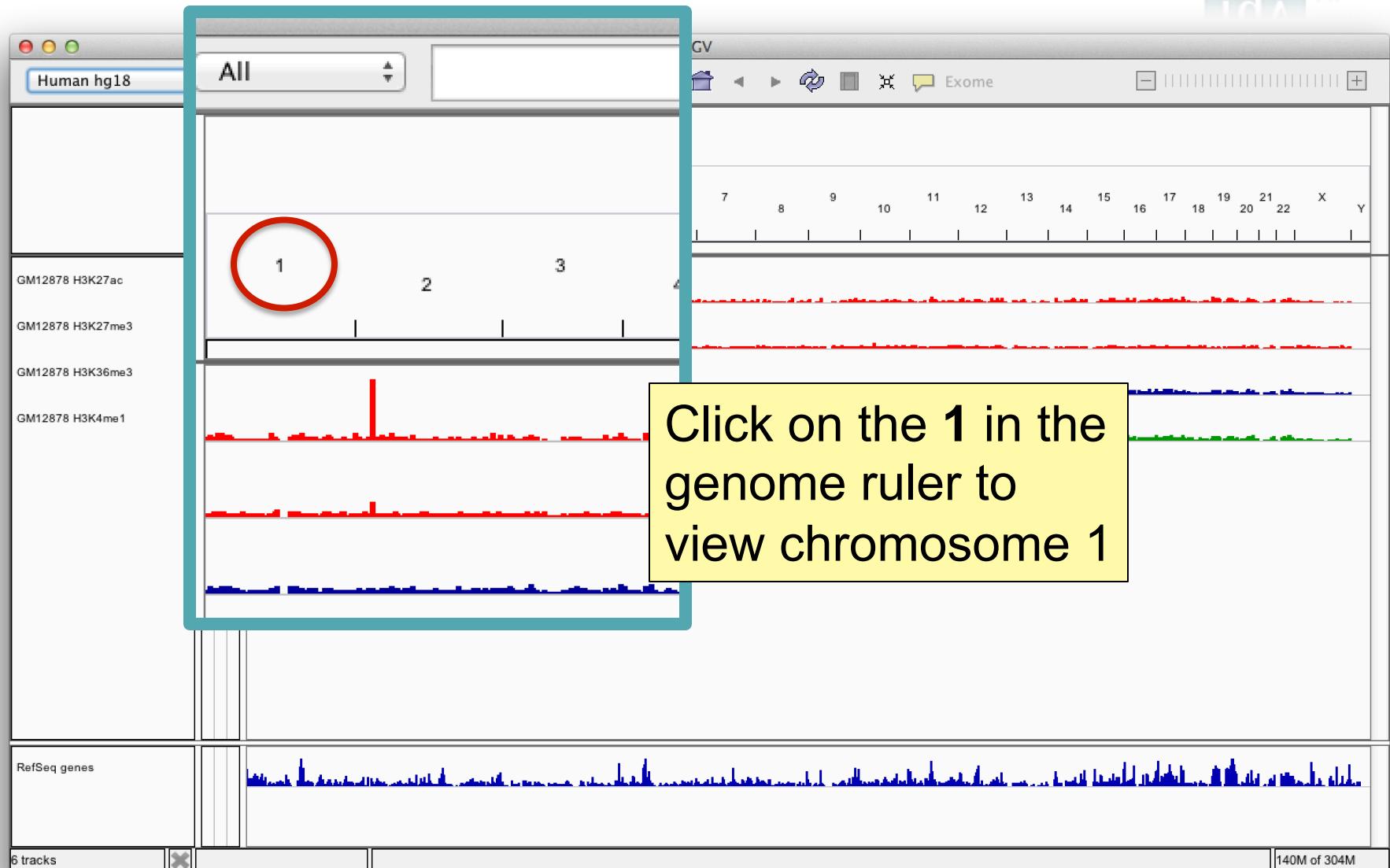
Screen layout



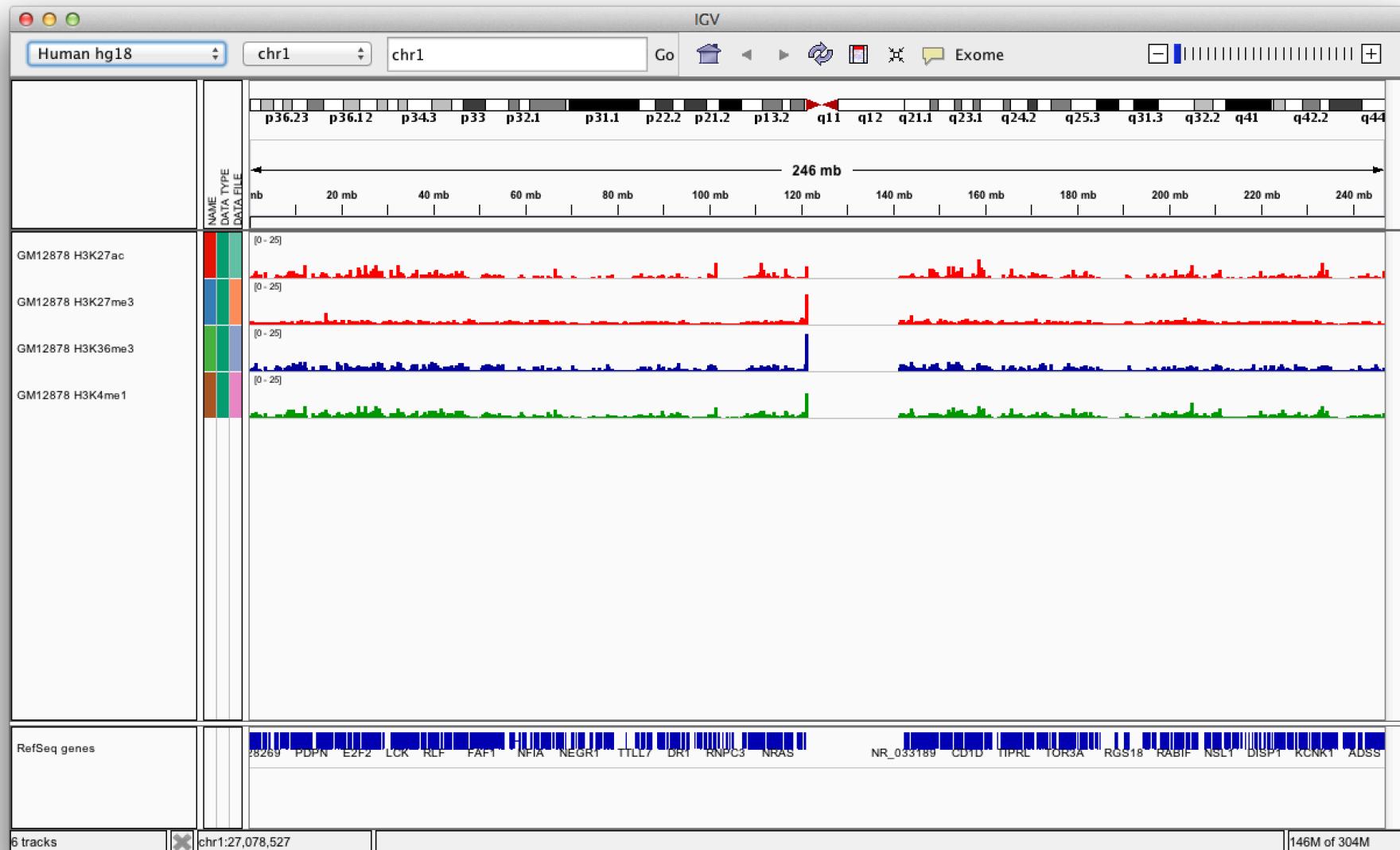
Screen layout



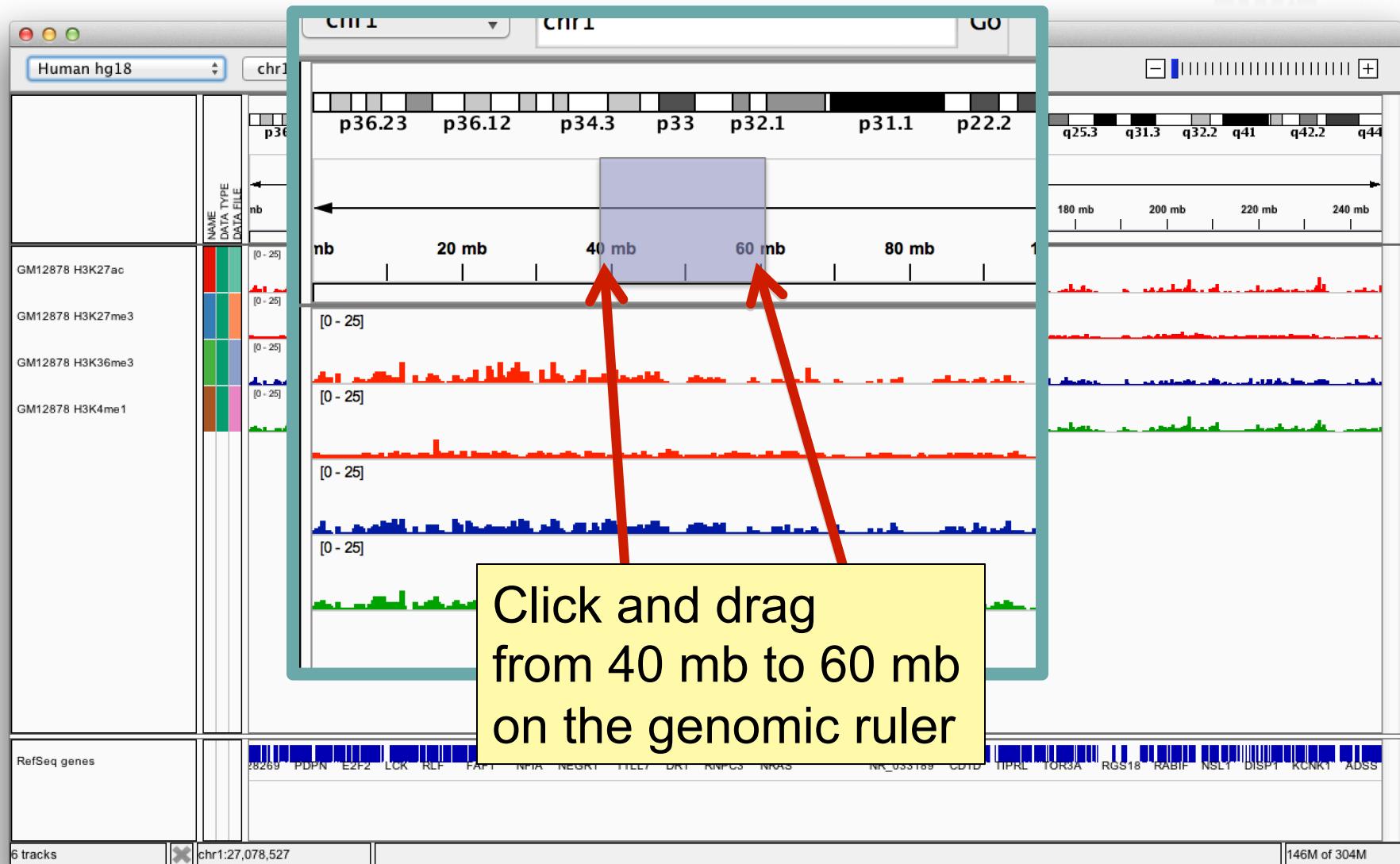
Navigate



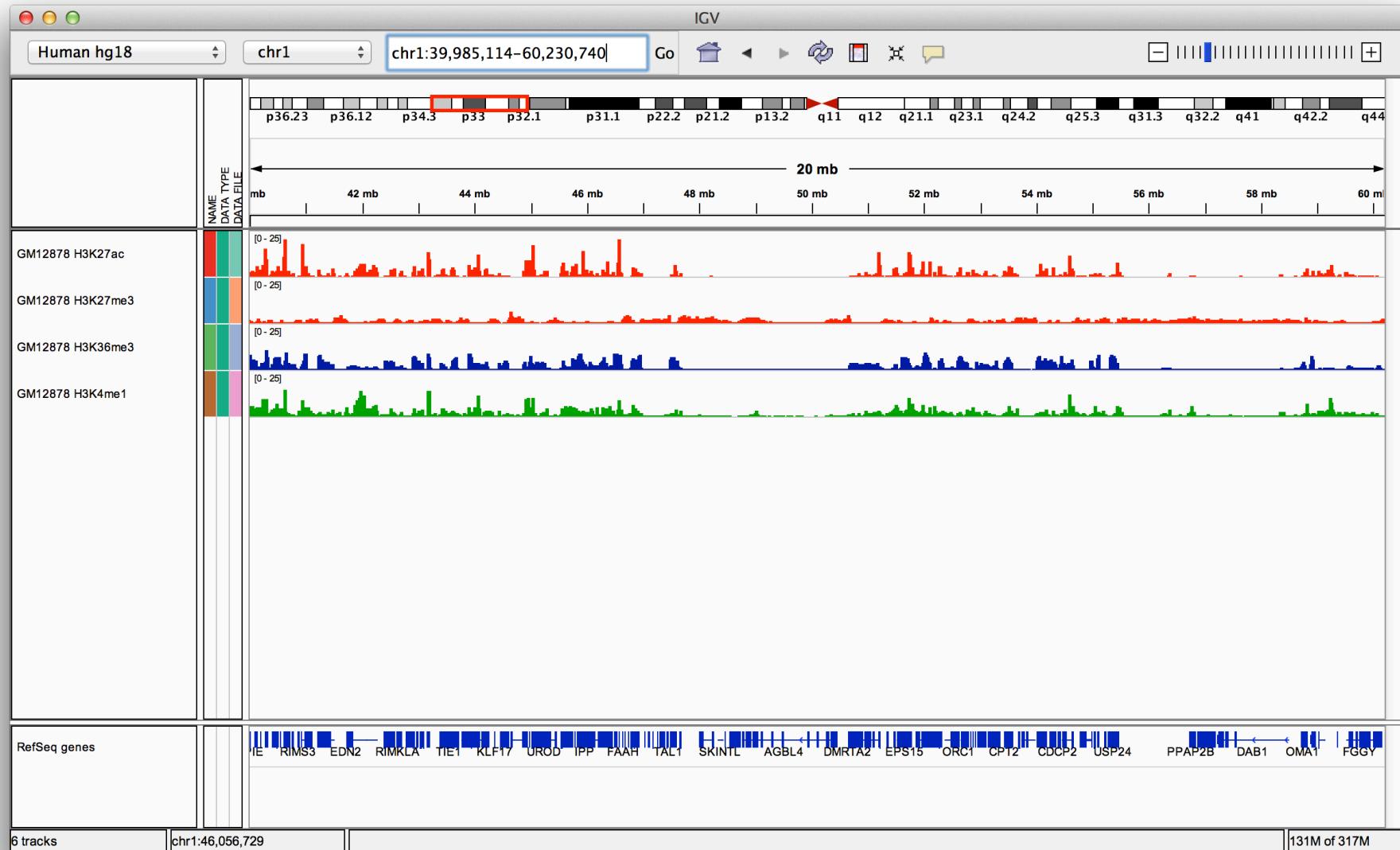
Navigate



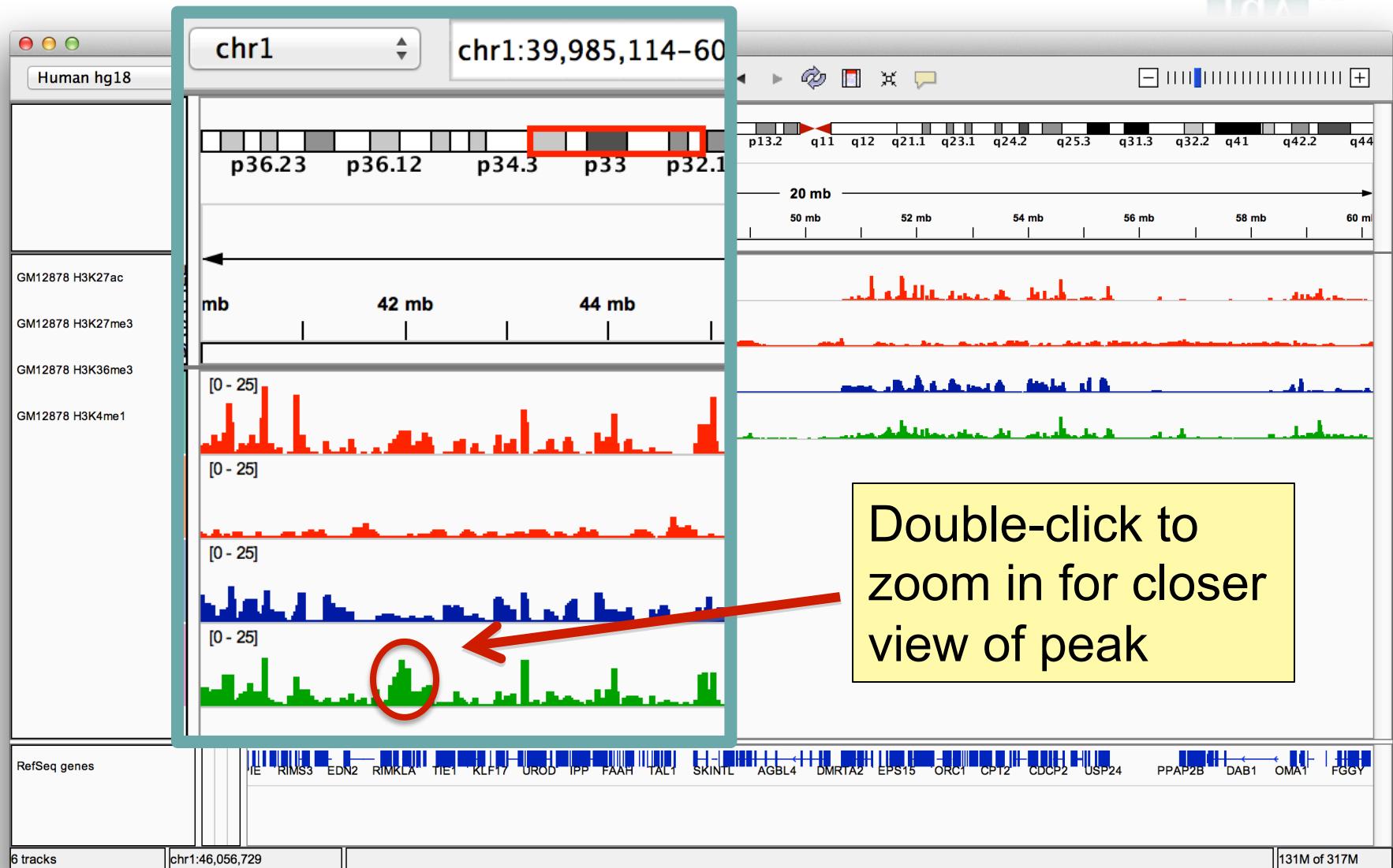
Navigate



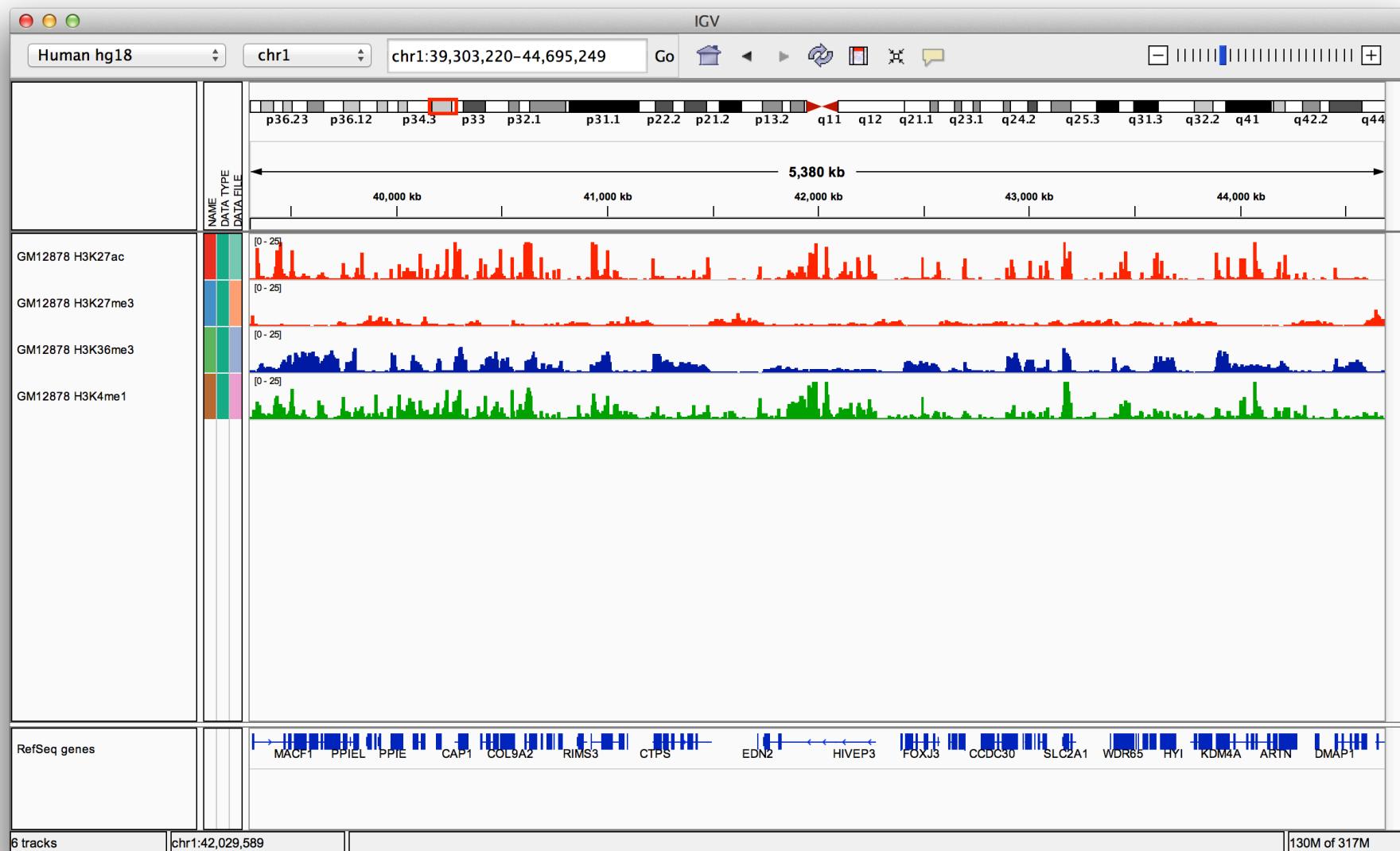
Navigate



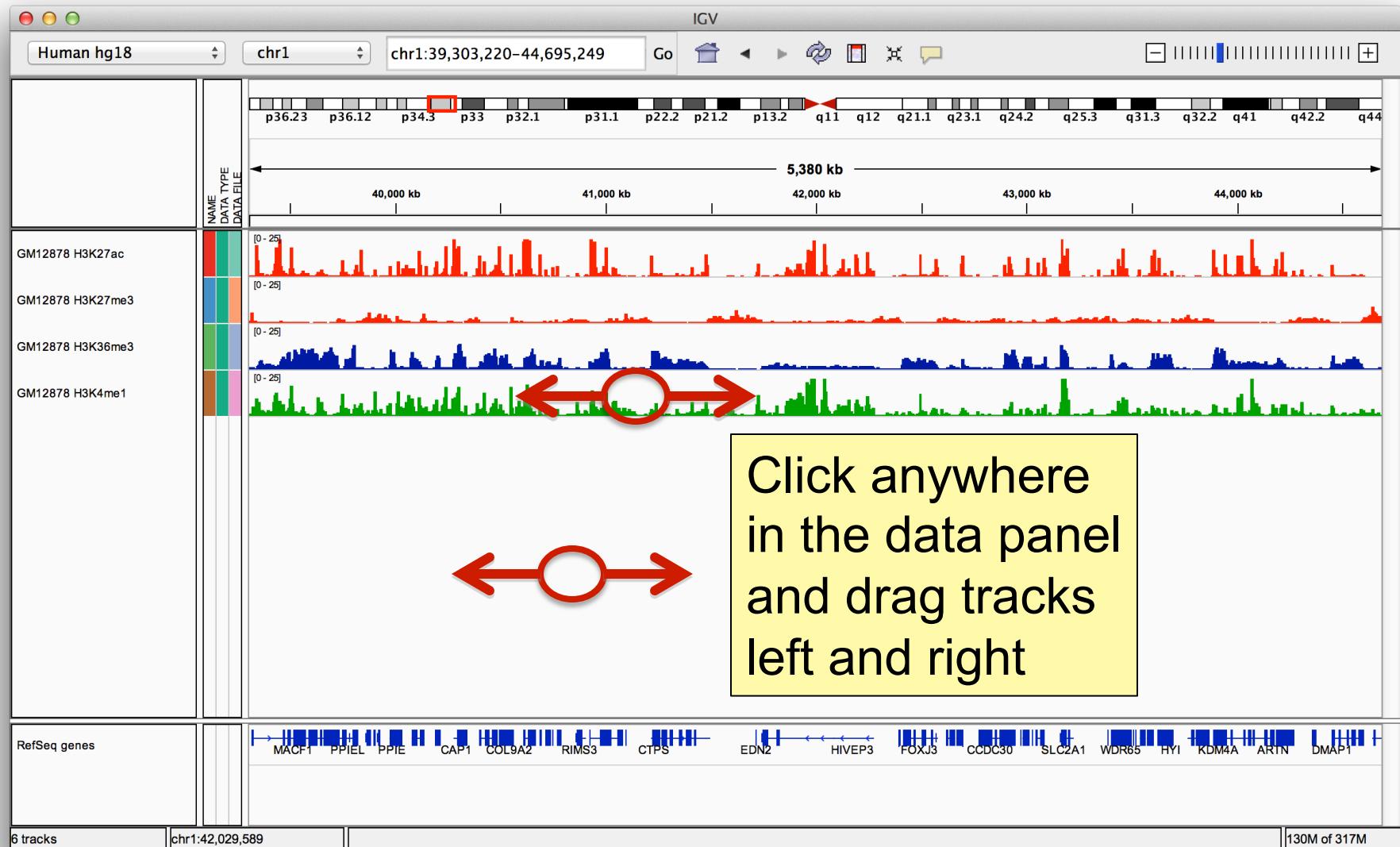
Navigate



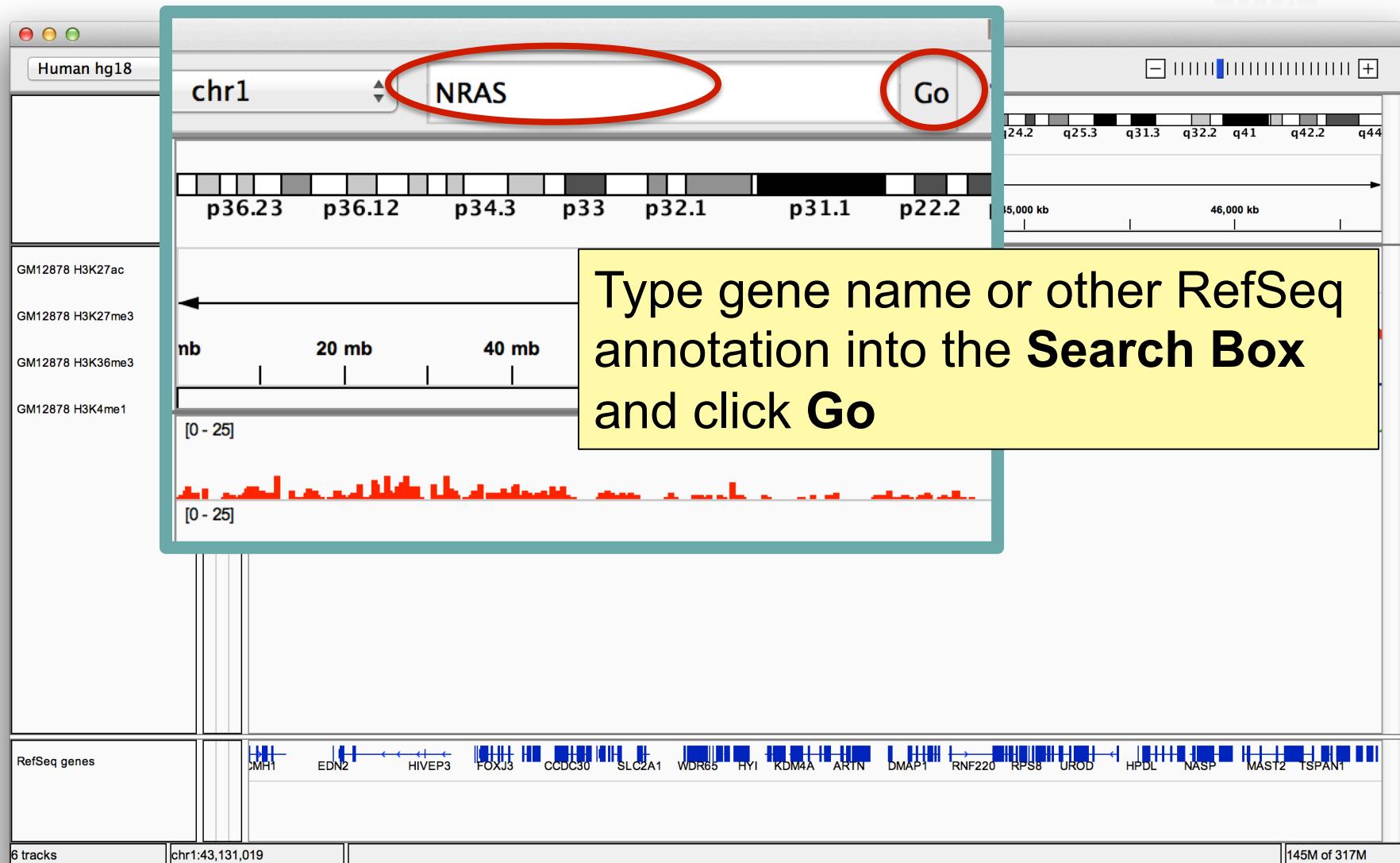
Navigate



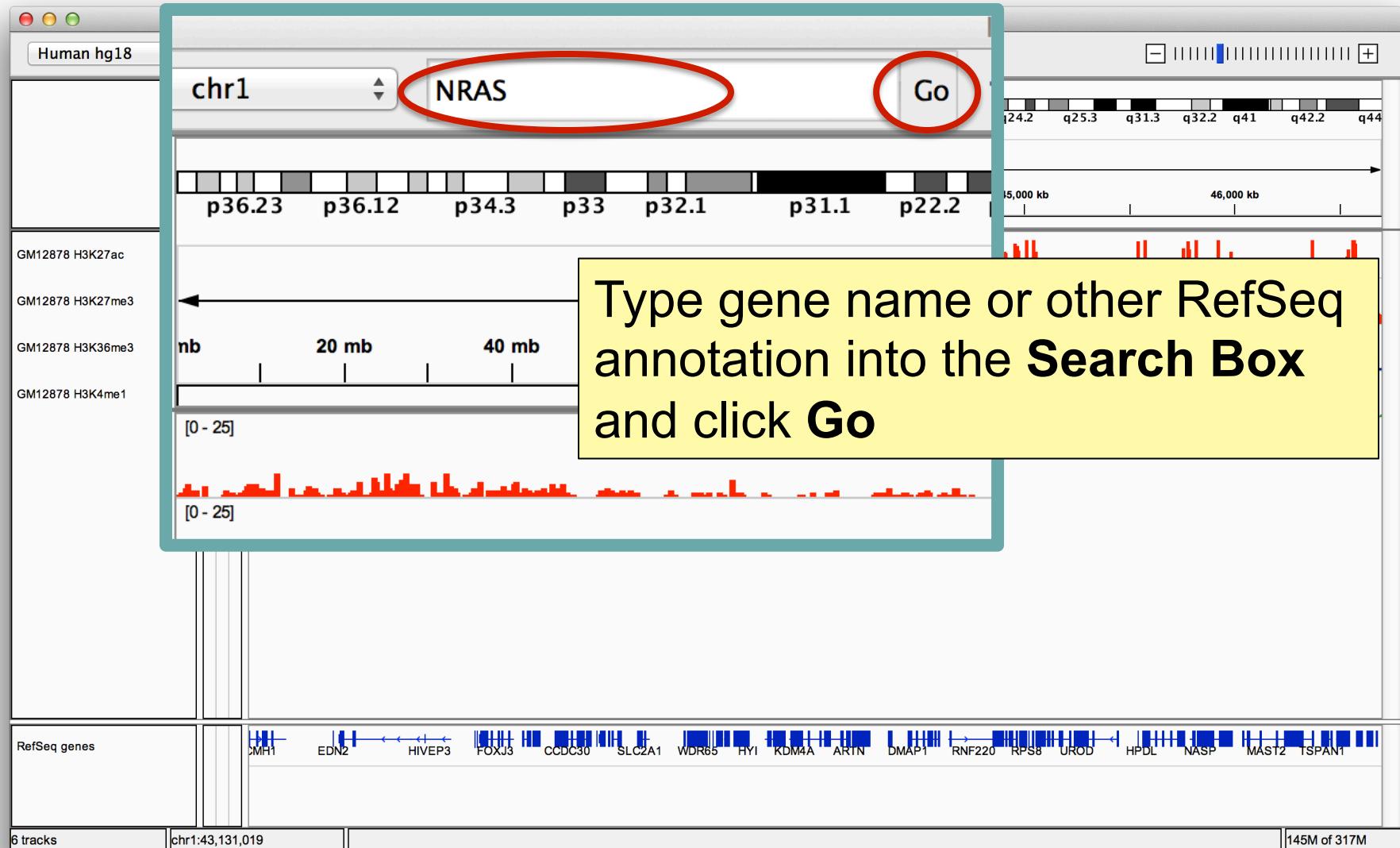
Navigate



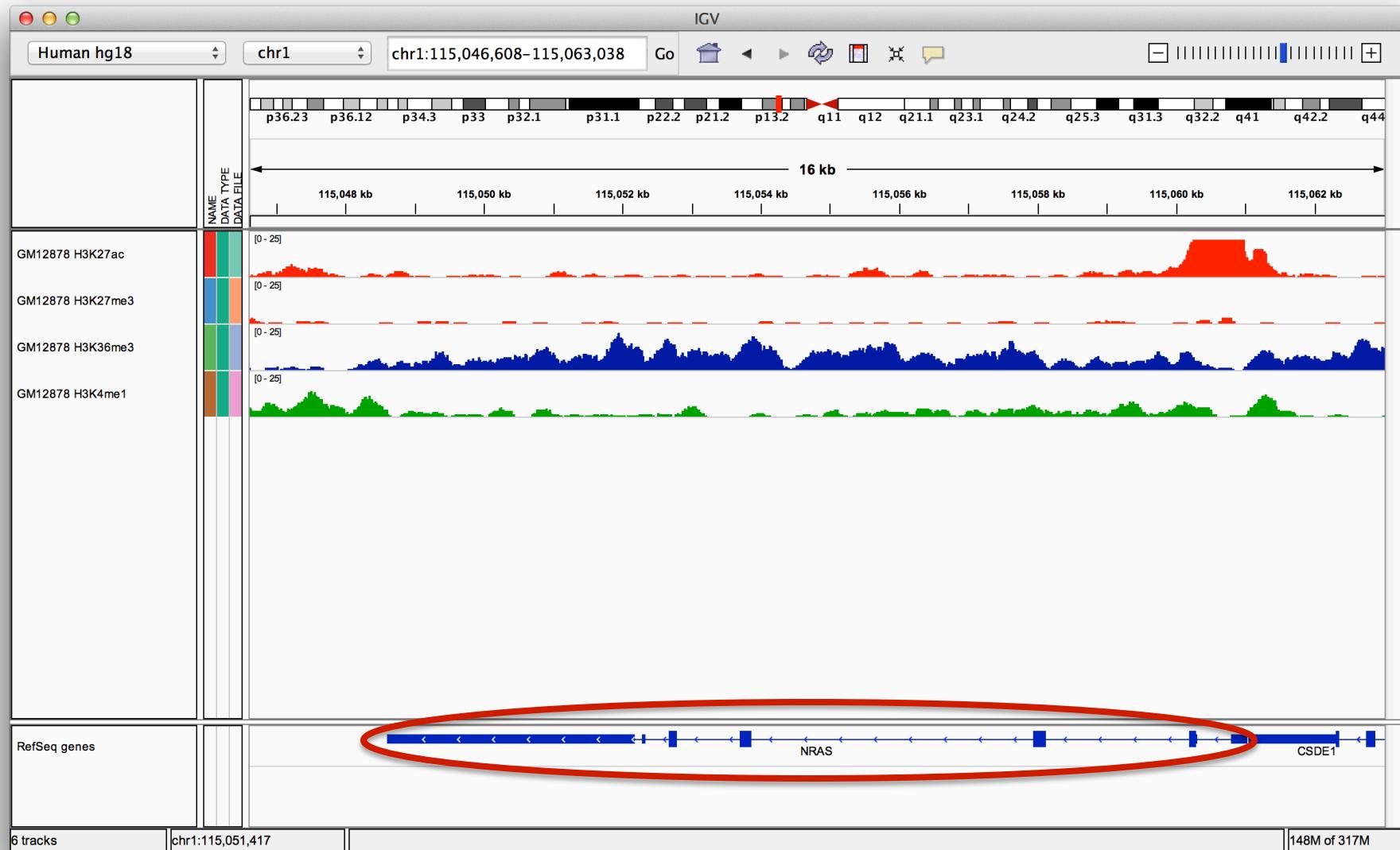
Navigate



Navigate



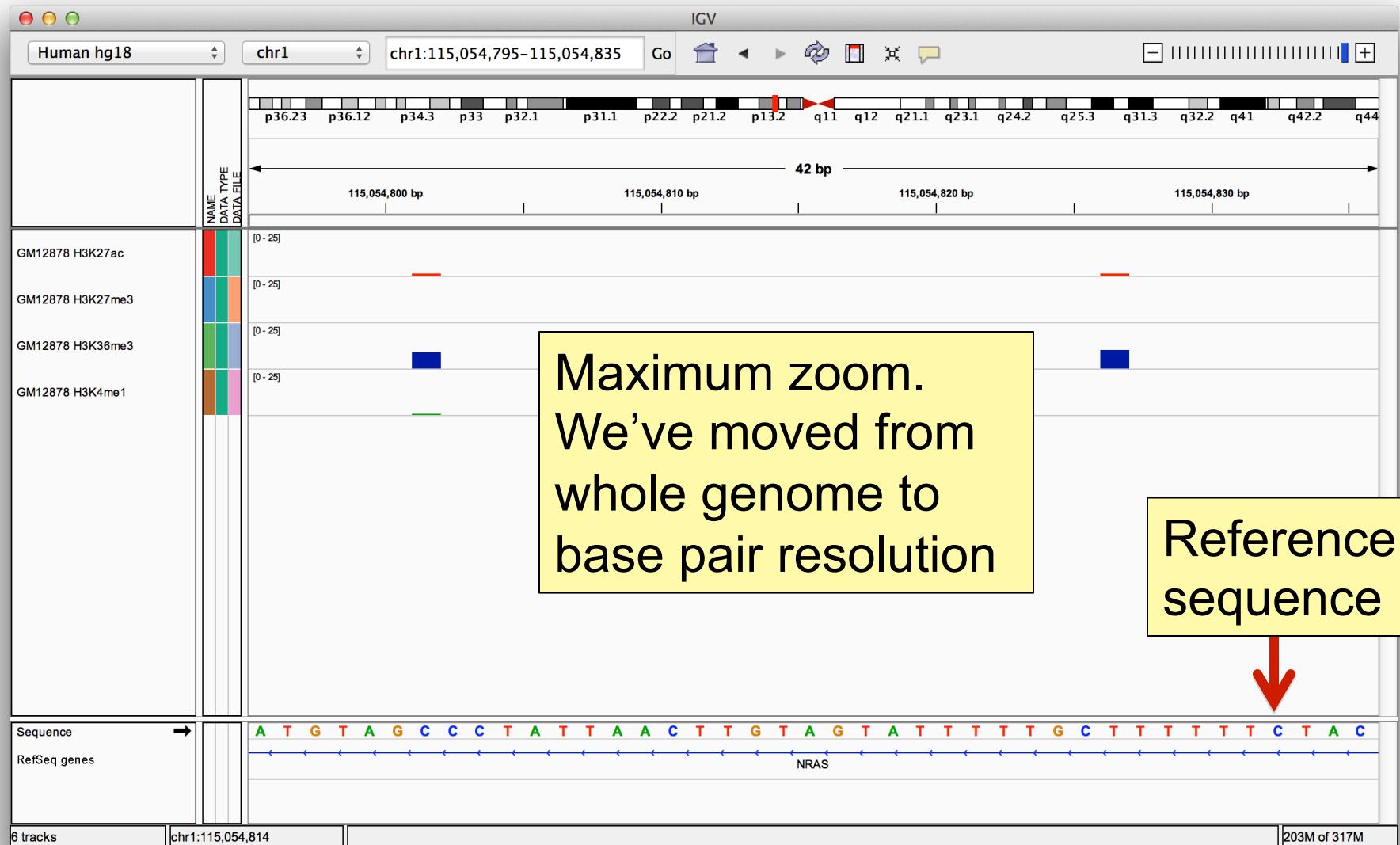
Navigate



Navigate



Navigate

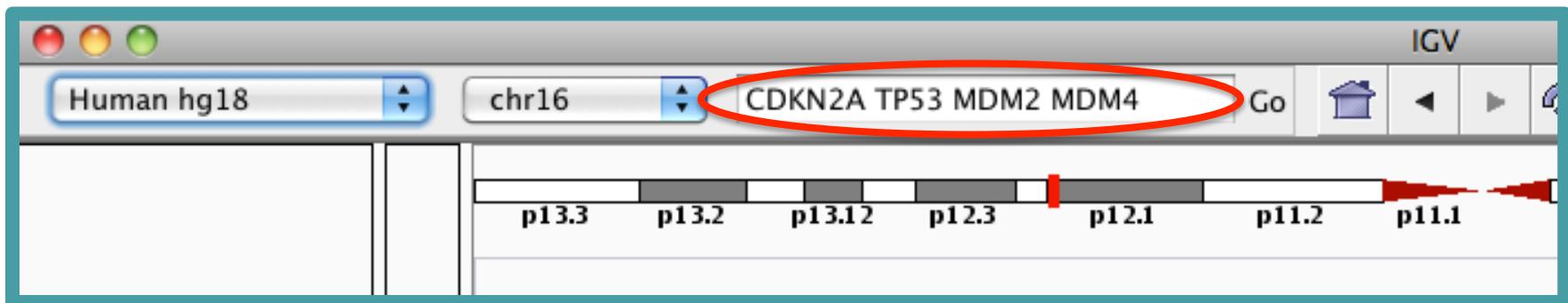


Viewing multiple regions



- **Search box**

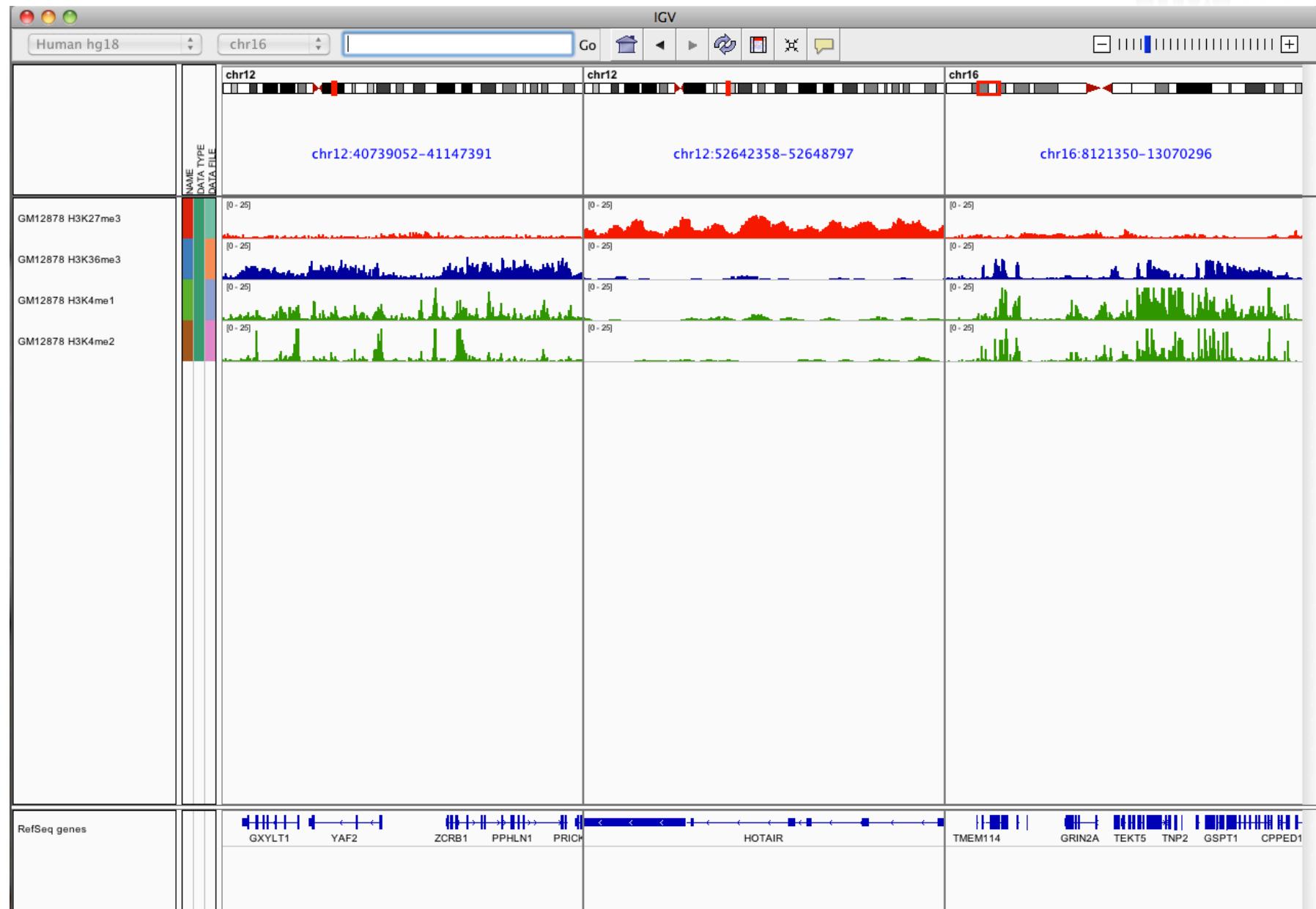
Enter multiple loci or features in the search box



- **Regions > Gene Lists...**

Select from a number of pre-defined gene lists, or
Create your own persistent list

Viewing multiple regions



Viewing multiple regions

To go back to the standard, single-region view:

- *double-click* on a region label – or –
- *right-click* and select “Switch to standard view”



File formats and track types

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

File formats and track types

- The **file format** defines the track type.
- The **track type** determines the display options
- IGV supports many different file formats.

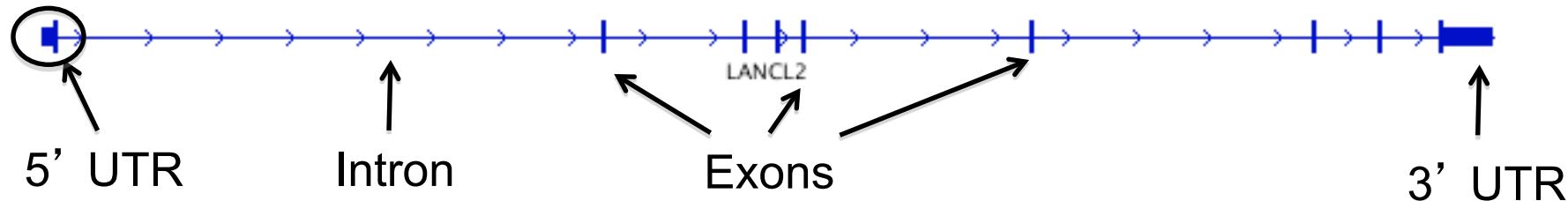
- [BAM](#)
- [BED](#)
- [BedGraph](#)
- [bigBed](#)
- [bigWig](#)
- [Birdsuite Files](#)
- [broadPeak](#)
- [CBS](#)
- [CN](#)
- [Cufflinks Files](#)
- [Custom File Formats](#)
- [Cytoband](#)
- [FASTA](#)
- [GCT](#)
- [genePred](#)
- [GFF](#)
- [GISTIC](#)
- [Goby](#)
- [GWAS](#)
- [IGV](#)
- [LOH](#)
- [MAF \(Multiple Alignment Format\)](#)
- [MAF \(Mutation Annotation Format\)](#)
- [Merged BAM File](#)
- [MUT](#)
- [narrowPeak](#)
- [PSL](#)
- [RES](#)
- [SAM](#)
- [Sample Information](#)
- [SEG](#)
- [SNP](#)
- [TAB](#)
- [TDF](#)
- [Track Line](#)
- [Type Line](#)
- [VCF](#)
- [WIG](#)

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

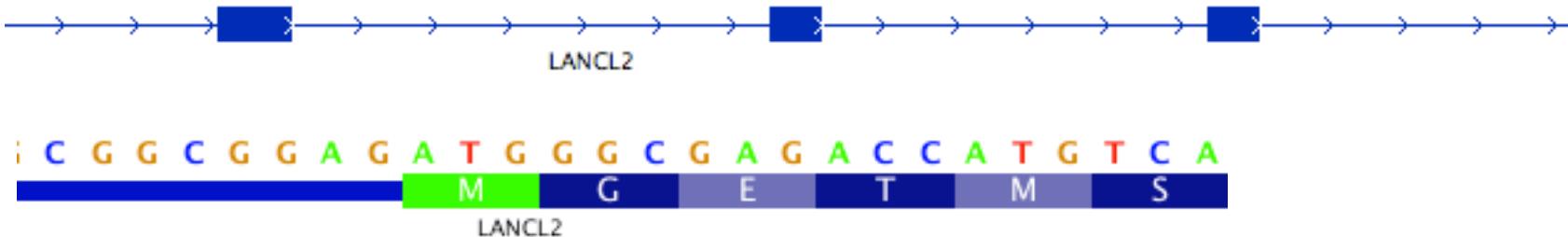
Genome annotation track



UCSC style gene representation



Zoomed in views



Zoomed out views

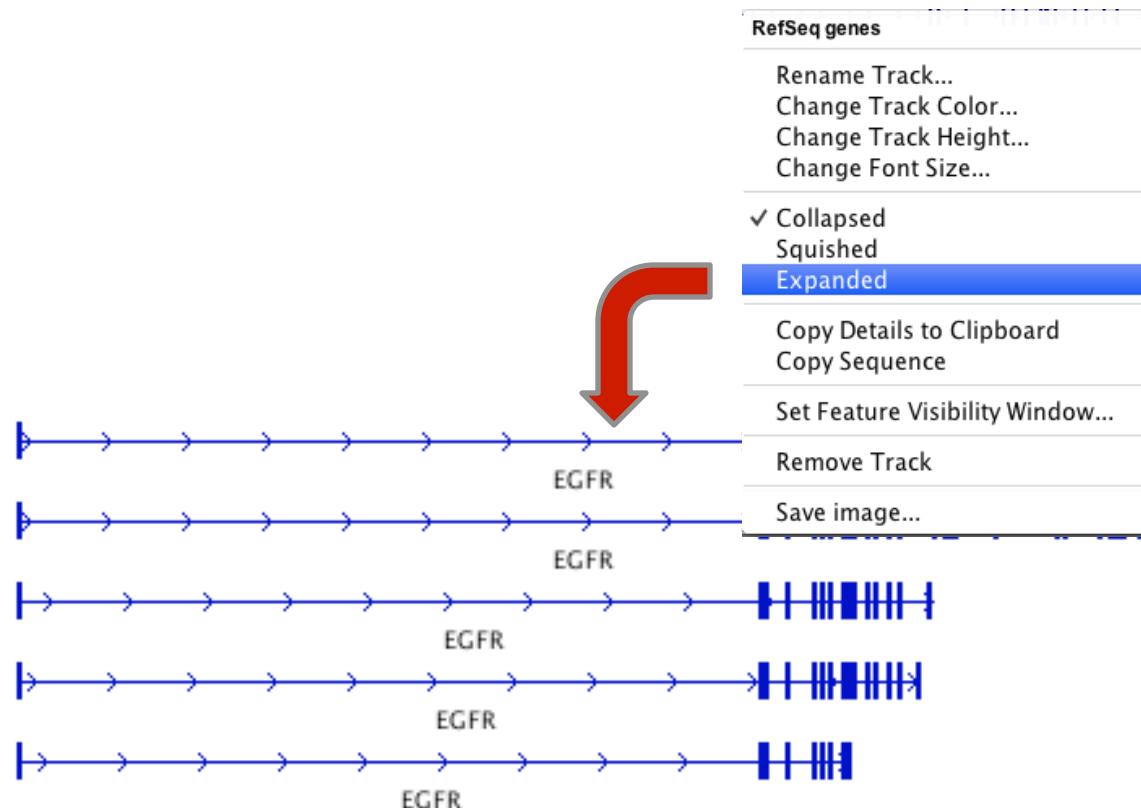


Annotation display mode

1. Features are drawn in a single row, by default

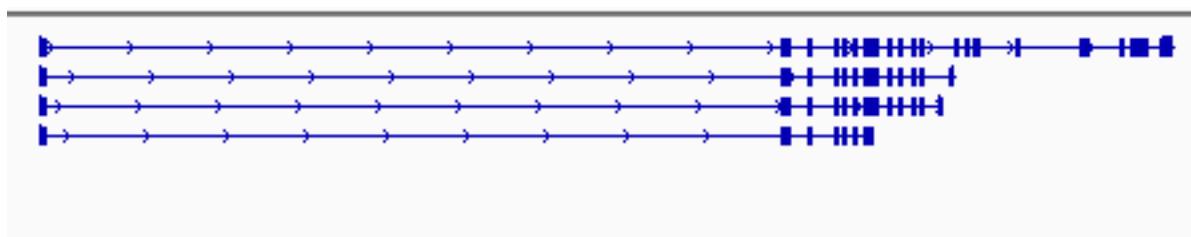
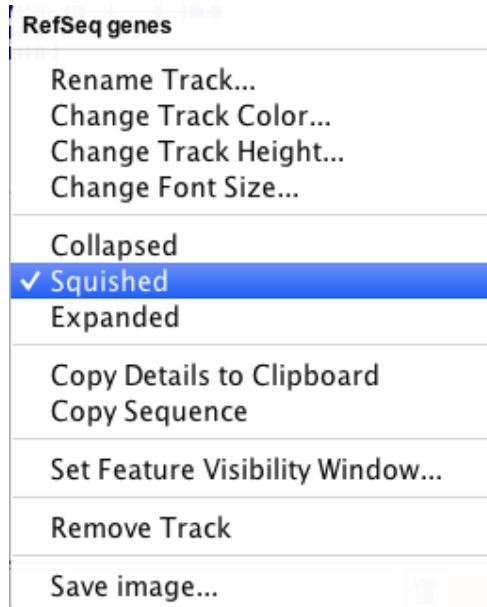


2. Expand the track using the popup menu



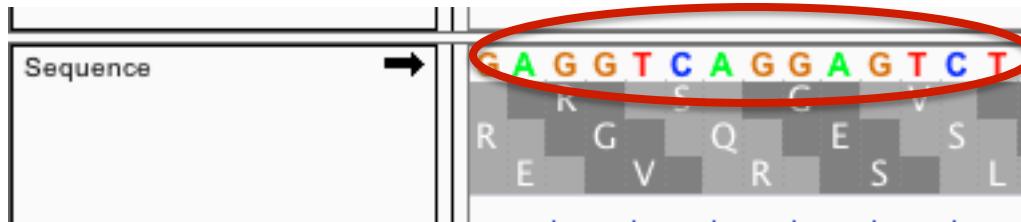
Annotation display mode

3. For a compact view of all variants use “Squished”



Reference sequence

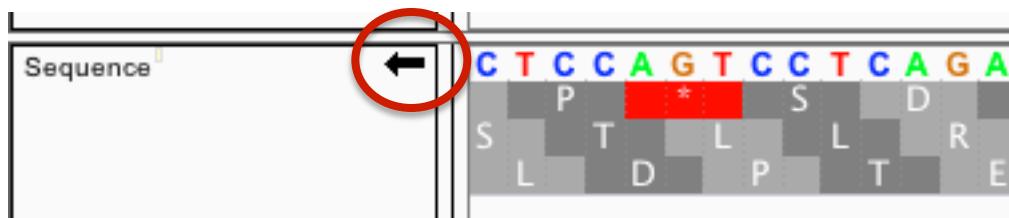
Click anywhere on the sequence to see a 3 frame translation.



By default the sequence for the forward strand is shown.



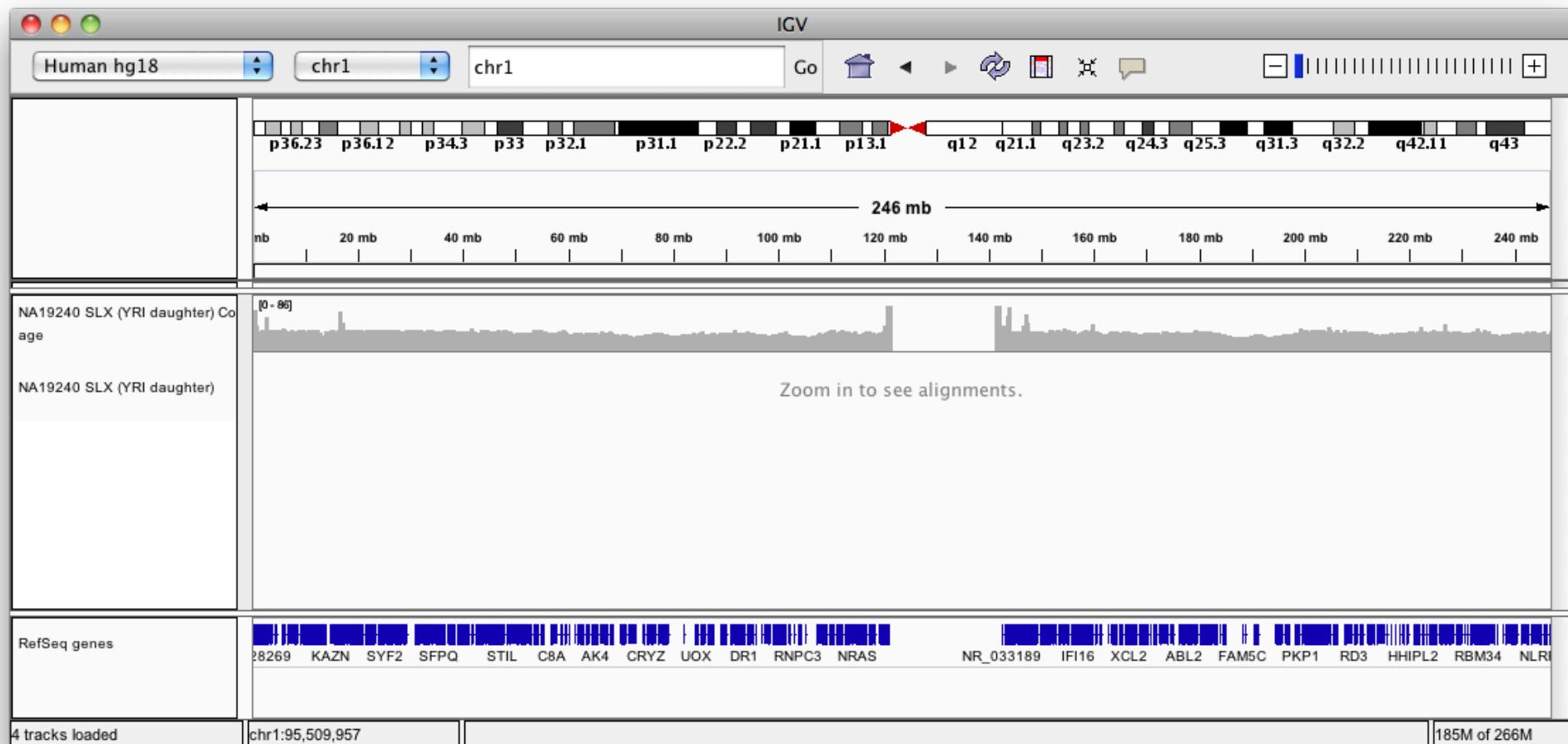
Click the arrow on the left to reverse the strand.



Viewing NGS Data

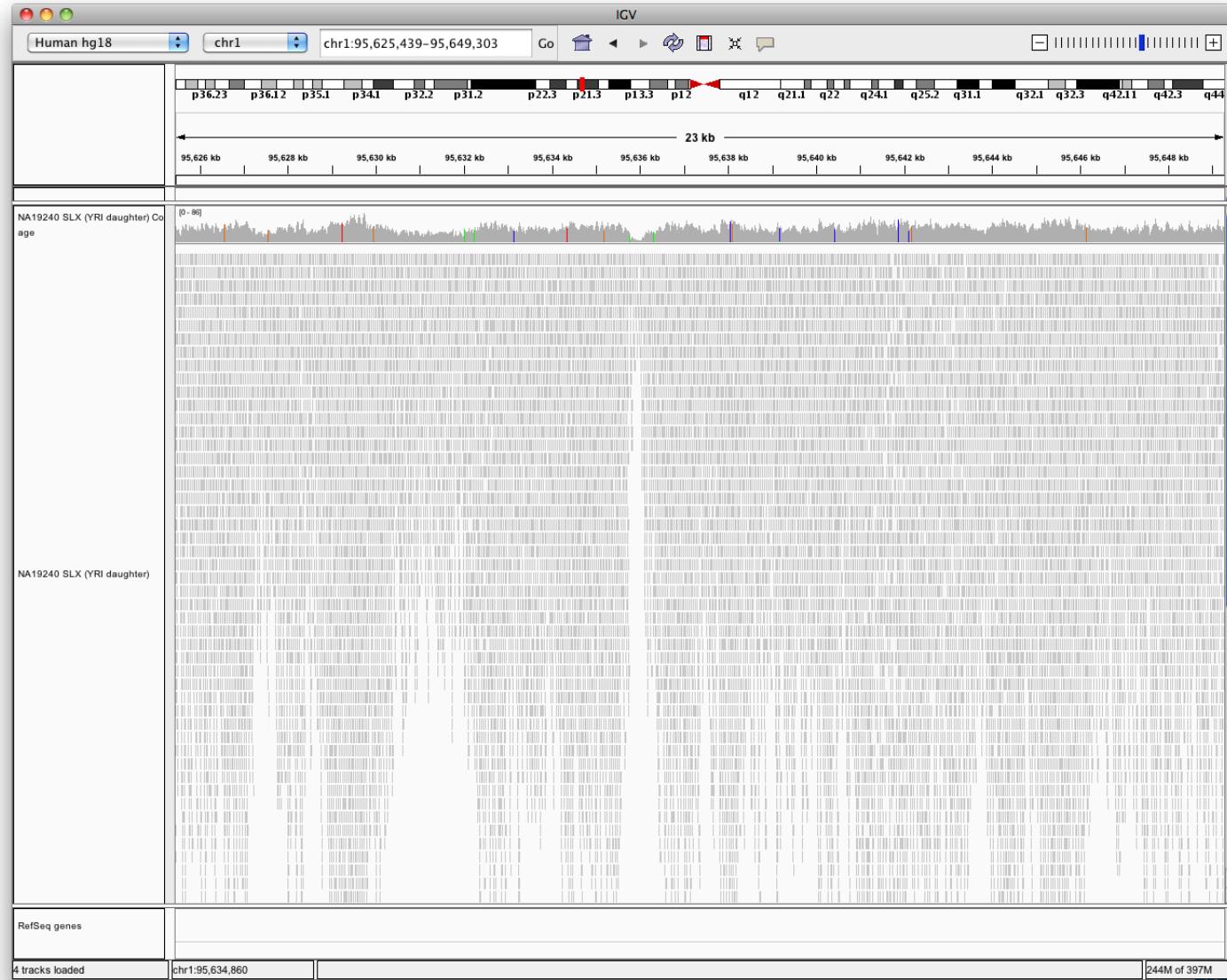
Viewing alignments

Whole chromosome view



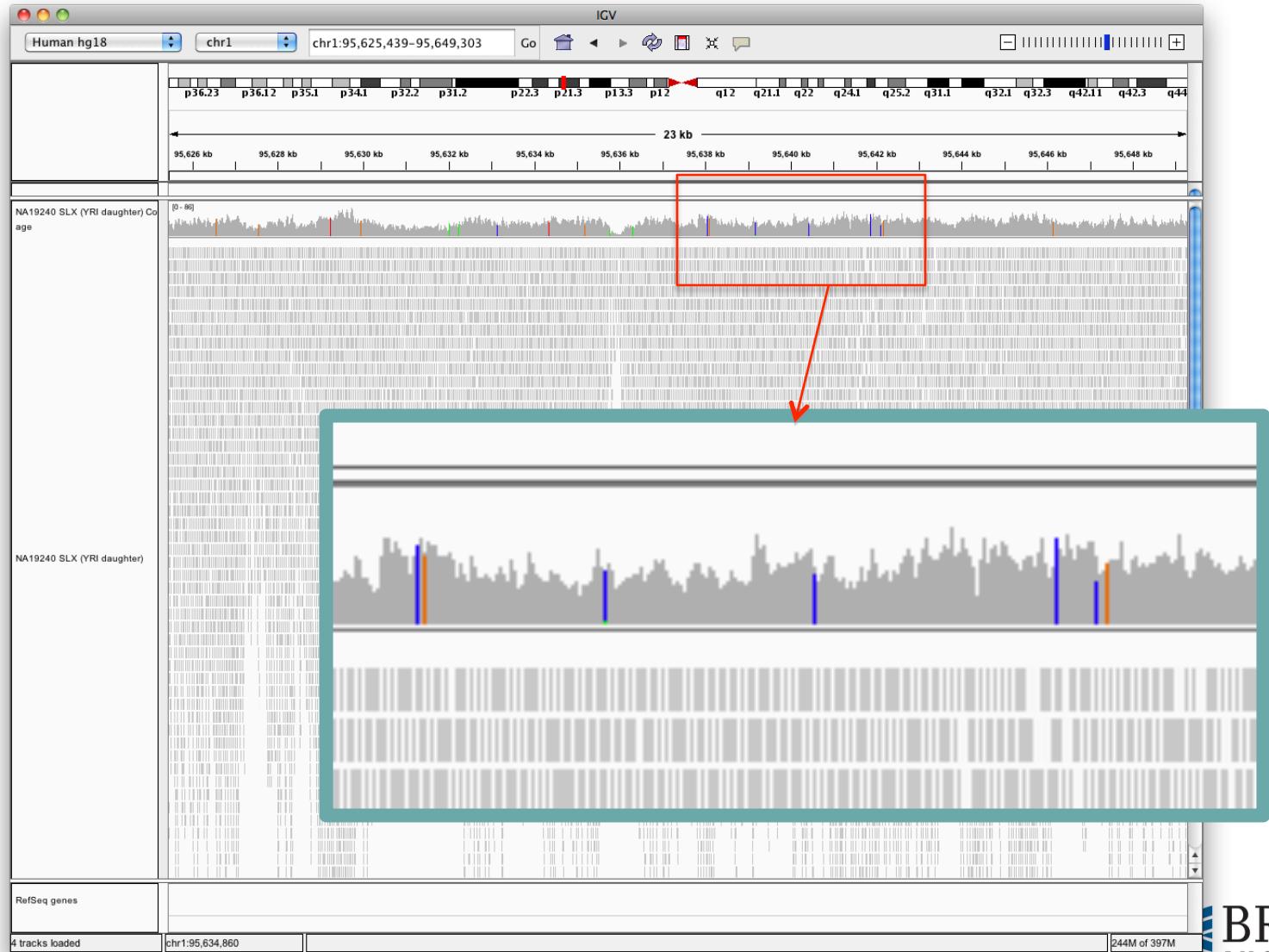
Viewing alignments

Zoom in to view alignments



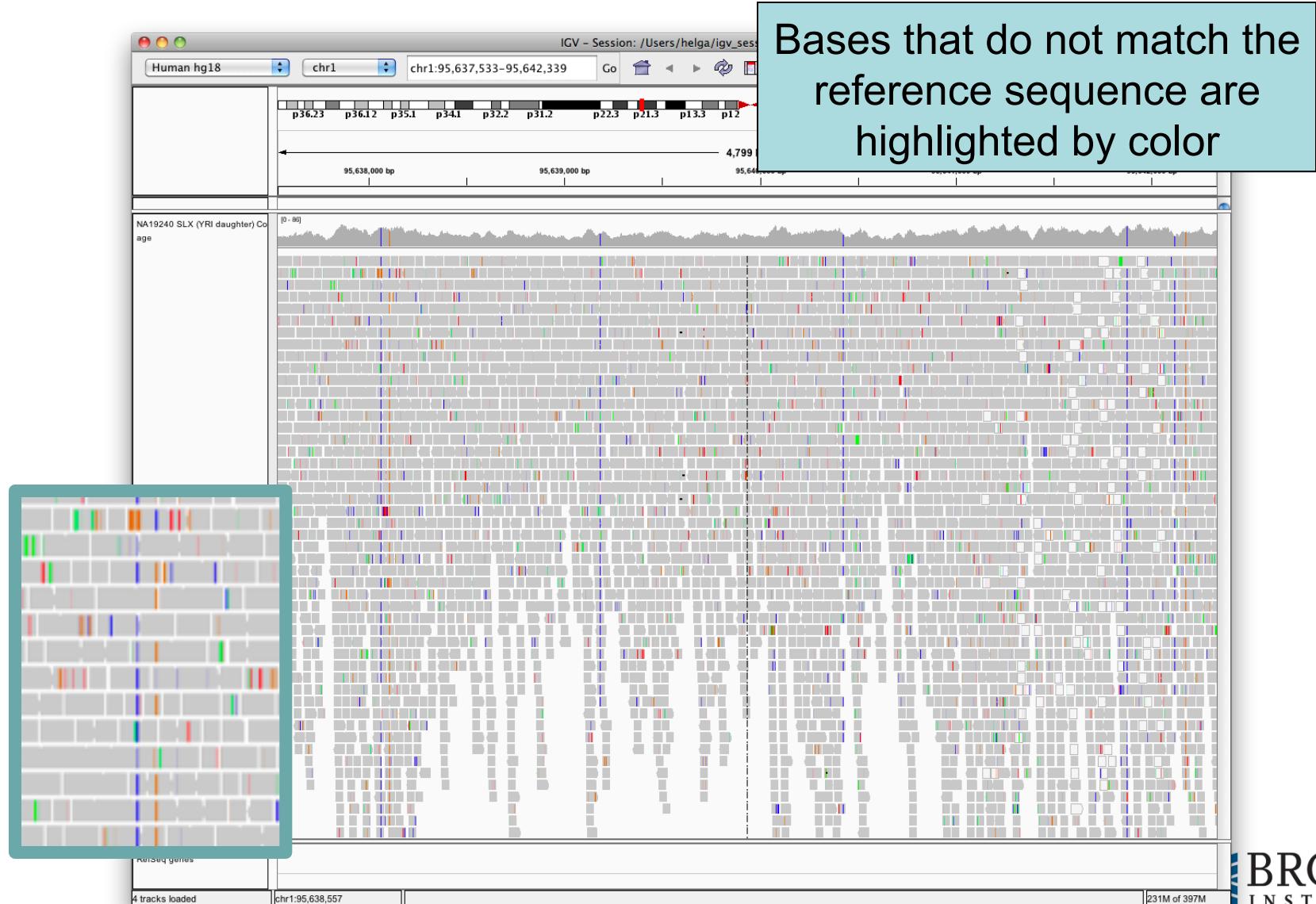
Viewing alignments

Coverage track now has more detail



Viewing alignments

Zoom in to see more detail



Viewing alignments

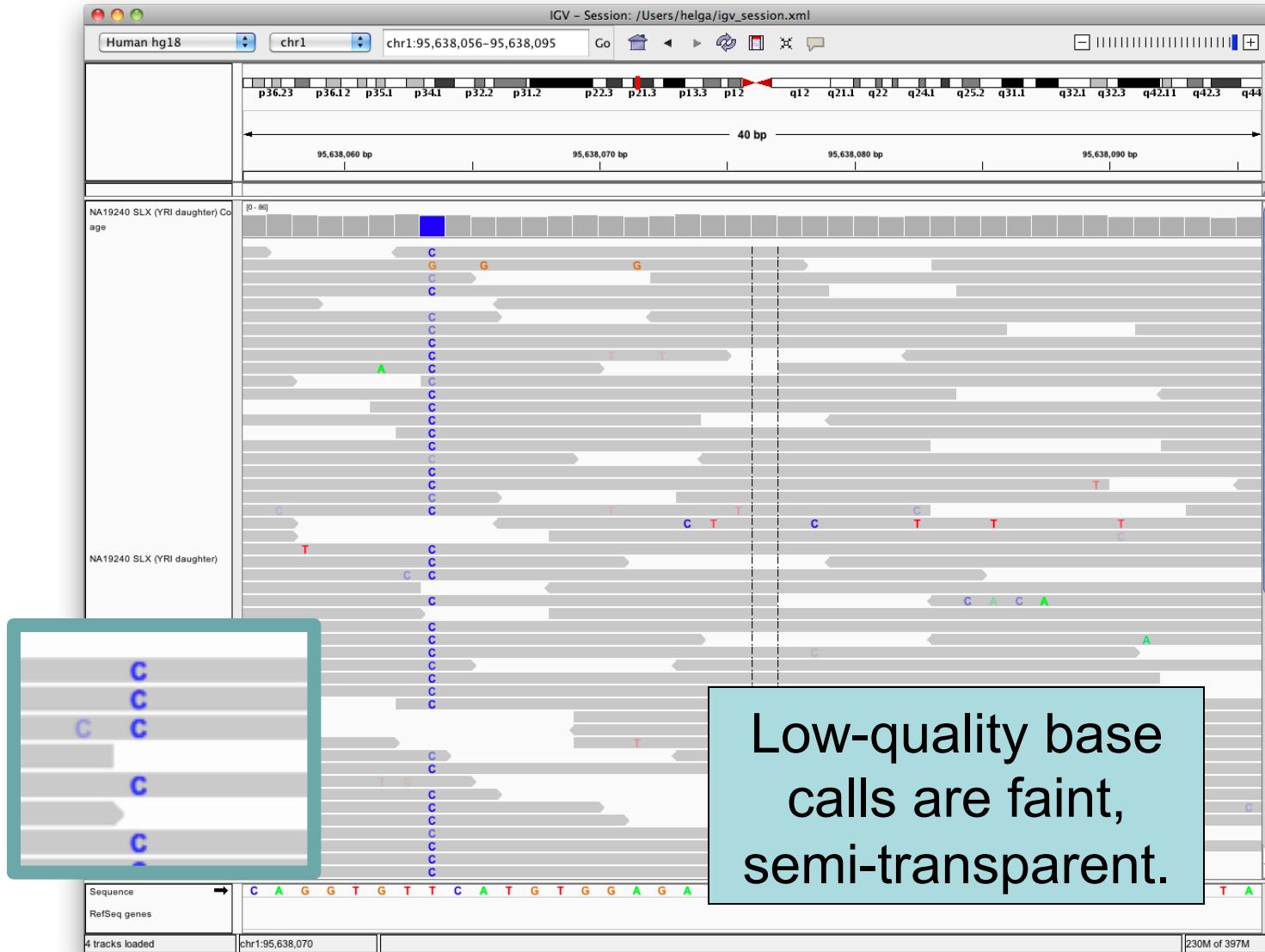
Zoom in to see more detail



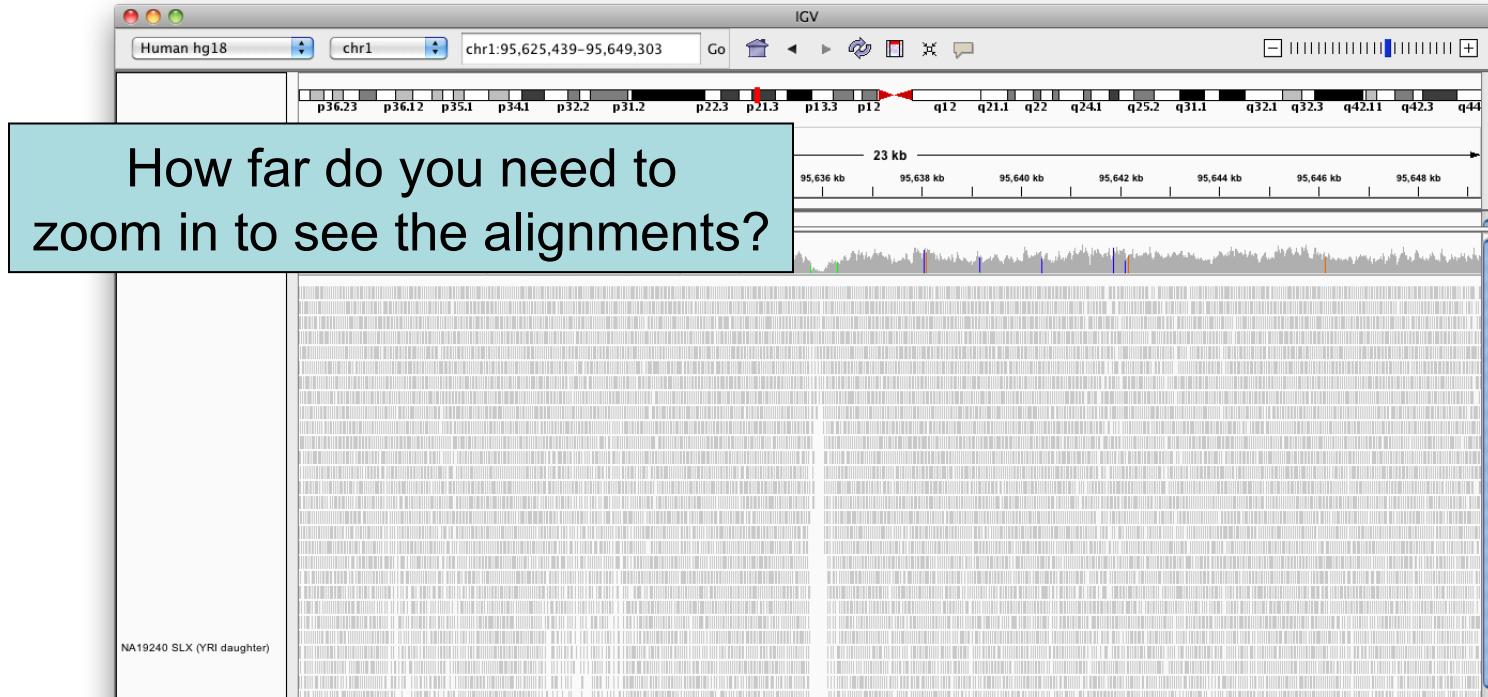
Viewing alignments



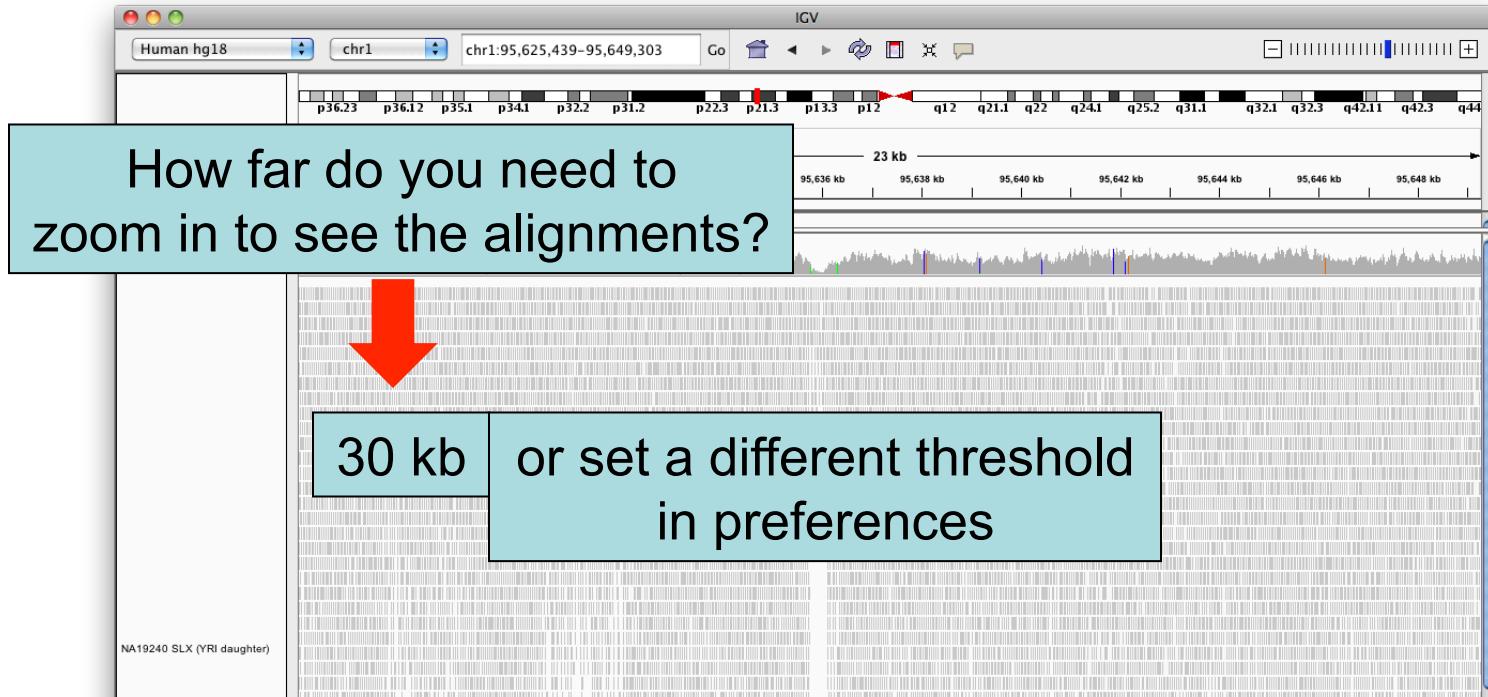
Zoom in to see more detail



Viewing alignments

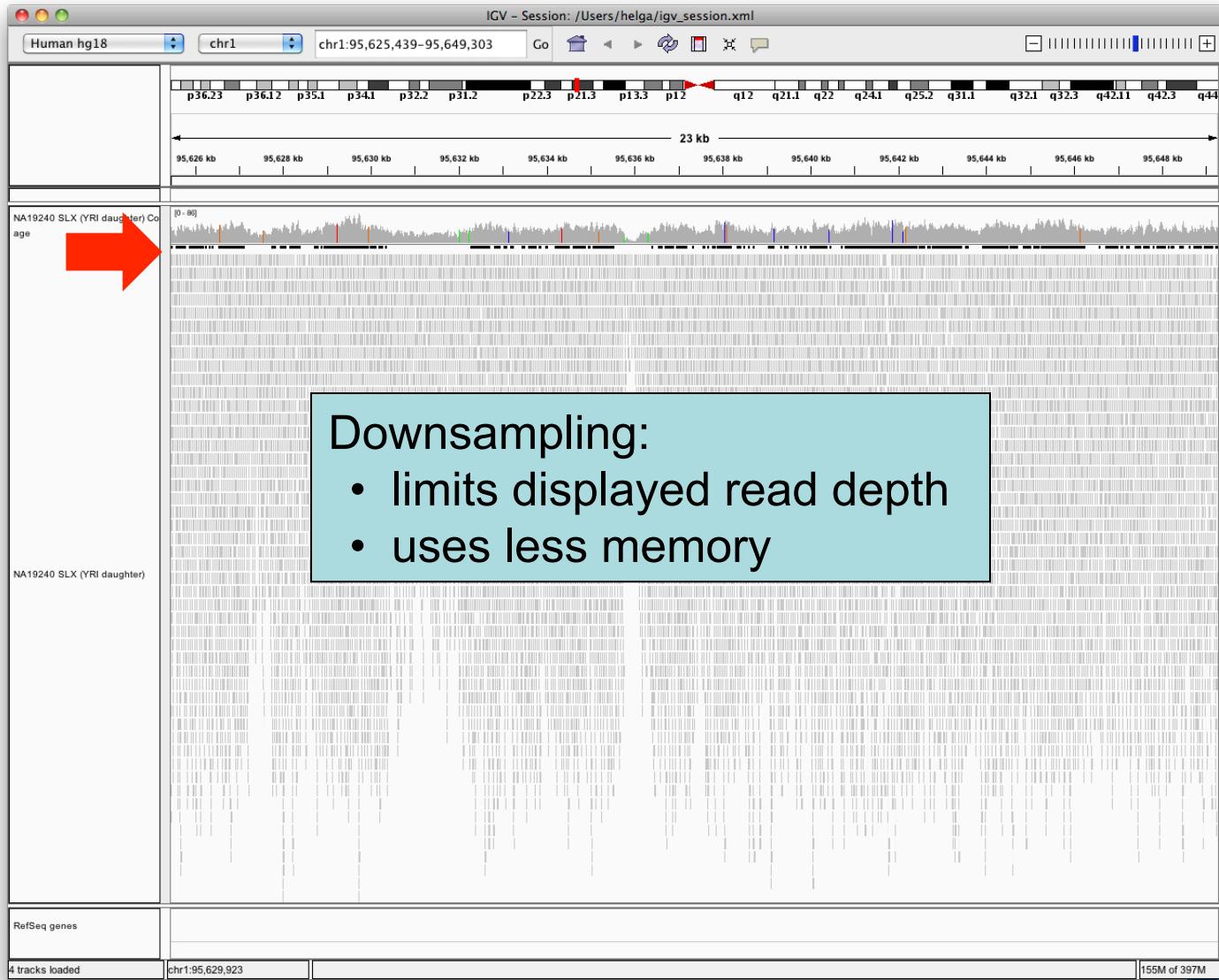


Viewing alignments



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

Viewing alignments



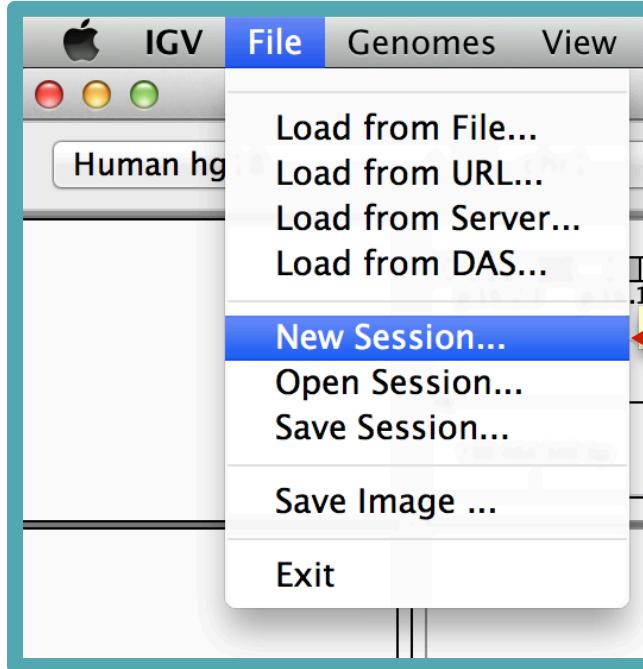
Viewing SNPs



Hands-on exercise

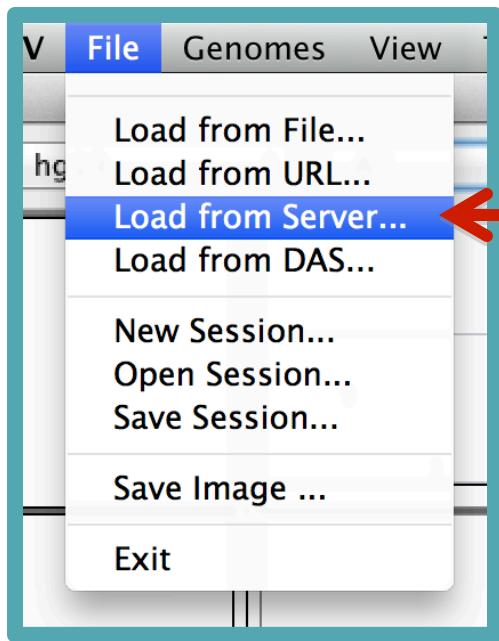
- Load alignments from whole genome sequencing
- View sites where SNPs were called
- Sort and color to highlight patterns

Viewing SNPs



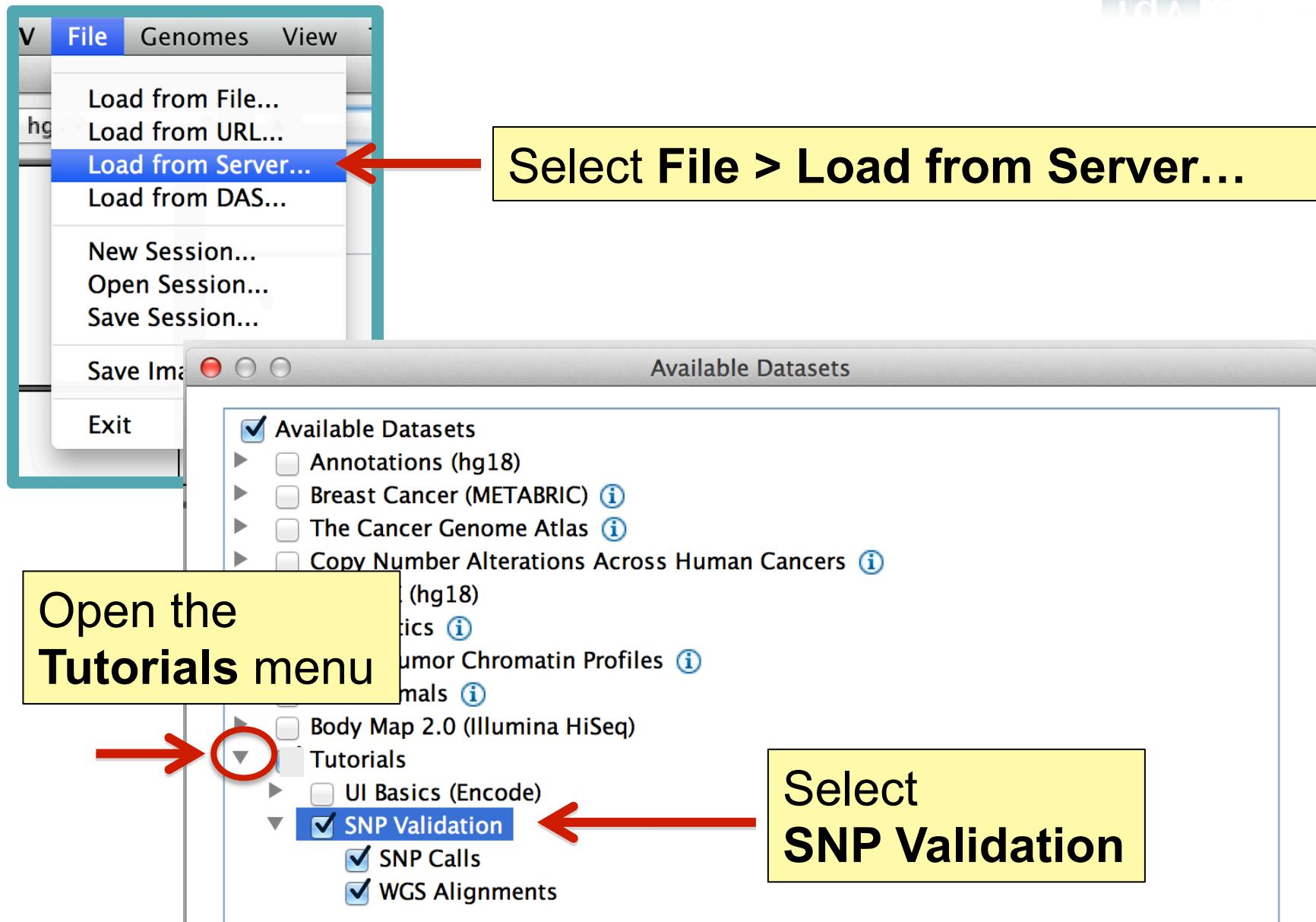
Before we start:
Select File > New Session
to clear IGV window

Viewing SNPs

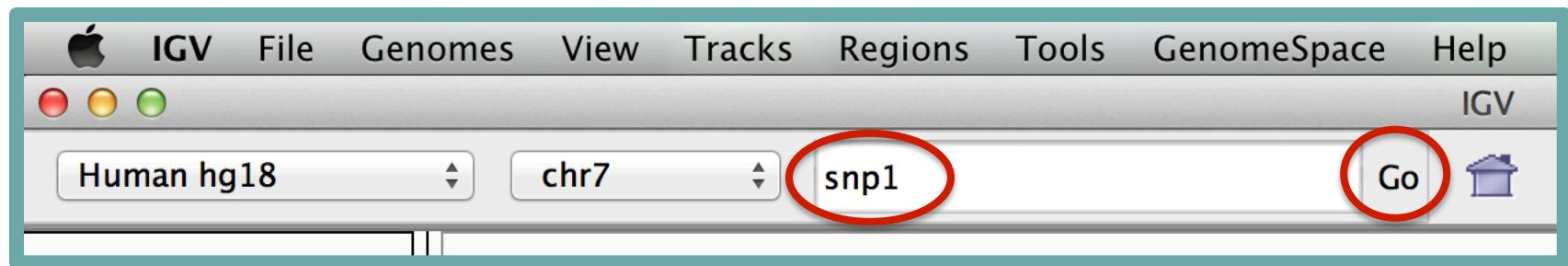


Select File > Load from Server...

Viewing SNPs

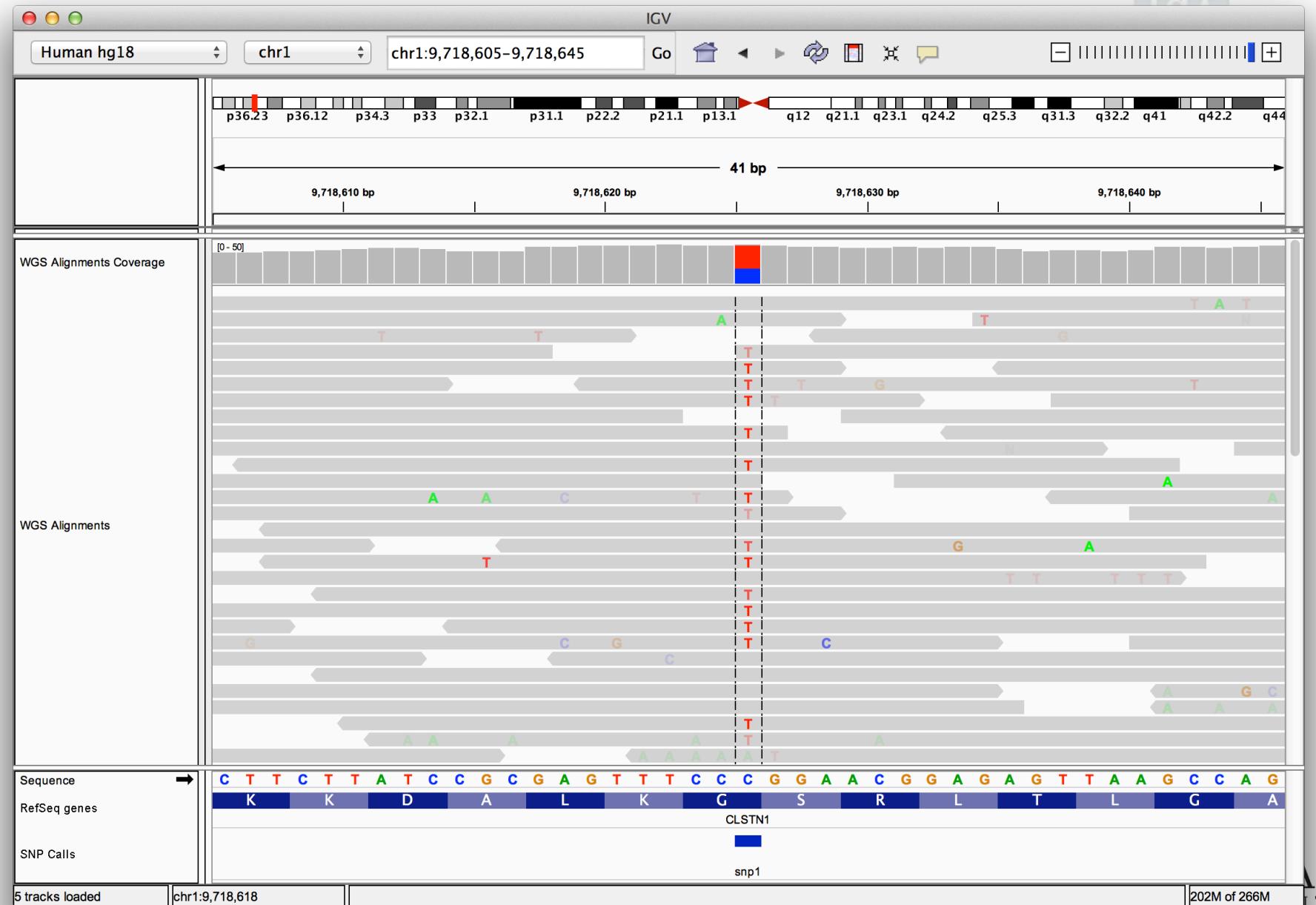


Viewing SNPs

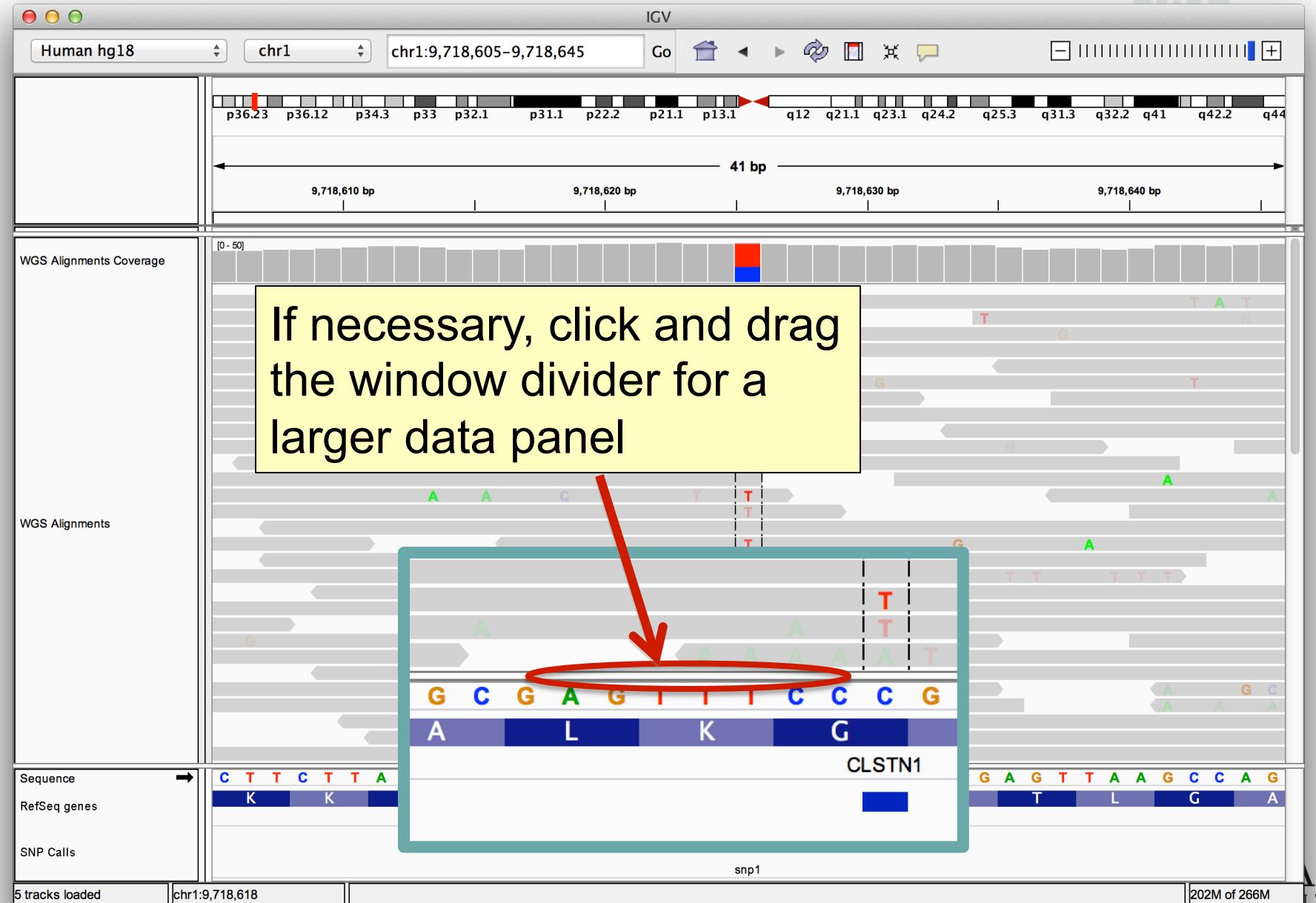


Type “snp1” in the **Search Box**
and click **Go**

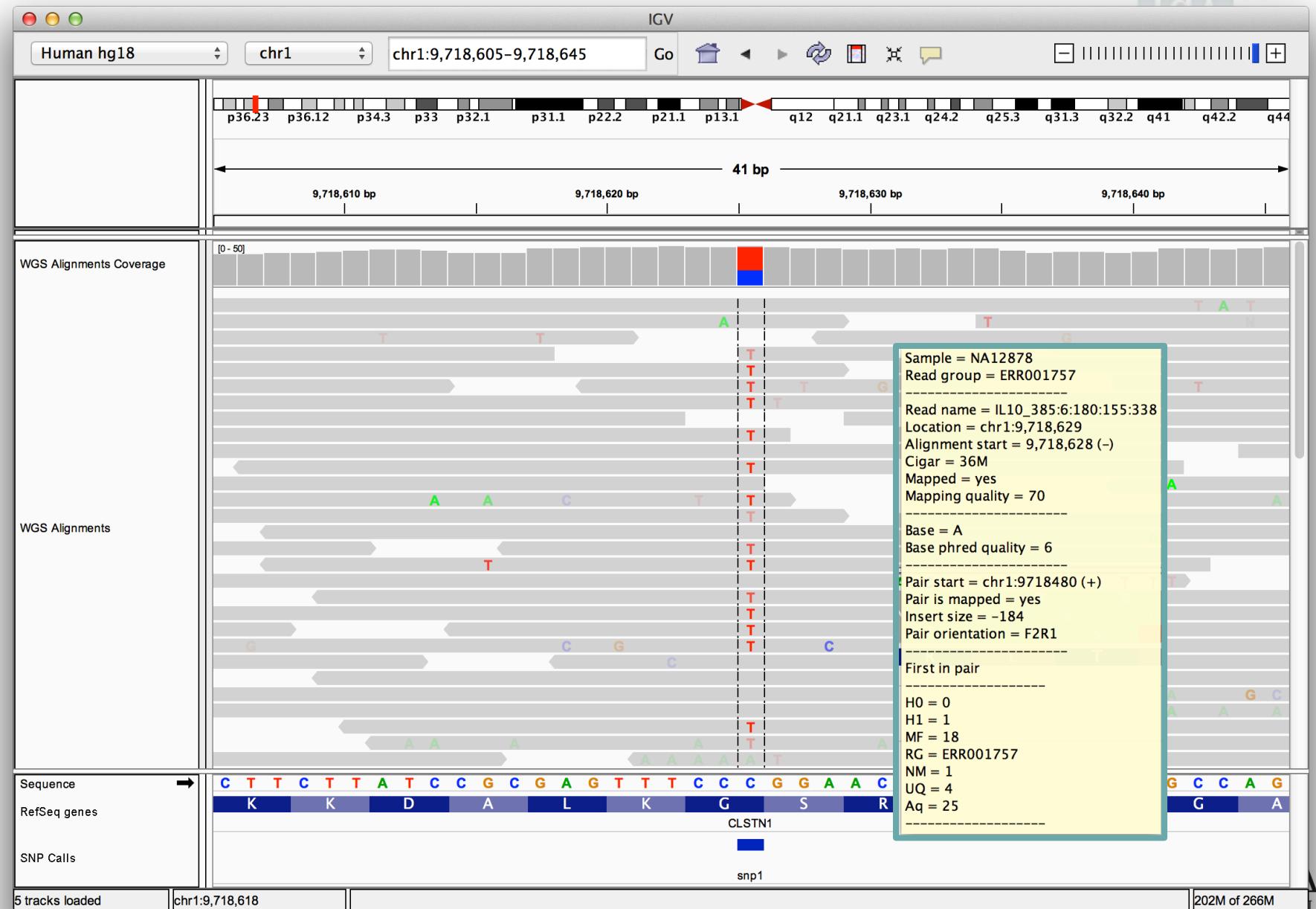
Viewing SNPs



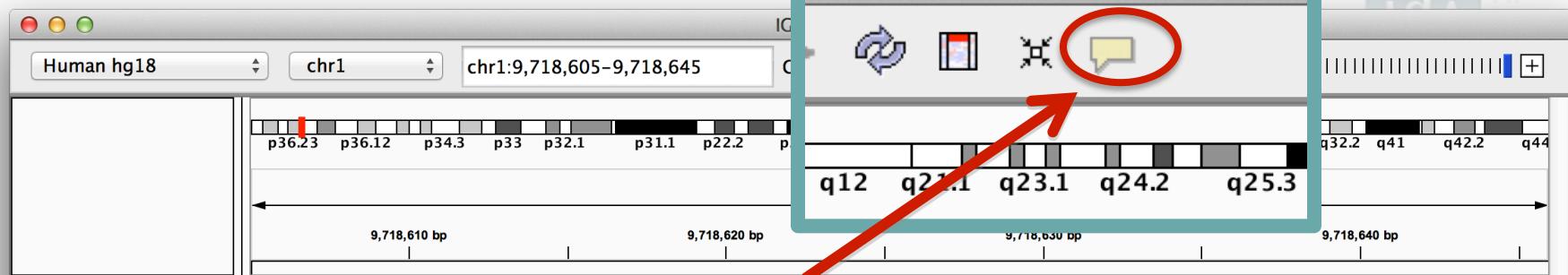
Viewing SNPs



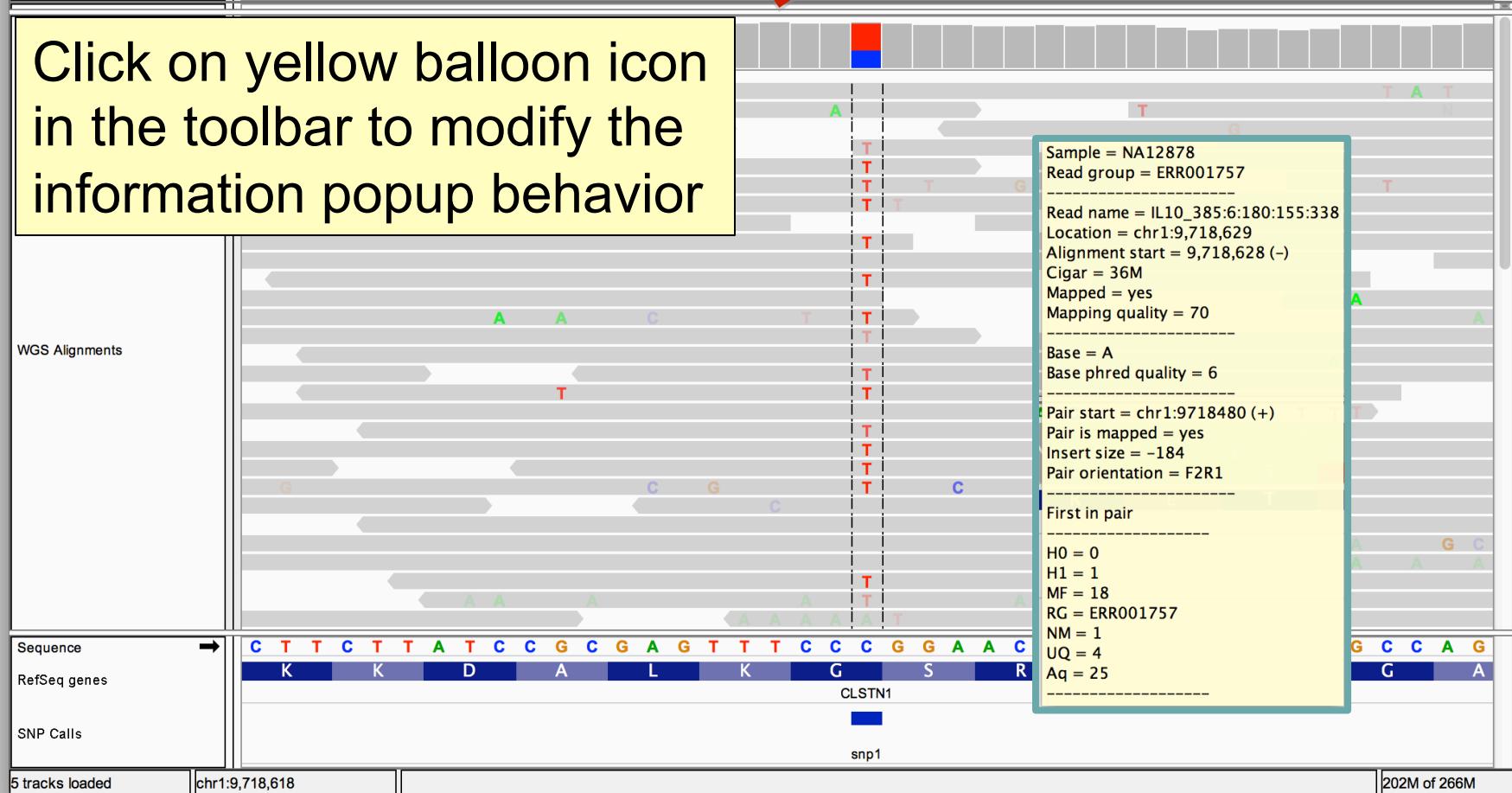
Viewing SNPs



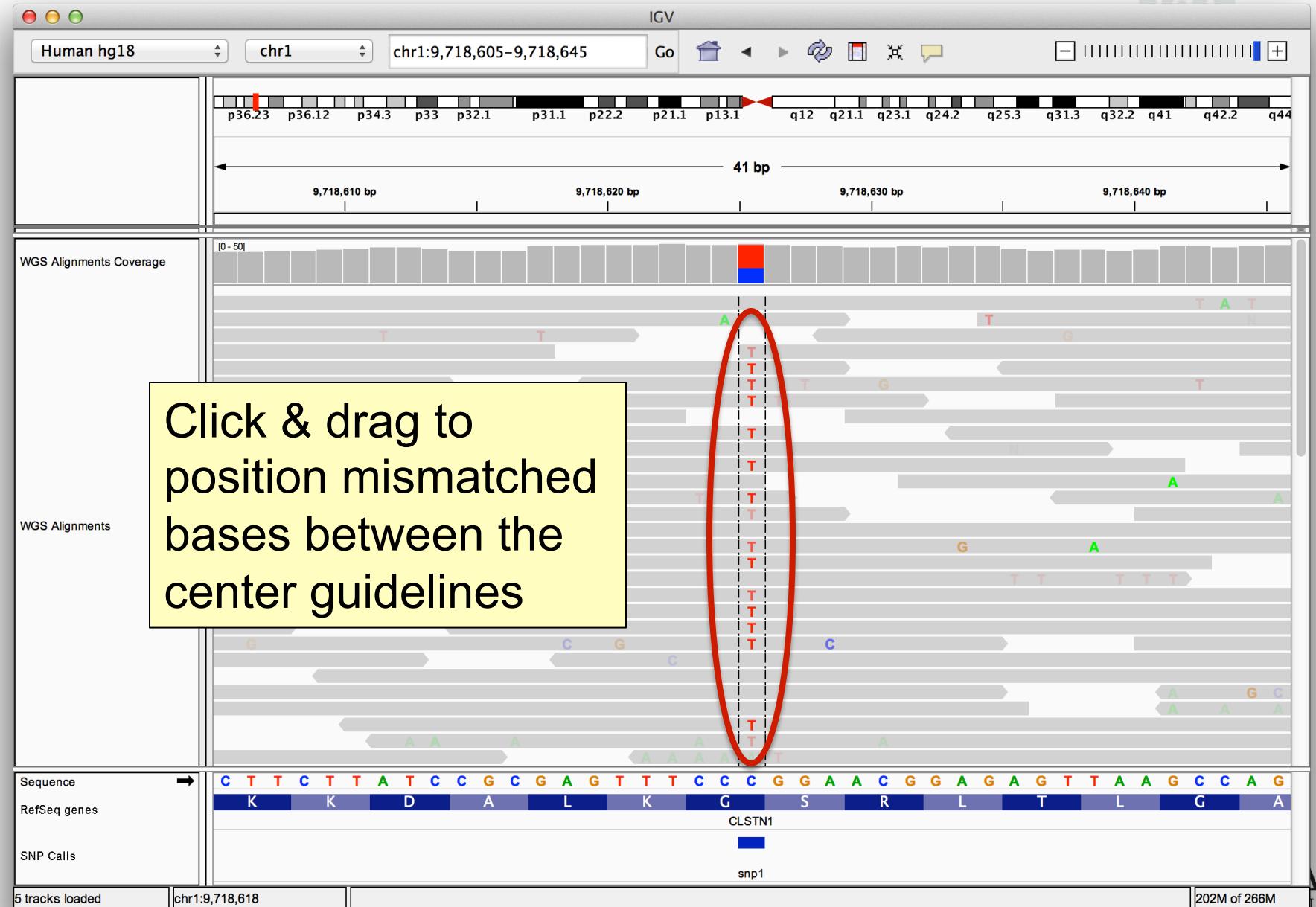
Viewing SNPs



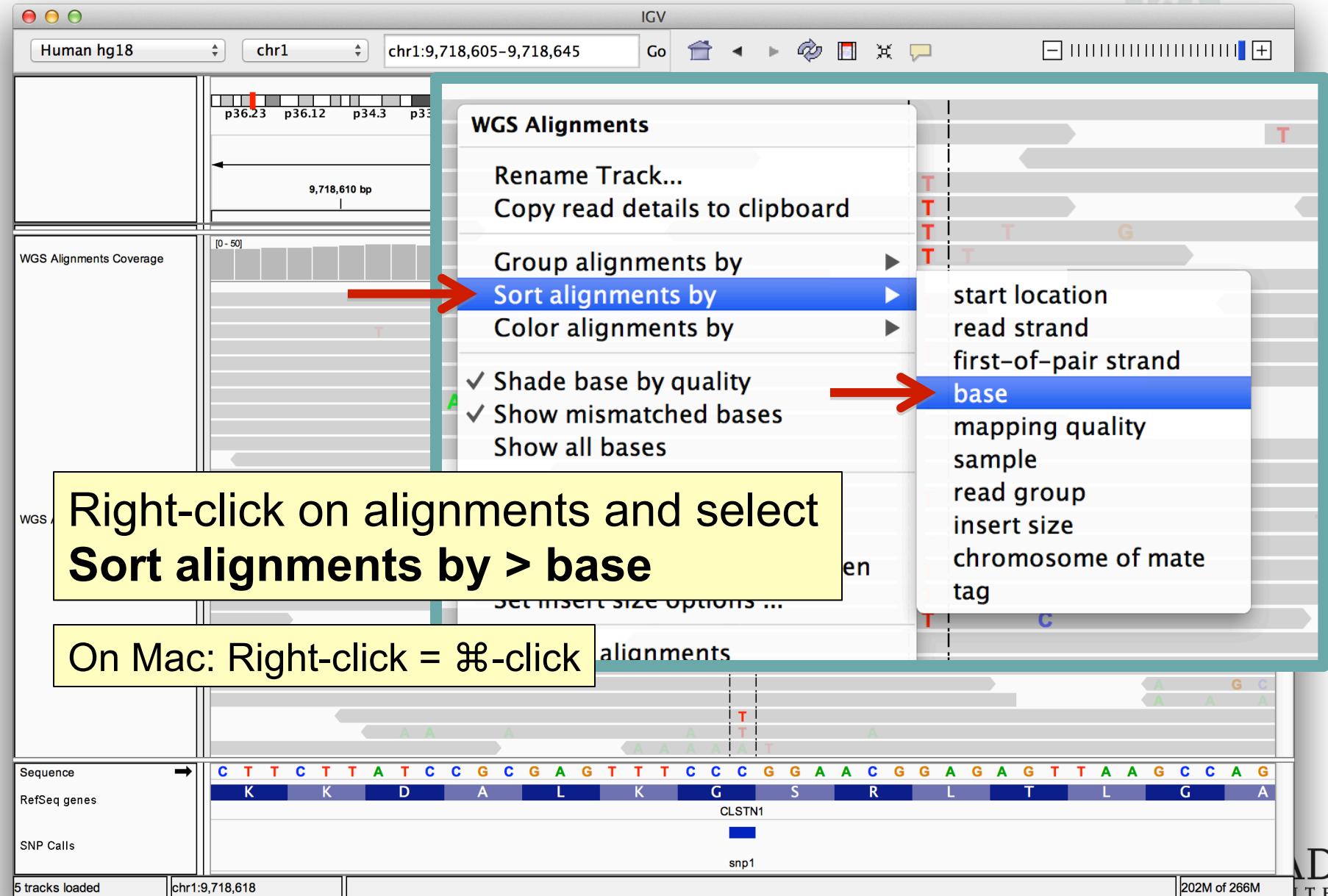
Click on yellow balloon icon in the toolbar to modify the information popup behavior



Viewing SNPs



Viewing SNPs



IGV

Human hg18 chr1 chr1:9,718,605–9,718,645 Go

WGS Alignments Coverage

WGS

Right-click on alignments and select **Sort alignments by > base**

On Mac: Right-click = ⌘-click

Sequence RefSeq genes SNP Calls

CLSTN1

snp1

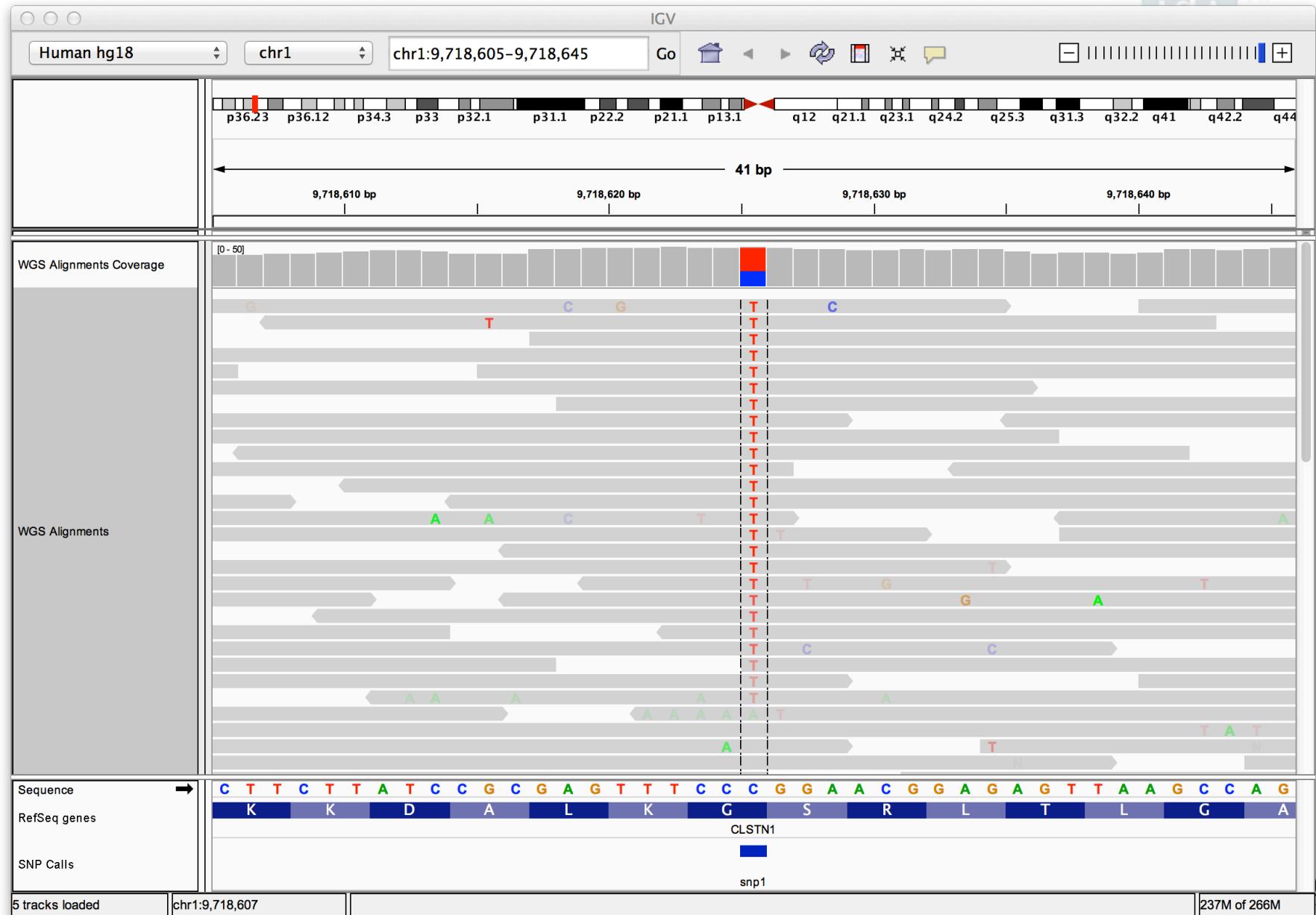
5 tracks loaded chr1:9,718,618 202M of 266M

WGS Alignments

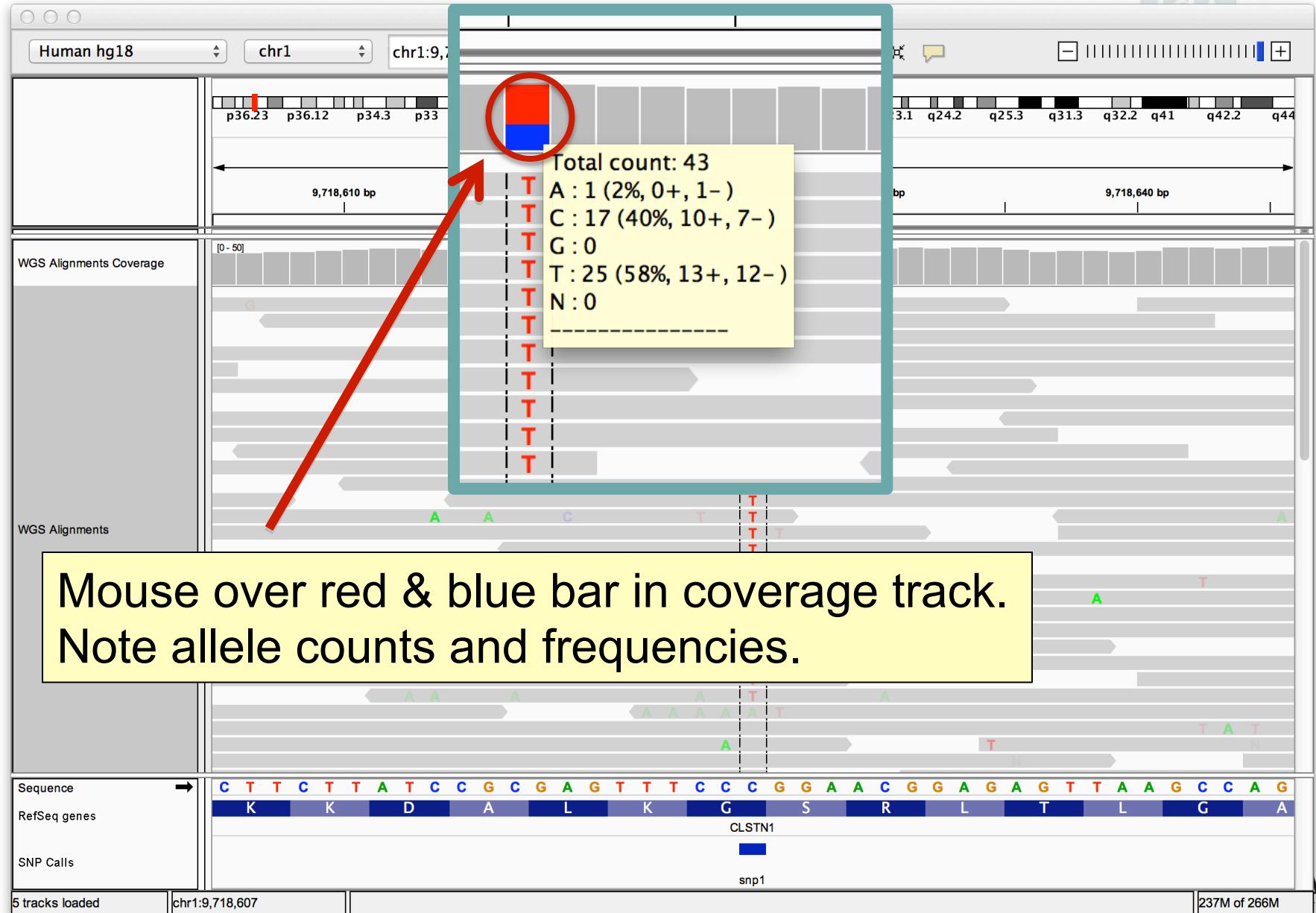
- Rename Track...
- Copy read details to clipboard
- Group alignments by ▶
- Sort alignments by ▶**
- Color alignments by ▶
- Shade base by quality
- Show mismatched bases
- Show all bases

- start location
- read strand
- first-of-pair strand
- base**
- mapping quality
- sample
- read group
- insert size
- chromosome of mate
- tag

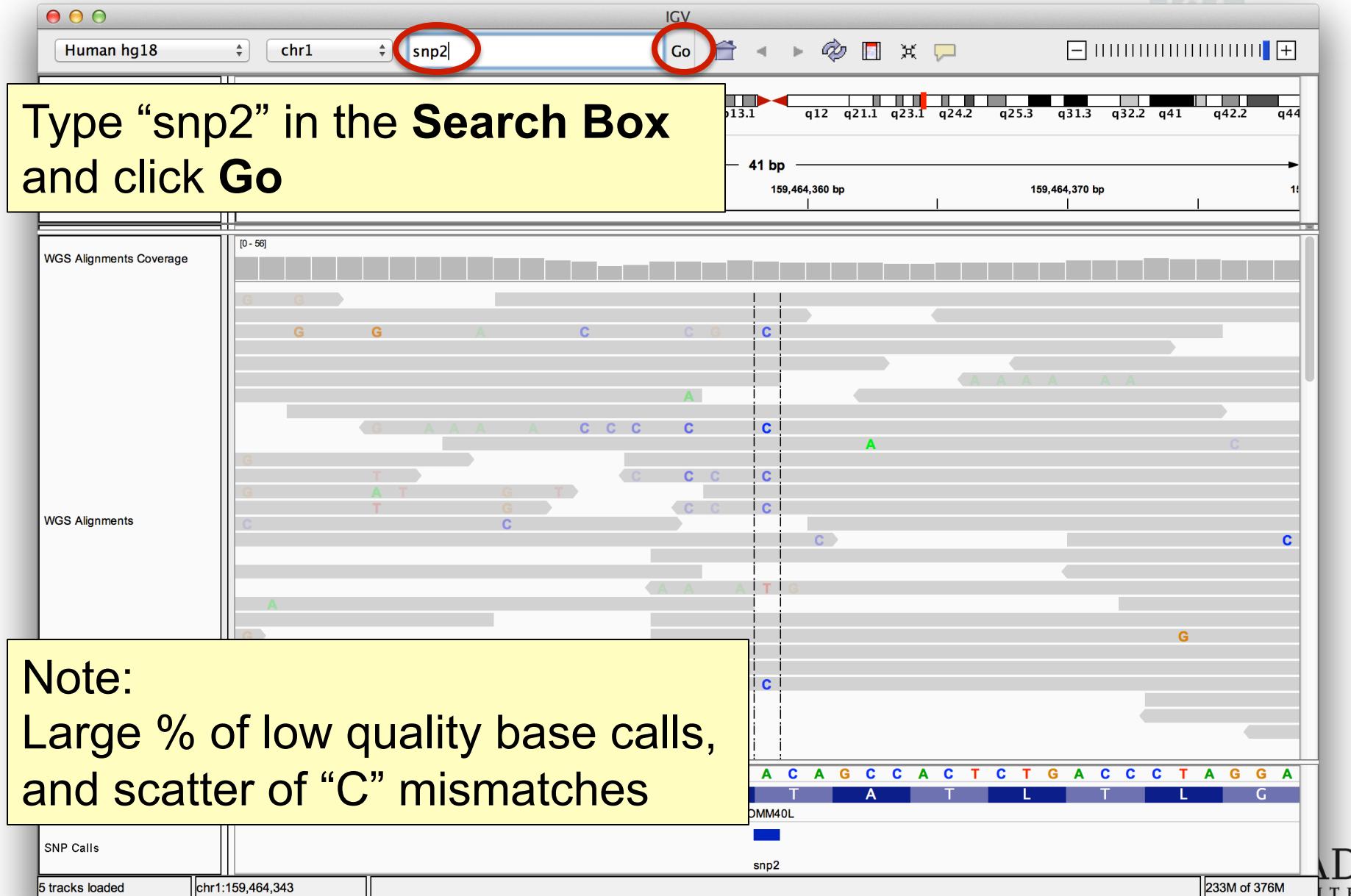
Viewing SNPs



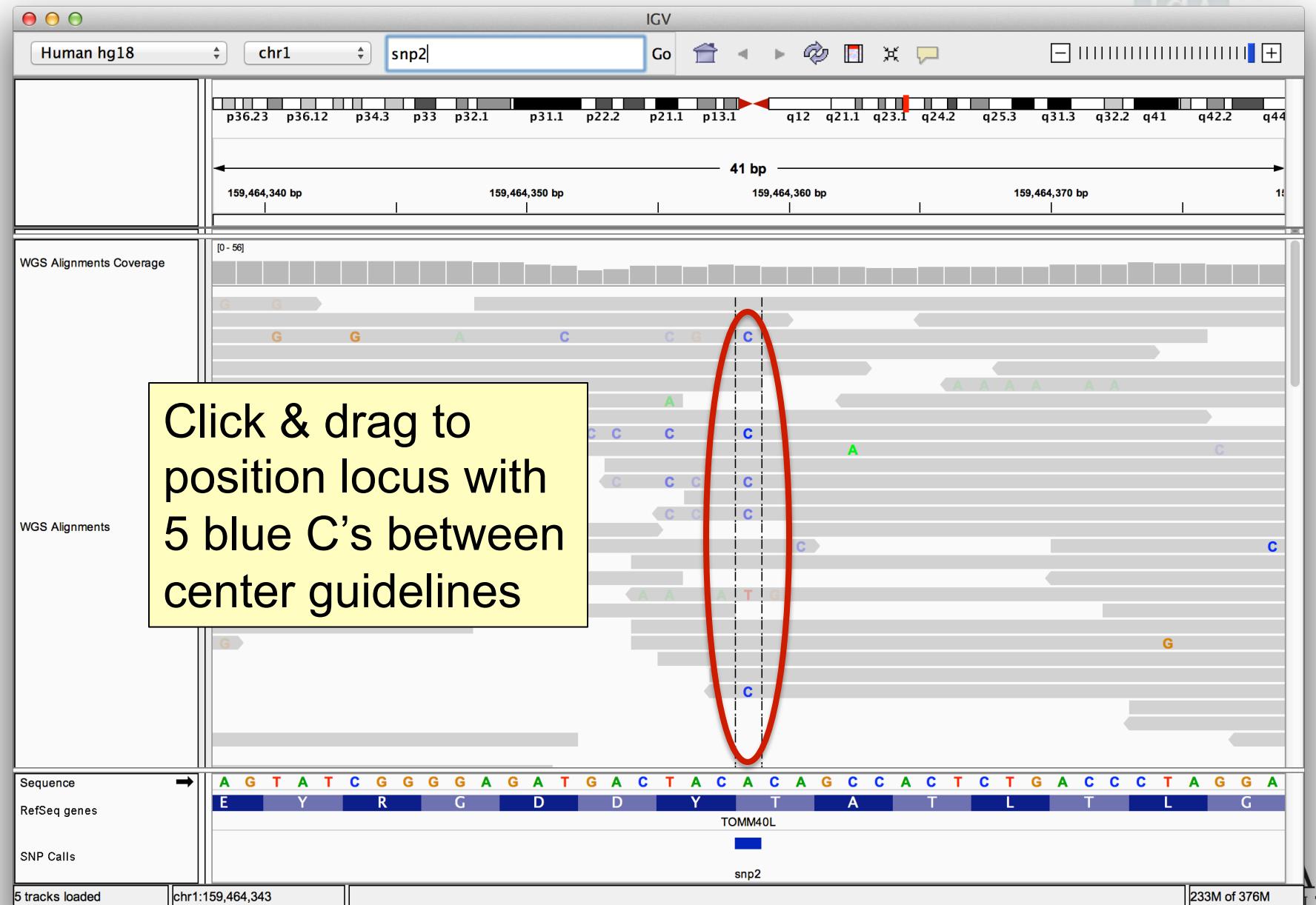
Viewing SNPs



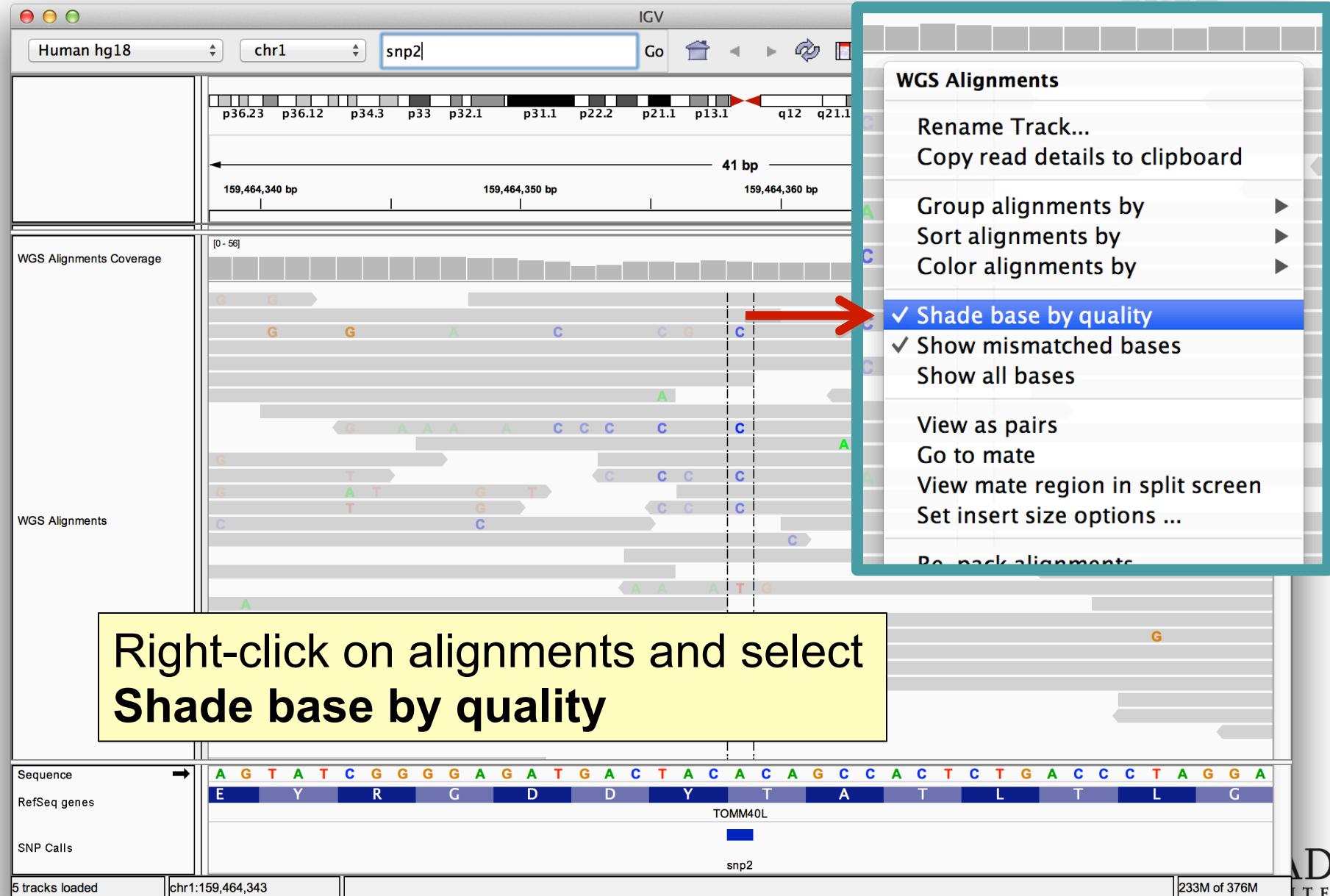
Viewing SNPs



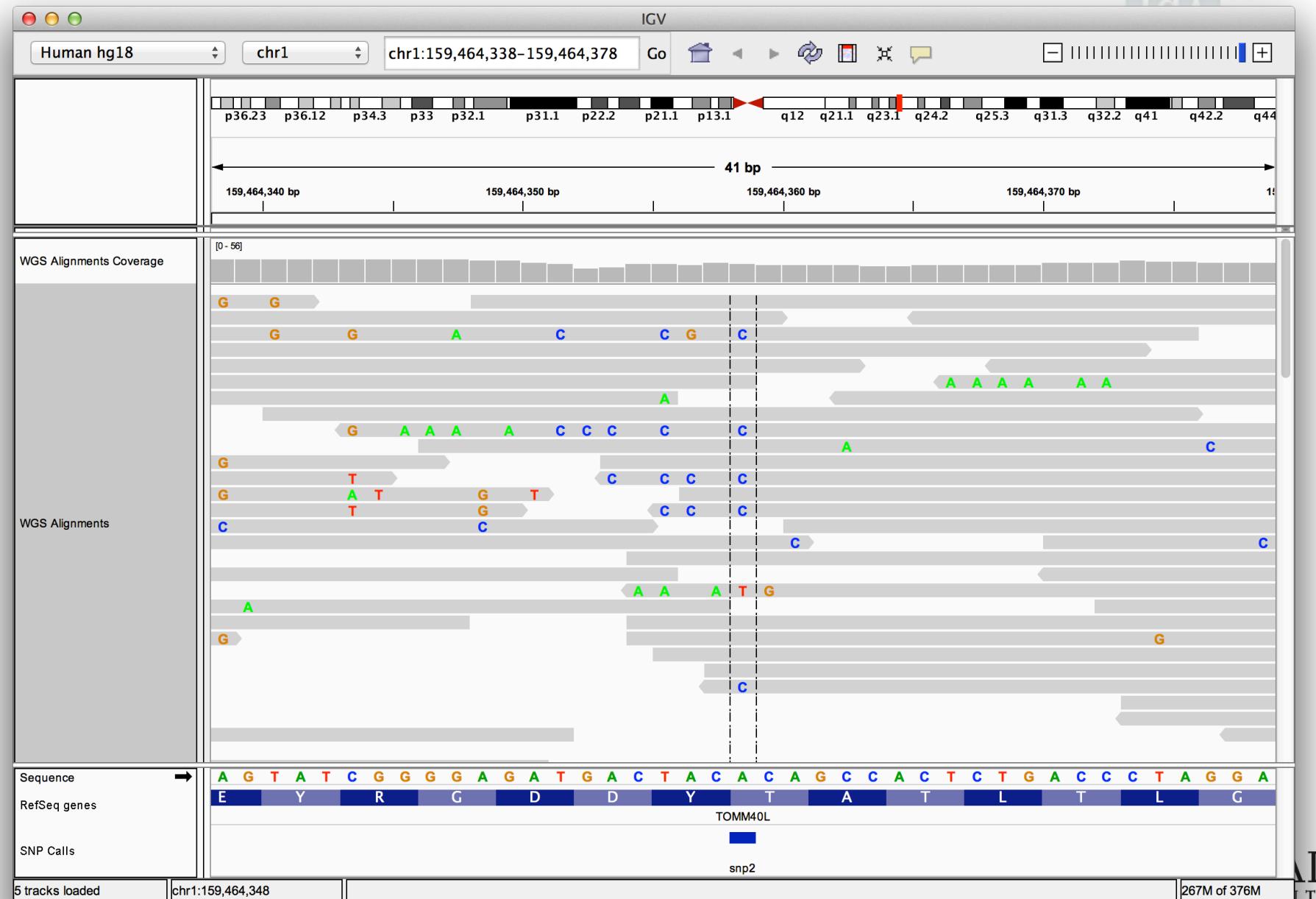
Viewing SNPs



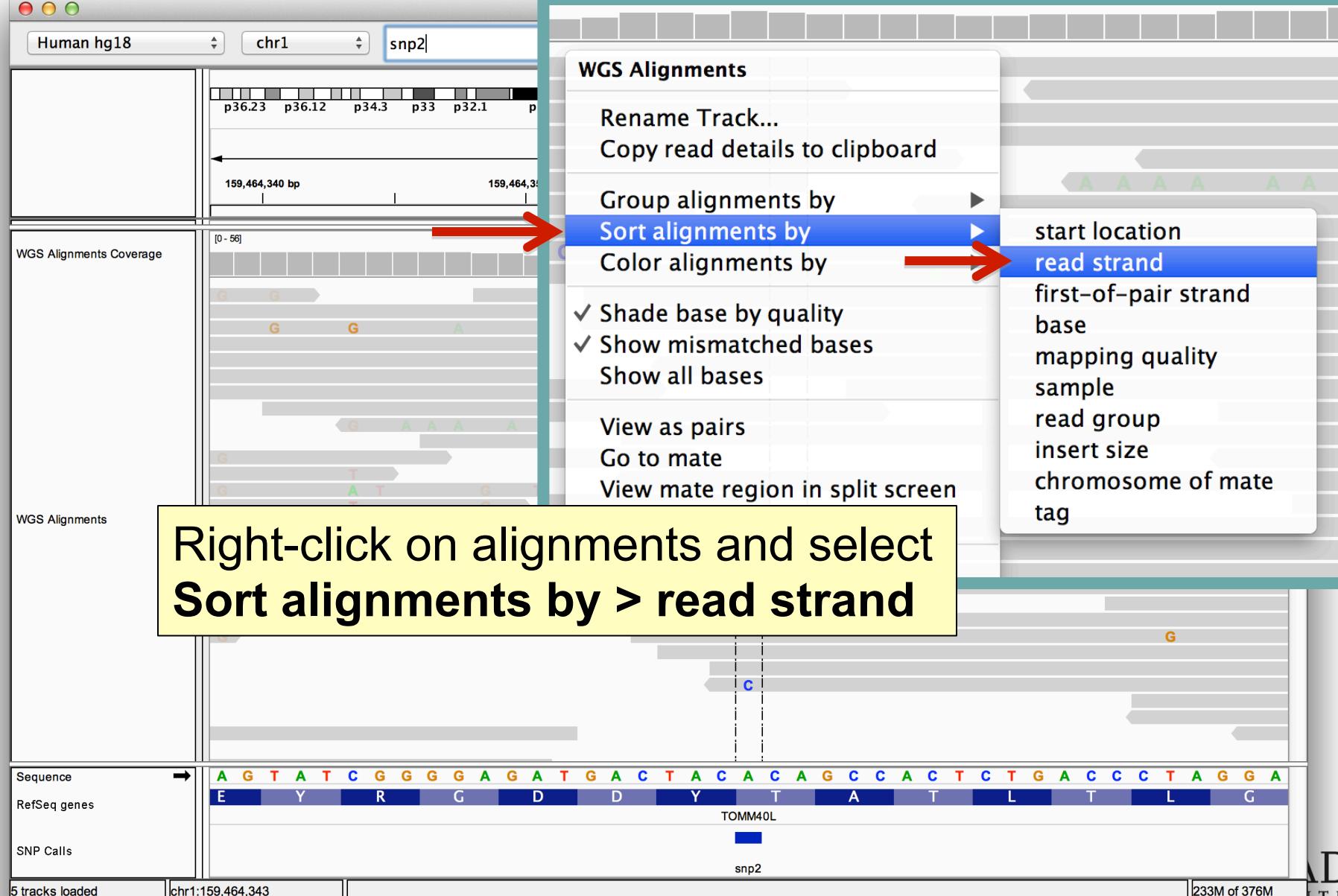
Viewing SNPs



Viewing SNPs



Viewing SNPs



The screenshot shows the IGV interface with a SNP track labeled "snp2" selected. A context menu is open over the alignments, with the "Sort alignments by" option highlighted. A secondary dropdown menu is open under "read strand", also with "read strand" highlighted. A yellow callout box contains the instructions: "Right-click on alignments and select Sort alignments by > read strand".

WGS Alignments

- Rename Track...
- Copy read details to clipboard
- Group alignments by
- Sort alignments by**
- Color alignments by
- ✓ Shade base by quality
- ✓ Show mismatched bases
- Show all bases
- View as pairs
- Go to mate
- View mate region in split screen

start location

read strand

- first-of-pair strand
- base
- mapping quality
- sample
- read group
- insert size
- chromosome of mate tag

Right-click on alignments and select Sort alignments by > read strand

Sequence →

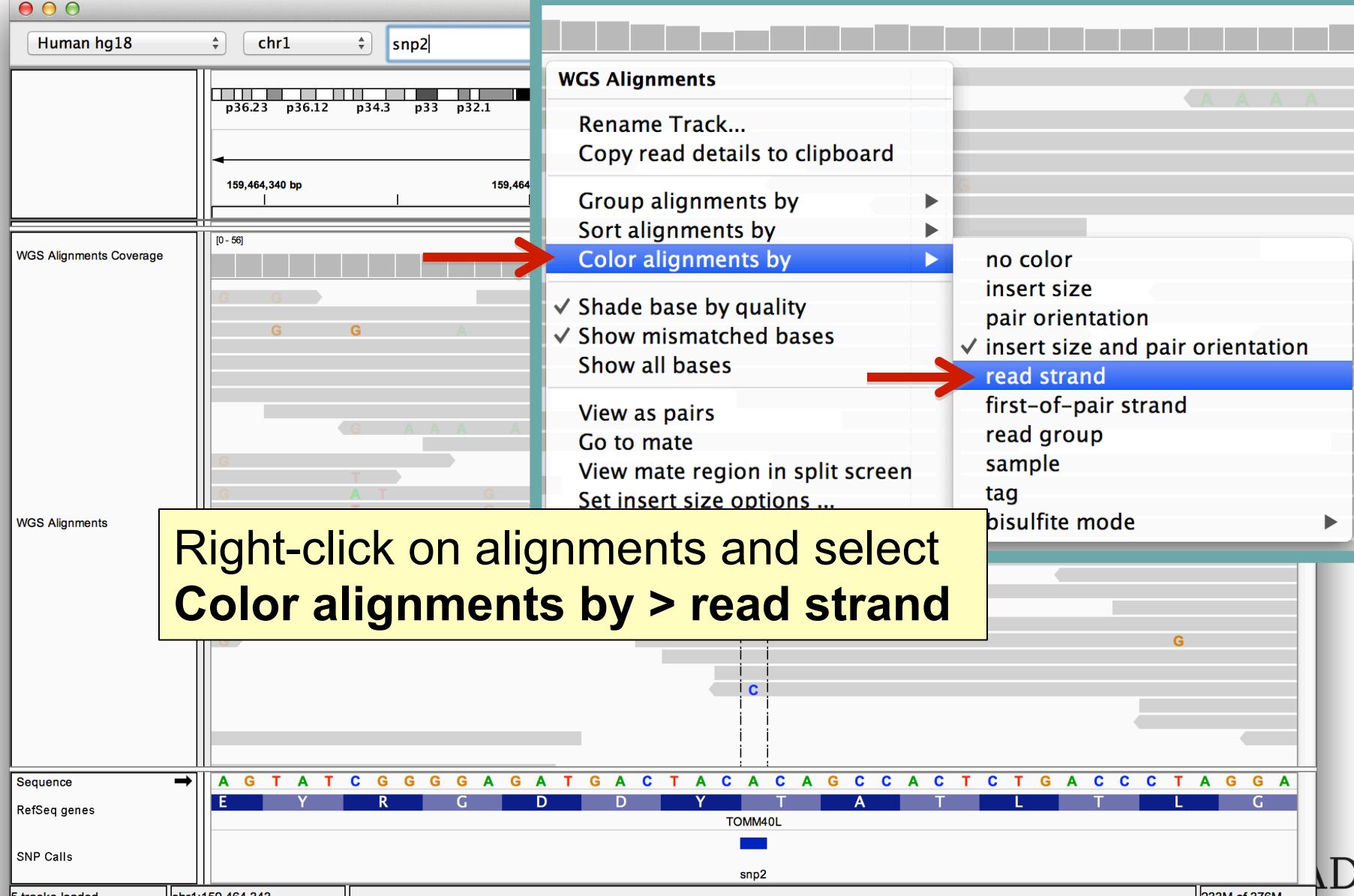
A	G	T	A	T	C	G	G	G	A	G	A	T	G	A	C	T	A	C	A	G	C	C	A	C	T	C	T	G	A	C	C	C	T	A	G	G
E	Y	R	G	D	D	D	Y	T	A	T	A	T	L	T	L	T	L	T	L	G																

RefSeq genes

SNP Calls

5 tracks loaded chr1:159,464,343 233M of 376M

Viewing SNPs



The screenshot shows the IGV interface with the following details:

- Top Left:** Human hg18 genome browser view showing chromosome chr1 from position 159,464,340 bp to 159,464.
- Top Right:** A context menu is open over a WGS Alignments track. The menu items include: Rename Track..., Copy read details to clipboard, Group alignments by, Sort alignments by, **Color alignments by**, Shade base by quality, Show mismatched bases, Show all bases, View as pairs, Go to mate, View mate region in split screen, and Set insert size options ...
- Bottom Left:** A yellow callout box contains the text: "Right-click on alignments and select Color alignments by > read strand".
- Bottom Right:** A sequence track for the TOMM40L gene, showing the DNA sequence: A G T A T C G G G G A G A T G A C T T A C A C A G C C A C T C T G A C C C T A G G A. Below the sequence, RefSeq genes are listed: E Y R G D D Y T A T L T L G. The gene name TOMM40L is centered above the sequence, and the SNP identifier.snp2 is at the bottom.

Viewing SNPs



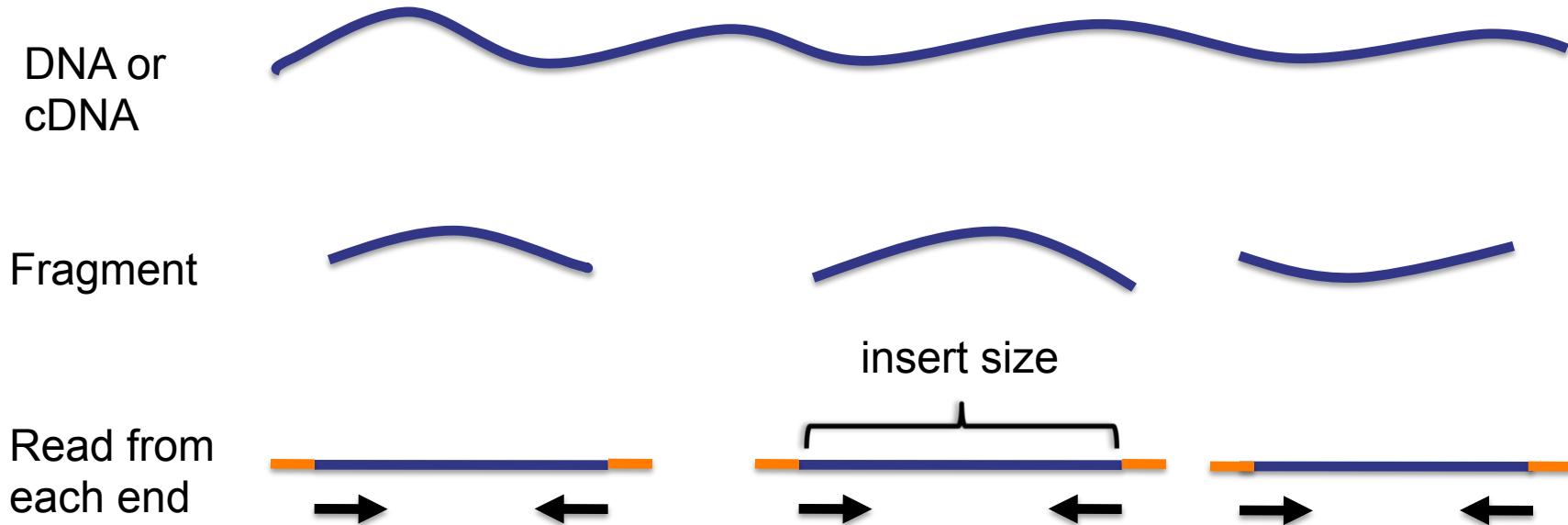
Viewing Structural Events

Structural events

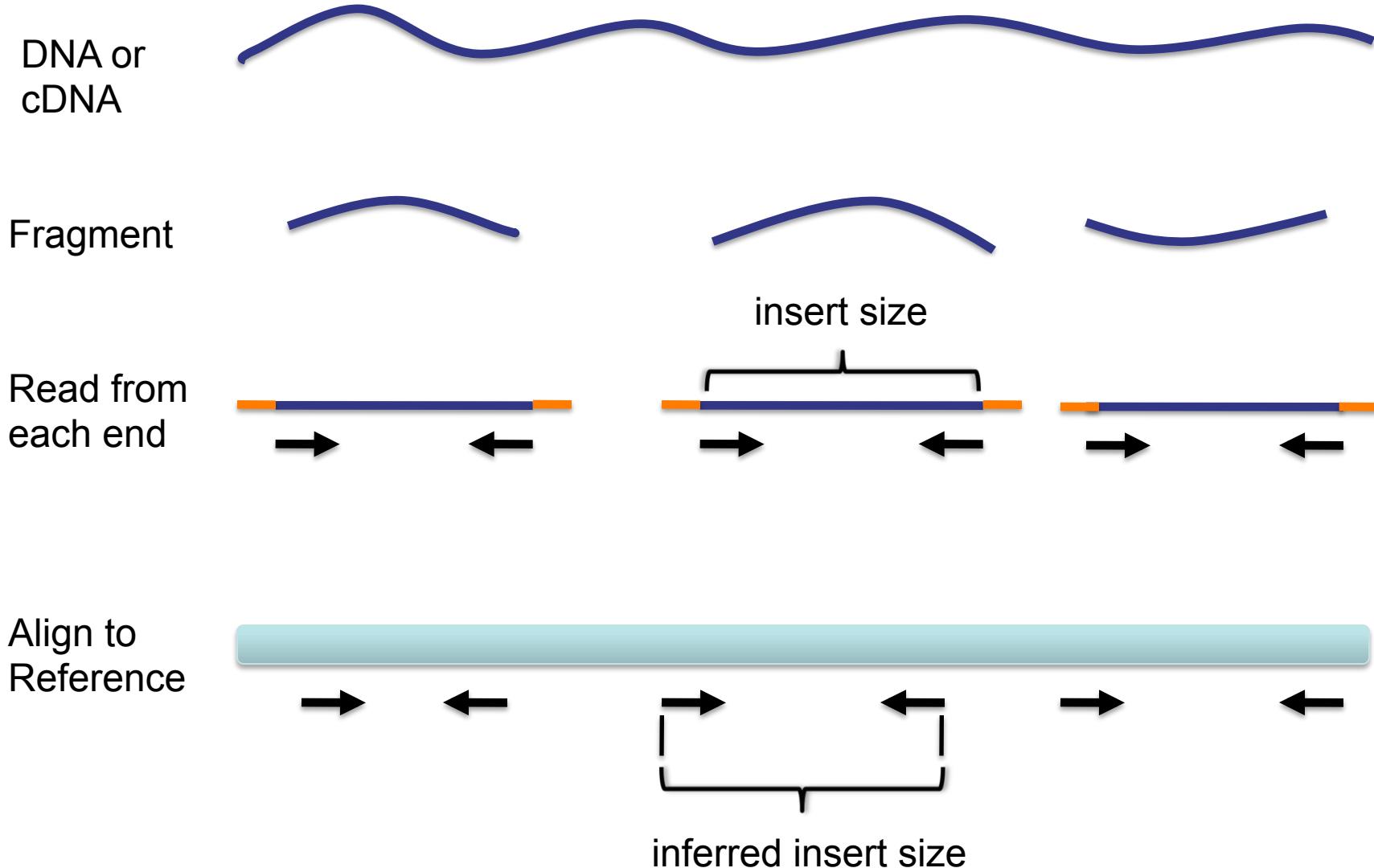


- Paired reads can yield evidence for genomic “structural events”, such as deletions, translocations, and inversions.
- Alignment coloring options help highlight these events based on:
 - Inferred insert size (template length)
 - Pair orientation (relative strand of pair)

Paired-end sequencing



Paired-end sequencing



Interpreting Insert Size

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



Deletion



Reference
Genome



Subject



Deletion

Reference
Genome



Subject



Deletion



Reference
Genome



Subject

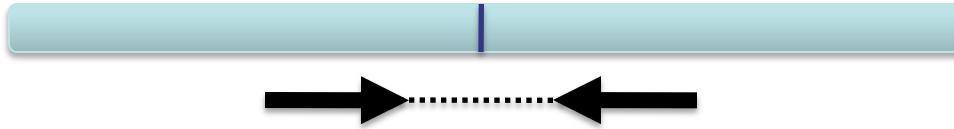


Deletion

Reference
Genome



Subject



Deletion

Reference
Genome



Subject



Deletion

Reference
Genome

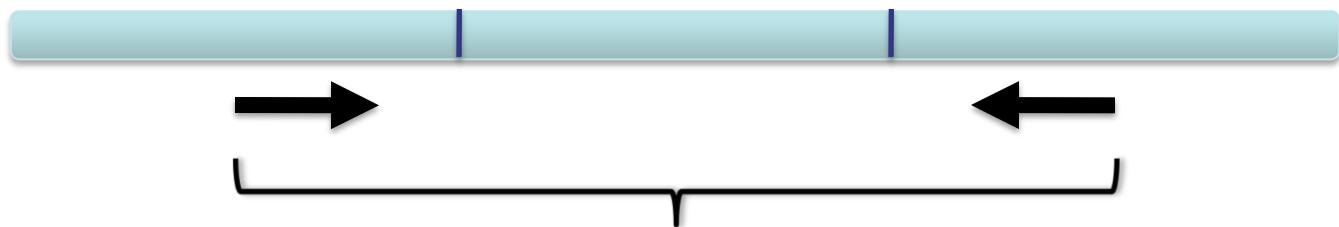


Subject



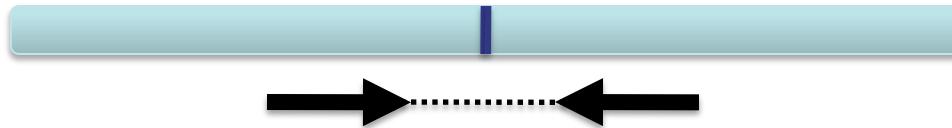
Deletion

Reference
Genome



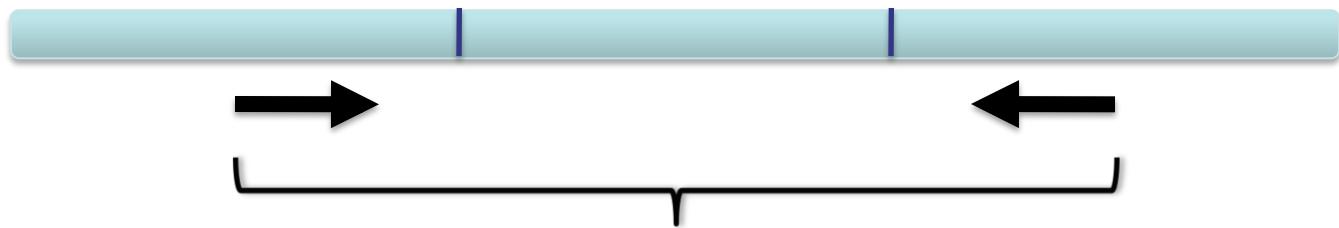
inferred insert size

Subject



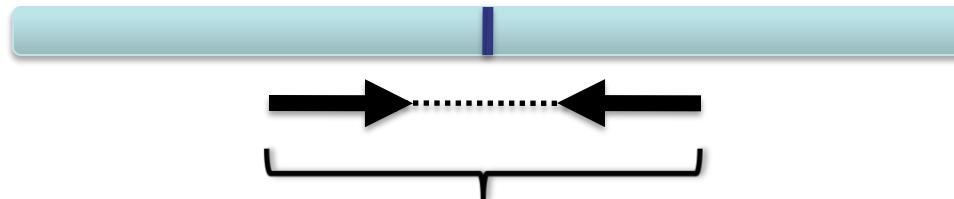
Deletion

Reference
Genome



inferred insert size

Subject

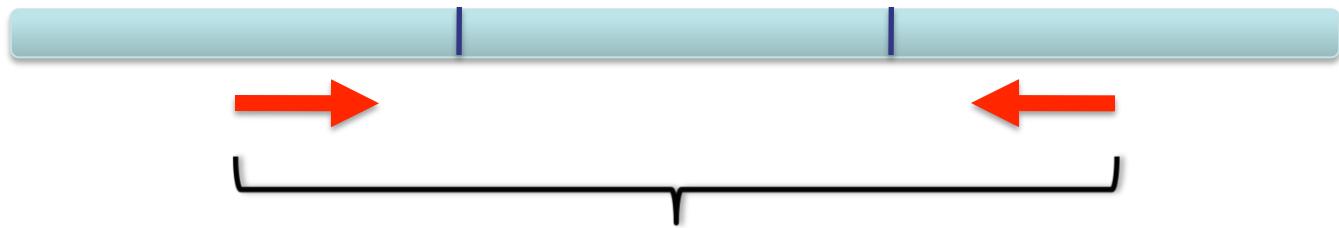


expected insert size

Deletion

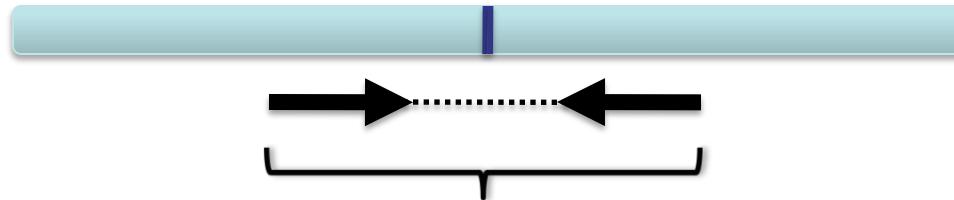
Inferred insert size is > expected value

Reference
Genome



inferred insert size

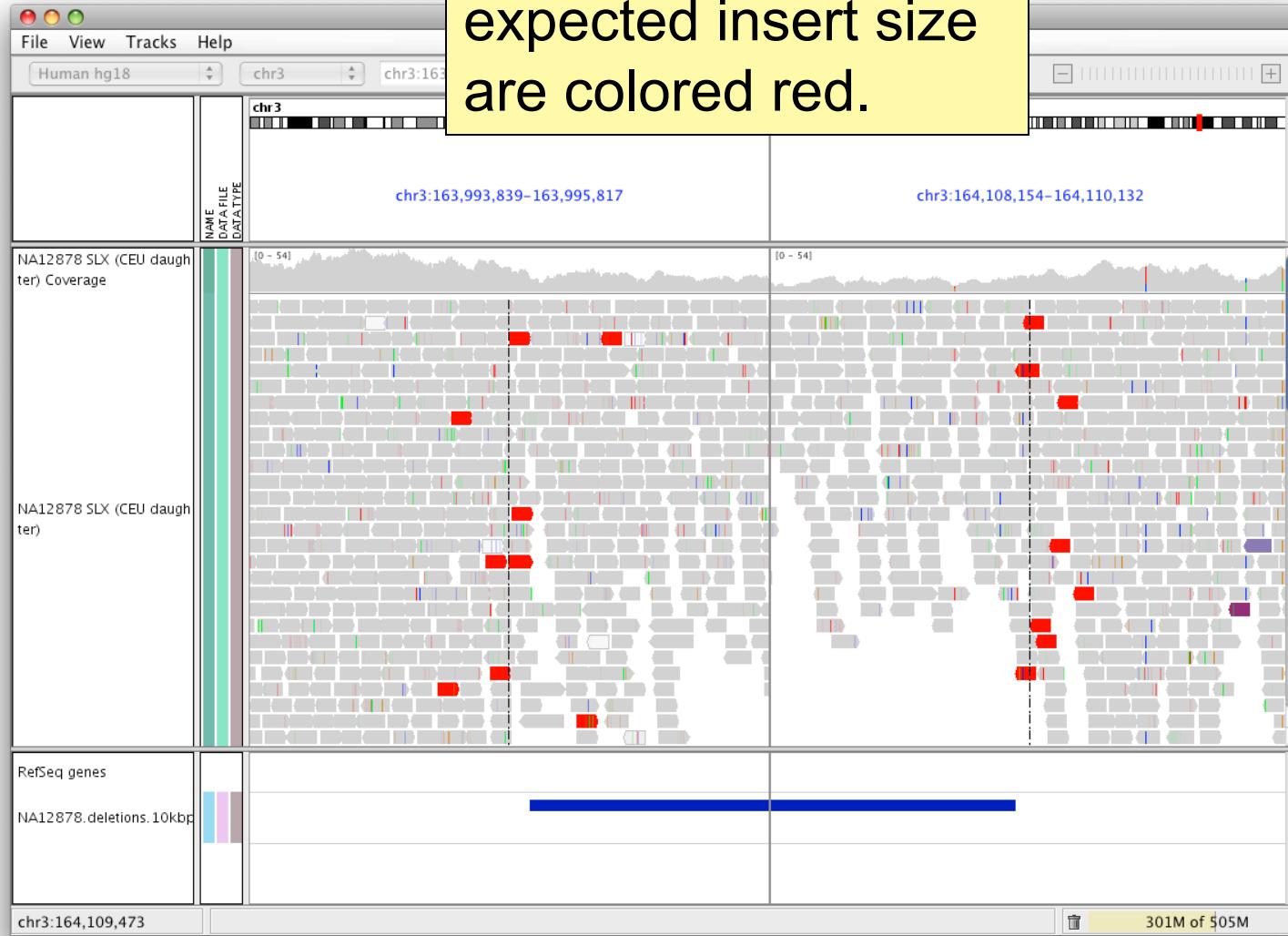
Subject



expected insert size

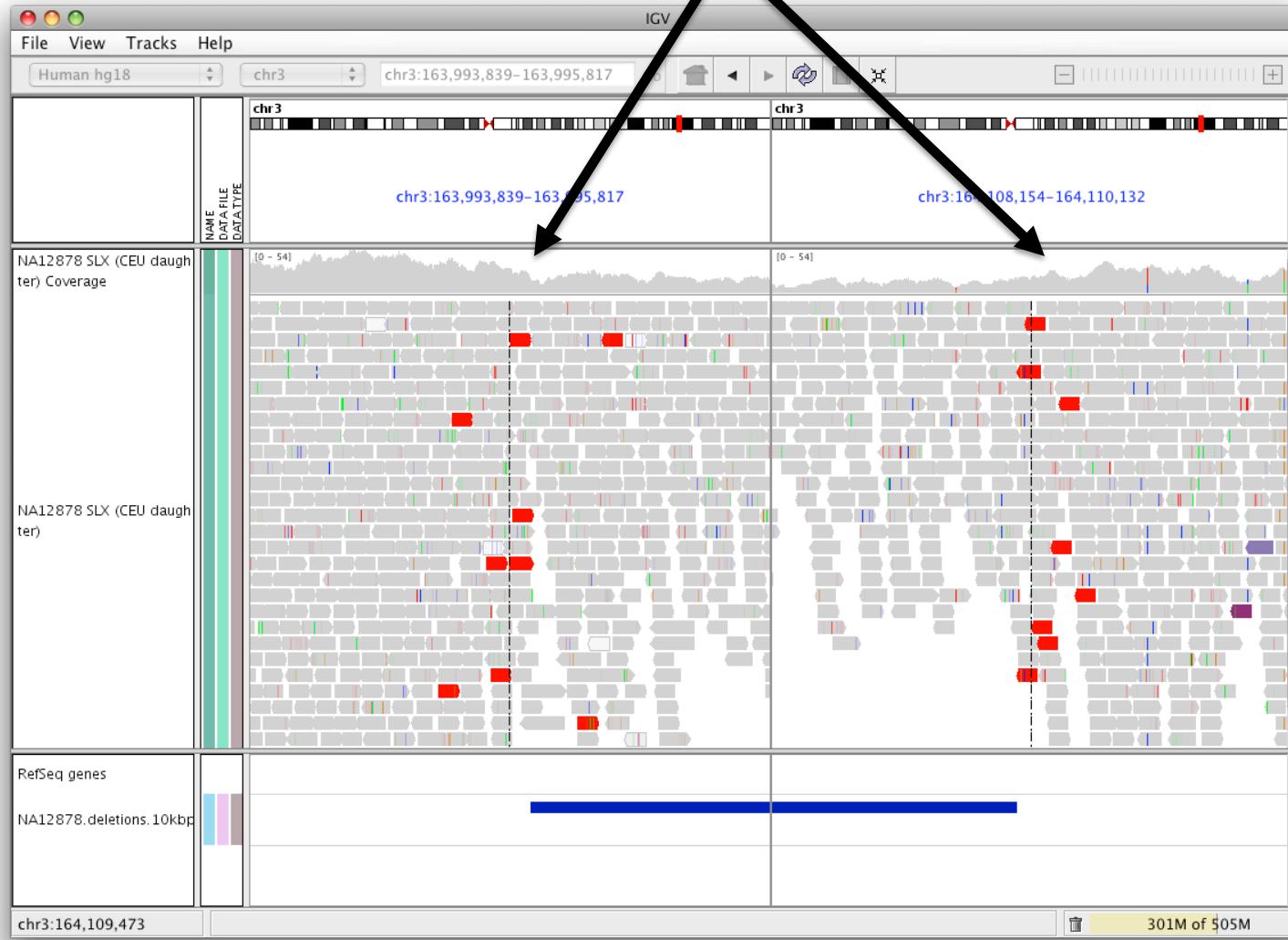
Deletion

Pairs with larger than expected insert size are colored red.



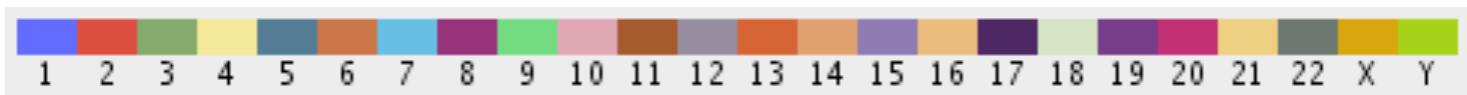
Deletion

Note drop in coverage

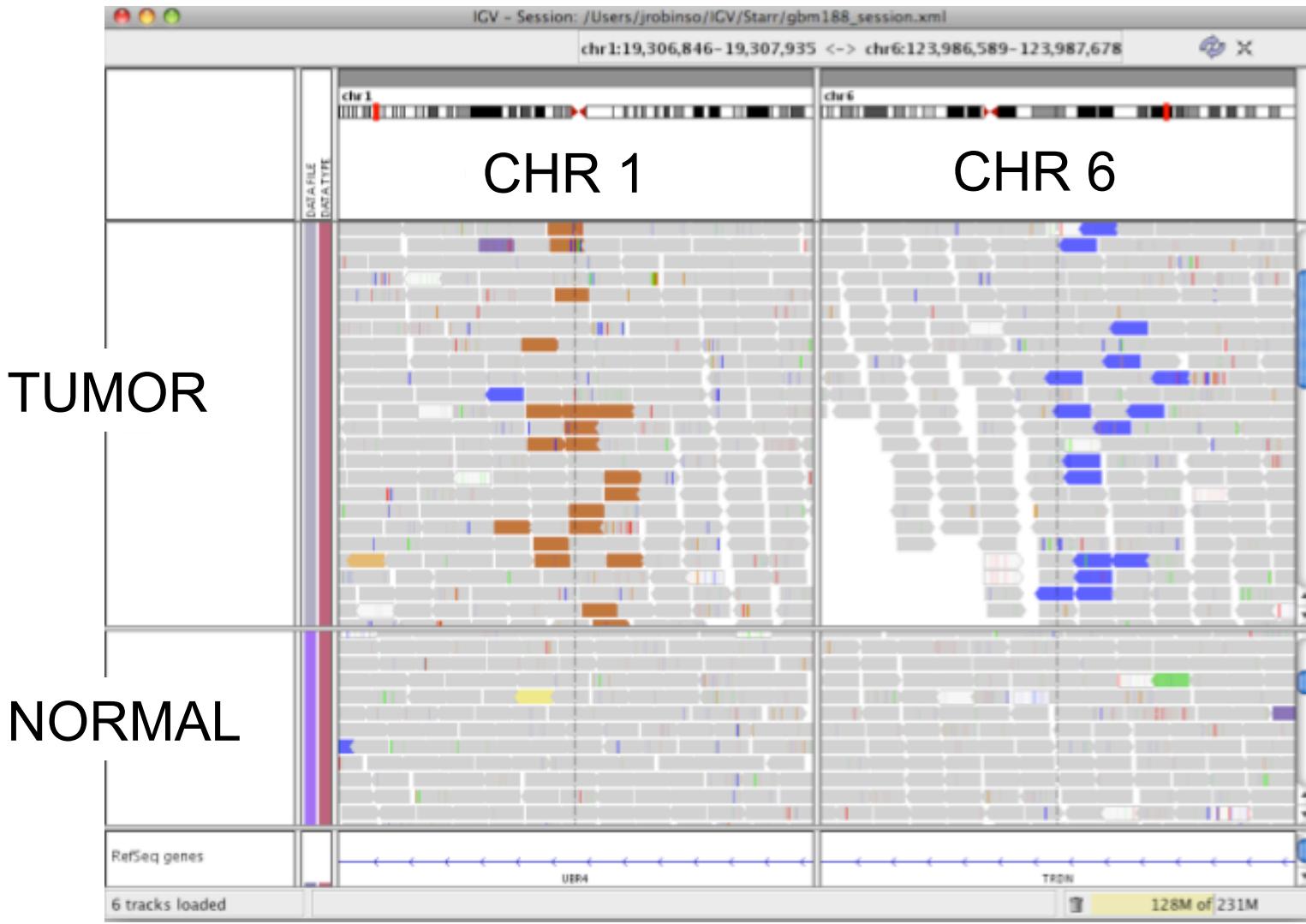


Insert size color scheme

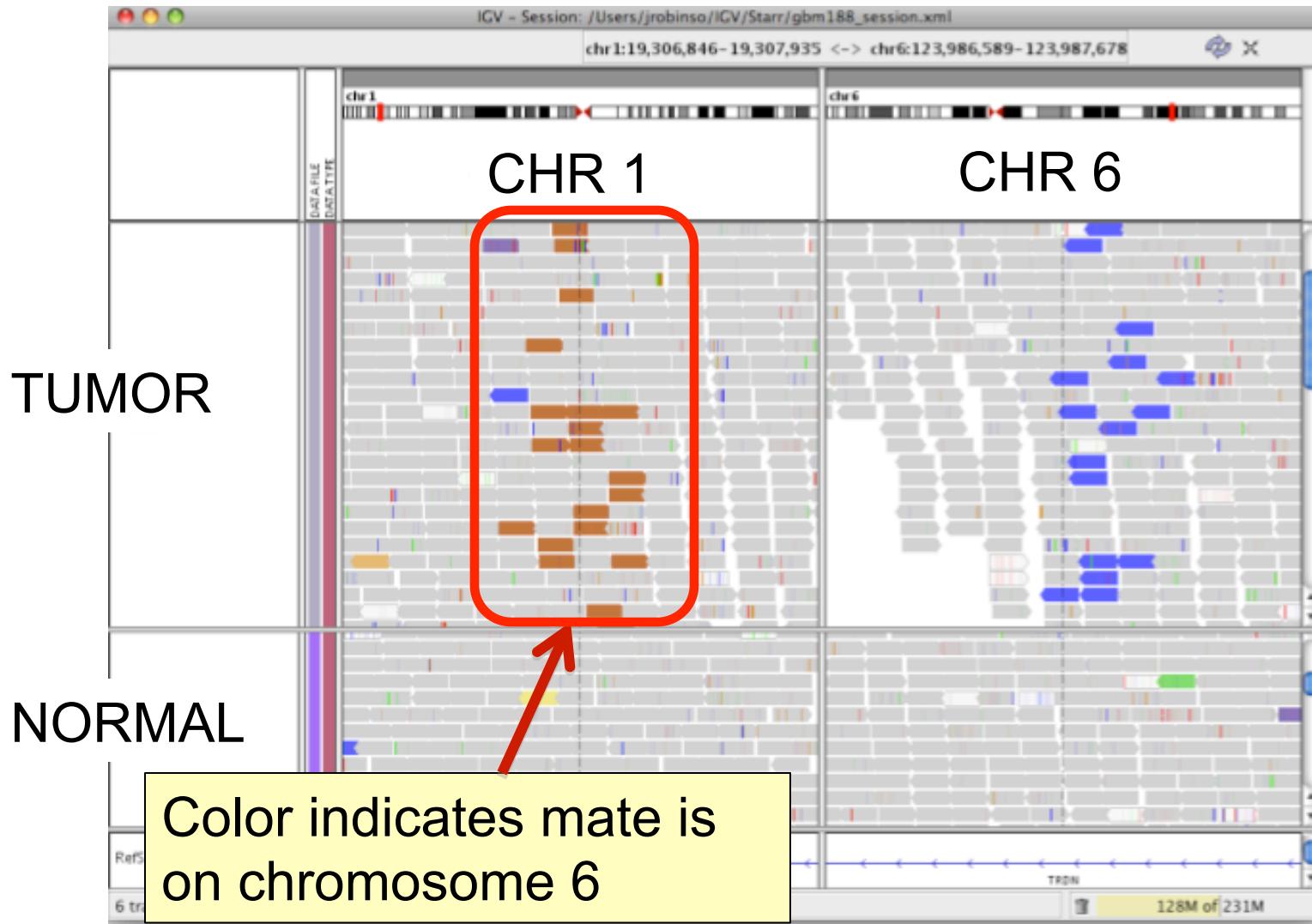
- Smaller than expected insert size: 
- Larger than expected insert size: 
- Pairs on different chromosomes
Each end colored by chromosome of its mate



Rearrangement



Rearrangement



Interpreting Pair Orientations

Interpreting pair orientations



Orientation of paired reads can reveal structural events, including:

- inversions
- duplications
- translocations

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

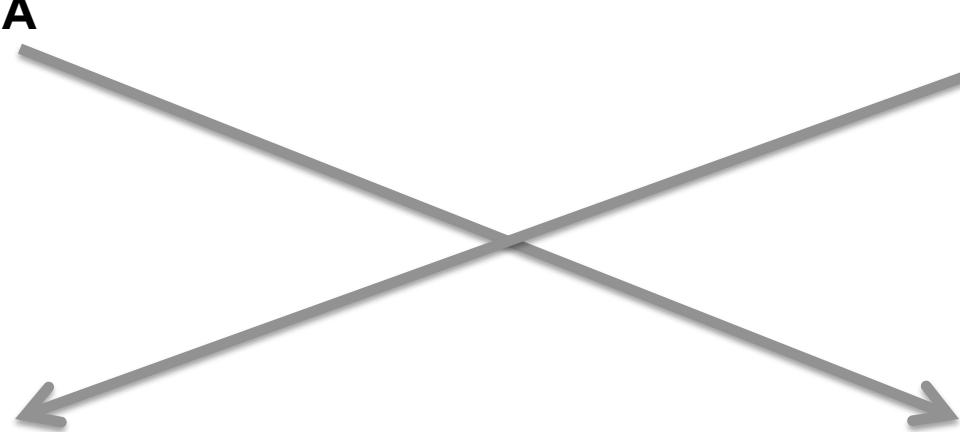
B

Subject



B

A



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



A

B

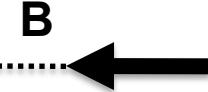


Subject

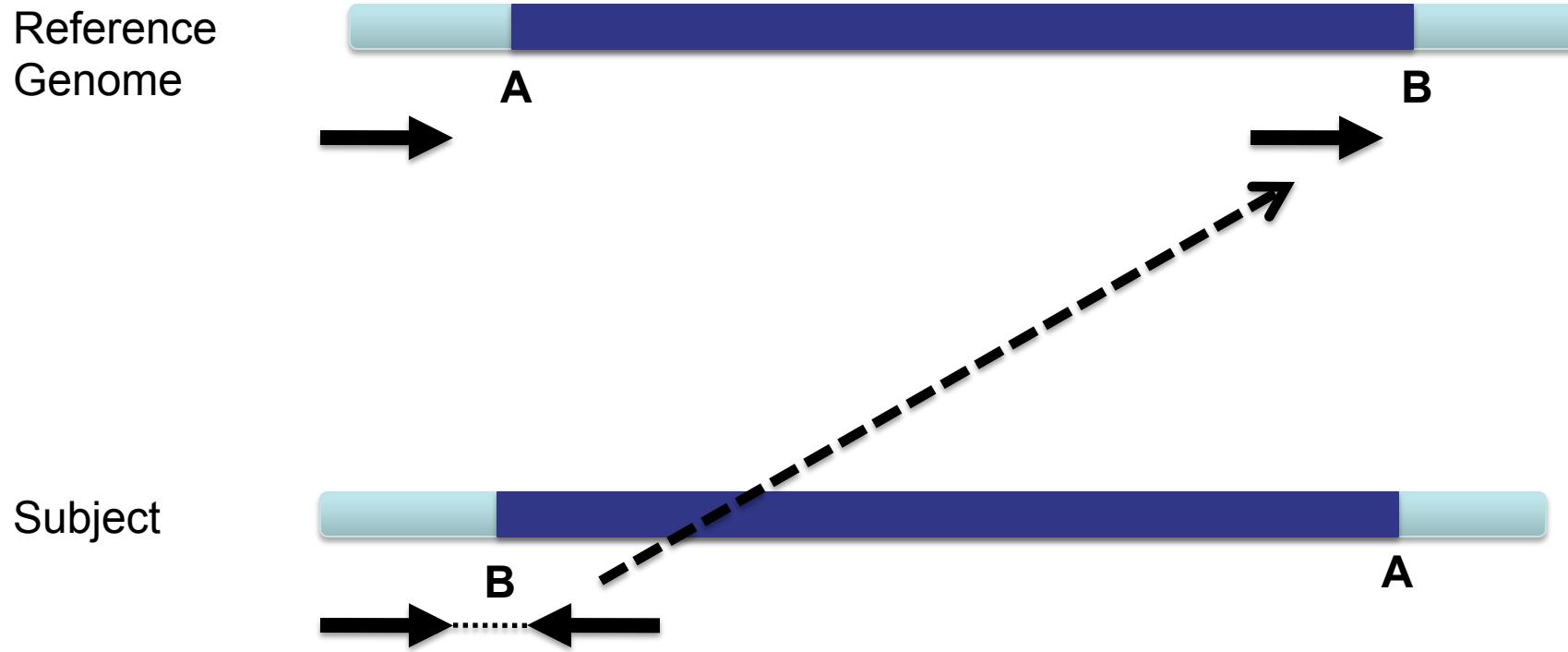


B

A



Inversion



Inversion

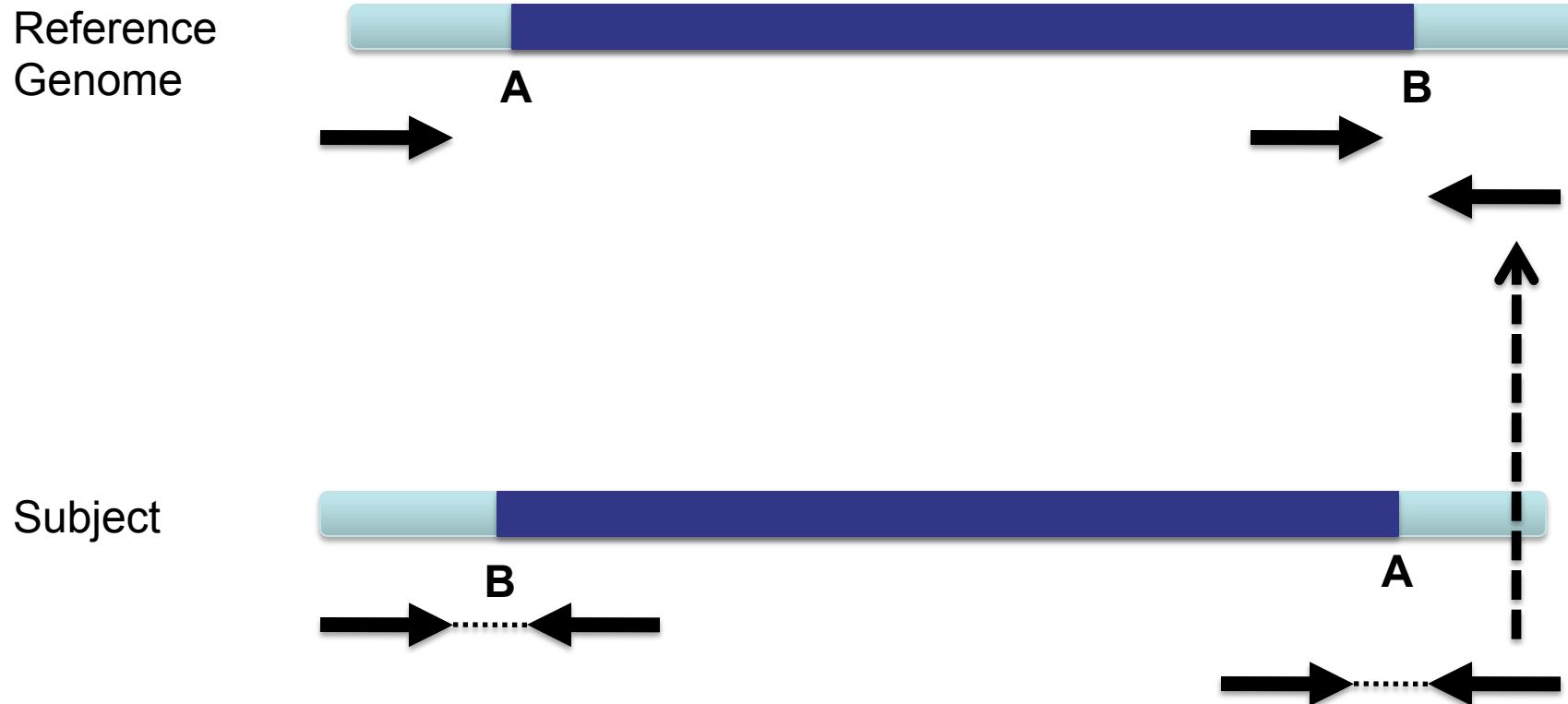
Reference
Genome



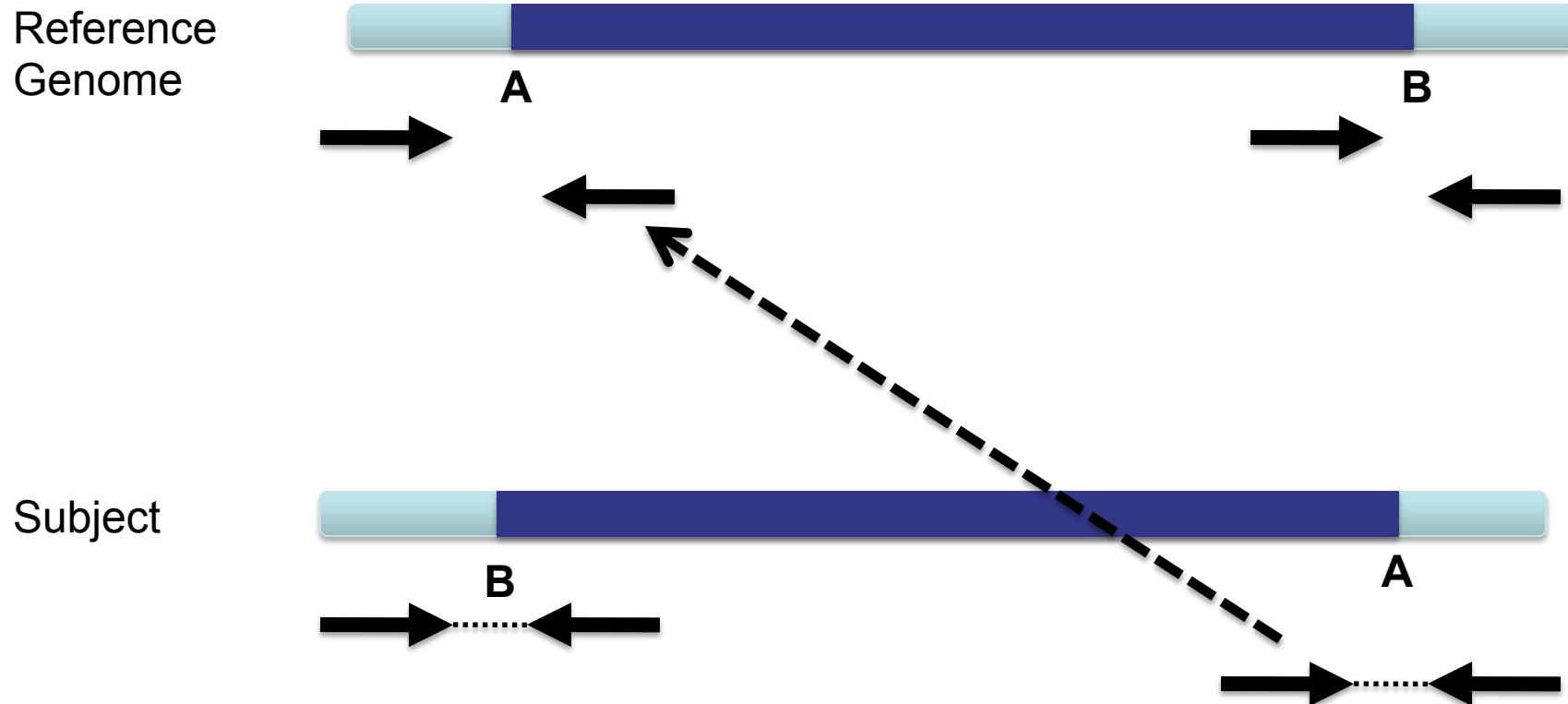
Subject



Inversion

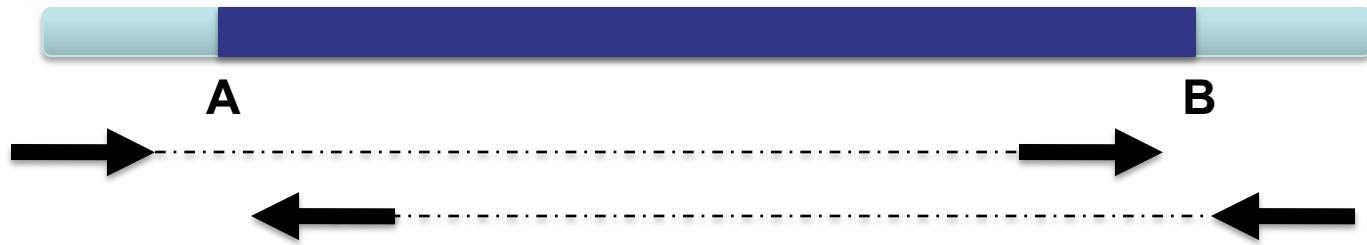


Inversion



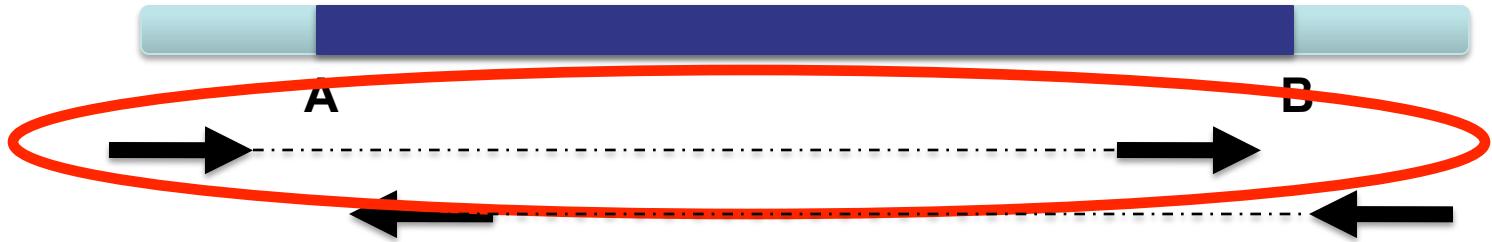
Inversion

Reference
Genome



Inversion

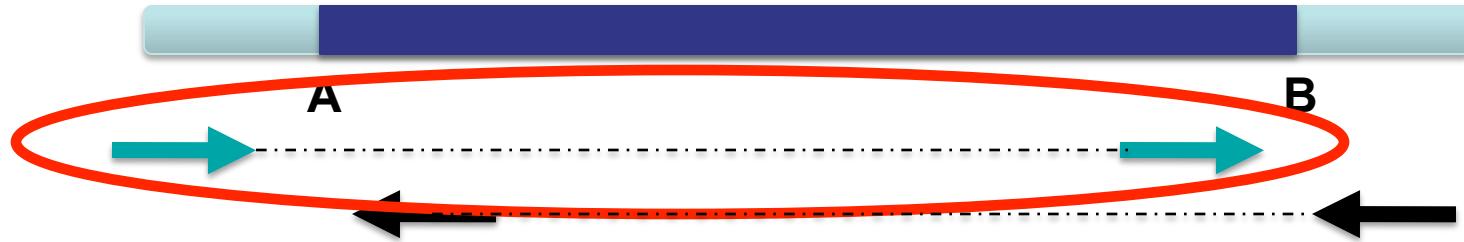
Reference
Genome



Anomaly –
Expected pair orientation is
inward facing (→ ←)

Inversion

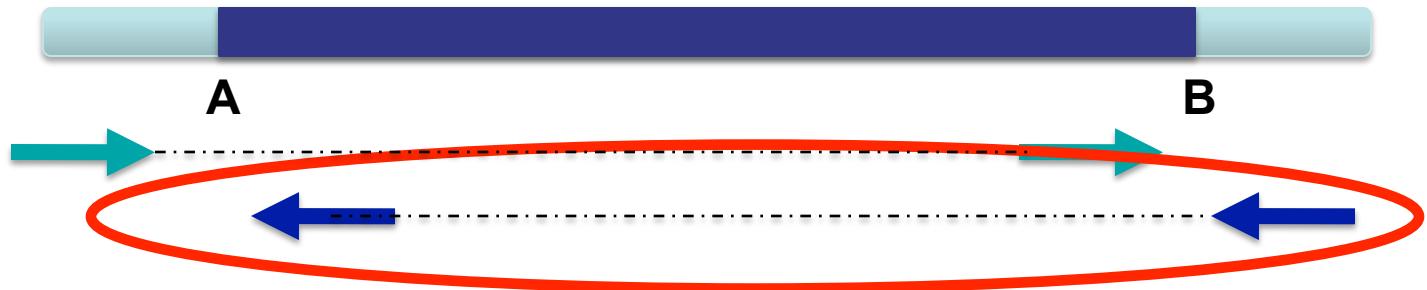
Reference
Genome



“Left” side pair

Inversion

Reference
Genome



“Right” side pair

Color by pair orientation



NA12878 WGS

- Rename Track...
- Copy read details to clipboard
- Group alignments by ►
- Sort alignments by ►
- Color alignments by ►**

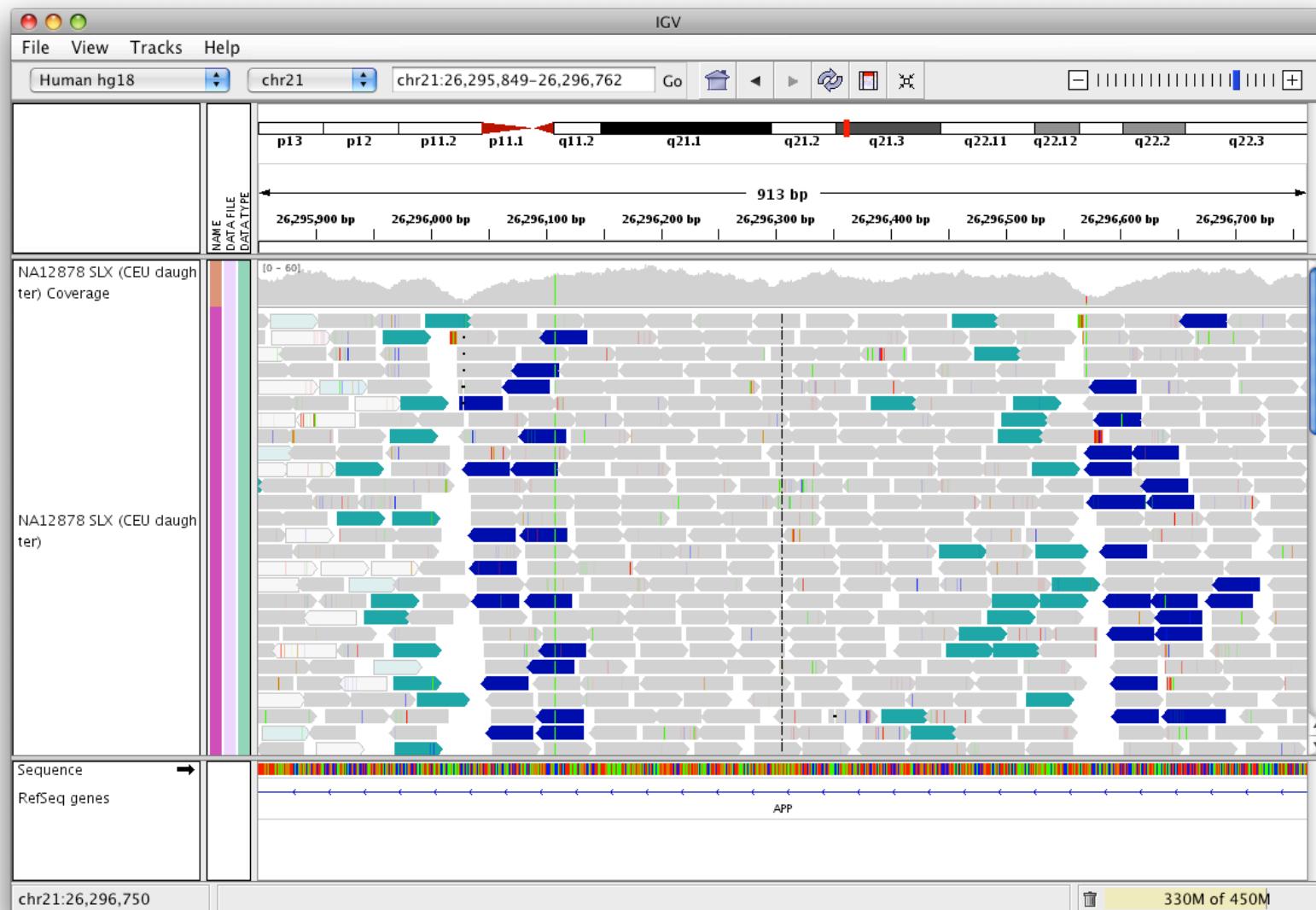
- ✓ Shade base by quality
- ✓ Show mismatched bases
- Show all bases

- View as pairs
- Go to mate
- View mate region in split screen
- Set insert size options ...

- Re-pack alignments

- no color
- insert size
- ✓ pair orientation**
- insert size and pair orientation
- read strand
- first-of-pair strand
- read group
- sample
- tag
- bisulfite mode

Inversion

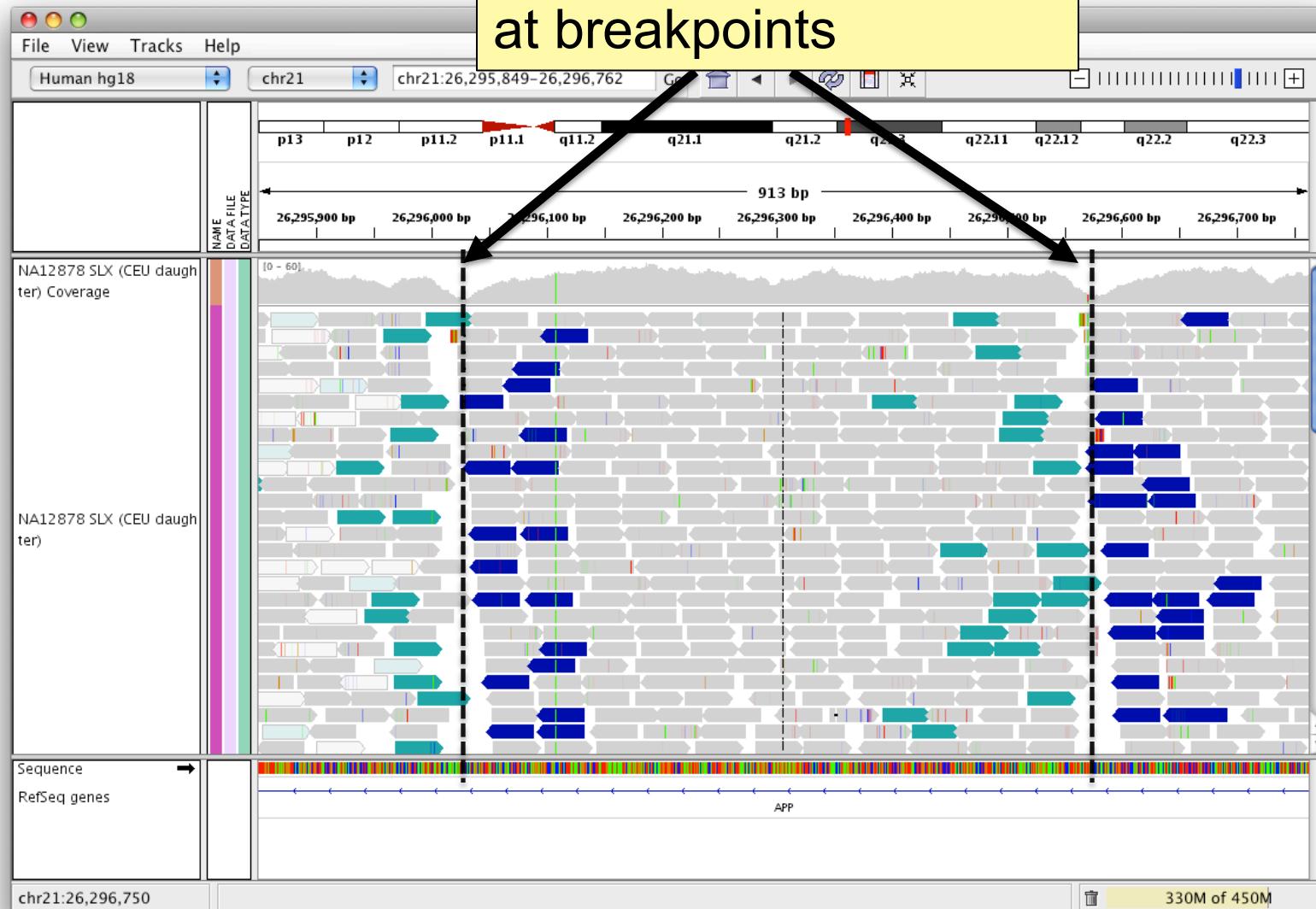


Inversion

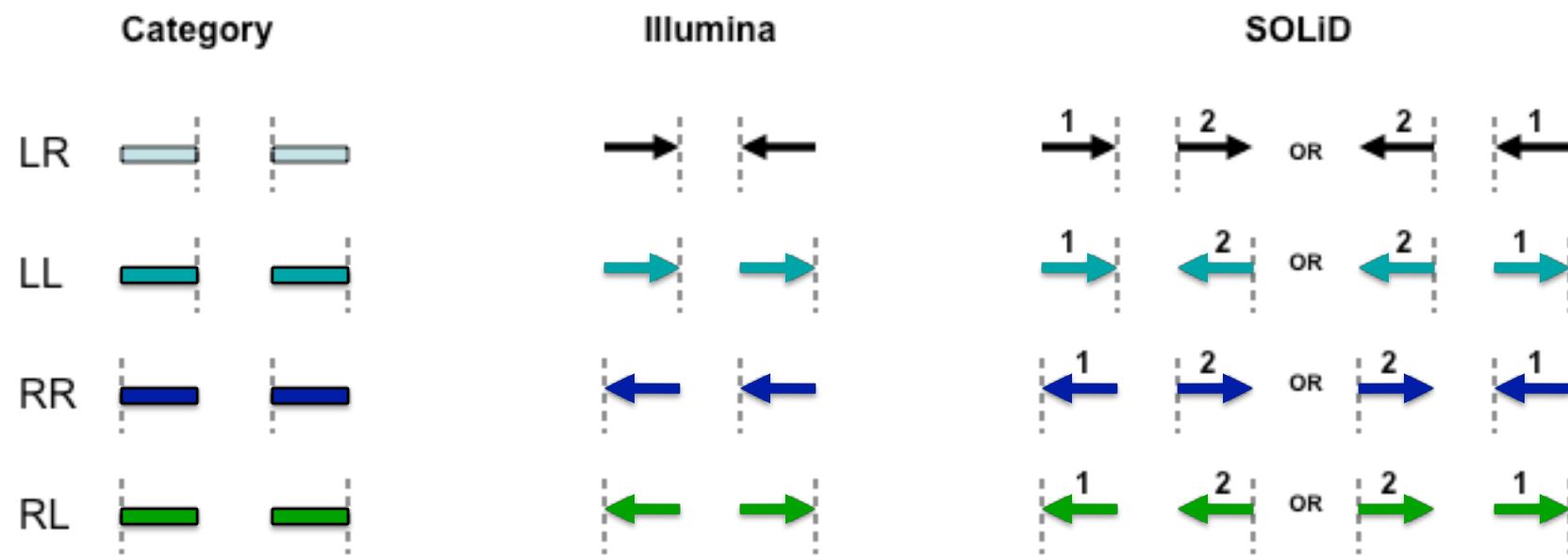


Integrative
Genomics
Viewer

Note drop in coverage
at breakpoints



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.

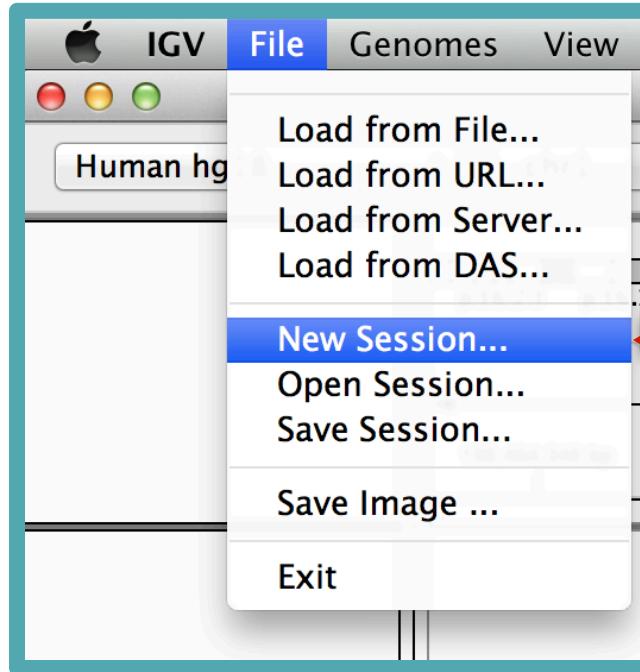
These categories only apply to reads where both mates map to the same chromosome.

Figure courtesy of Bob Handsaker

Hands-on exercise

- Examine tissue-specific alternative splicing.
- Data: Illumina BodyMap 2.0

http://www.illumina.com/science/data_library.ilmn



Before we start:
Select File > New Session
to clear IGV window

RNA-Seq Setup

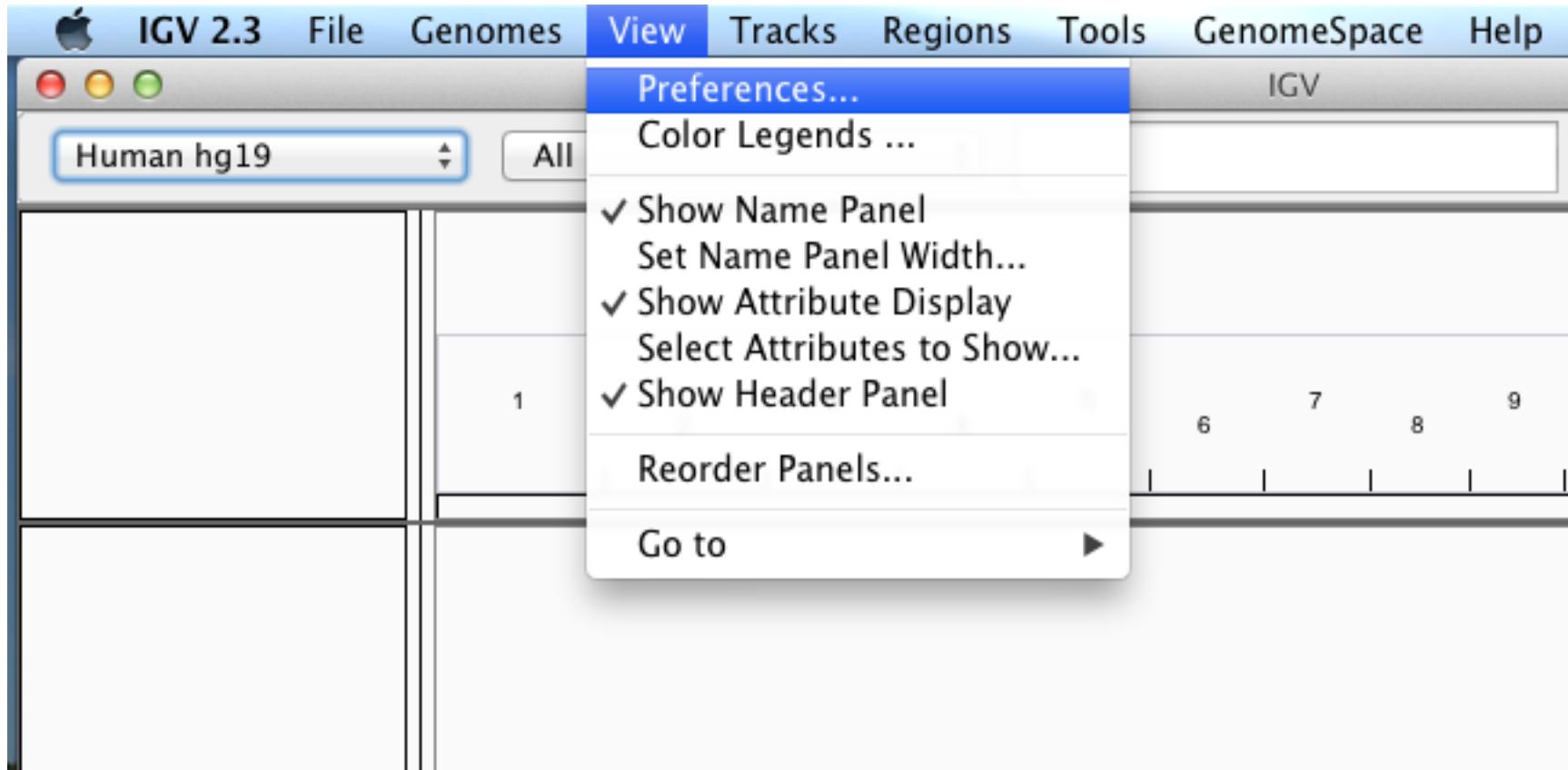
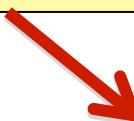


- Step 1: Tune settings for RNA.

RNA-seq alignments



Select View > Preferences...



RNA-seq alignments



Click Alignments tab

The screenshot shows the IGV (Integrative Genomics Viewer) software interface. The title bar says "Human hg18". The top navigation bar has tabs: General, Tracks, Mutations, Chars, Alignments (which is highlighted with a red box and a red arrow pointing to it), Probes, Proxy, Advanced, and IonTorrent. Below the tabs are several configuration sections:

- Visibility range threshold (kb):** 30 (Nominal window size at which alignments become visible)
- Downsample reads:** checked, Max read count: 100, per window size (bases): 50
- Filter and shading options:**
 - Coverage allele-freq threshold: 0.2
 - Mapping quality threshold: 0
 - Filter duplicate reads (checked)
 - Filter vendor failed reads (checked)
 - Filter secondary alignments (unchecked)
 - Flag unmapped pairs (unchecked)
 - Shade mismatched bases by quality: 5 to 20 (checked)
 - Show center line (checked)
 - Show coverage track (checked)
 - Show soft-clipped bases (unchecked)
 - Flag zero-quality alignments (checked)
- Splice Junction Track Options:**
 - Show junction track (unchecked)
 - Min flanking width: 0
 - Min junction coverage: 1
 - Show flanking regions (checked)
- Insert Size Options:**

These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.

Defaults	Minimum (bp): 50	Compute	Minimum (percentile): 0.5
	Maximum (bp): 1000		Maximum (percentile): 99.5

At the bottom, there are "OK" and "Cancel" buttons. The status bar at the bottom left says "5 tracks loaded" and "chr1:159,464,348". The status bar at the bottom right says "386M of 866M".

RNA-seq alignments



The screenshot shows the IGV software interface for Human hg18. The main window displays a genomic track for chromosome 15, showing a sequence of T A G G A with a G highlighted. The left sidebar shows tracks for 'Sequence' and 'RefSeq genes'. The top menu bar includes General, Tracks, Mutations, Charts, Alignments (selected), Probes, Proxy, Advanced, and IonTorrent. The Alignments tab contains several configuration options:

- Visibility range threshold (kb): 500 (Nominal window size at which alignments become visible)
- Downsample reads: checked, Max read count: 100, per window size (bases): 50
- Filter and shading options:
 - Coverage allele-freq threshold: 0.2
 - Mapping quality threshold: 0
 - Filter duplicate reads (checked)
 - Filter vendor failed reads (checked)
 - Filter secondary alignments (unchecked)
 - Flag unmapped pairs (unchecked)
 - Shade mismatched bases by quality: 5 to 20 (checked)
 - Show center line (checked)
 - Show coverage track (checked)
 - Show soft-clipped bases (unchecked)
 - Flag zero-quality alignments (checked)
- Splice Junction Track Options:
 - Show junction track (checked, highlighted with a red box and arrow)
 - Show flanking regions (unchecked)
- Insert Size Options:

These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.

Defaults	Minimum (bp): 50	Compute	Minimum (percentile): 0.5
	Maximum (bp): 1000		Maximum (percentile): 99.5

At the bottom, status bars indicate: 5 tracks loaded, chr1:159,464,348, and 386M of 866M.

Select Show junction track

RNA-seq alignments



IGV

Human hg18

General | Tracks | Mutations | Charts | Alignments **Alignments** | Probes | Proxy | Advanced | IonTorrent

Visibility range threshold (kb): 500 Nominal window size at which alignments become visible

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

Filter and shading options

Coverage allele-freq threshold: 0.2 Mapping quality threshold: 0

Filter duplicate reads Show center line

Filter vendor failed reads Show coverage track

Filter secondary alignments Show soft-clipped bases

Flag unmapped pairs Flag zero-quality alignments

Shade mismatched bases by quality: 5 to 20

Flag insertions larger than: bases

Filter alignments by read group URL or path to filter file

Splice Junction Track Options

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

Show flanking regions

Insert Size Options

These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.

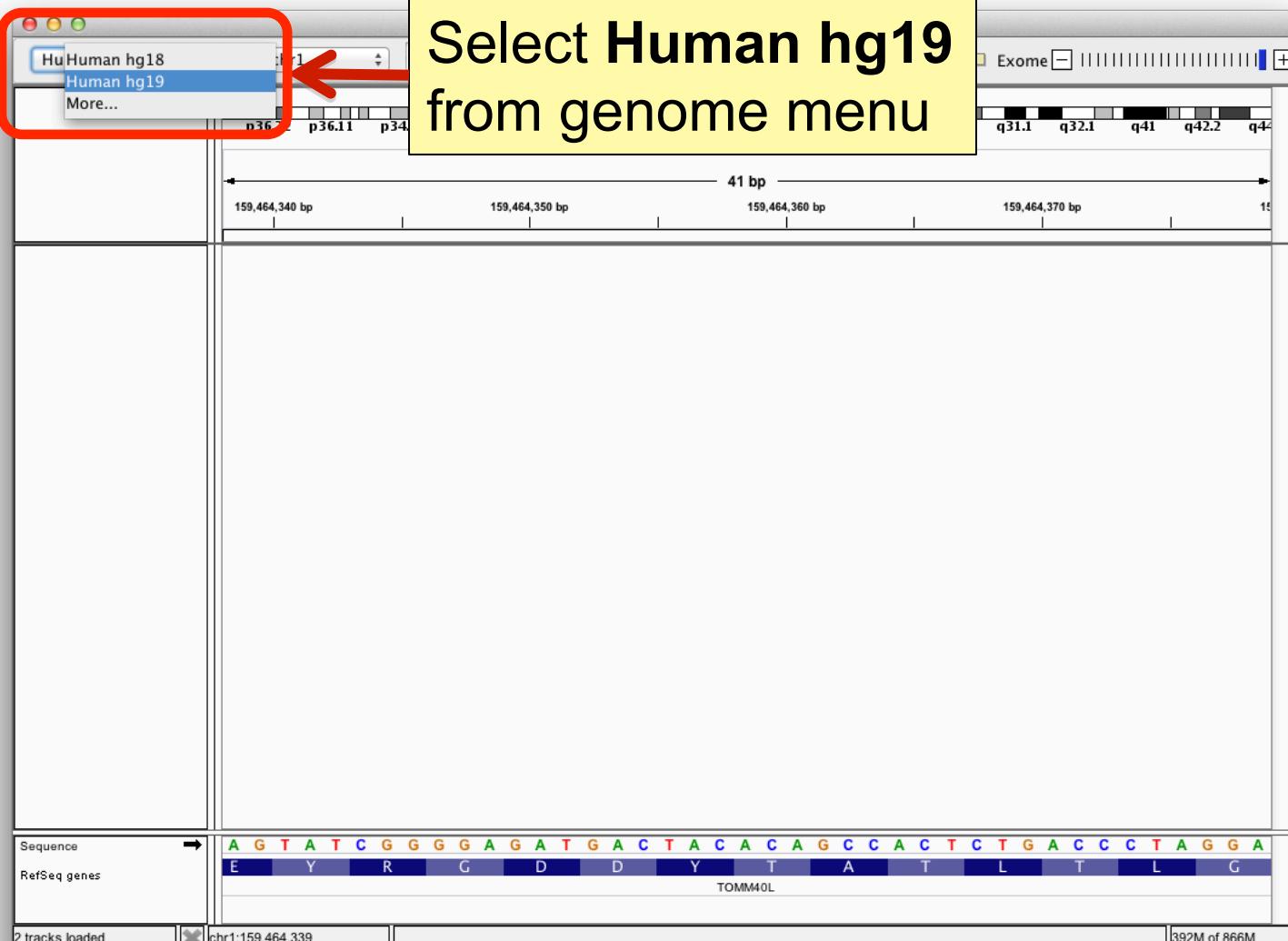
Defaults Minimum (bp): 50 Compute Minimum (percentile): 0.5
Maximum (bp): 1000 Maximum (percentile): 99.5

OK Cancel

5 tracks loaded 386M of 866M

Click OK to save changes

RNA-seq alignments

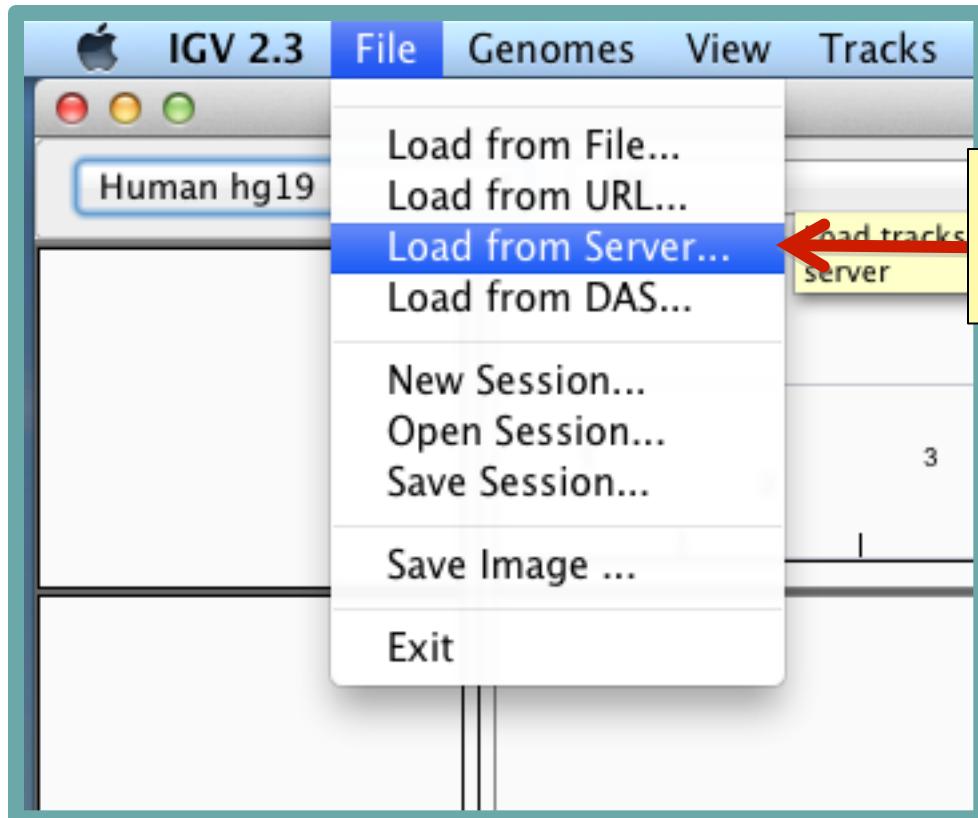


Select Human hg19 from genome menu

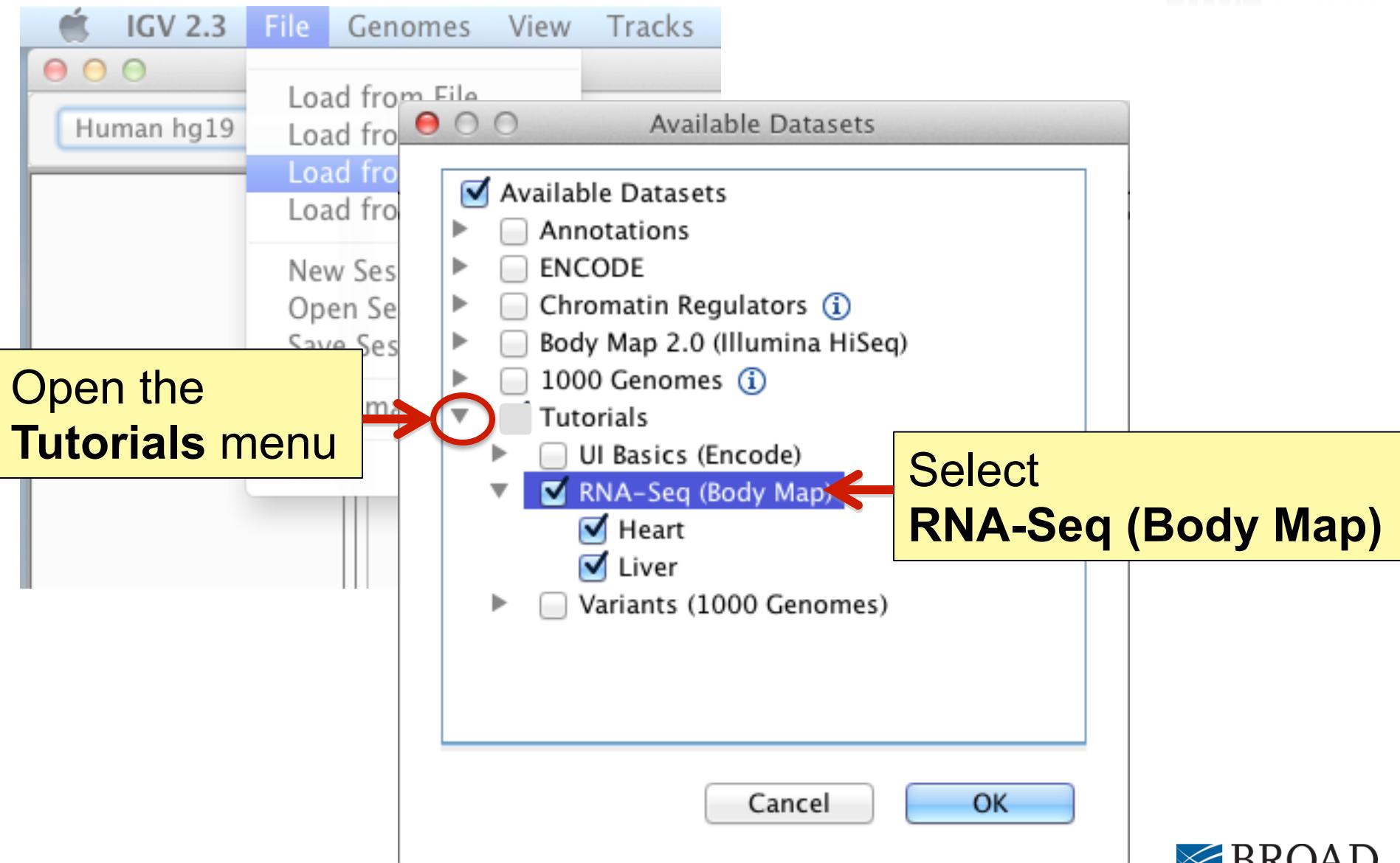
The screenshot shows the IGV interface with the following details:

- Genome Selection:** A red box highlights the genome menu at the top left, with "Human hg19" selected (indicated by a blue background). A yellow box with a red arrow points to this selection.
- Chromosome View:** The main panel displays chromosome 15 with a zoomed-in view of the p36 band. The scale bar indicates 41 bp. Reference coordinates are 159,464,340 bp, 159,464,350 bp, 159,464,360 bp, and 159,464,370 bp.
- Exome View:** A track labeled "Exome" is visible on the right side of the main panel.
- Sequence View:** At the bottom, a sequence track for the TOMM40L gene is shown. The sequence is: A G T A T C G G G A G A T G A C T A C A C A G C C A C T C T G A C C C T A G G A. Below the sequence, the gene name "TOMM40L" is labeled.
- Status Bar:** The bottom status bar shows "2 tracks loaded" and the genomic coordinate "chr1:159,464,339". It also indicates the total indexed data size as "392M of 866M".

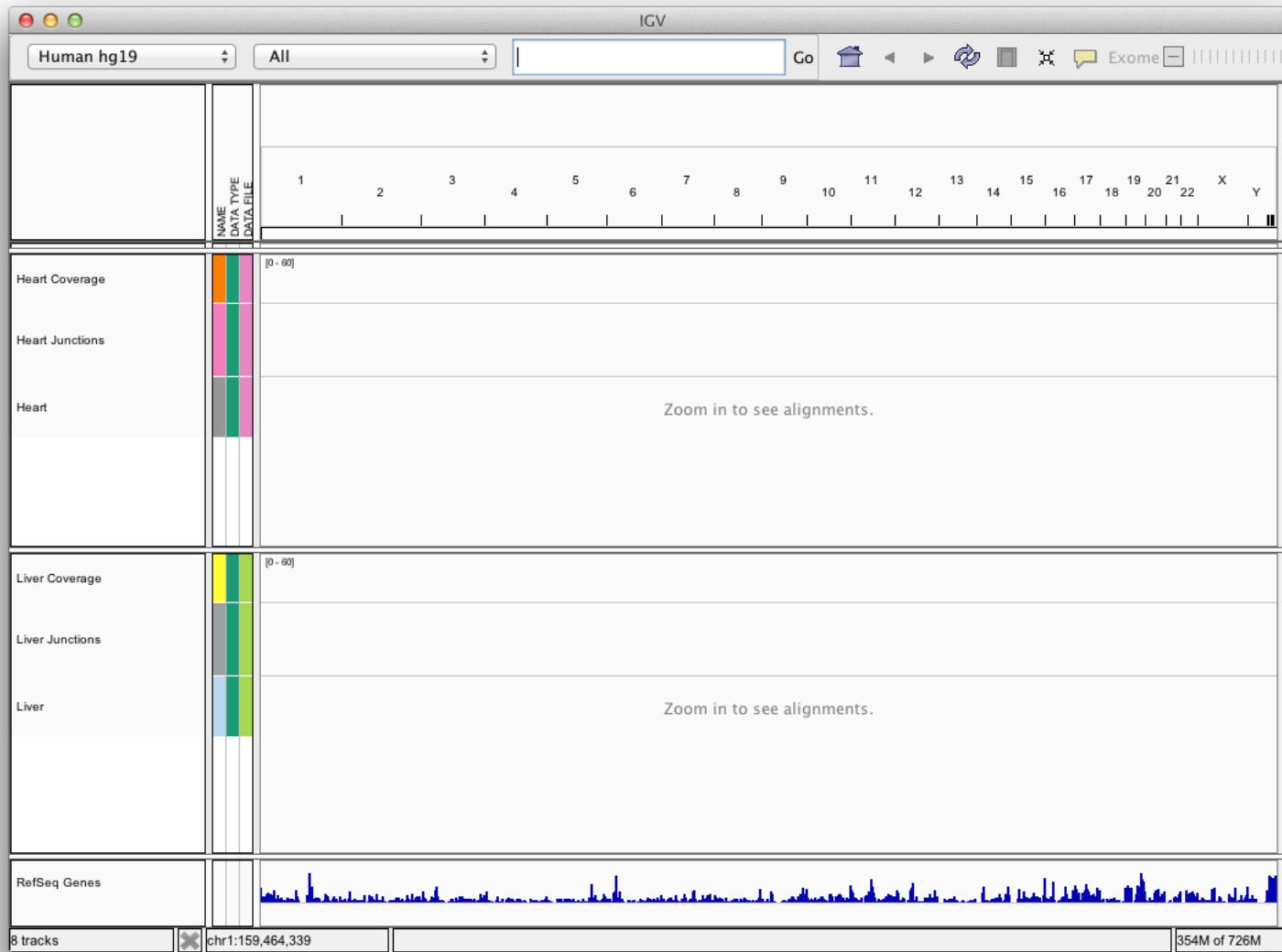
RNA-seq alignments



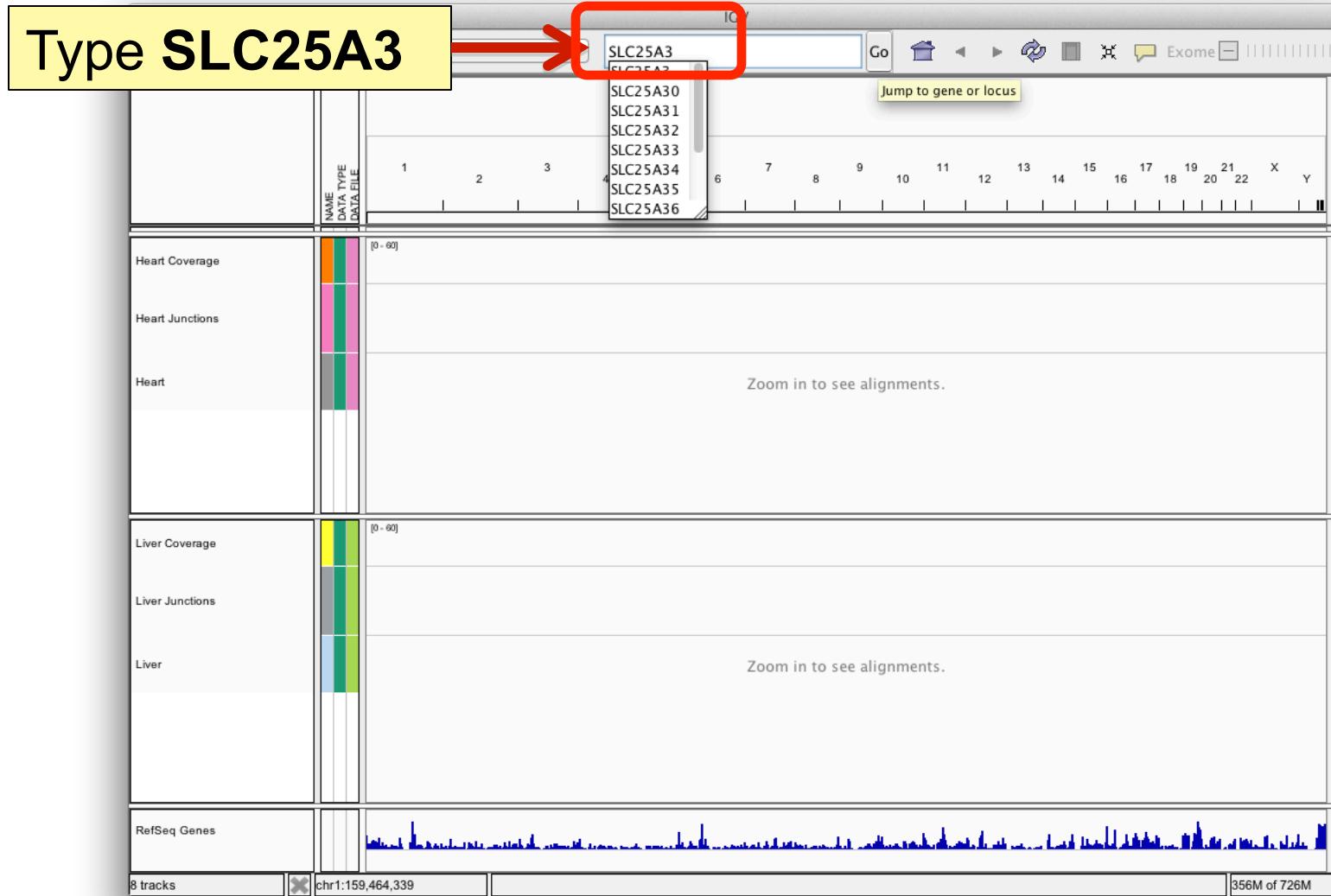
RNA-seq alignments



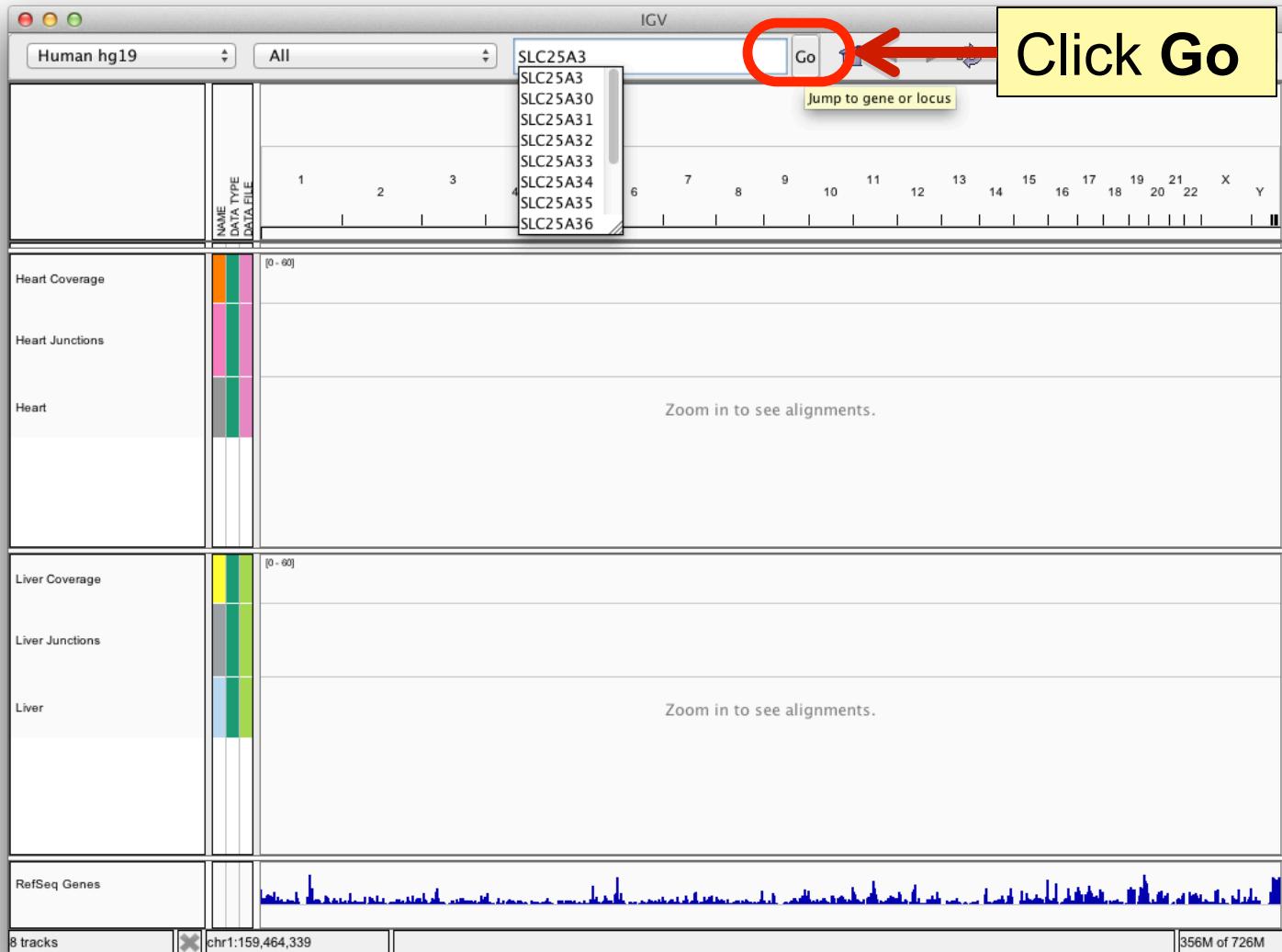
RNA-seq alignments



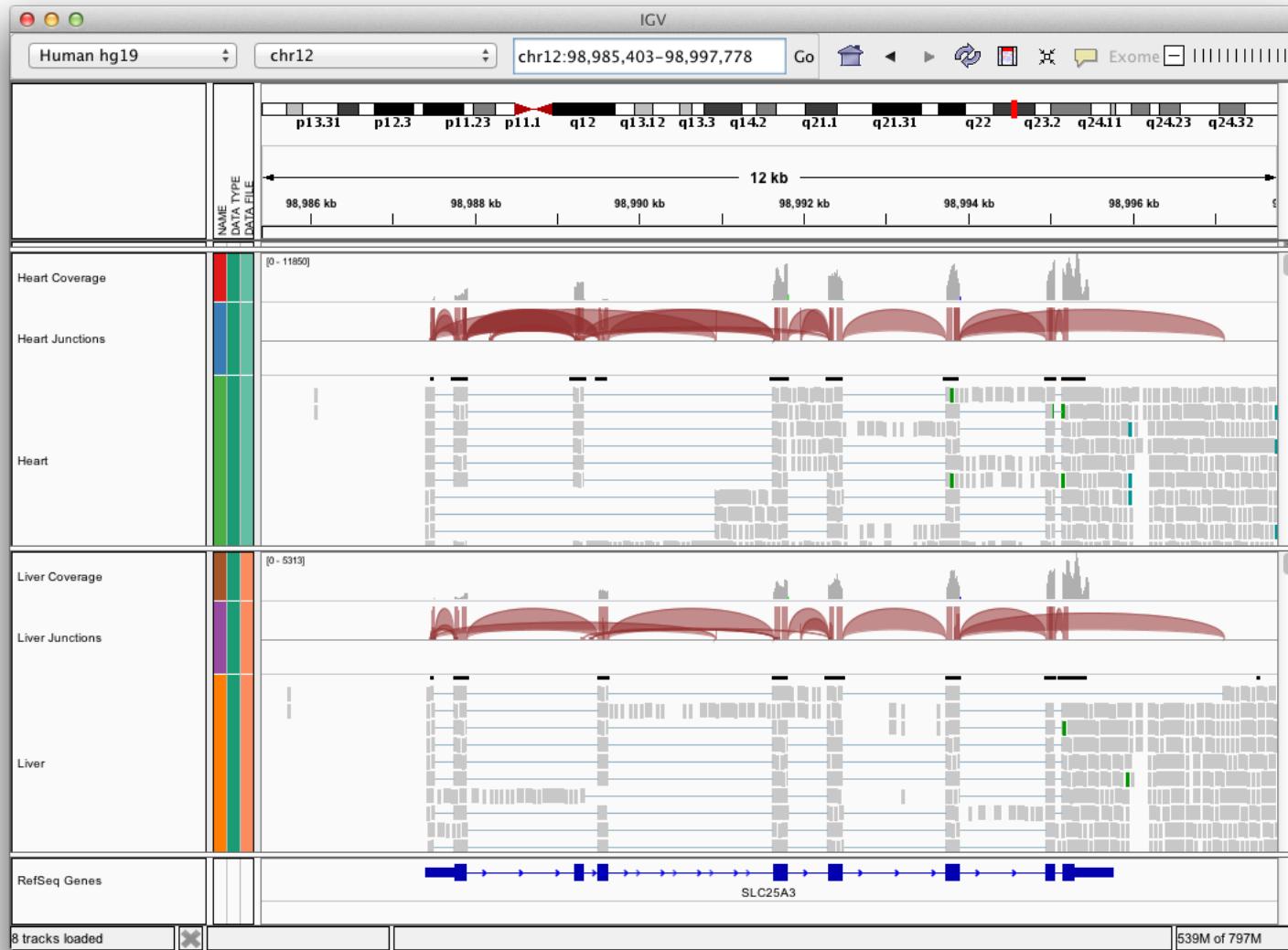
RNA-seq alignments



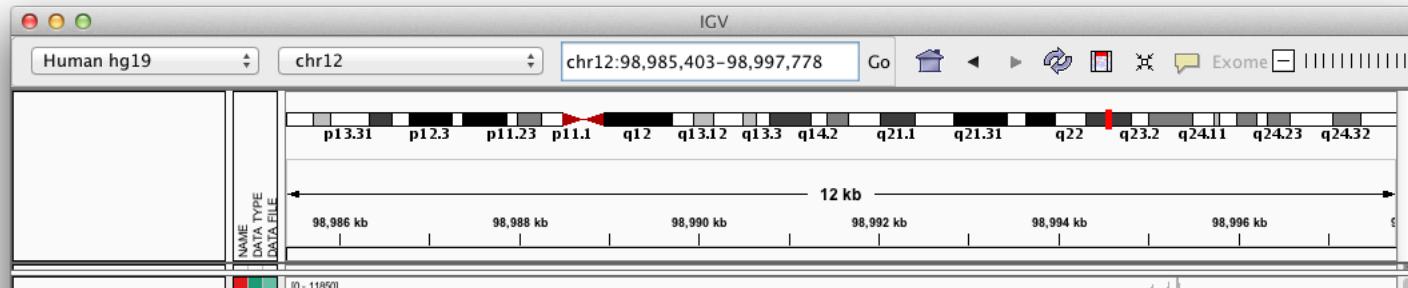
RNA-seq alignments



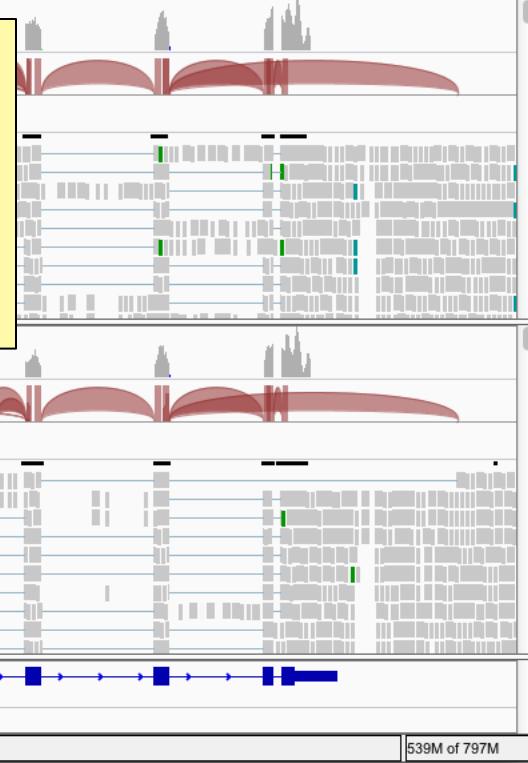
RNA-seq alignments



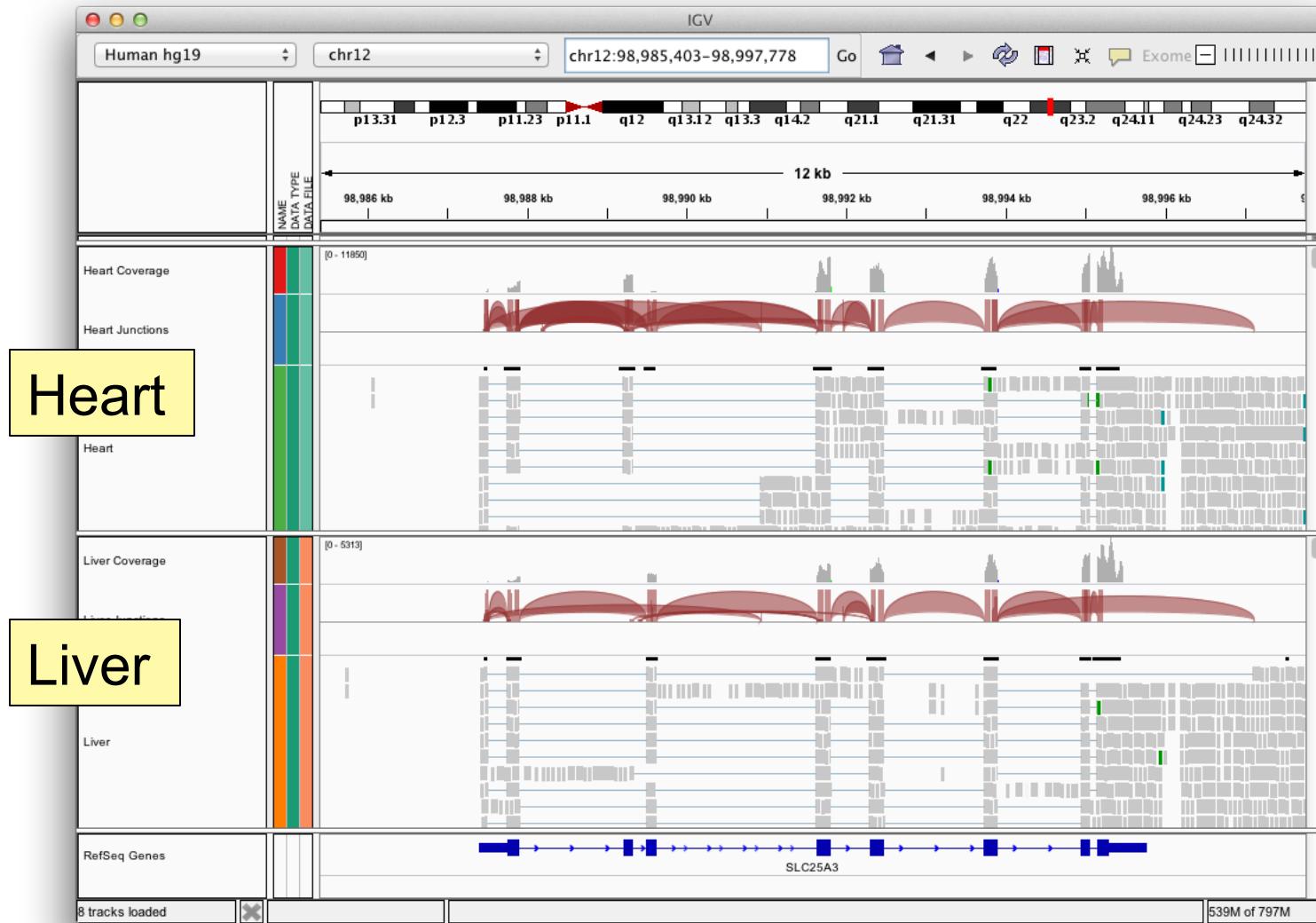
RNA-seq alignments



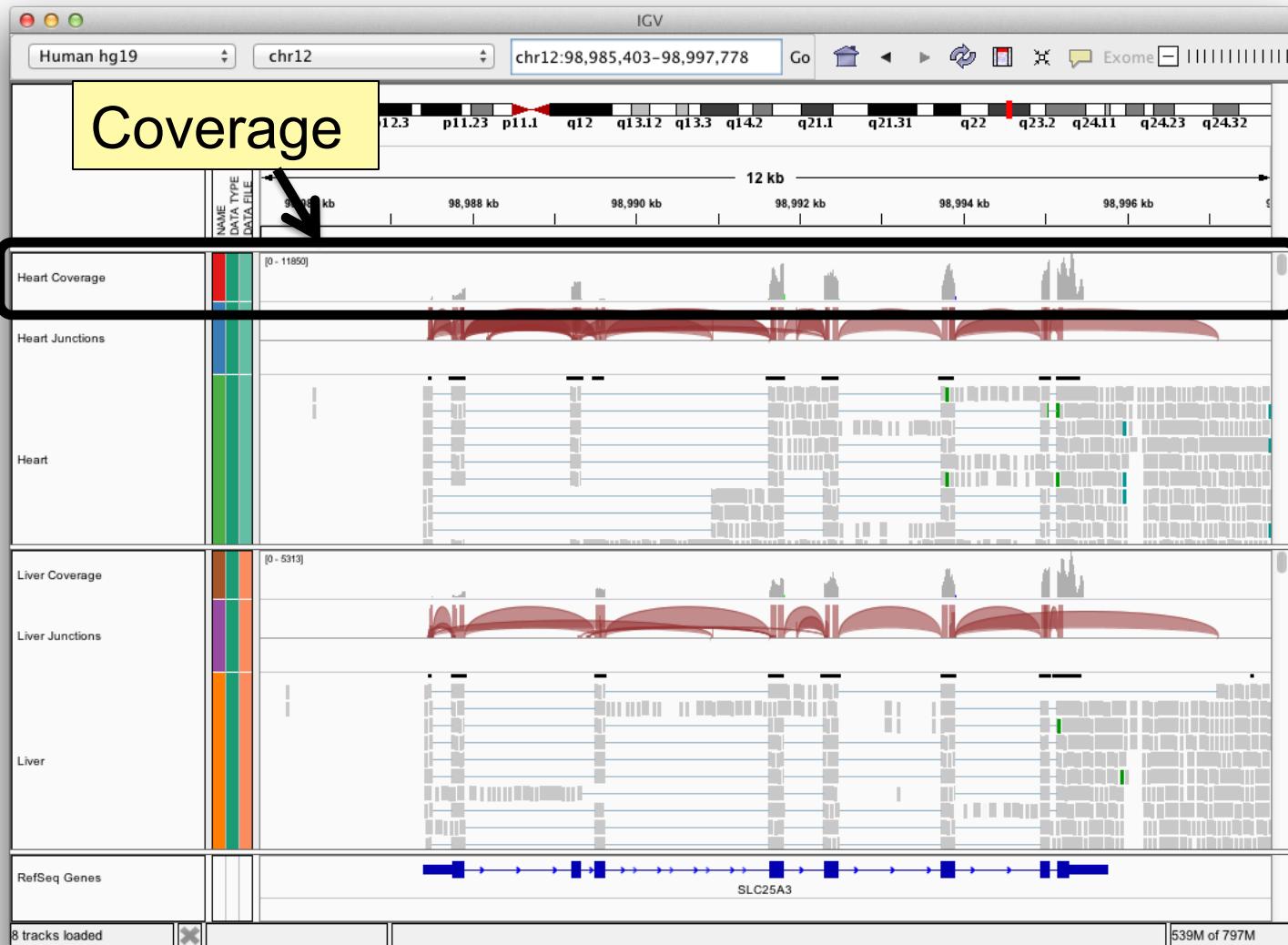
If reads are still blue & red from the settings for the last exercise, then right-click and select **Color alignments by > no color**



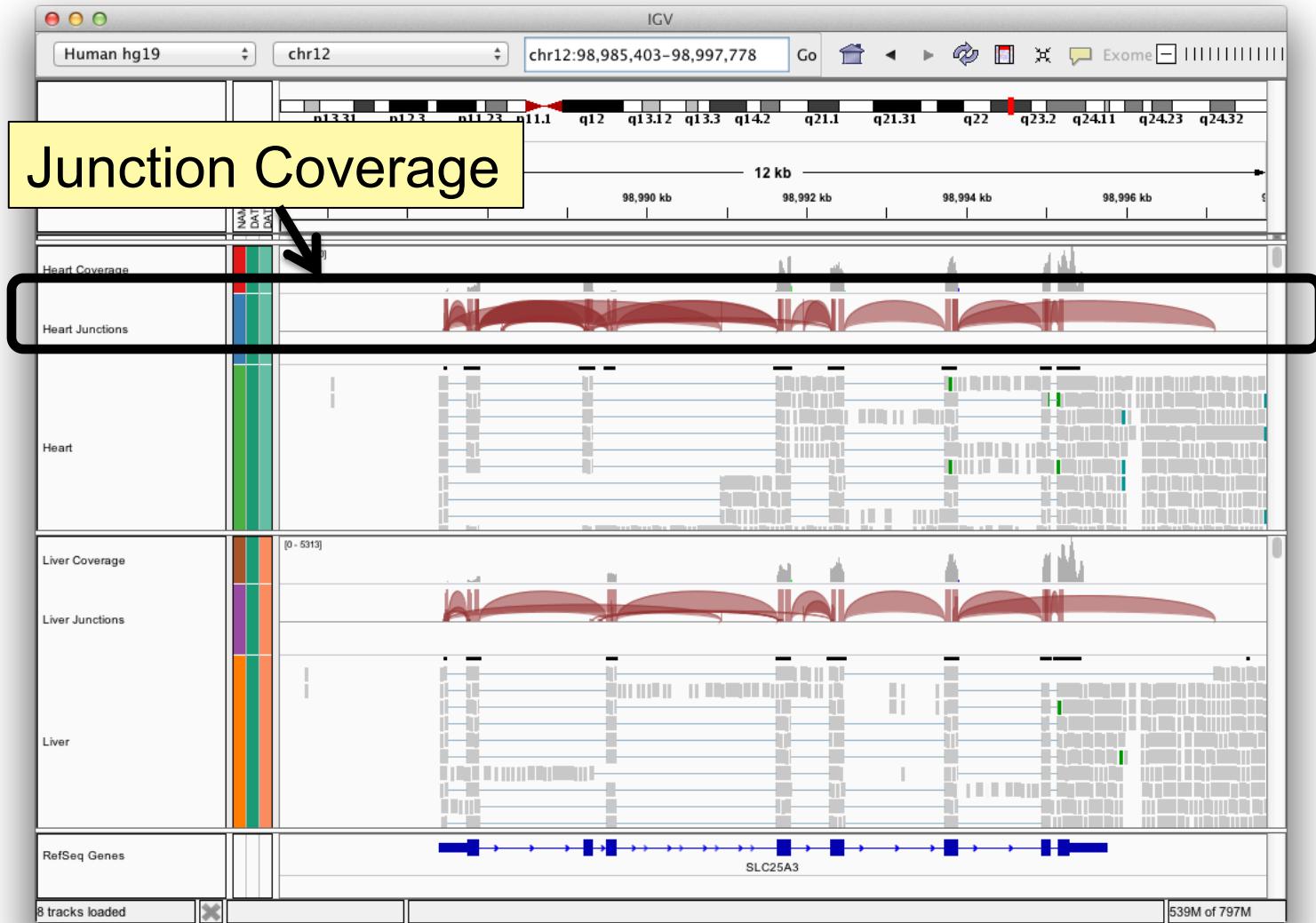
RNA-seq alignments



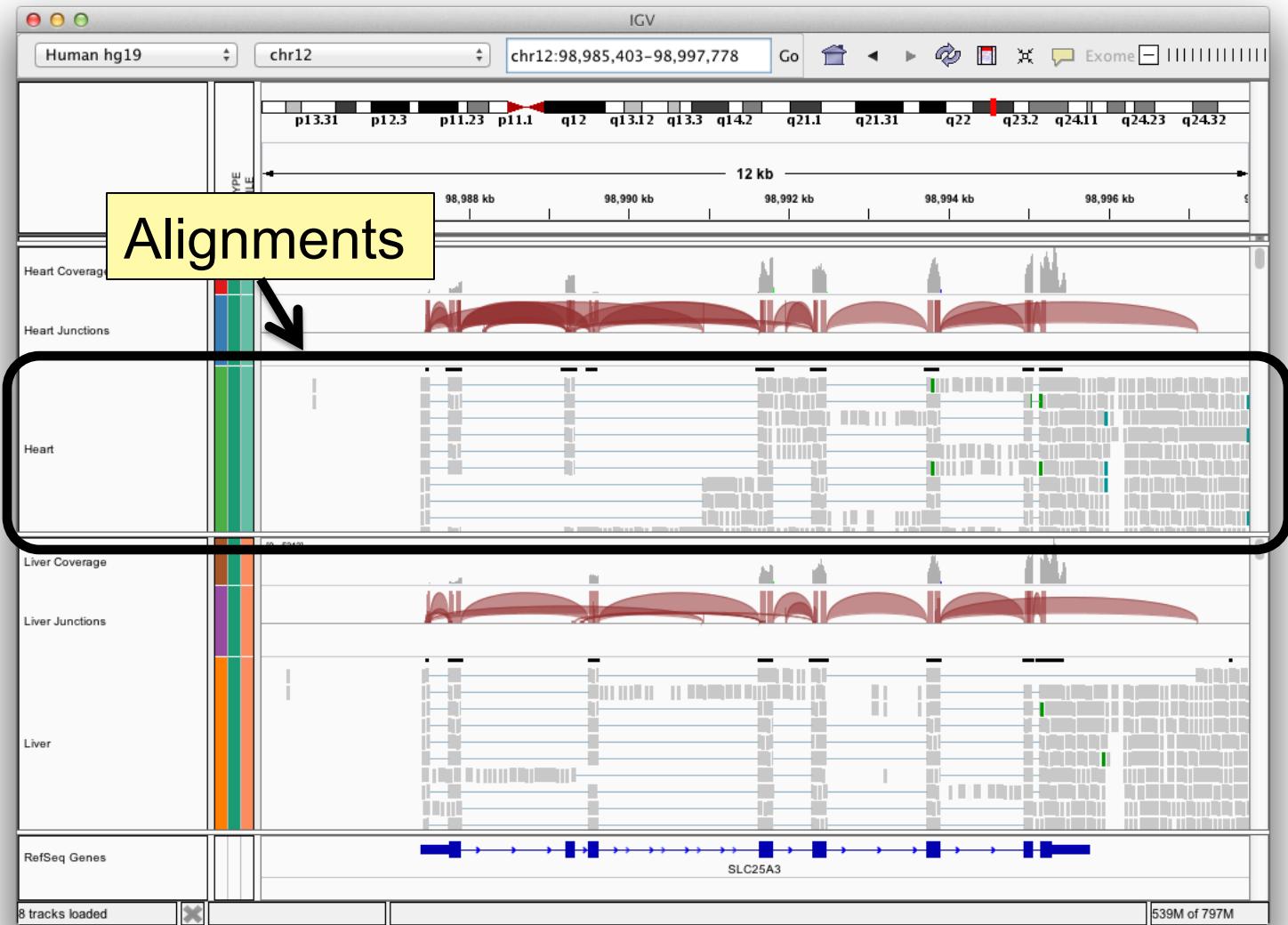
RNA-seq alignments



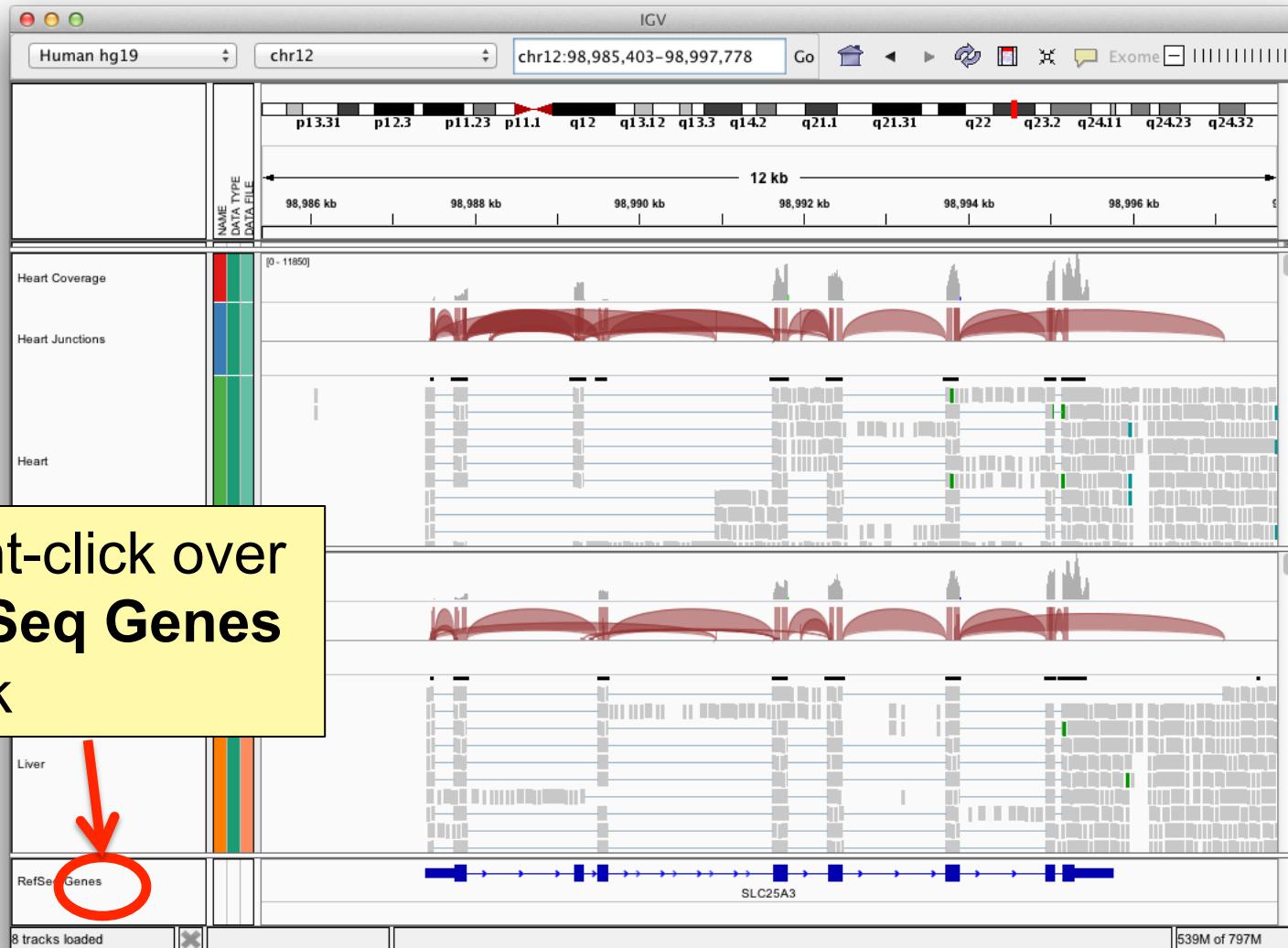
RNA-seq alignments



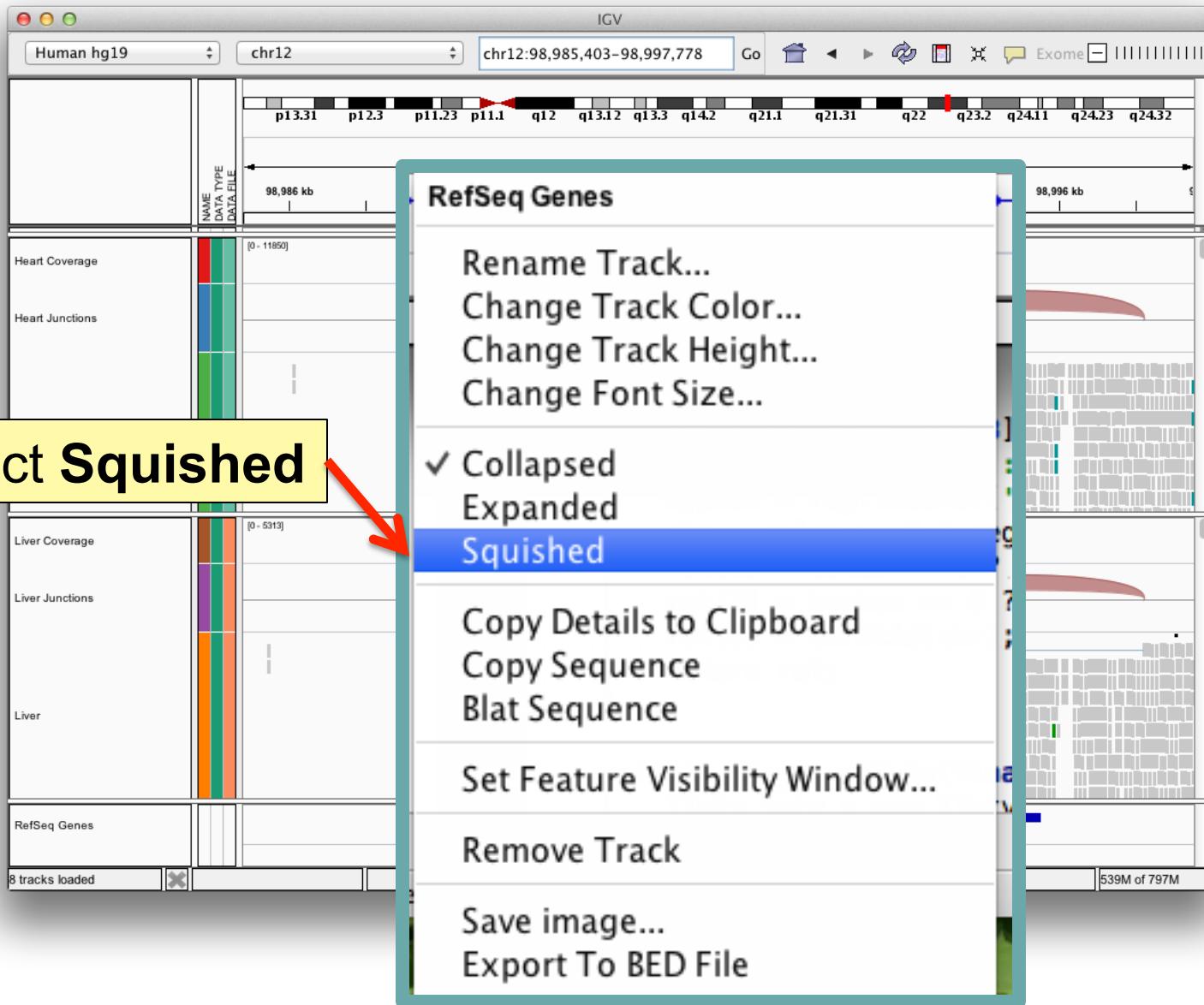
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments

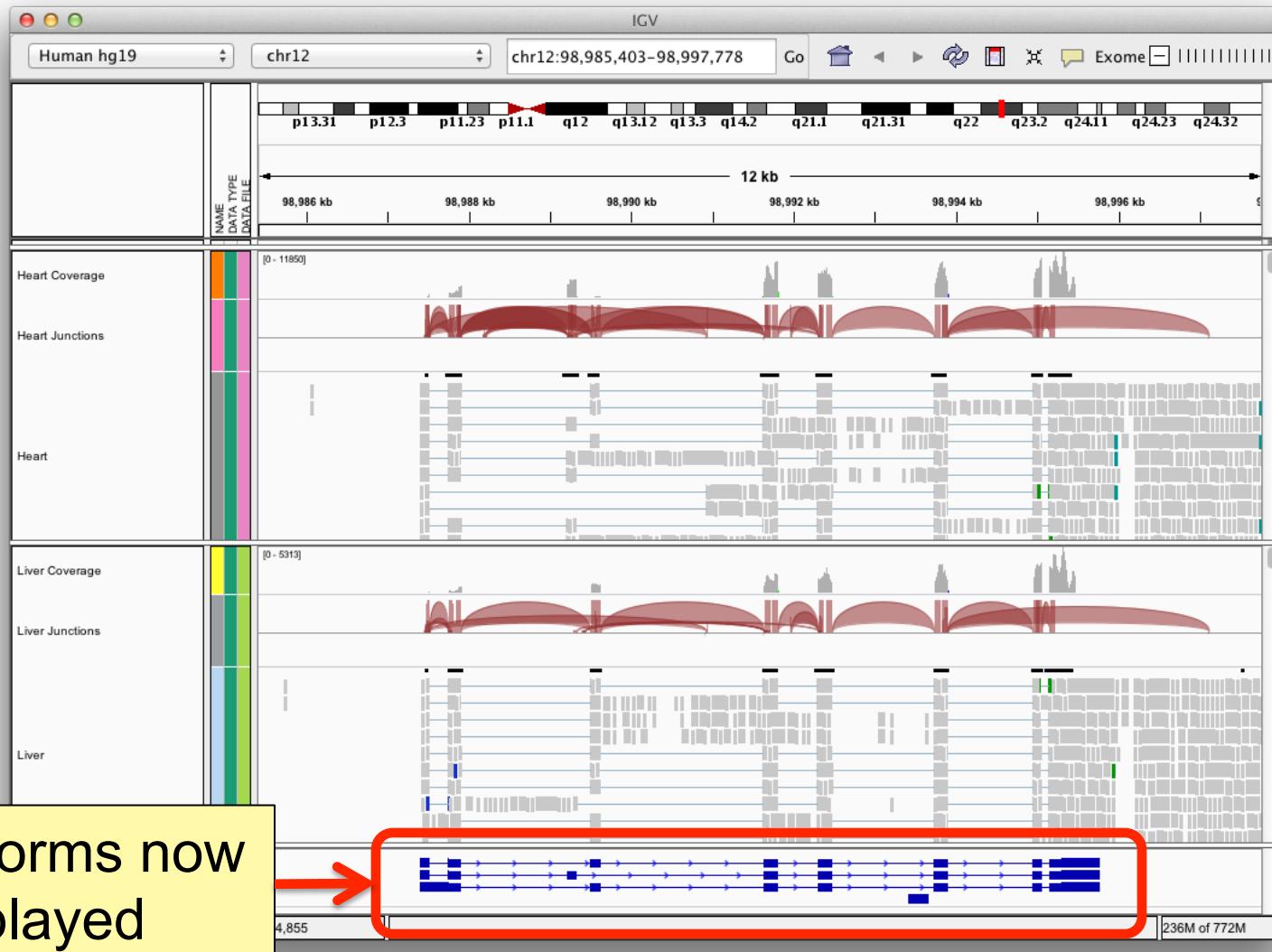


Select Squished

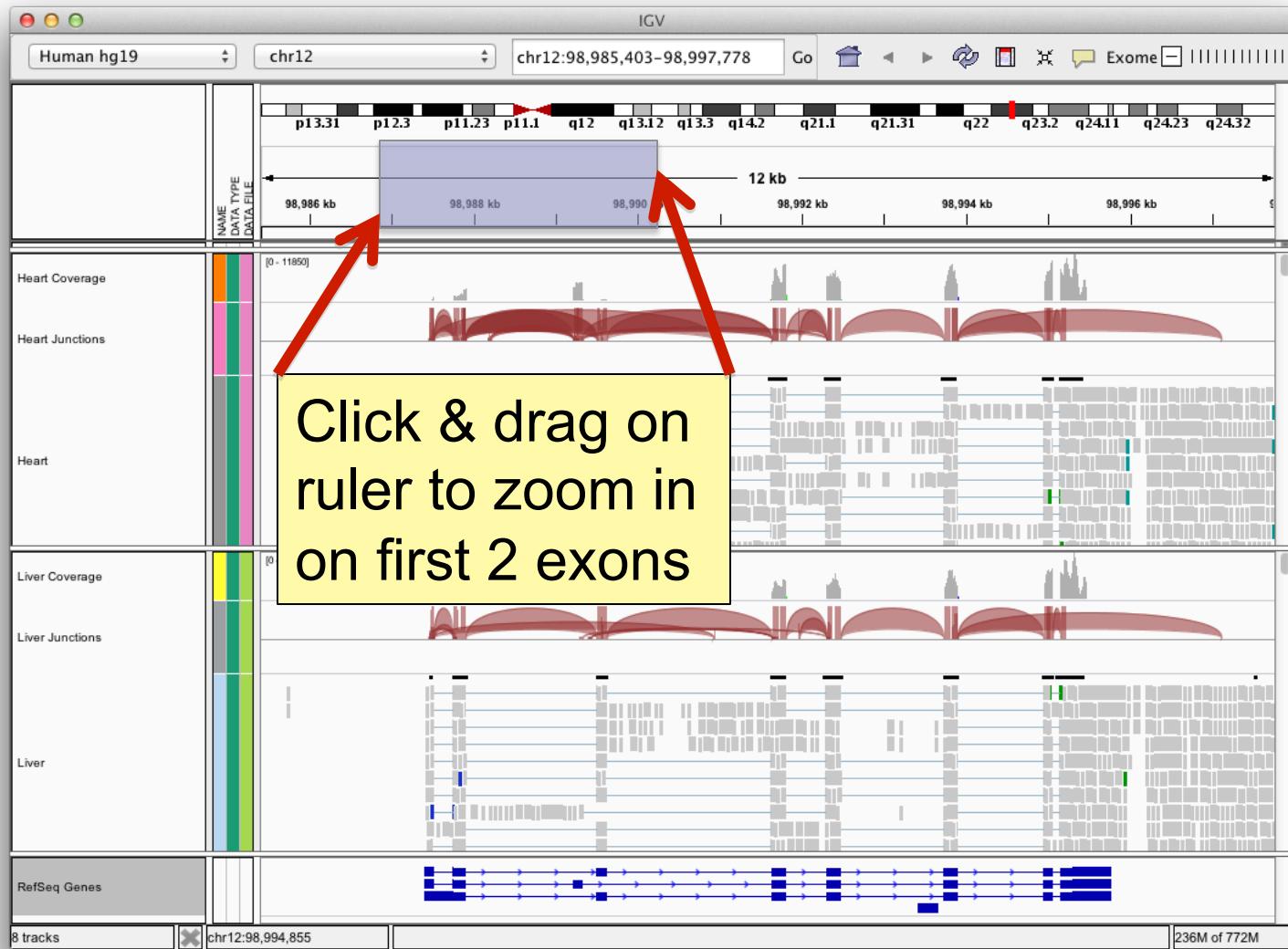
RefSeq Genes

- Rename Track...
- Change Track Color...
- Change Track Height...
- Change Font Size...
- Collapsed
- Expanded
- Squished**
- Copy Details to Clipboard
- Copy Sequence
- Blat Sequence
- Set Feature Visibility Window...
- Remove Track
- Save image...
- Export To BED File

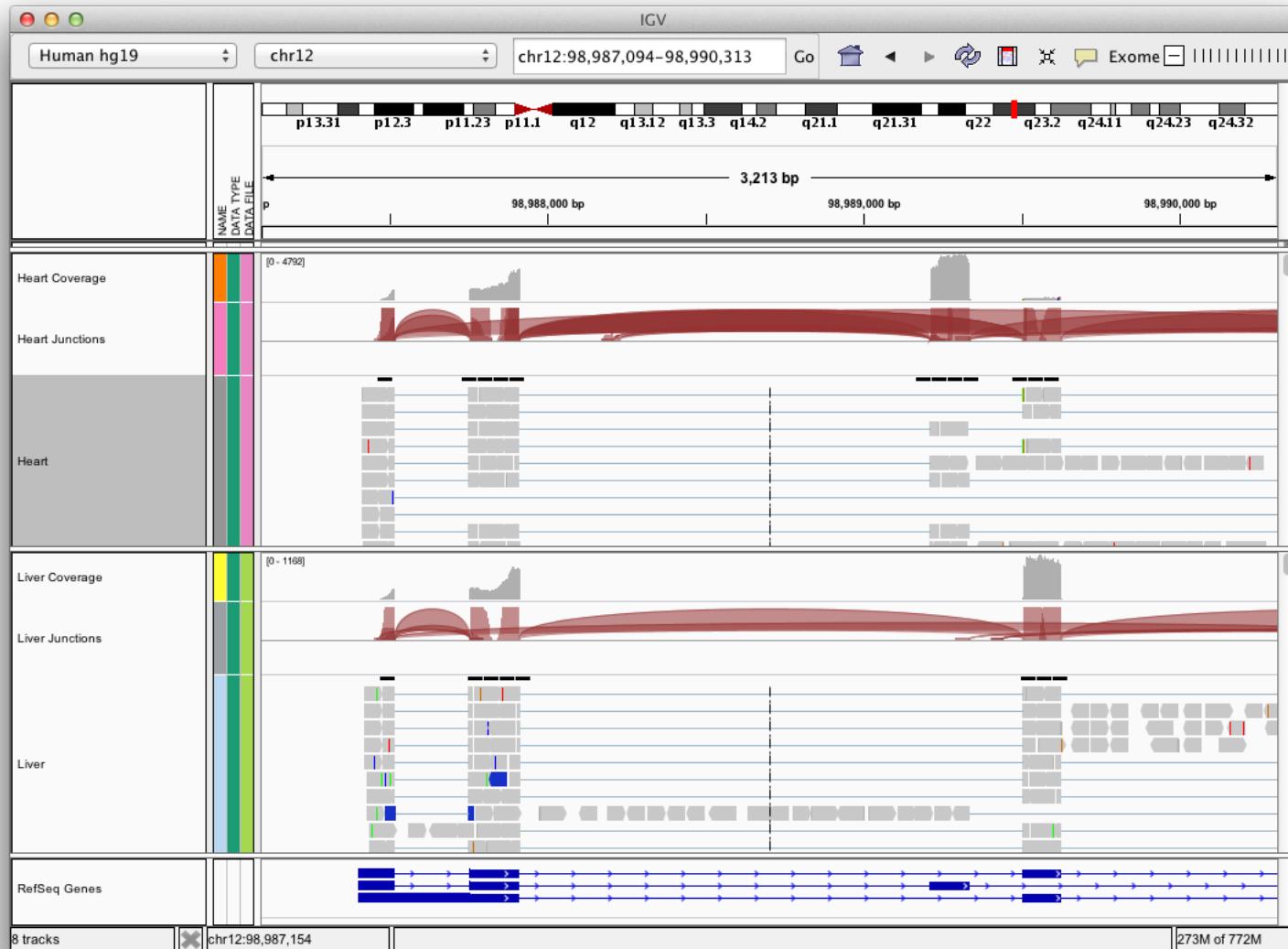
RNA-seq alignments



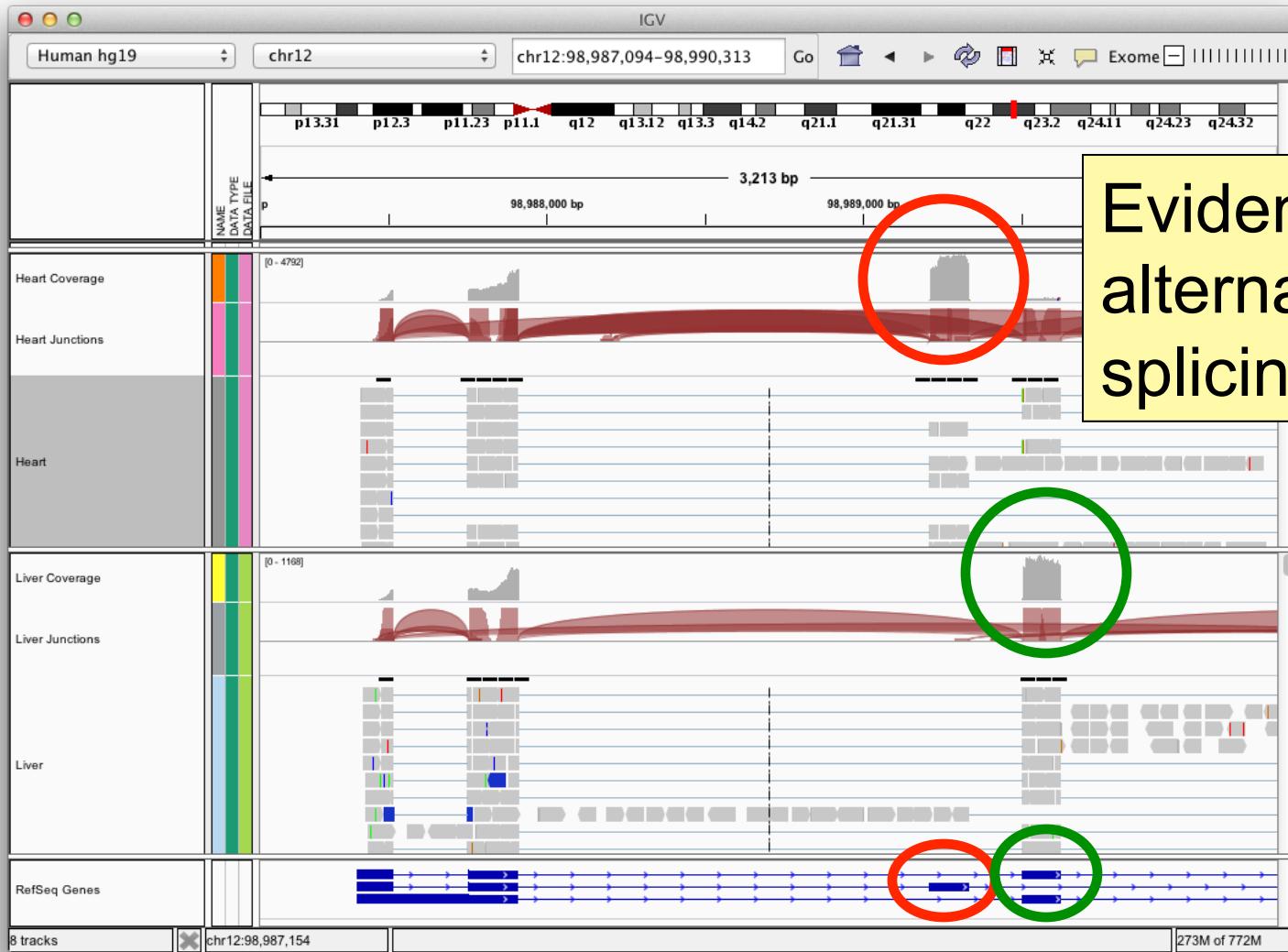
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments



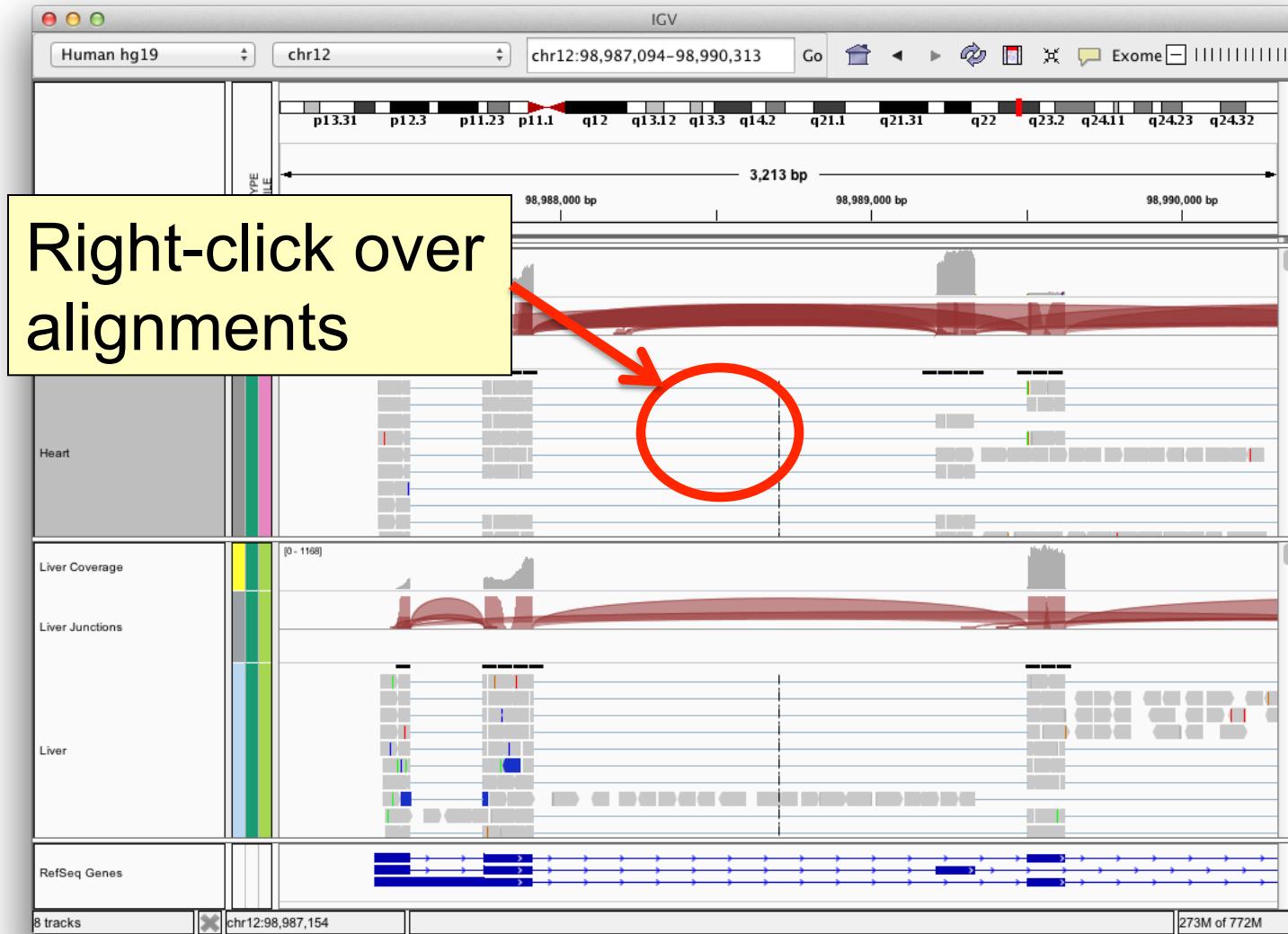
Sashimi plot

Viewing RNA splicing with Sashimi Plots

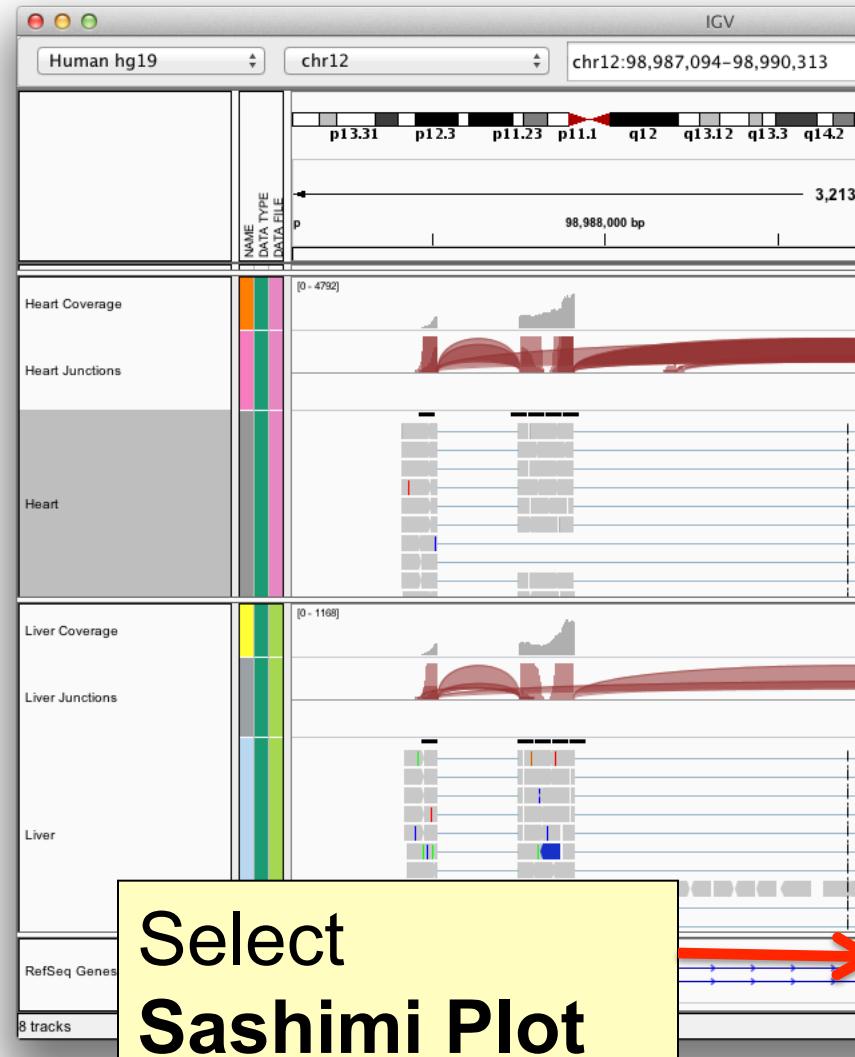
Reference: Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airoldi EM, Burge, CB.

Sashimi plots: Quantitative visualization of RNA sequencing read alignments. arXiv:1306.3466 [q-bio.GN], 2013

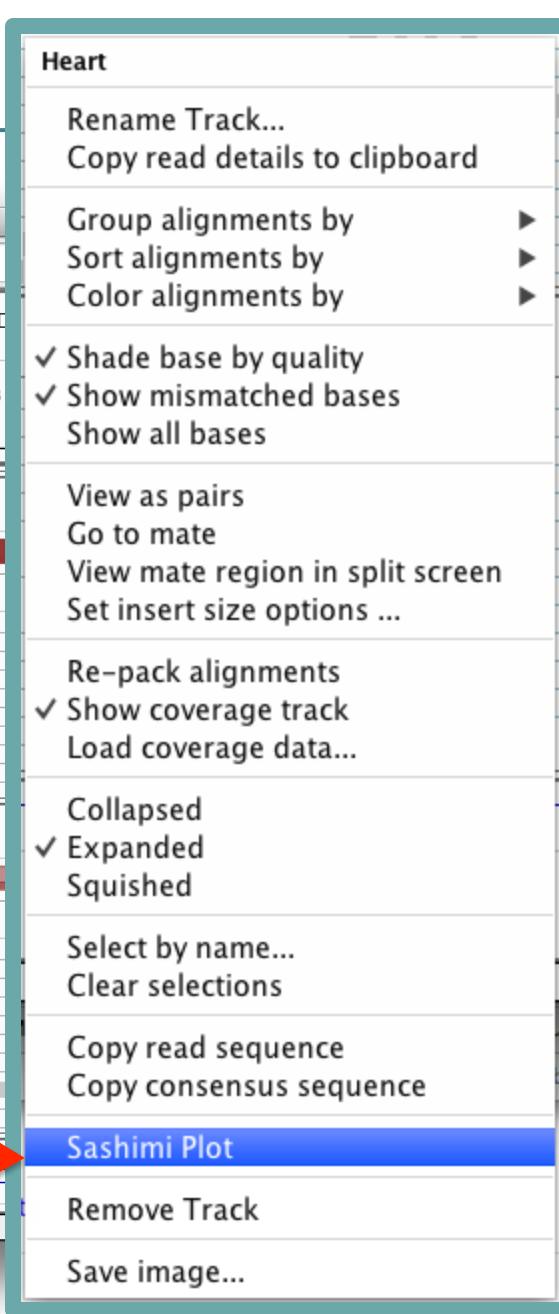
RNA-seq alignments



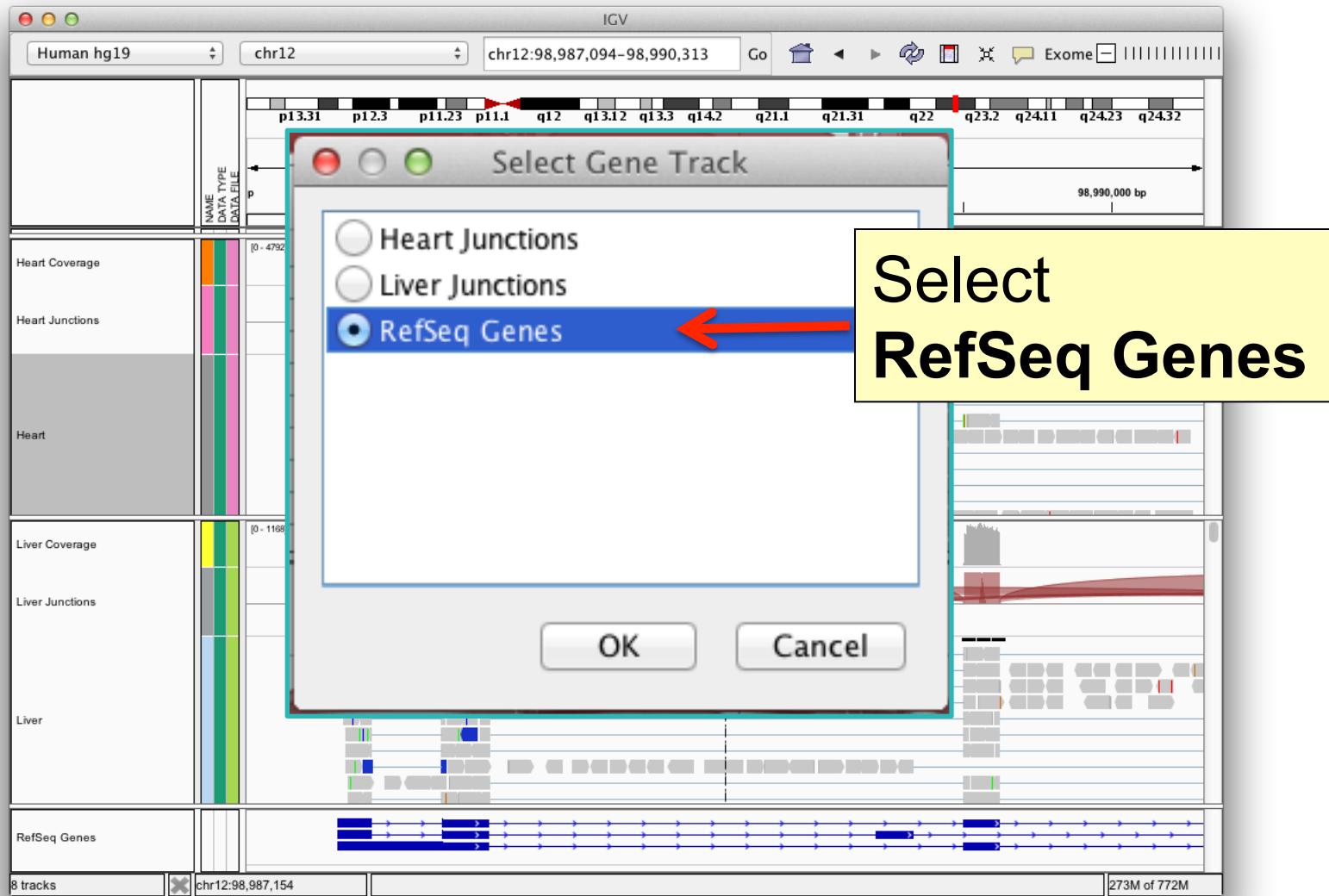
RNA-seq alignments



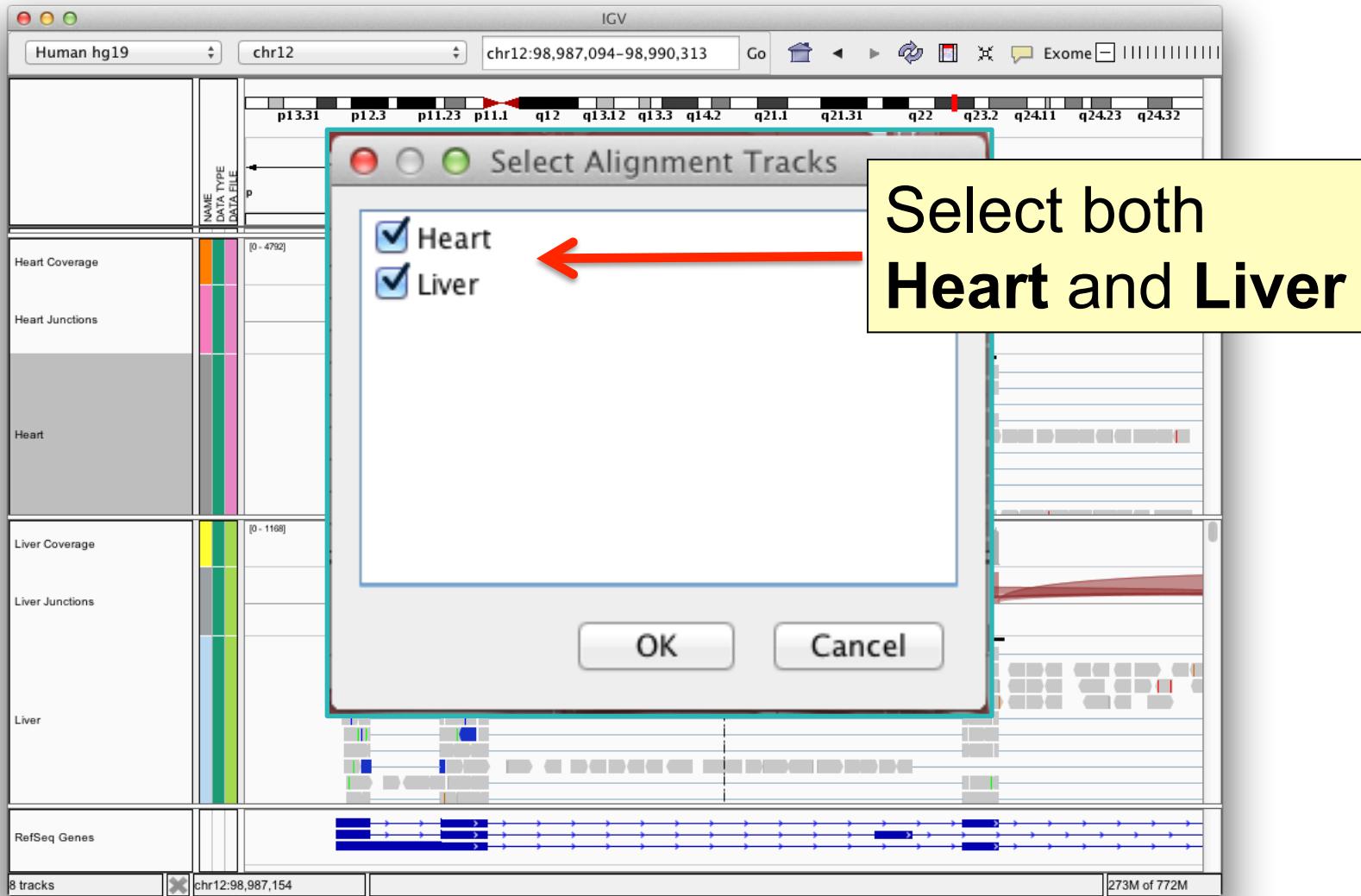
Select
Sashimi Plot



RNA-seq alignments



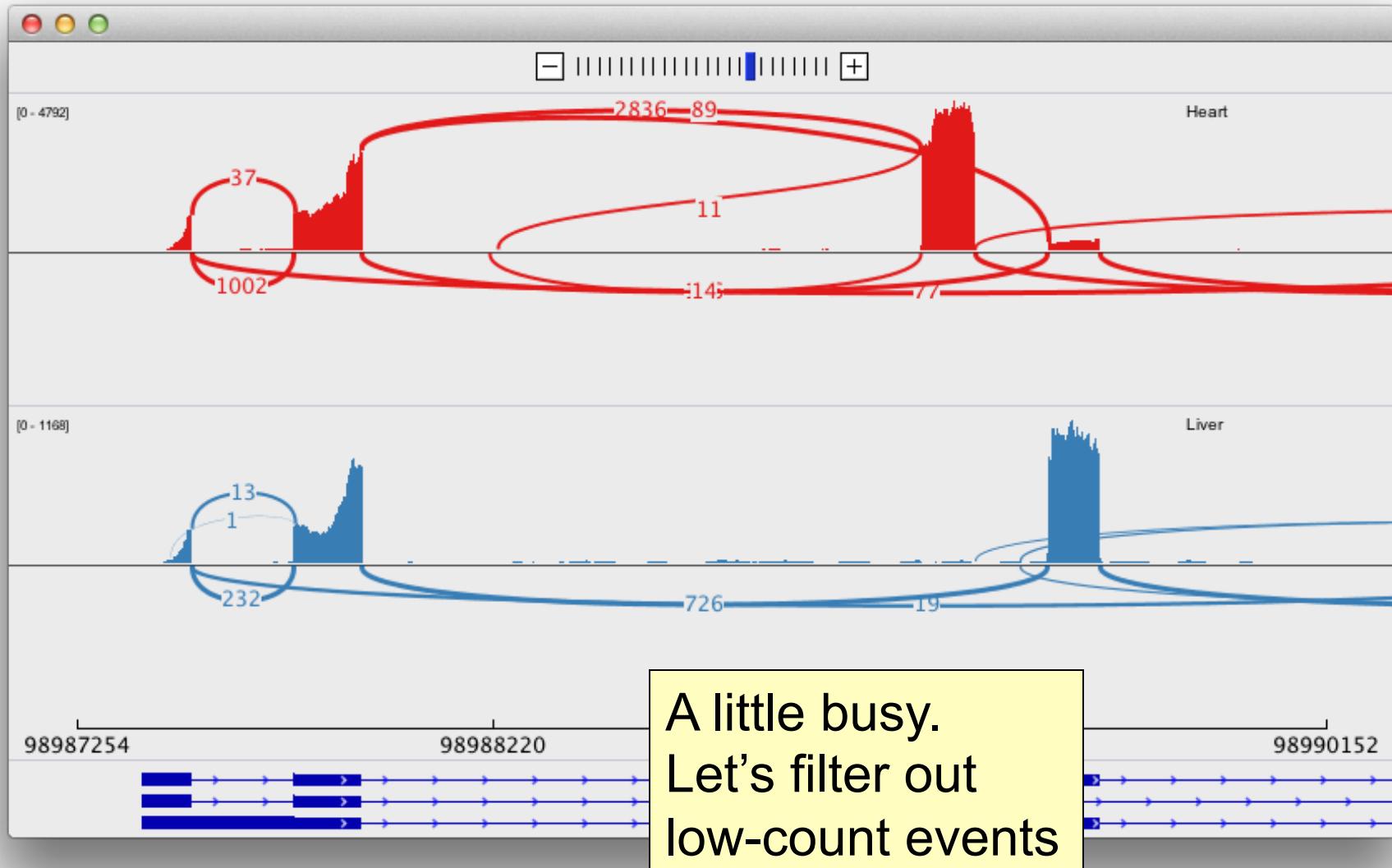
RNA-seq alignments



RNA-seq alignments



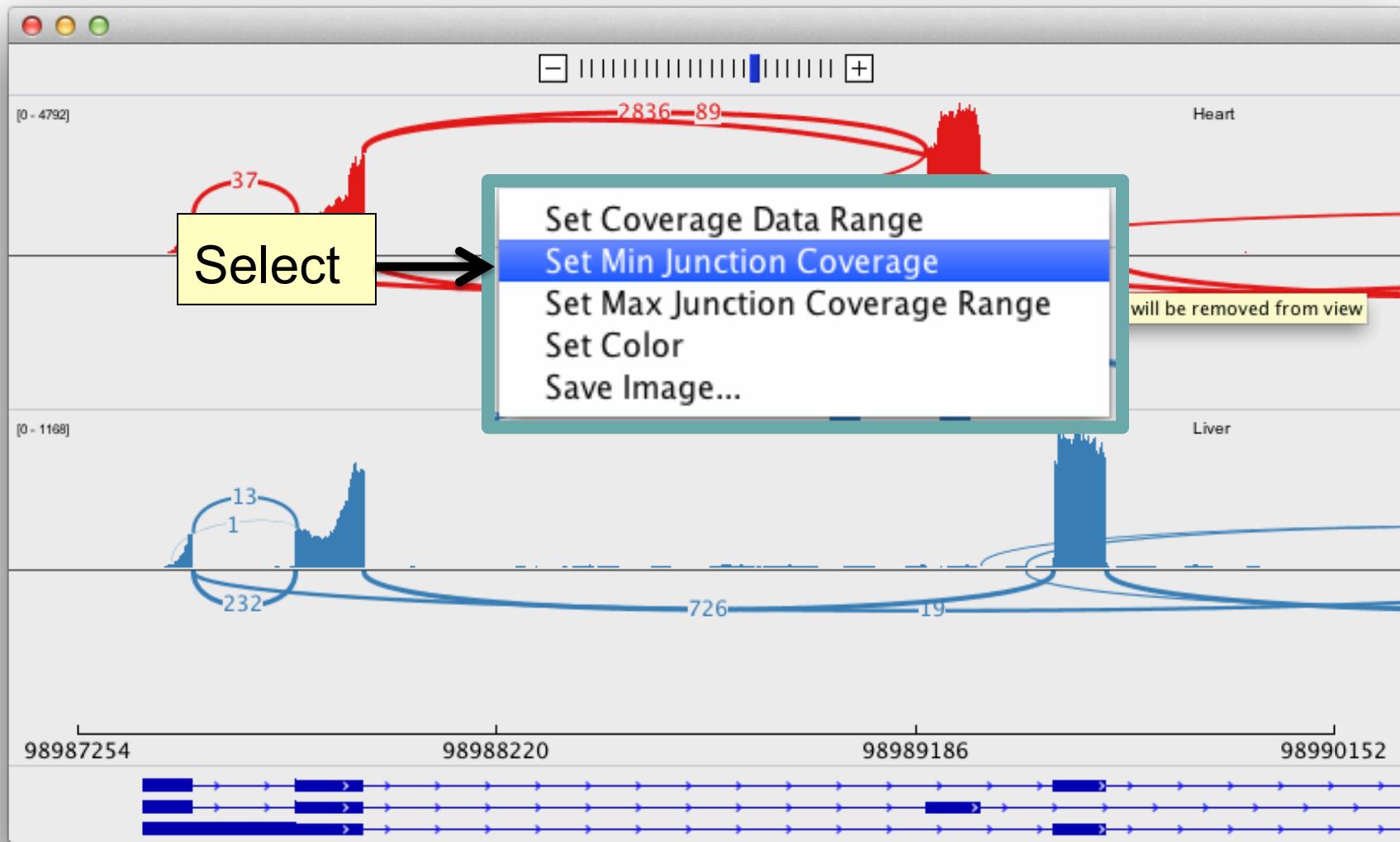
RNA-seq alignments



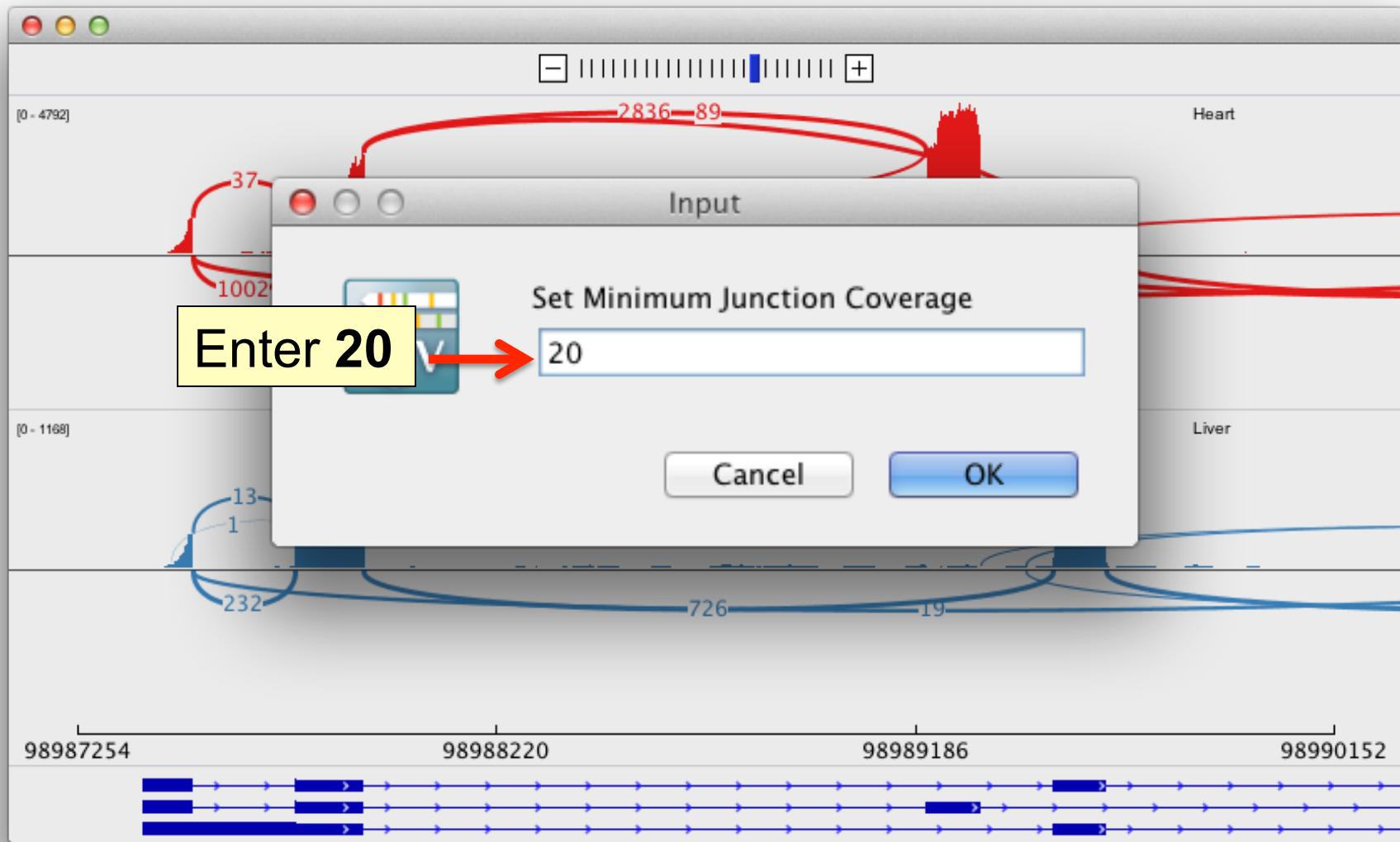
RNA-seq alignments



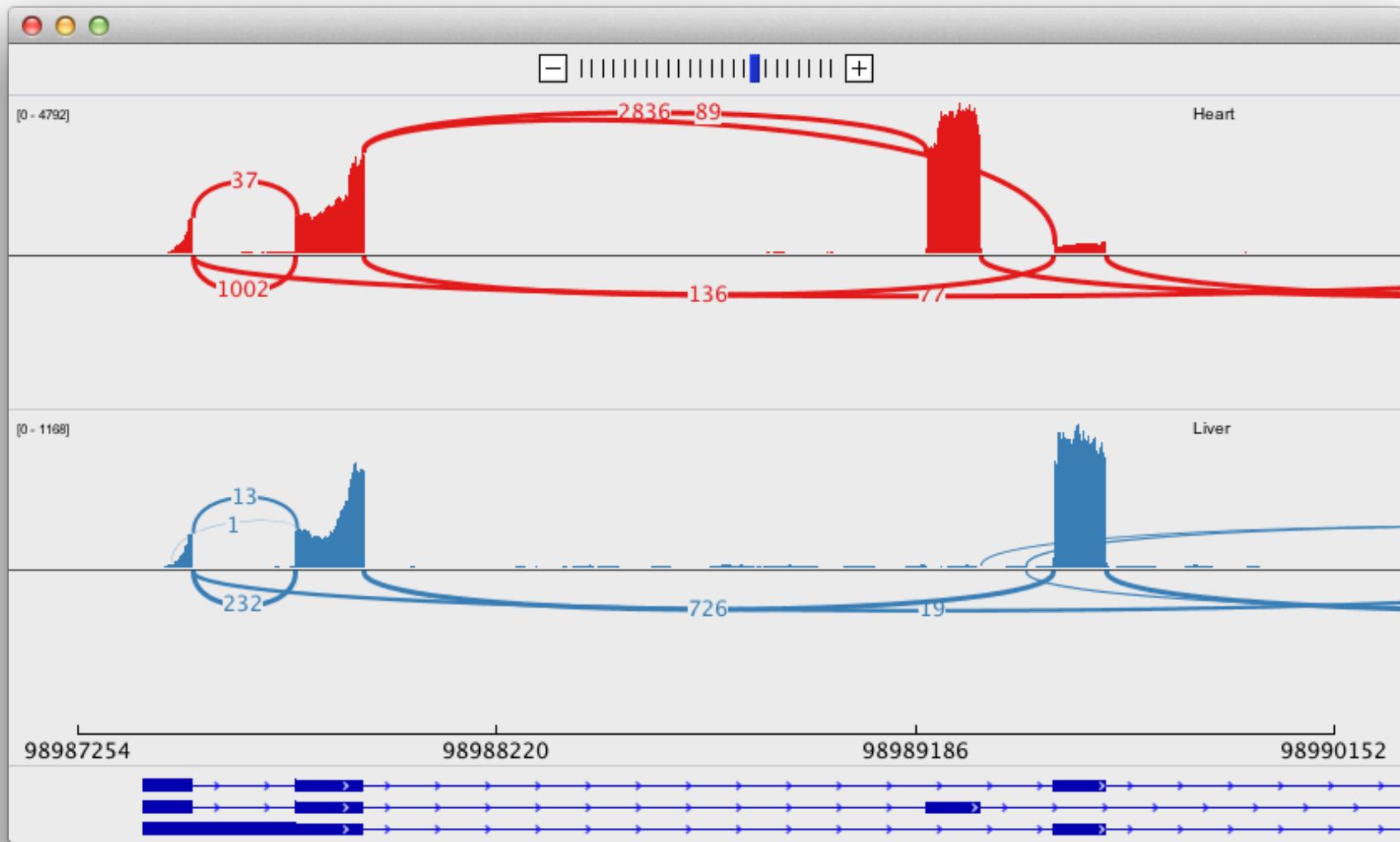
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments



RNA-seq alignments

A screenshot of the IGV software interface. The window title is "Human hg18". The menu bar includes General, Tracks, Mutations, Charts, Alignments (selected), Probes, Proxy, Advanced, and IonTorrent. The main panel shows several genomic tracks. On the left, there are tracks for "Sequence" and "RefSeq genes". On the right, there are tracks for chromosomes q42.2 and q42. A yellow callout box with a red arrow points to the "Show junction track" checkbox in the "Splice Junction Track Options" section. The callout box contains the text "Un-Check Show junction track".

Human hg18

General | Tracks | Mutations | Charts | Alignments | Probes | Proxy | Advanced | IonTorrent

Visibility range threshold (kb): 500 Nominal window size at which alignments become visible

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

Filter and shading options

Coverage allele-freq threshold: 0.2 Mapping quality threshold: 0

Filter duplicate reads Show center line

Filter vendor failed reads Show coverage track

Filter secondary alignments Show soft-clipped bases

Flag unmapped pairs Flag zero-quality alignments

Shade mismatched bases by quality: 5 to 20

Flag insertions larger than: bases

Filter alignments by read group URL or path to filter file

Splice Junction Track Options

Show junction track flanking width:
 Show flanking regions

Insert Size Options

These options control the color coding of paired-end insert sizes. If "compute" is selected values are computed from the actual size distribution of each library.

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

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

OK Cancel

5 tracks loaded chr1:159,464,348 386M of 866M

igvtools

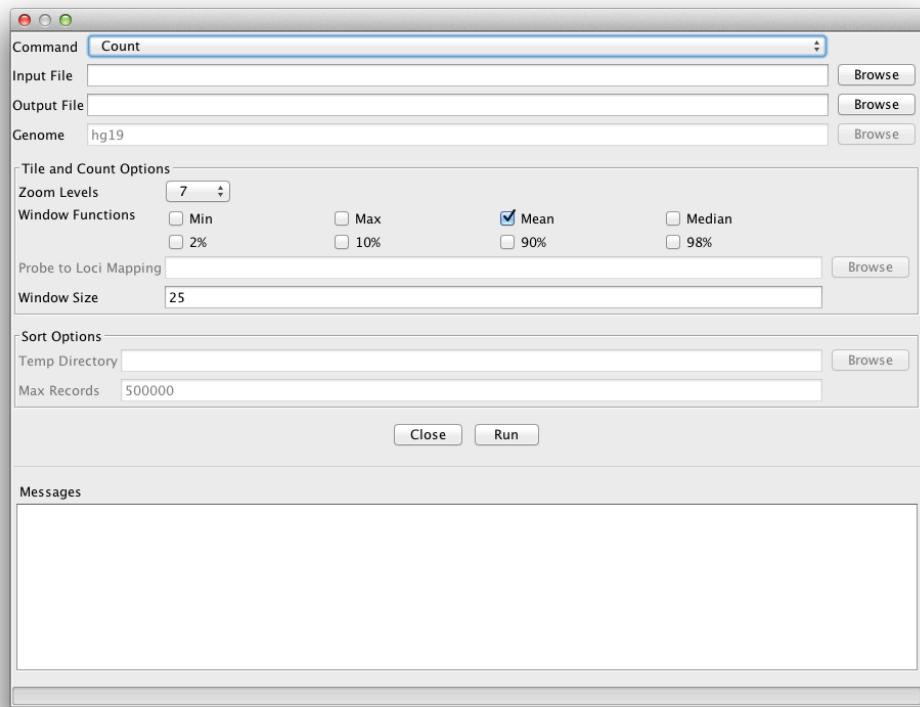
A set of utilities for preparing files for efficient display.

toTDF	<ul style="list-style-type: none">• Converts sorted data file to a binary tiled data file (TDF).• Supported file formats: .wig, .cn, .snp, .igv, .gct
count	<ul style="list-style-type: none">• Computes average alignment or feature density over a specified window size across the genome.• Supported file formats: .sam, .bam, .aligned, .sorted.txt, .bed
sort	<ul style="list-style-type: none">• Sorts file by genomic start position.• Supported file formats: .cn, .igv, .sam, .aligned, .bed.
index	<ul style="list-style-type: none">• Creates an index file for alignment or feature file.• Supported file formats: .sam, .aligned, .sorted.txt, .bed

igvtools



- Can be launched from the IGV user interface
File > Run igvtools...
- Or run from the command line



igvtools toTDF



The **toTDF** utility converts large ASCII data files into tiled data format (.tdf) files.

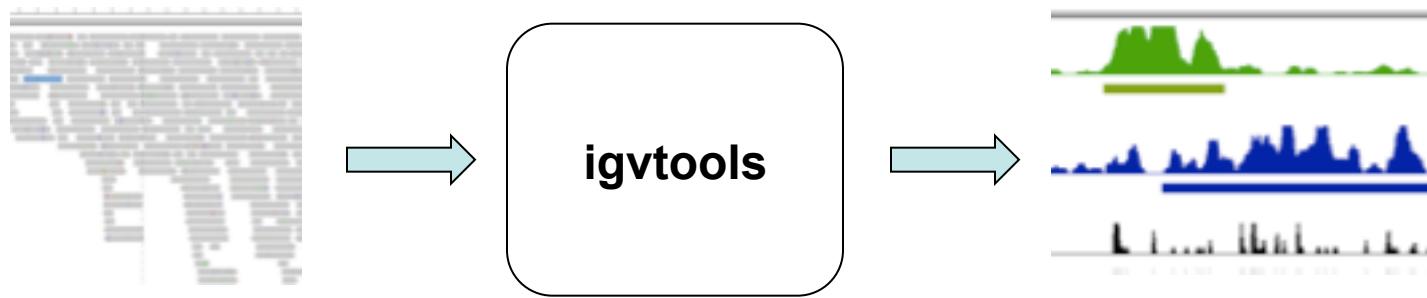
TDF files have the following advantages:

- Data is indexed for efficient retrieval.
- Data is preprocessed for zoomed out views.
- TDF files are web friendly – large data files can be shared over the web. Only small slices of the file are actually transferred as needed.

igvtools count



The **count** command is used to transform alignment files to read density TDF files, e.g. for ChIP-Seq, RNA-Seq, and similar alignment counting experiments.



Alignments

Alignments in bam/sam,
.aligned, or bed format

Read Density

TDF format, indexed and
optimized for fast retrieval at
multiple resolution scales

igvtools sort



- Sorts IGV-supported genomic formats by start position.
- The index command requires sorted files.

Example:

```
igvtools sort -m 1000000 -t ~/myTmpDir inputFile.sam  
outputFile.sorted.sam
```

- Uses combination of memory and disk to handle large files.
 - m = maximum # of lines to hold in memory. When this number is exceeded a temporary file is created.
 - t = directory used to create temporary files during sorting.

igvtools index



Creates an index file for viewing large files in bed, gff, or vcf formats.
An index is optional for bed or gff files, but required for vcf files.

An alternative indexing tool is “tabix”. Tabix both compresses and indexes genomic files. IGV can read either type of index (igvtools or tabix).

Example: igvtools index myFeatures.bed

The index file must remain in the same directory as the input file

Computing coverage: igvtools

Hands-on exercise

- Compute alignment coverage from a BAM file using igvtools count command.

Data source

Illumina BodyMap

Download data files required for this exercise from:
ftp://ftp.broadinstitute.org/pub/igv/CSH_2013/files.zip

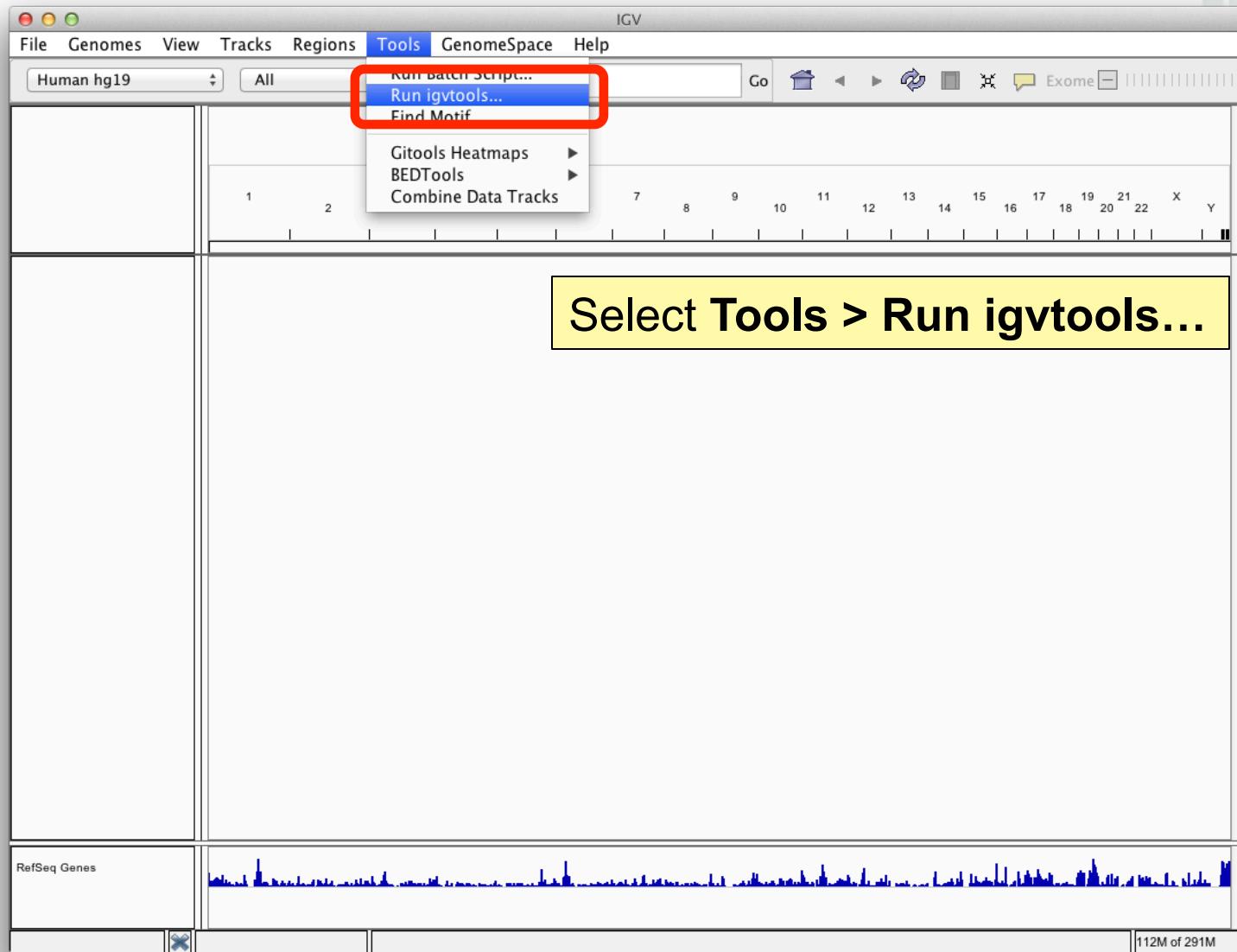
Files included in the zip:

heart.bodyMap.bam

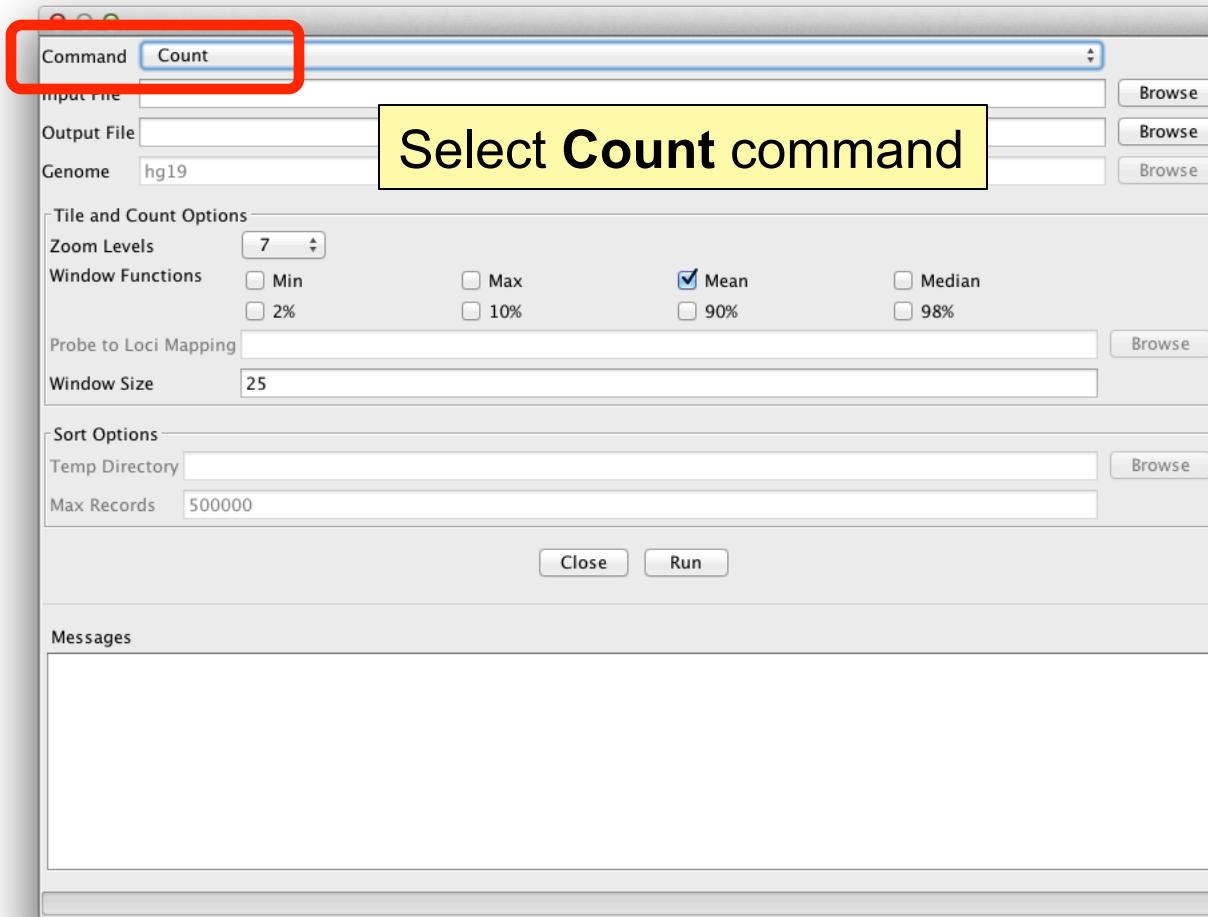
heart.bodyMap.bam.bai

sacCer3.fa (used in next exercise)

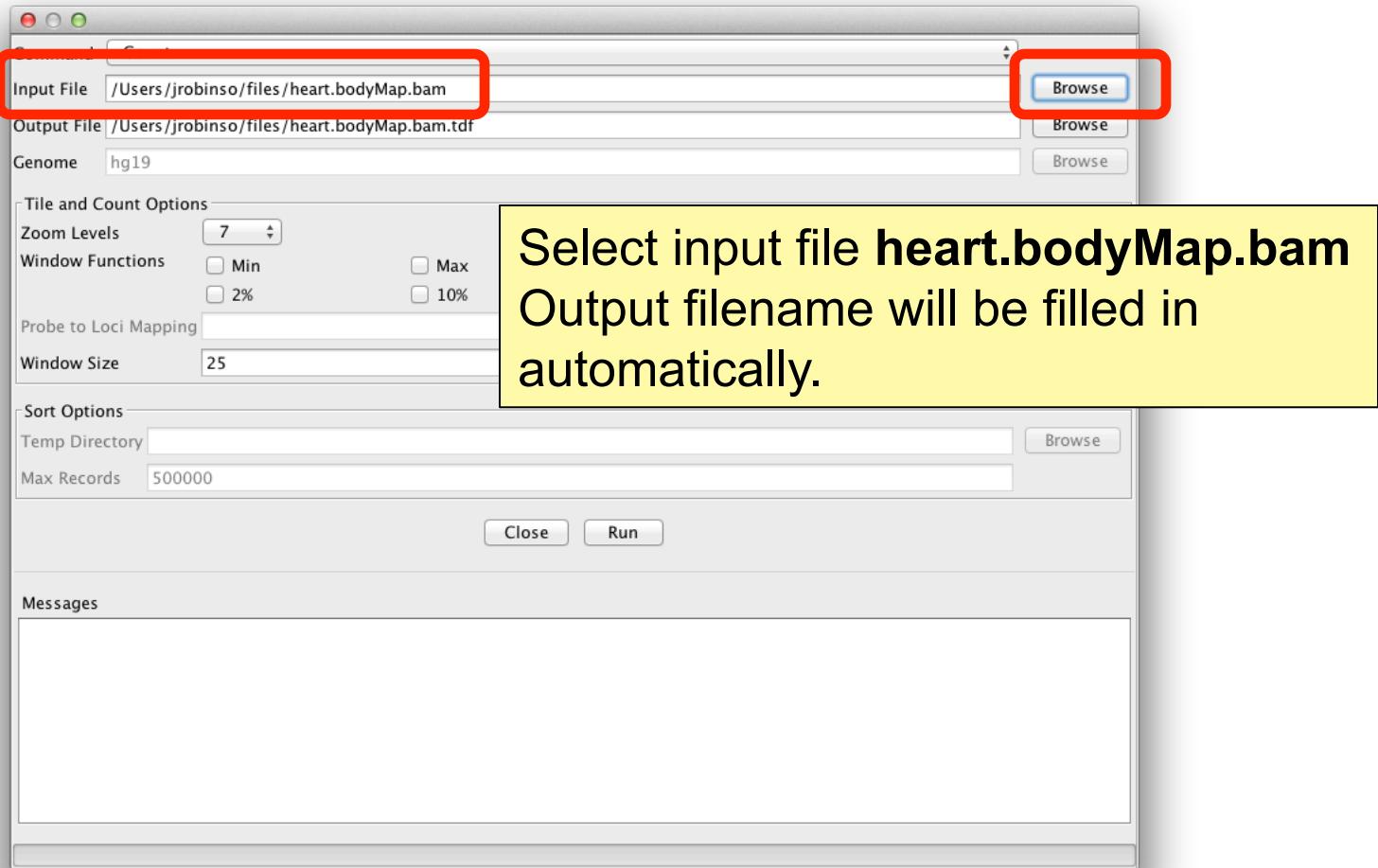
Computing coverage: igvtools



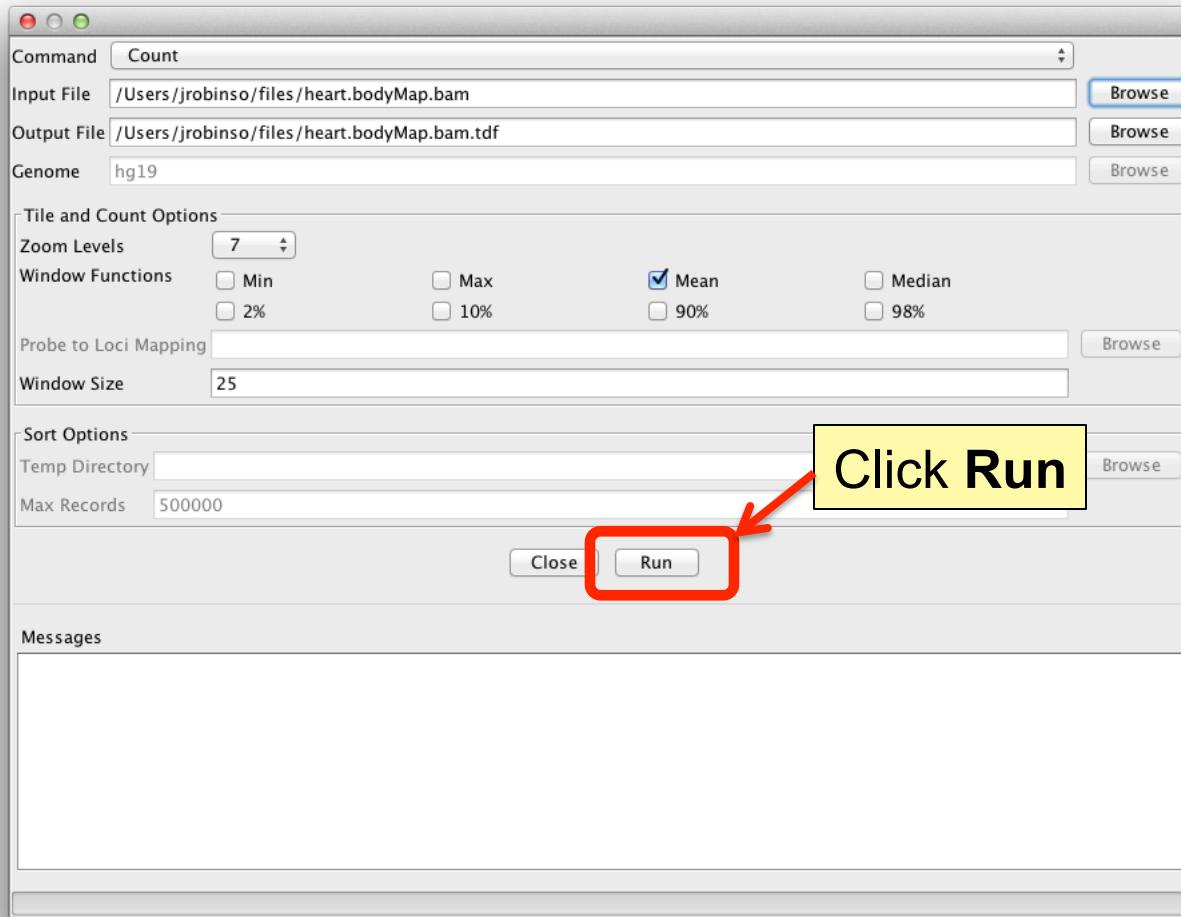
Computing coverage: igvtools



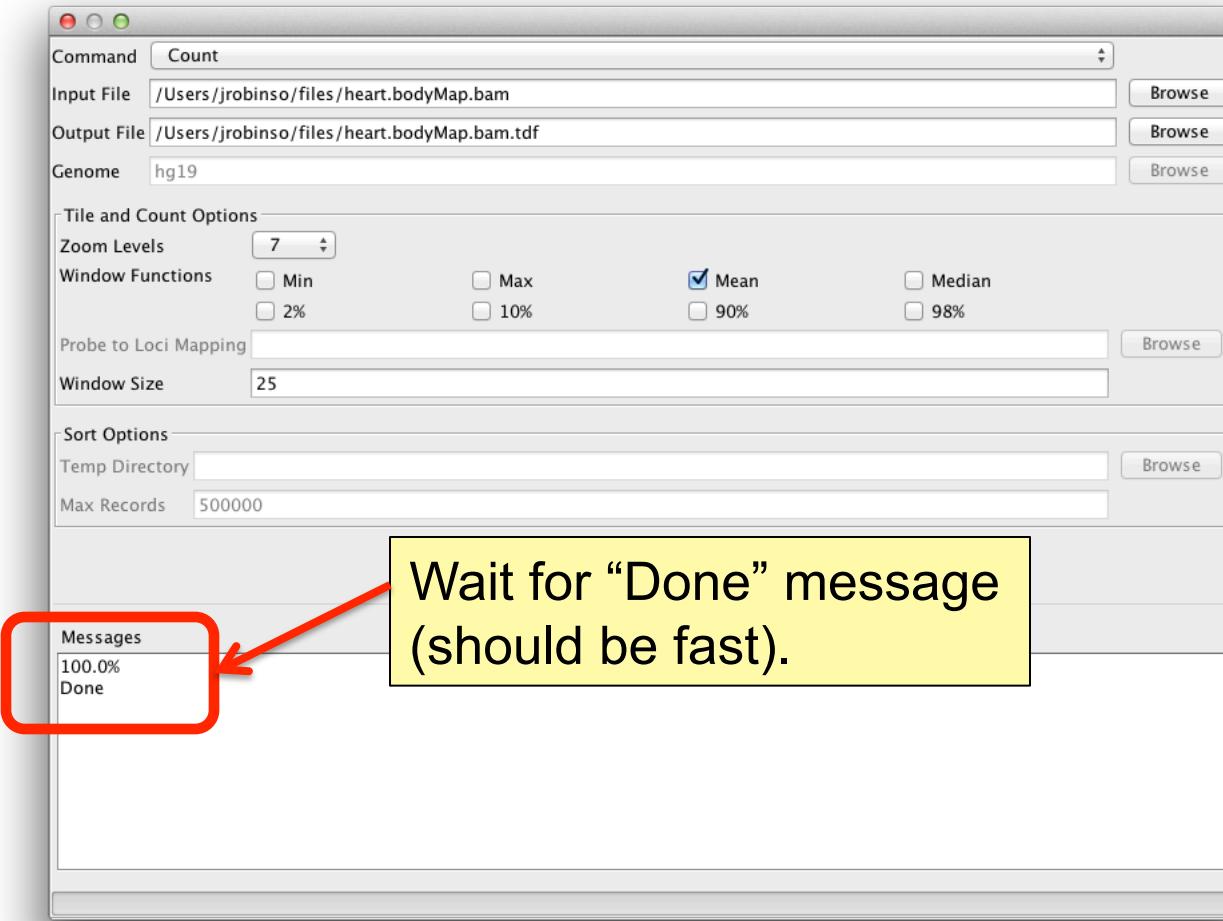
Computing coverage: igvtools



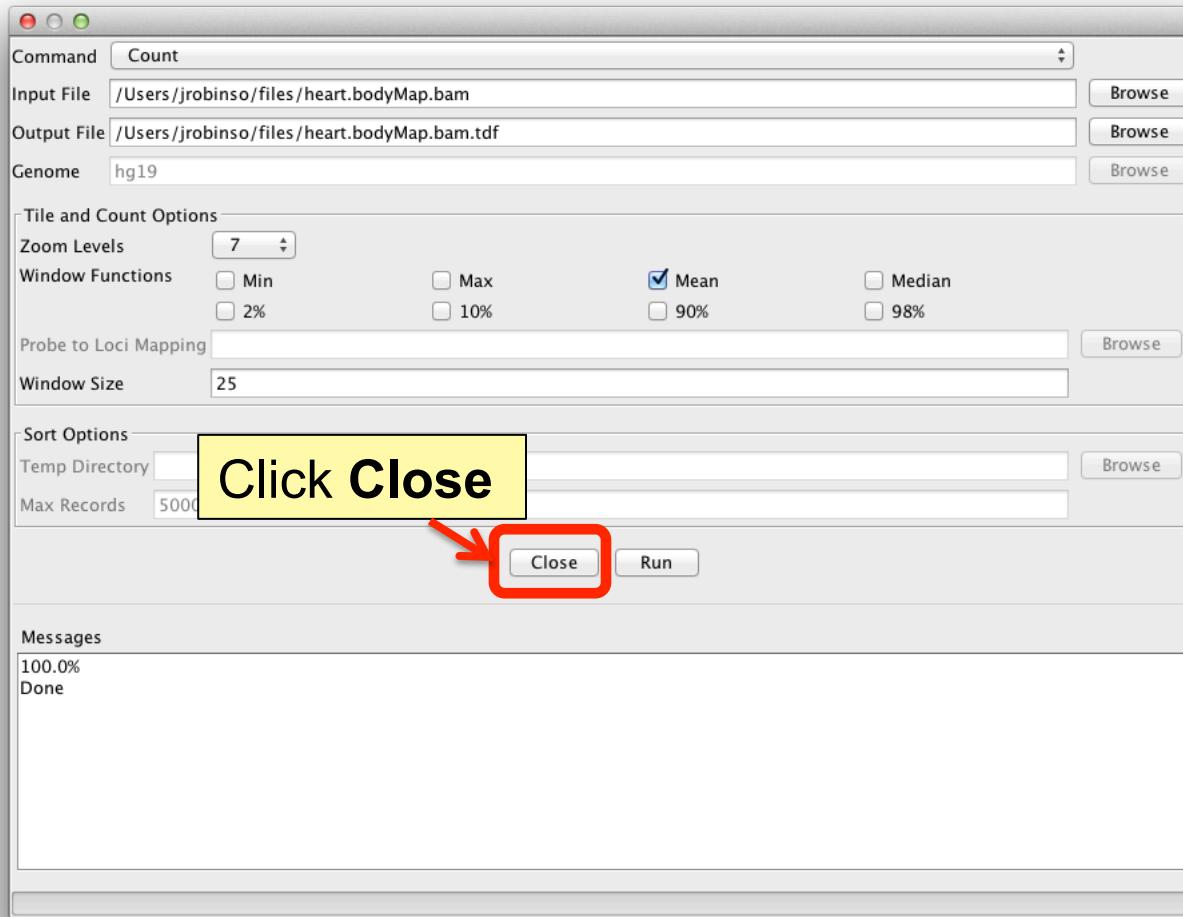
Computing coverage: igvtools



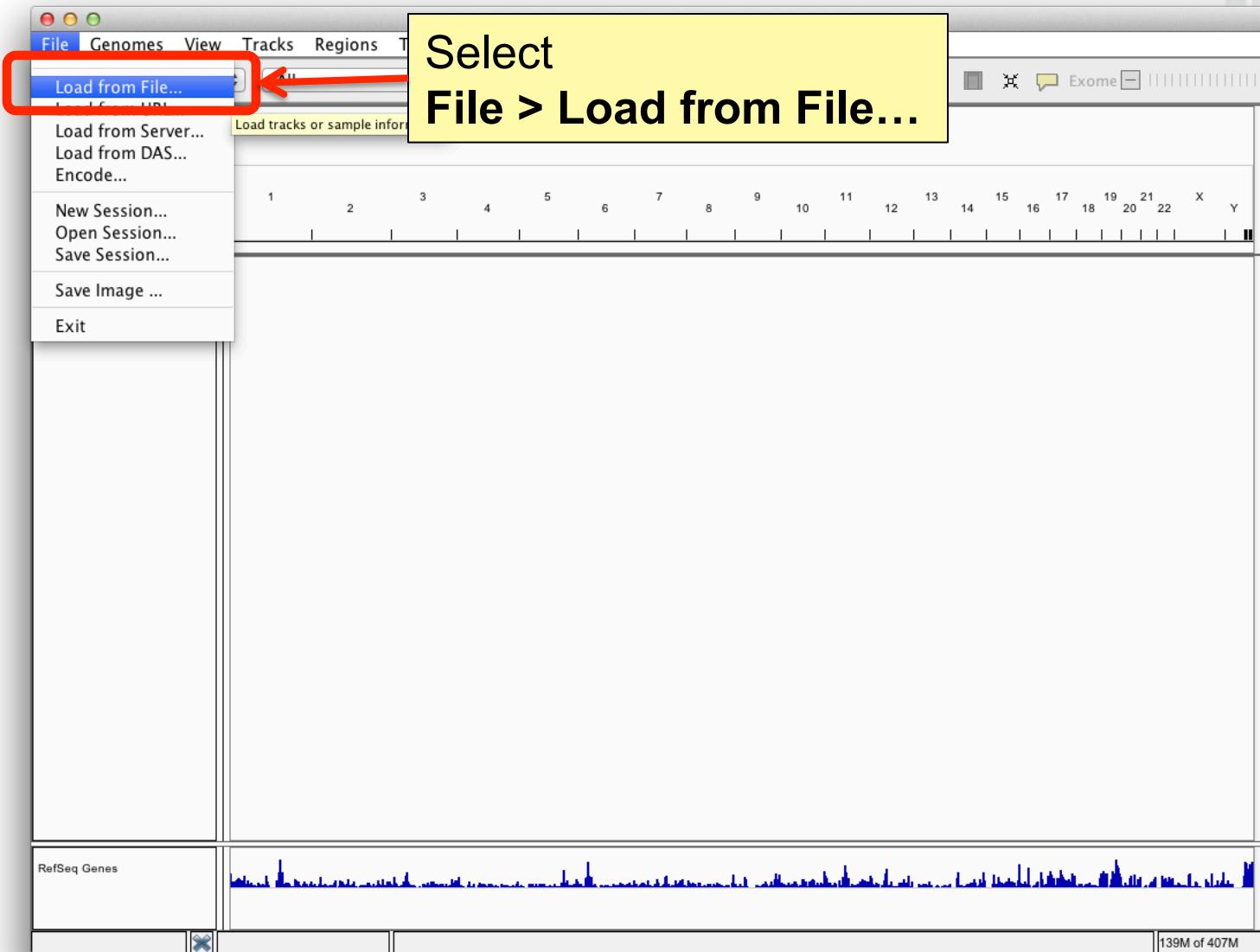
Computing coverage: igvtools



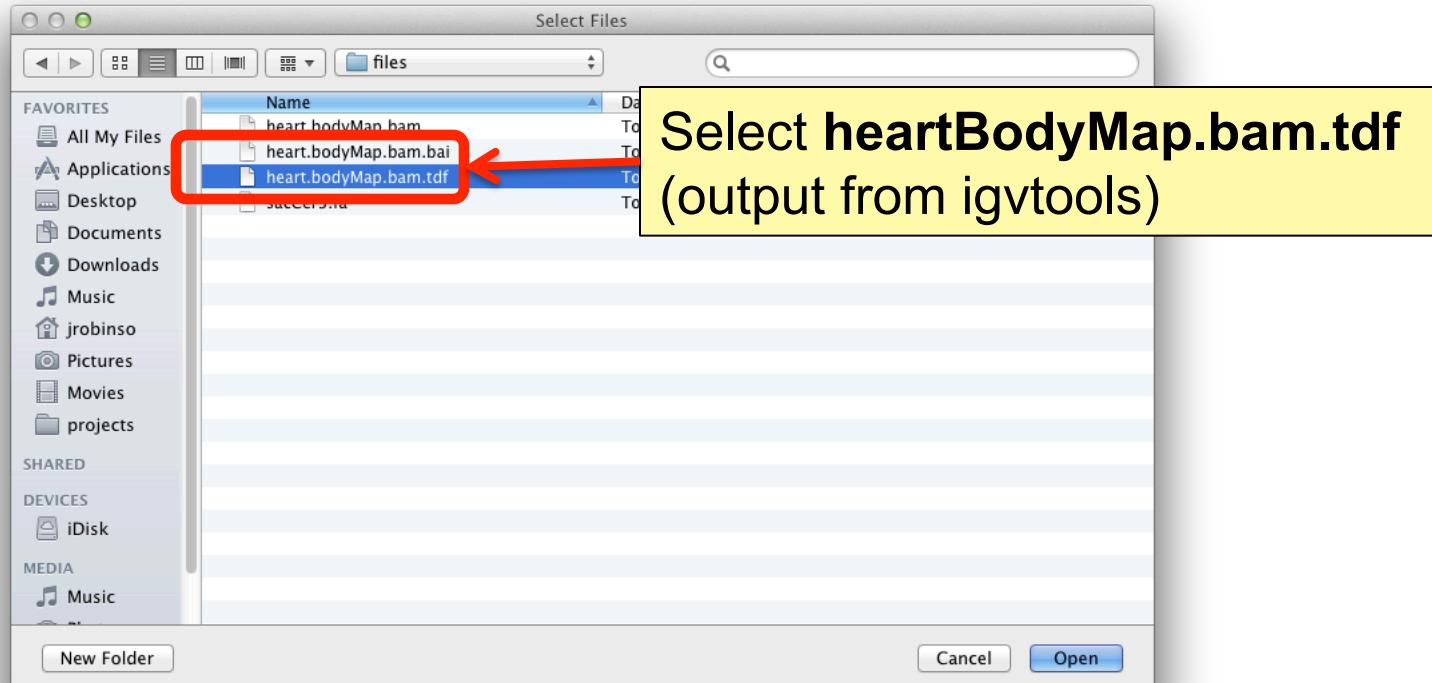
Computing coverage: igvtools



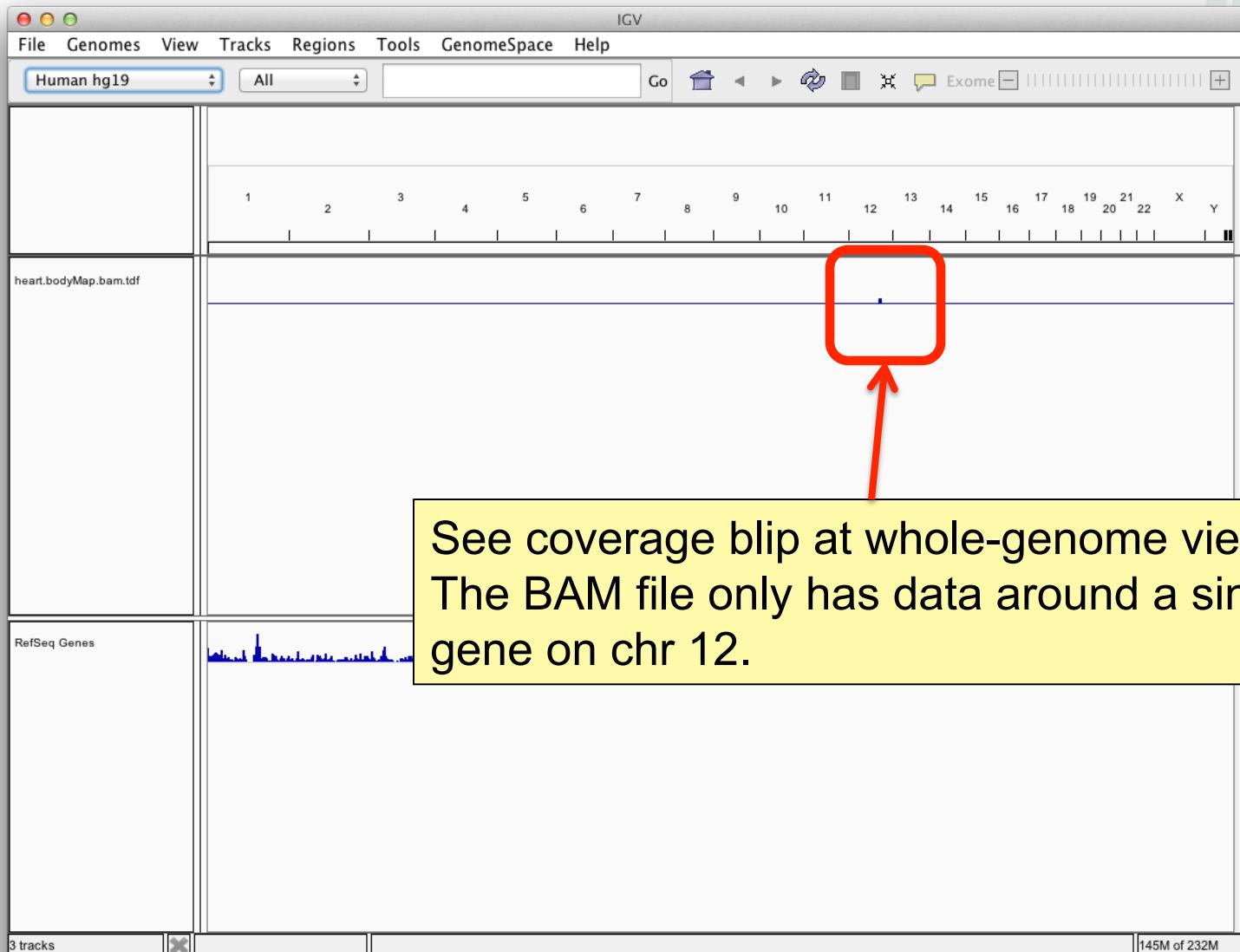
Computing coverage: igvtools



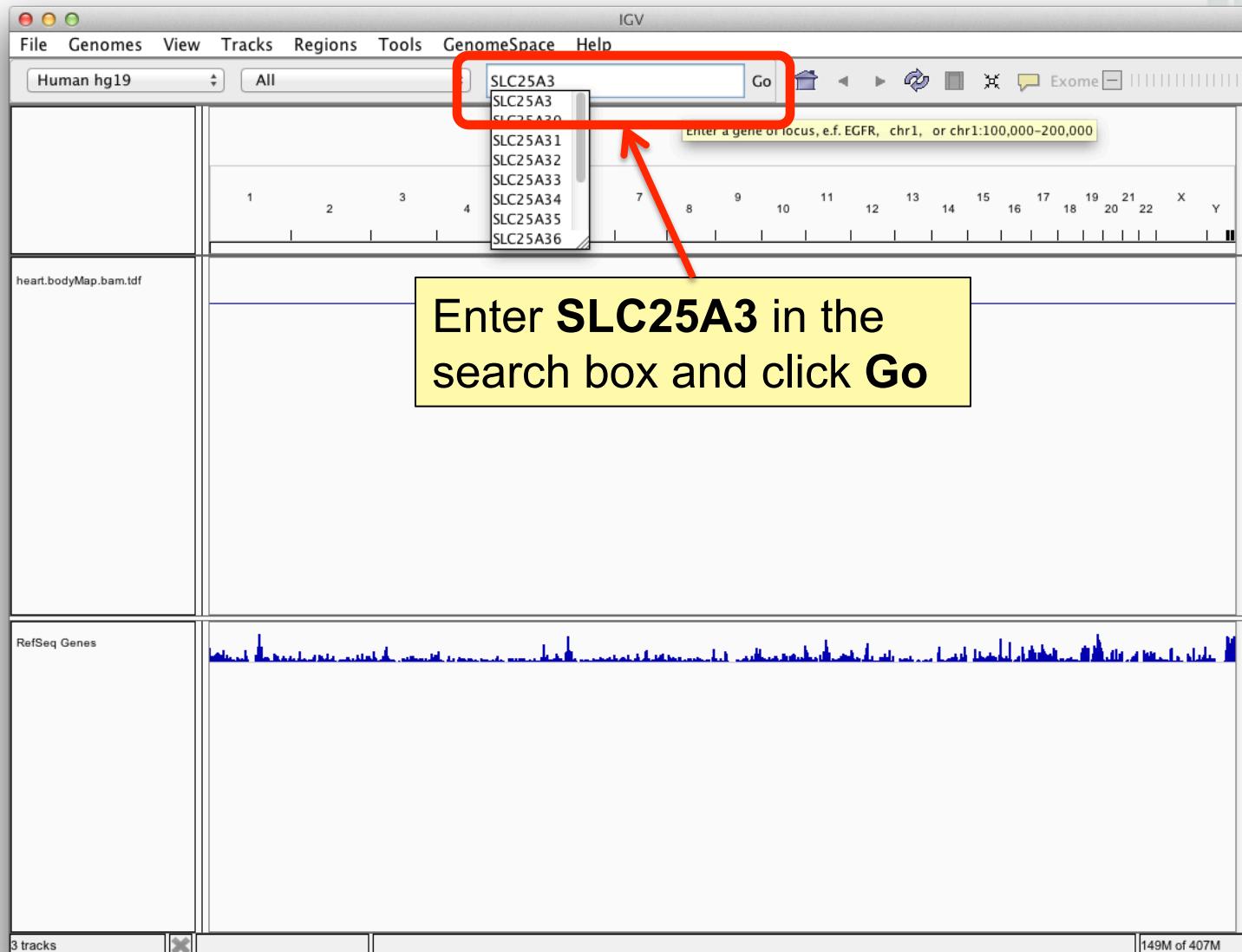
Computing coverage: igvtools



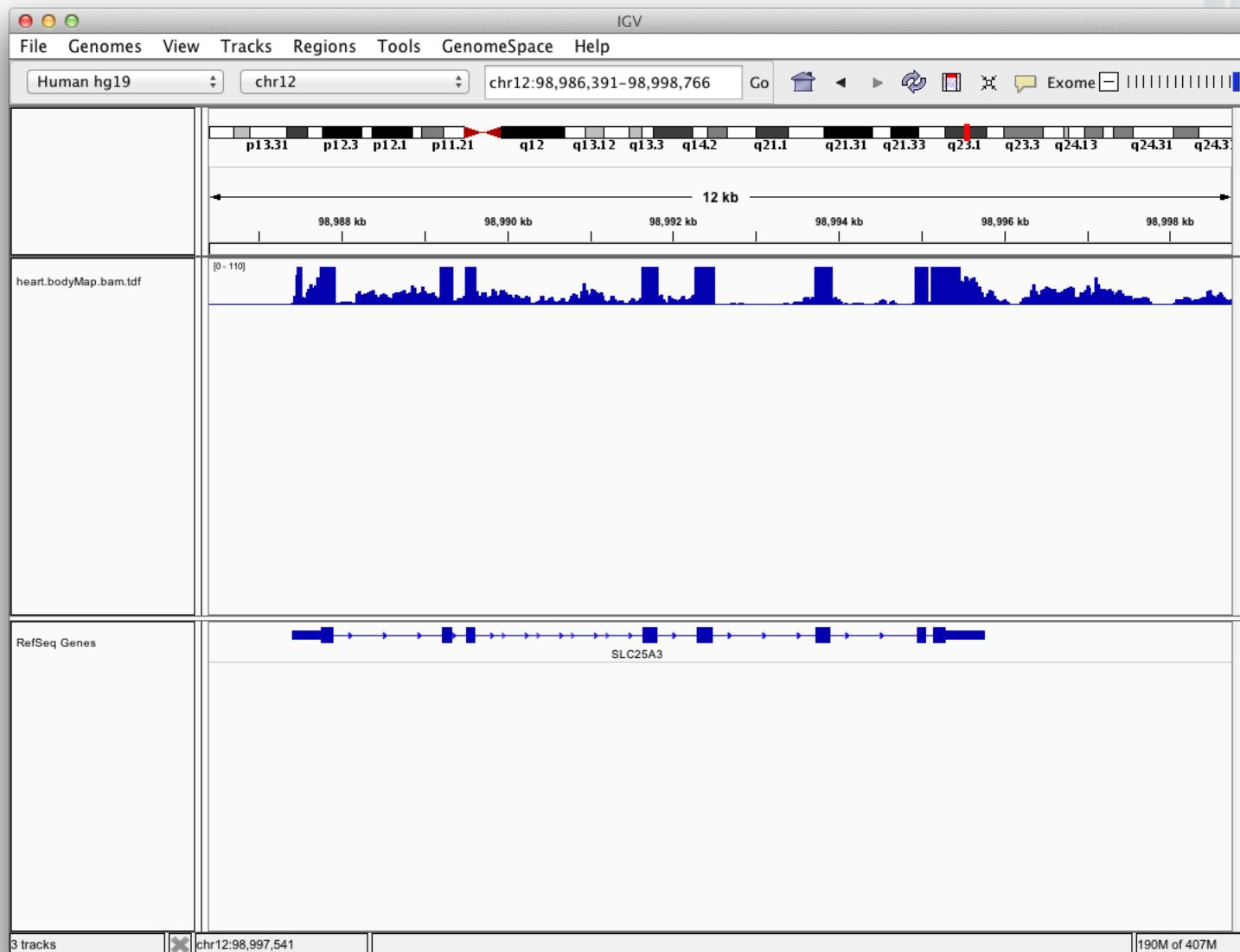
Computing coverage: igvtools



Computing coverage: igvtools



Computing coverage: igvtools



Exercises



- Computing total and strand specific coverage with igvtools
- IGV batch scripting
- Controlling igv from a web page

Acknowledgments



IGV Team

Jim Robinson, Helga Thorvaldsdóttir, Jill Mesirov (PI)

Funding

IGV development has been made possible with funding from:

- National Cancer Institute (NCI) <http://cancer.gov/>
- Starr Cancer Consortium <http://www.starrcancer.org/>
- National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health <http://www.nigms.nih.gov/>
- IGV participates in GenomeSpace <http://genomespace.org/>, which is funded by the the National Human Genome Research Institute (NHGRI) <http://www.genome.gov/>

For further information and help:

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

<http://groups.google.com/group/igv-help>

Cite:

Robinson et al.

Integrative Genomics Viewer.

Nature Biotechnology 29, 24–26 (2011).

Thorvaldsdóttir, Robinson, and Mesirov.

*Integrative Genomics Viewer (IGV):
high-performance genomics data
visualization and exploration.*

Briefings in Bioinformatics (2012).