

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

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

# Agenda

---

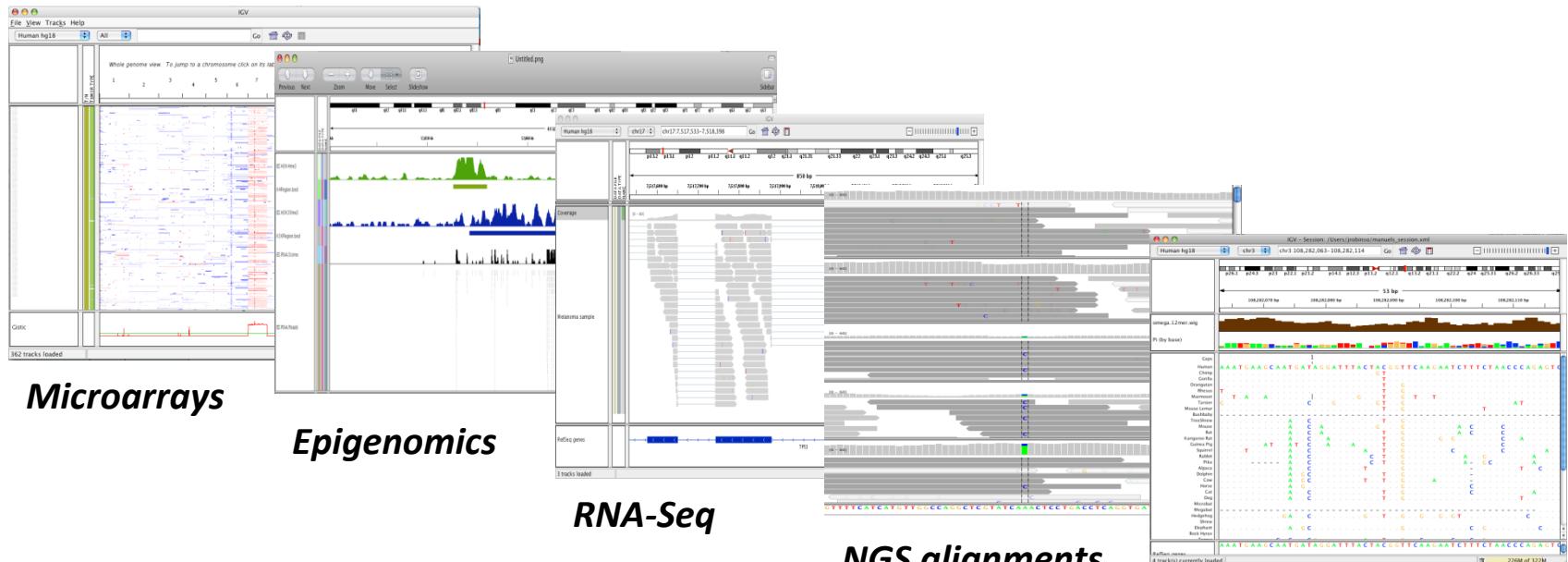


- Introduction
- Using IGV: The Basics
- Data Tracks and File Formats
- NGS Alignments
  - SNPs
  - Structural Events
  - RNA-seq
- igvtools
- Exercises

# What is IGV



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

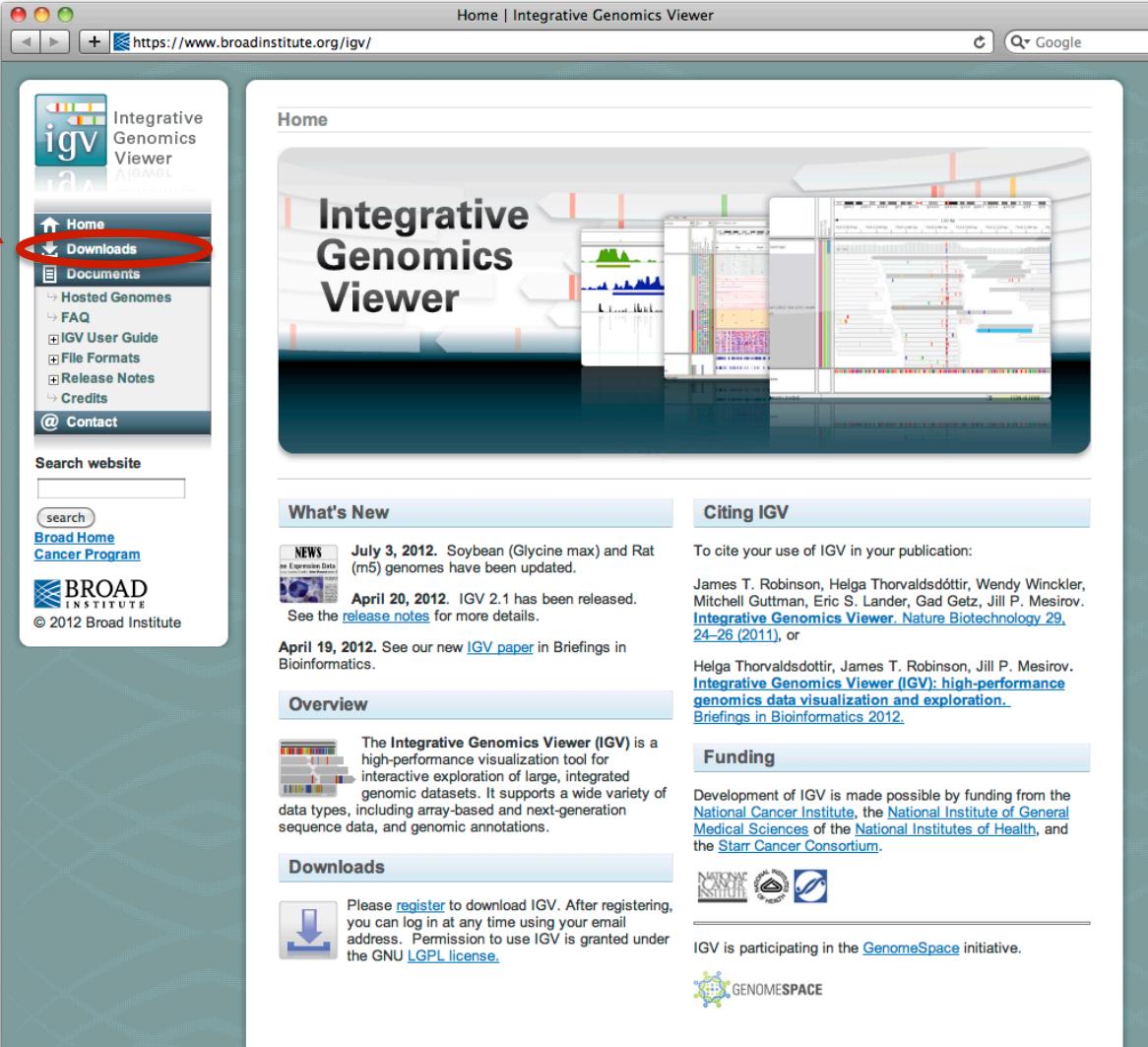


Comparative genomics

# 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 main content area features a large image of the IGV software interface, which includes tracks for genomic data like DNA sequence, RNA expression, and protein levels. Below this are sections for 'What's New', 'Citing IGV', 'Overview', 'Funding', and 'Downloads'. The 'Downloads' section contains text about registering to download the software and mentions the GNU GPL license. Logos for the National Cancer Institute, National Institute of General Medical Sciences, and Starr Cancer Consortium are visible in the footer.

Home | Integrative Genomics Viewer  
https://www.broadinstitute.org/igv/

Home Downloads Documents Hosted Genomes FAQ IGV User Guide File Formats Release Notes Credits Contact

Search website

search Broad Home Cancer Program

BROAD INSTITUTE © 2012 Broad Institute

## Home

# Integrative Genomics Viewer



### What's New

**NEWS** July 3, 2012. Soybean (*Glycine max*) and Rat (rn5) genomes have been updated.

**NEWS** 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 GPL 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](#).

NATIONAL CANCER INSTITUTE NATIONAL INSTITUTE OF GENERAL MEDICAL SCIENCES Starr Cancer Consortium

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

GENOME SPACE

# 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. The main content area shows instructions for installing IGV, starting with '1. Mac Application'. A red arrow points from a callout box to the 'Download Mac App' button. Below it, there are Java Web Start launch buttons for 750 MB, 1.2 GB, 2 GB, and 10 GB memory options. A section for 'Binary Distribution' is also visible at the bottom.

Downloads

Integrative Genomics Viewer (IGV) (Version 2.3)

Install IGV

Options for installing and running IGV:

- (Mac only) Download and run the Mac application; or
- (All systems) Use the Java Web Start buttons (Mac users: see below for limitations); or
- (All systems) Download the binary distribution and run IGV from the command line.

**1. Mac Application**

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 Java Web Start buttons below.**

**Download Mac App**

**2. Java Web Start**

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

**\*Mac Users:** The Java Web Start 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 with 750 MB  
Maximum usable memory for Windows OS with 32-bit Java.

**Launch**

Launch with 1.2 GB  
Maximum usable memory for 32-bit Mac OS.

**Launch**

Launch with 2 GB  
Maximum usable memory for 32-bit Mac OS.

**Launch**

Launch with 10 GB  
For large memory machines with 64-bit Java.

**3. Binary Distribution**

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

# Using IGV: The Basics

# Using IGV: the basics



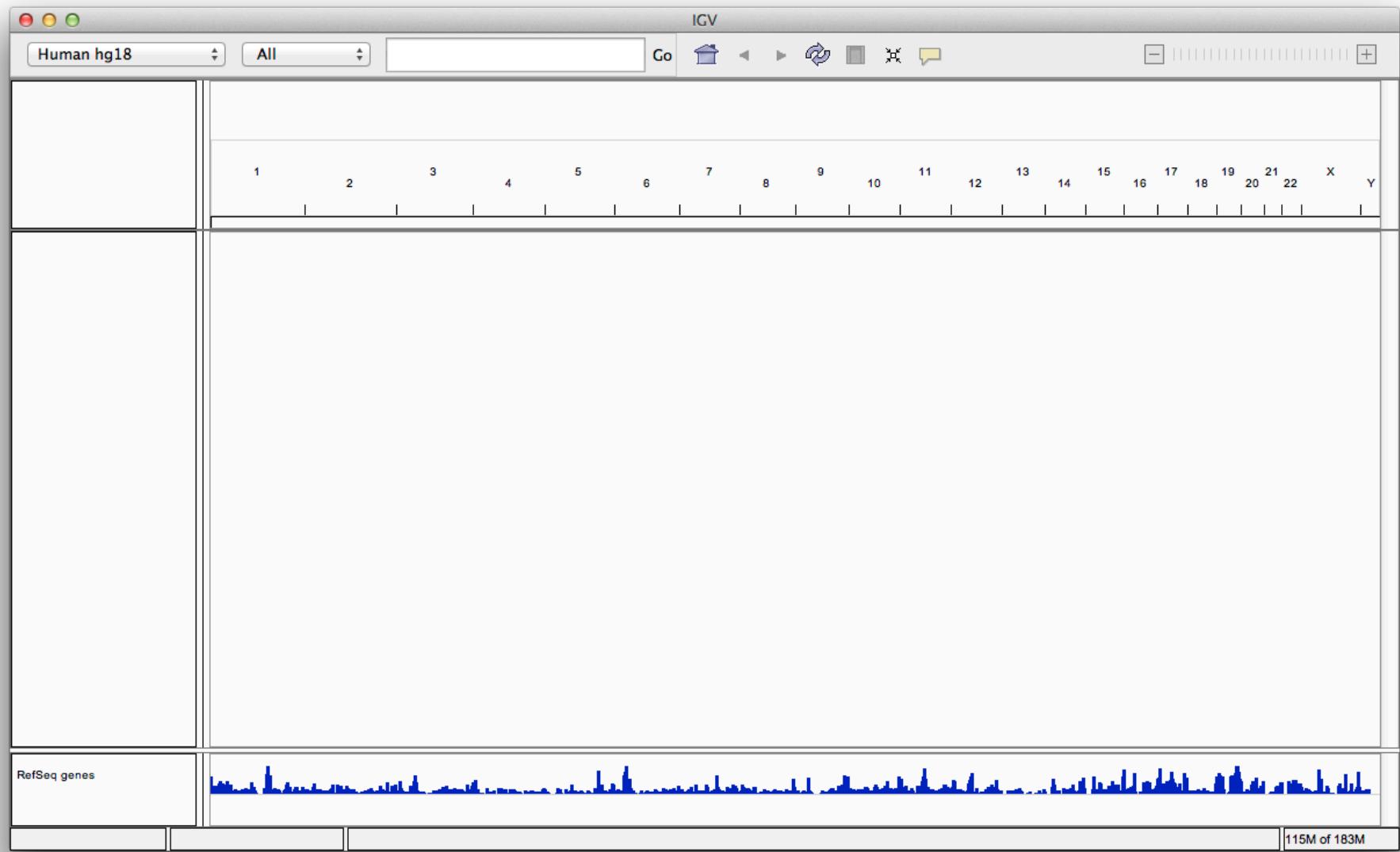
## Hands-on exercise

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

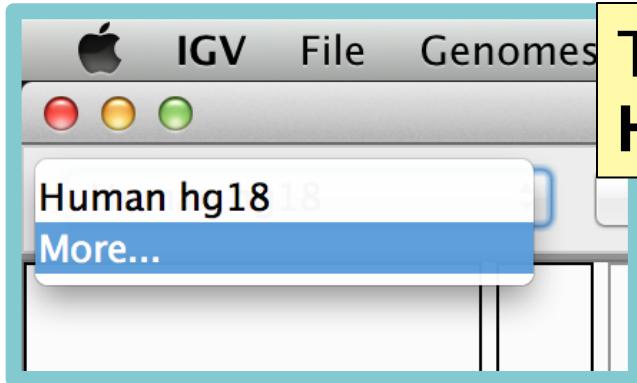
# Select the reference genome



A screenshot of the IGV software interface. At the top left, there is a drop-down menu labeled "Human hg18". Next to it is another button labeled "All". A yellow callout box with a black border and black text is positioned over the "Human hg18" button, containing the instruction "Select genome from the drop-down menu". The main window shows a genomic track for chromosome 14, with tracks for chromosomes 14 through 22, X, and Y visible at the top. Below the tracks, 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 displayed. The entire interface has a light gray background with a dark gray header bar.



# Select the reference genome

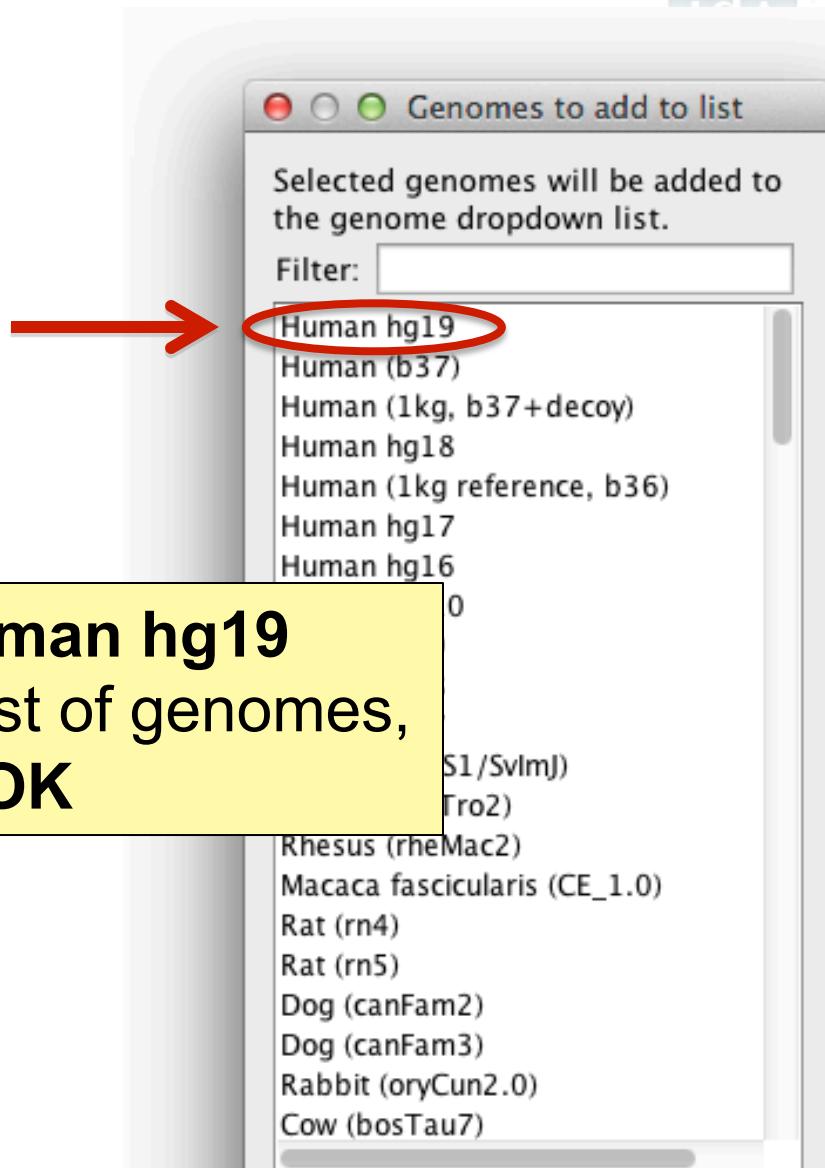
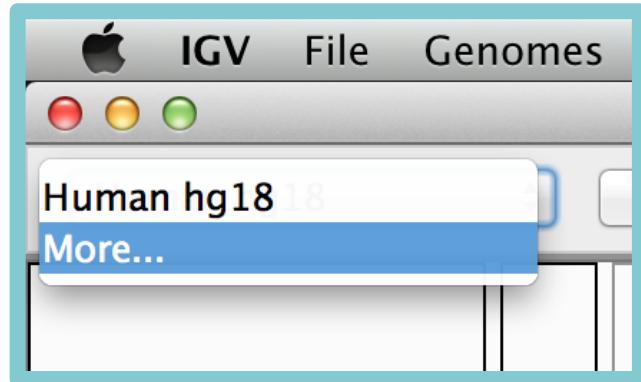


Today, we will use both  
**Human hg18 and hg19**



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

# Select the reference genome

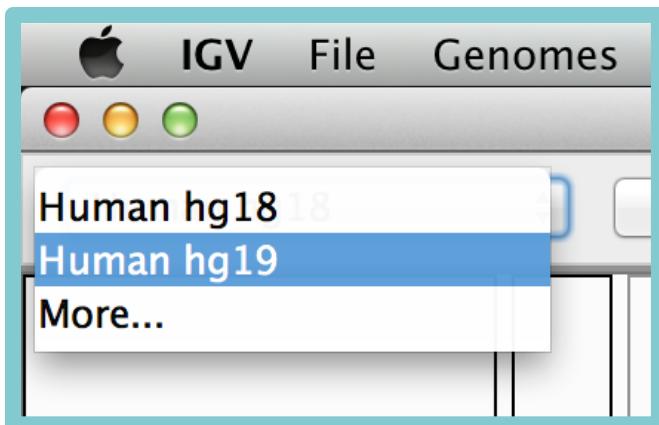
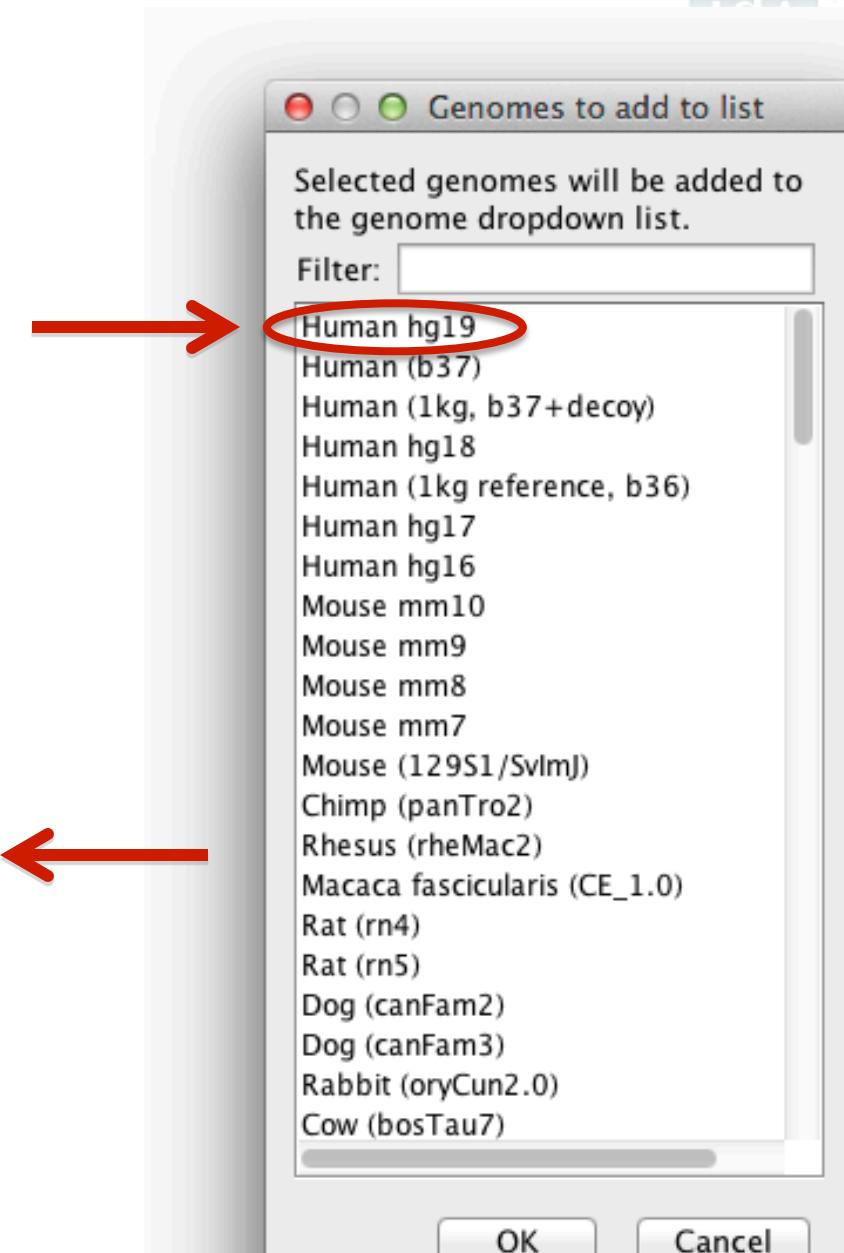
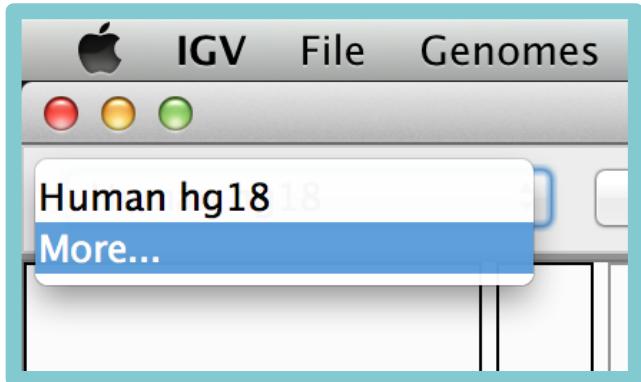


Select **Human hg19** from the list of genomes, and click **OK**

OK

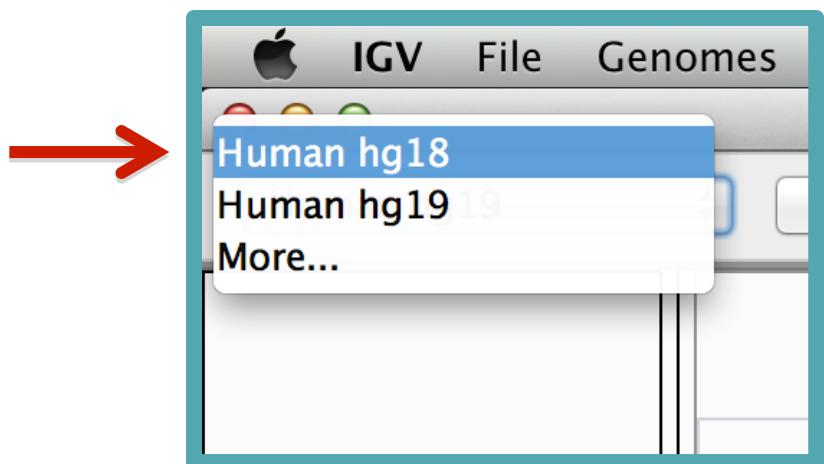
Cancel

# Select the reference genome



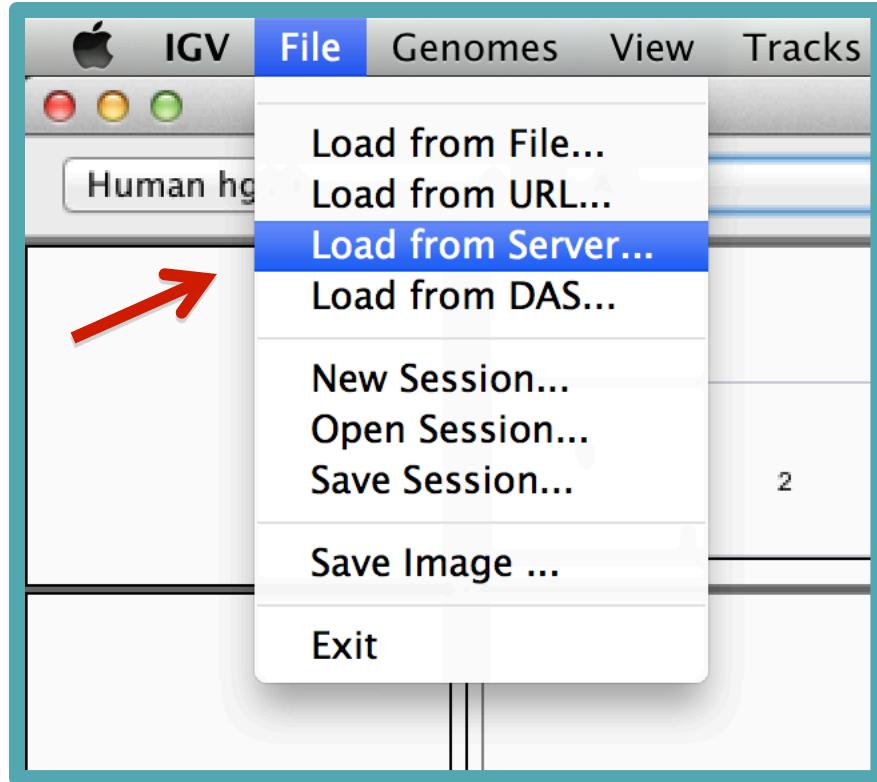
# Select the reference genome

Select Human hg18

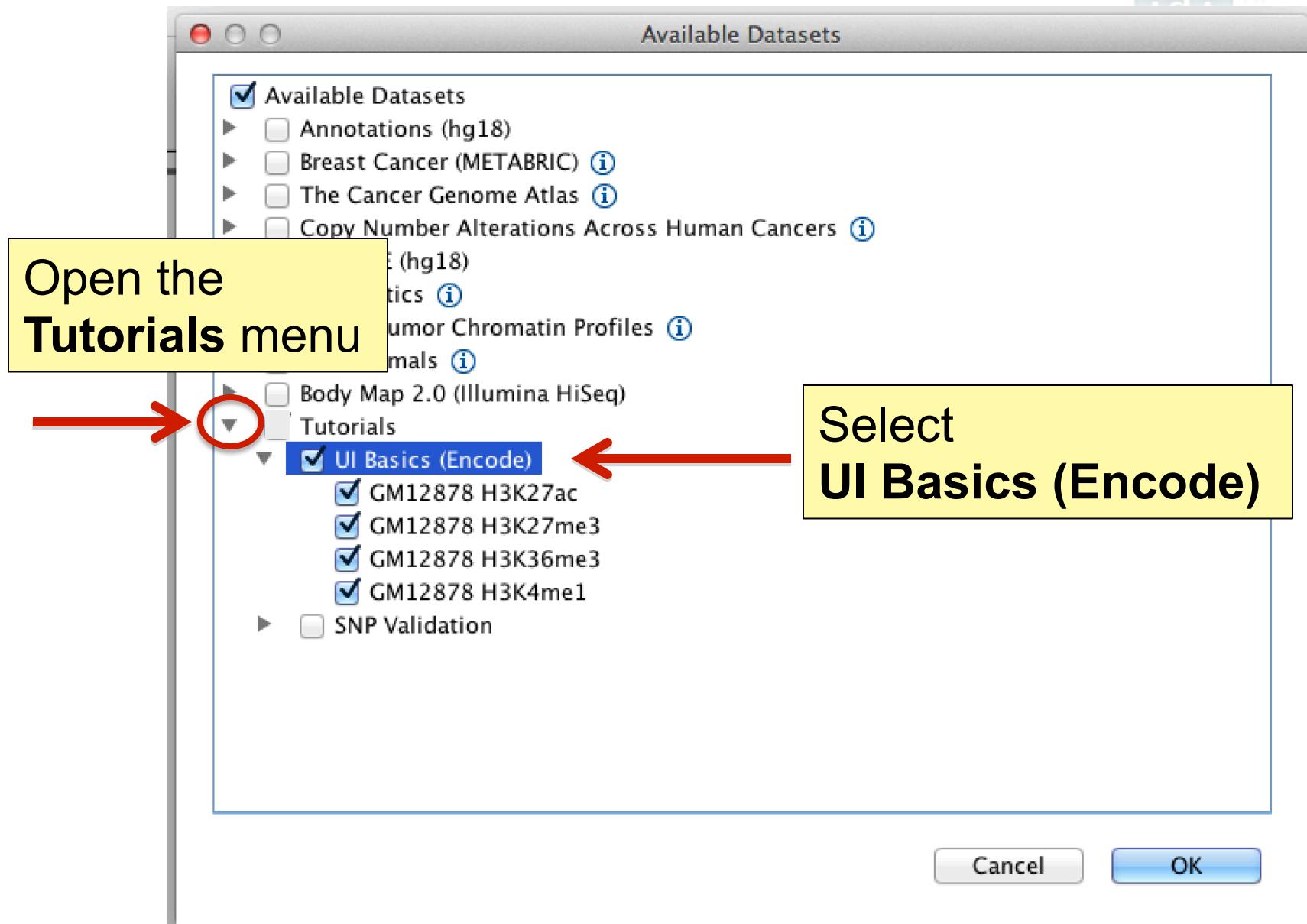


# Load data

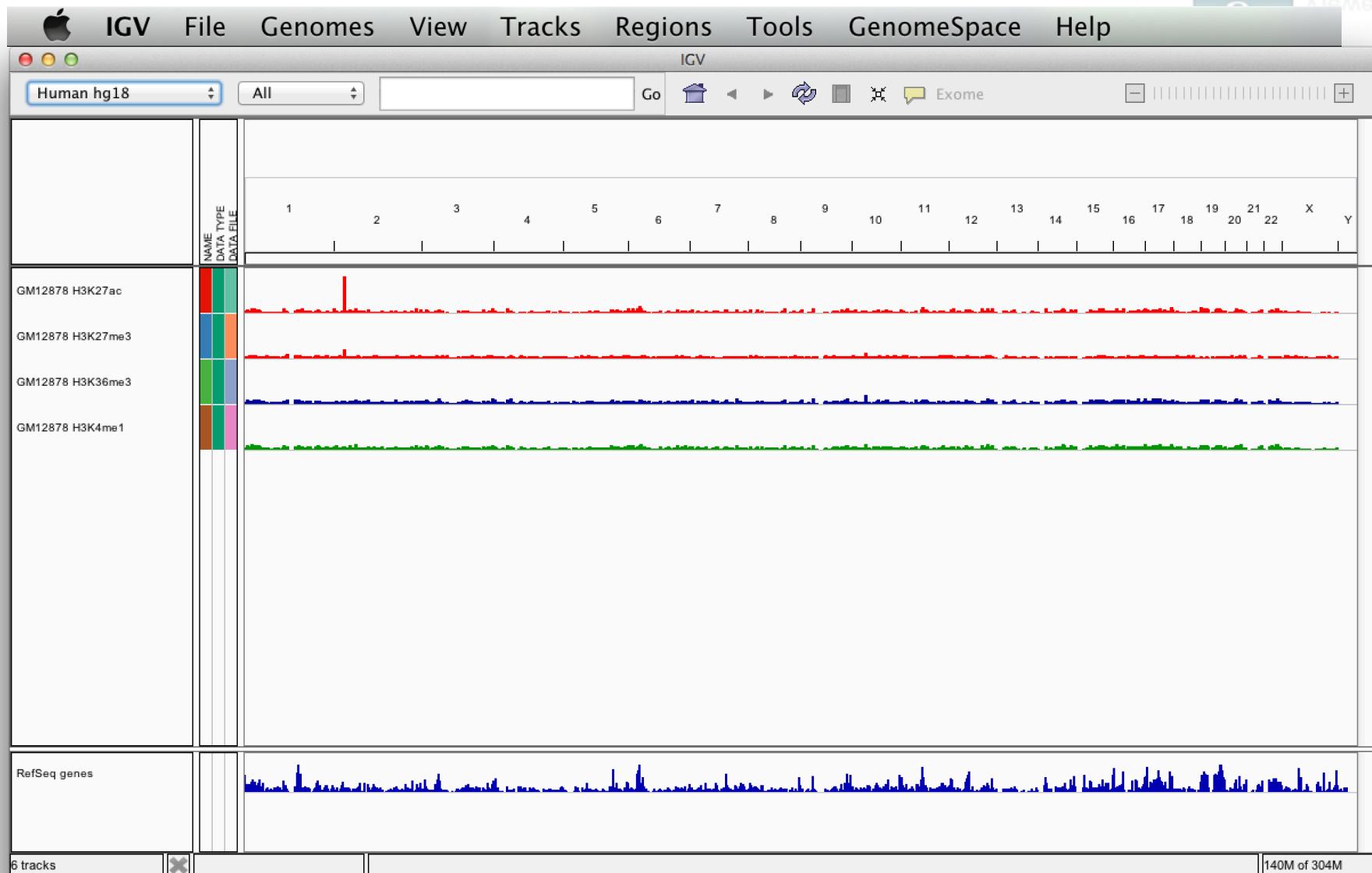
Select File > Load from Server...



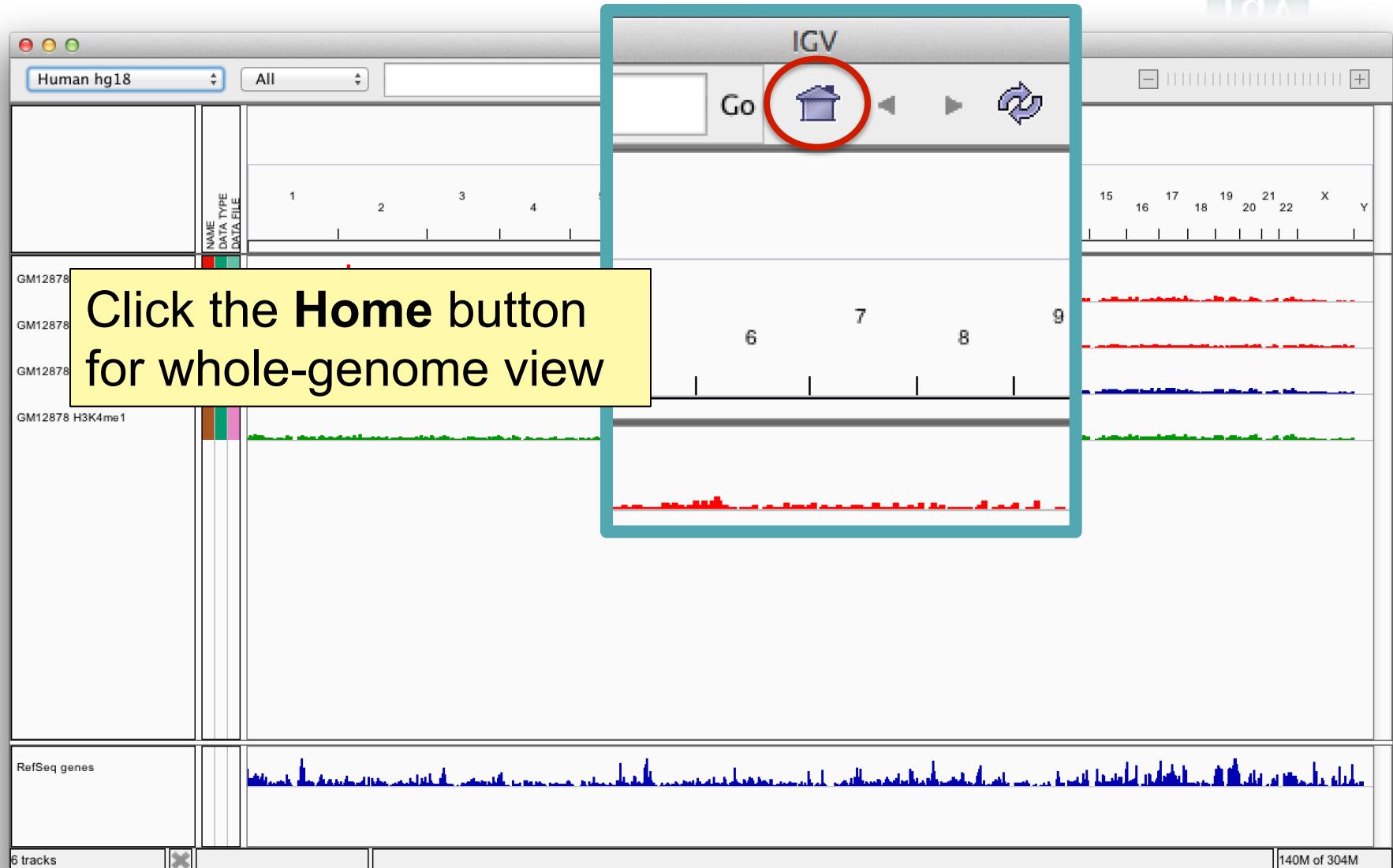
# Load data



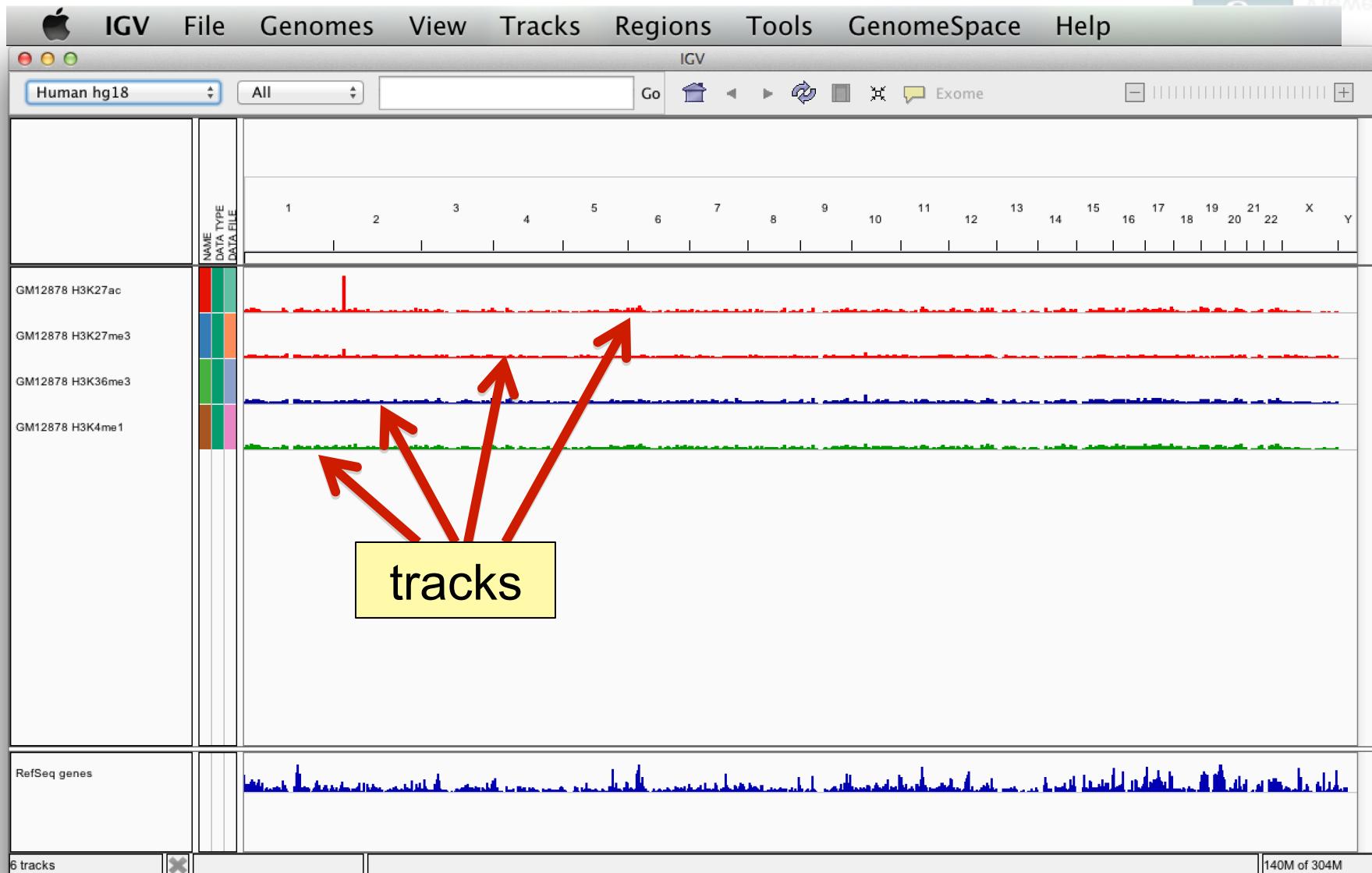
# Screen layout



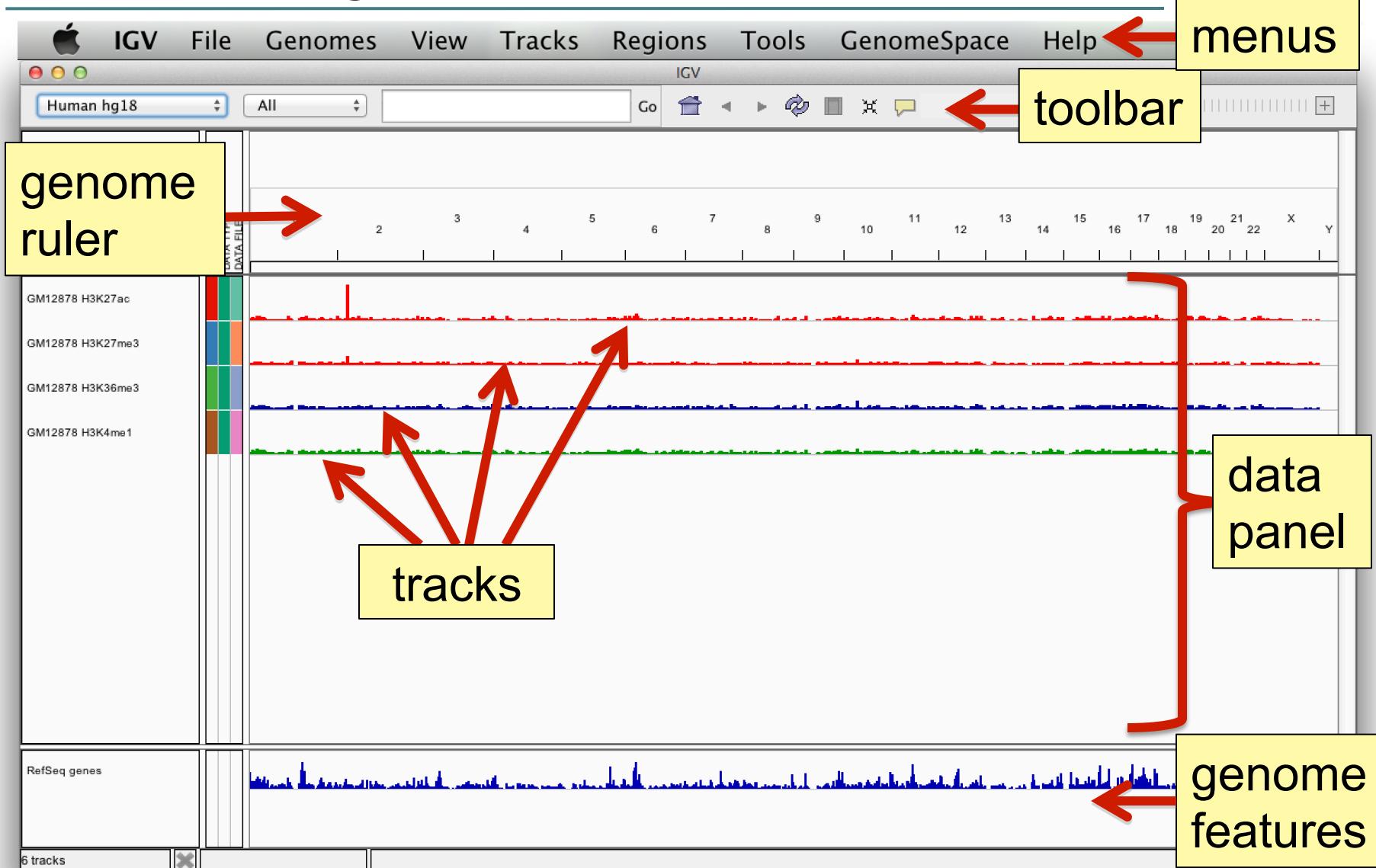
# Screen layout



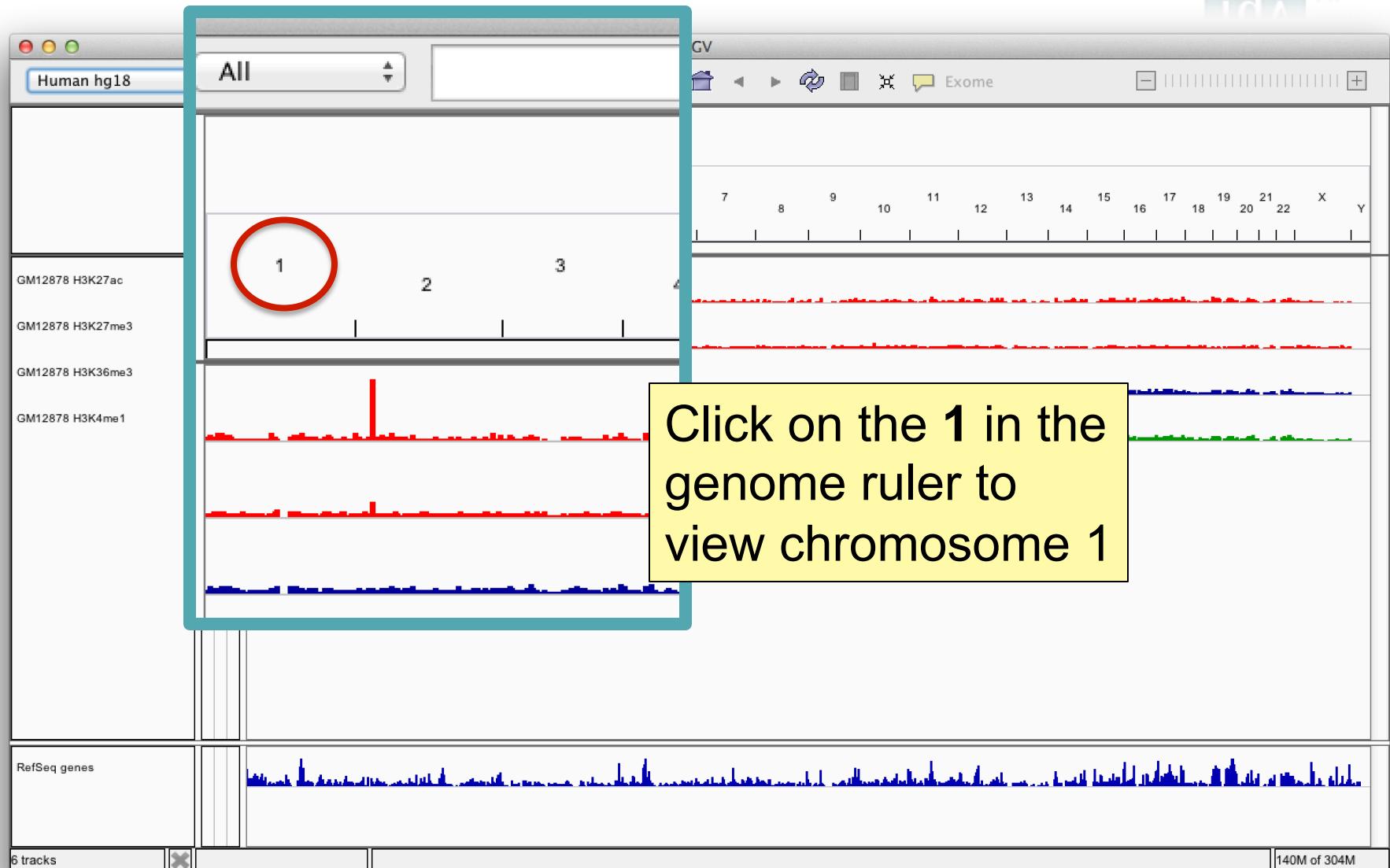
# Screen layout



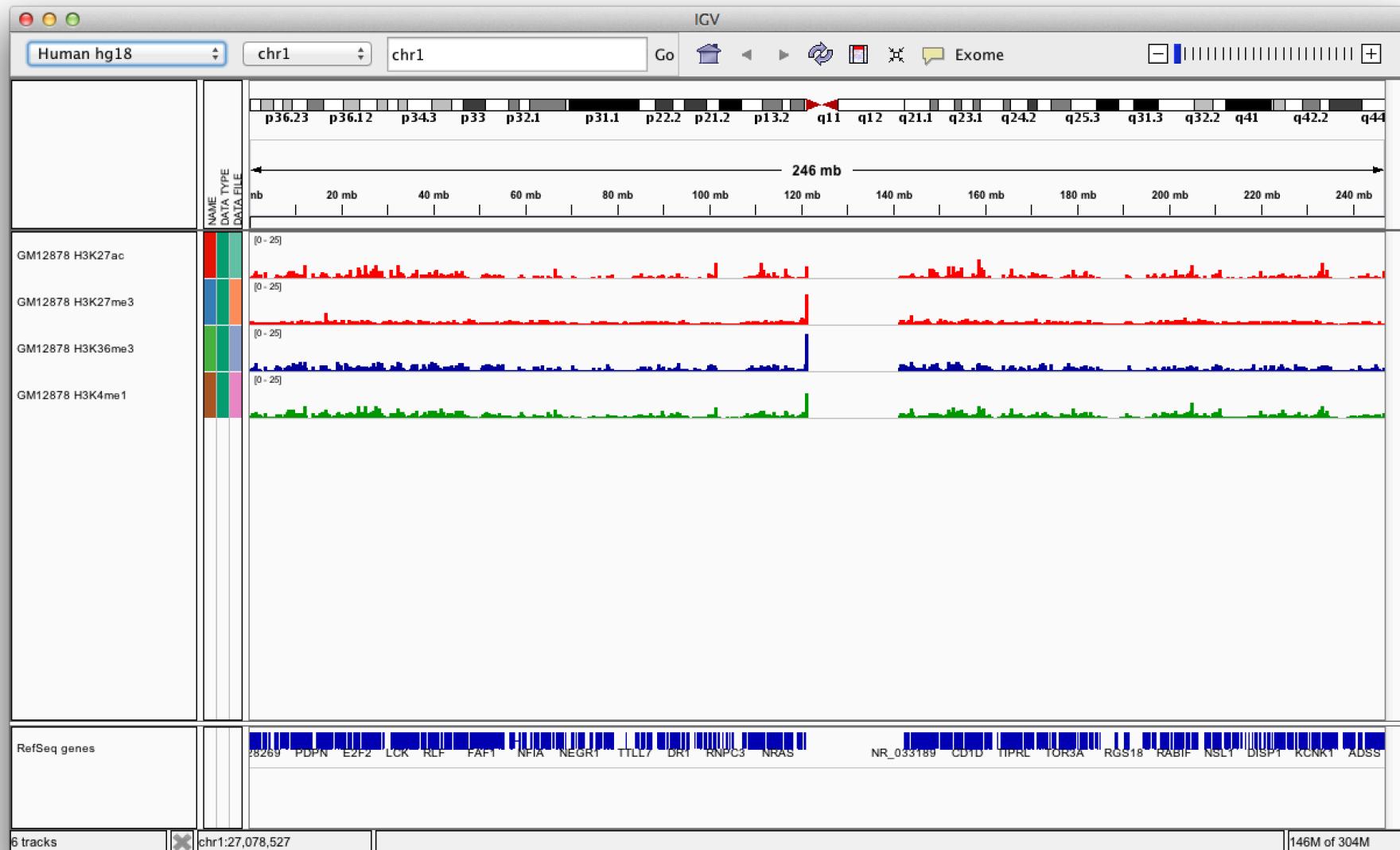
# Screen layout



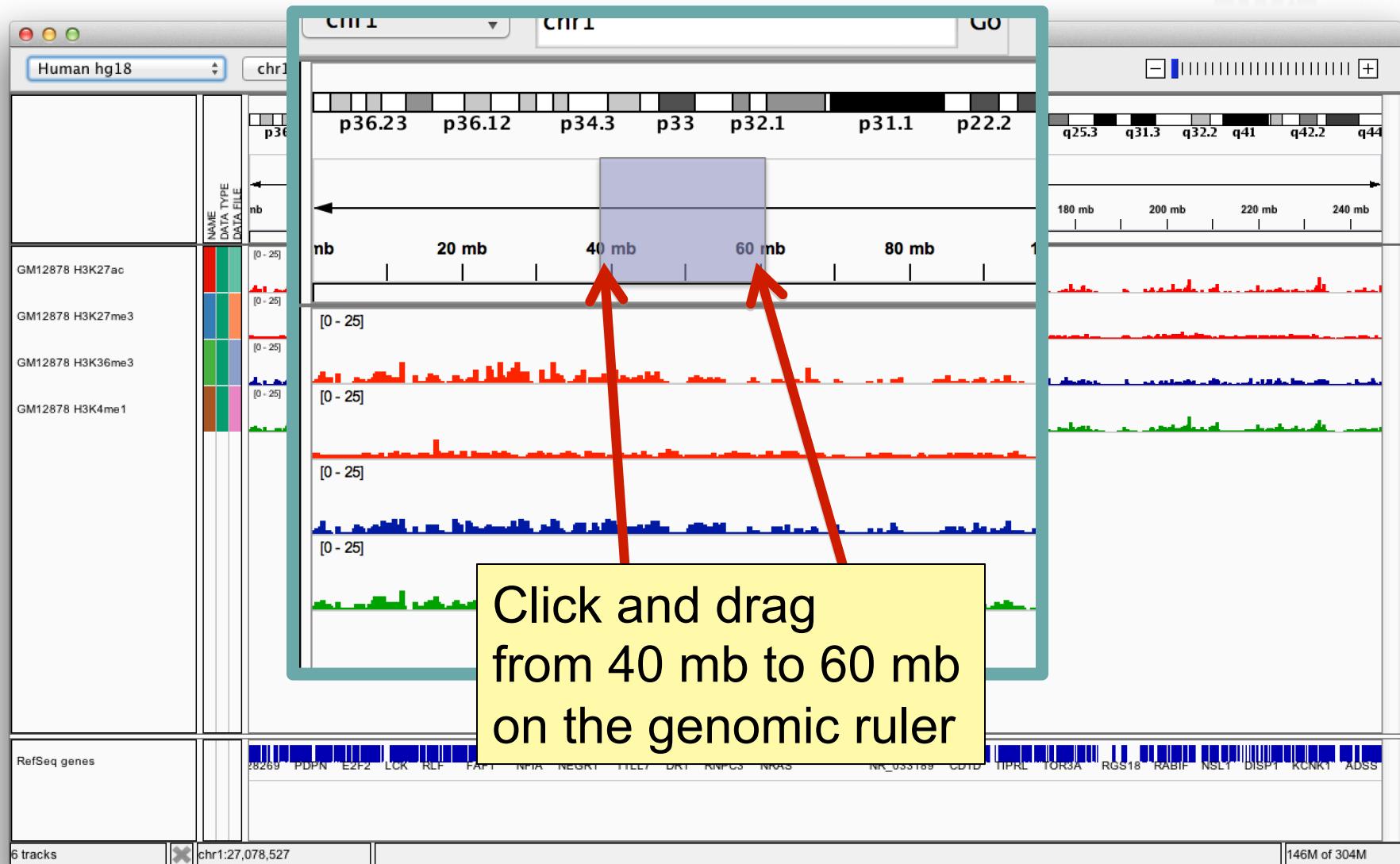
# Navigate



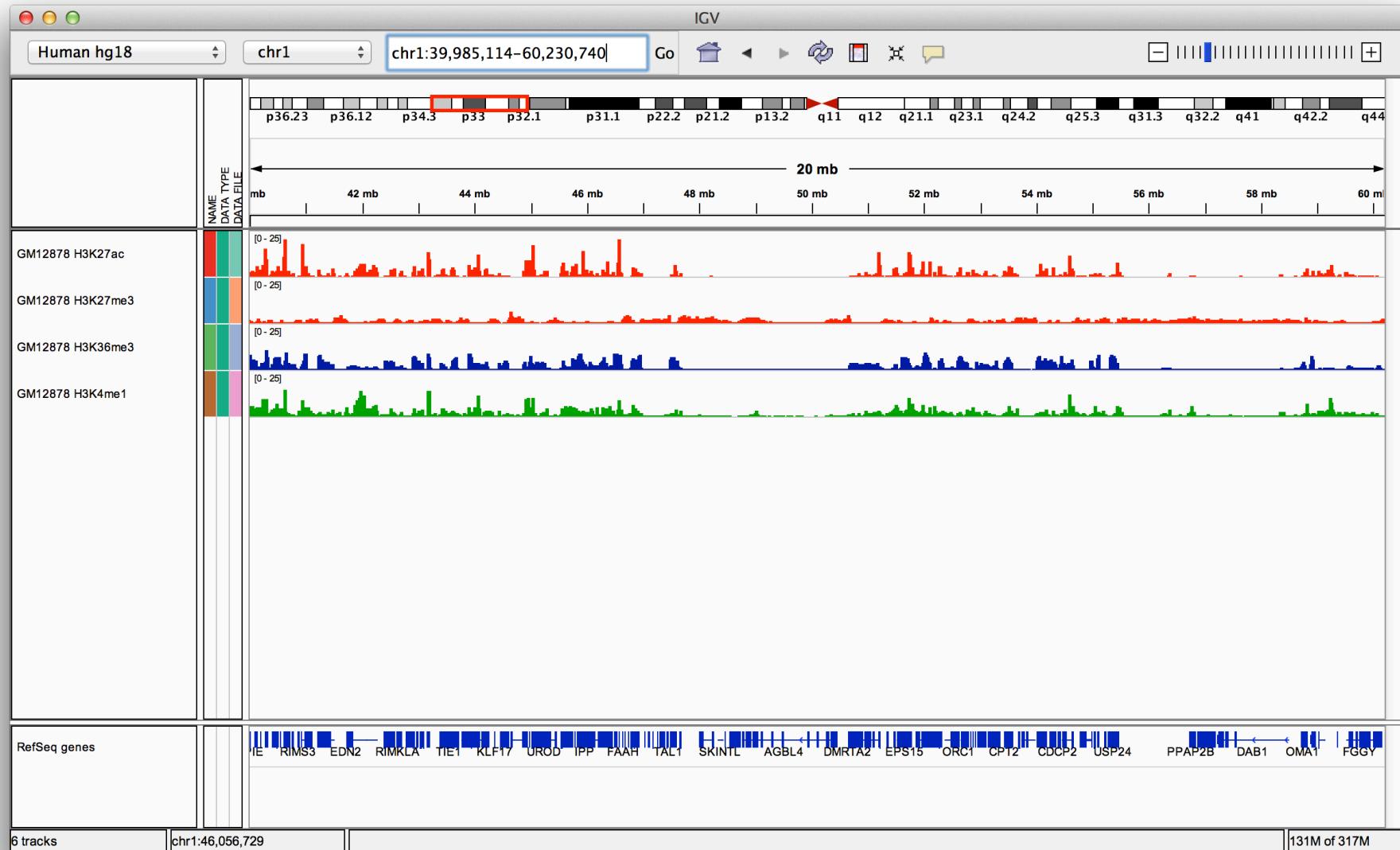
# Navigate



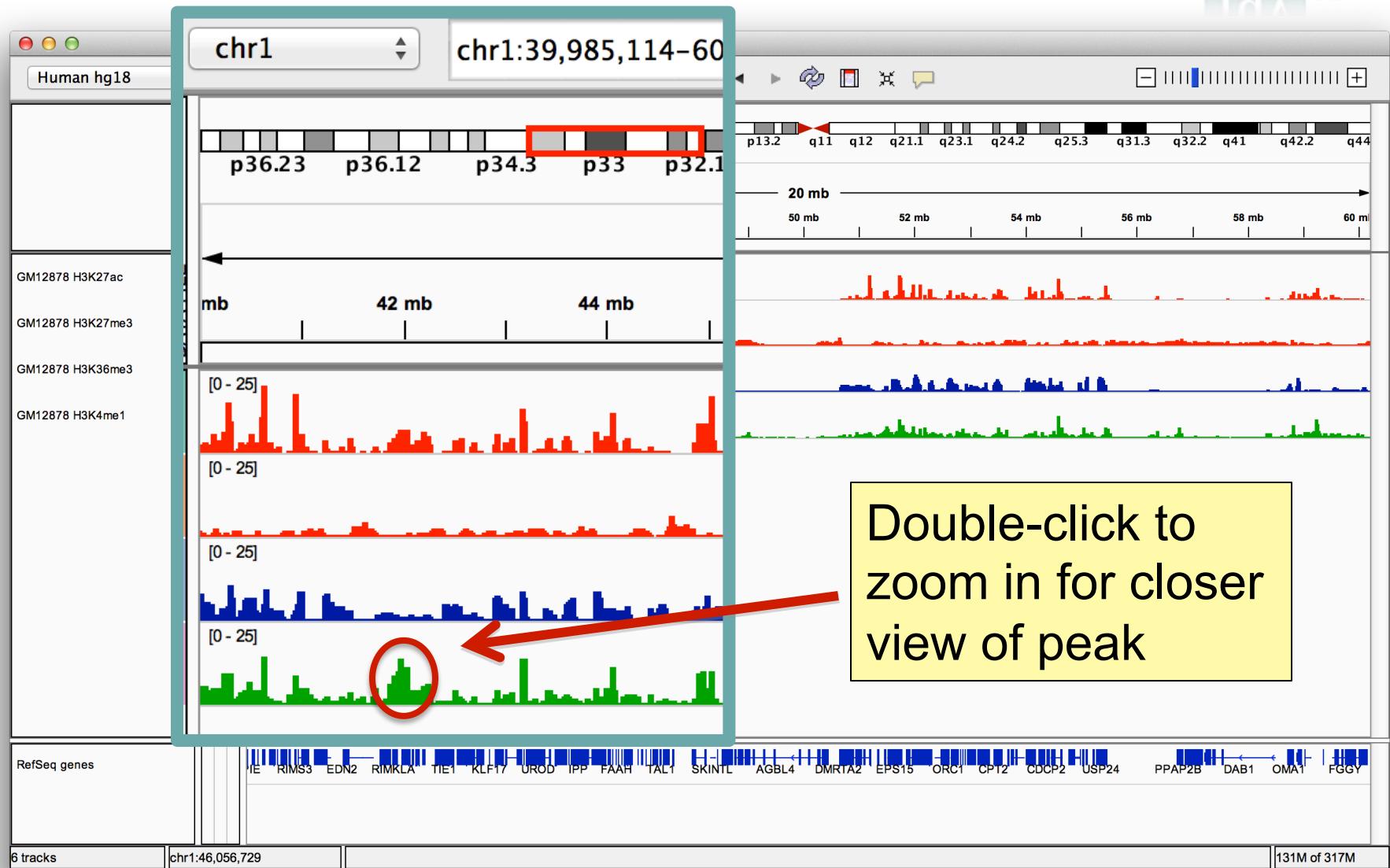
# Navigate



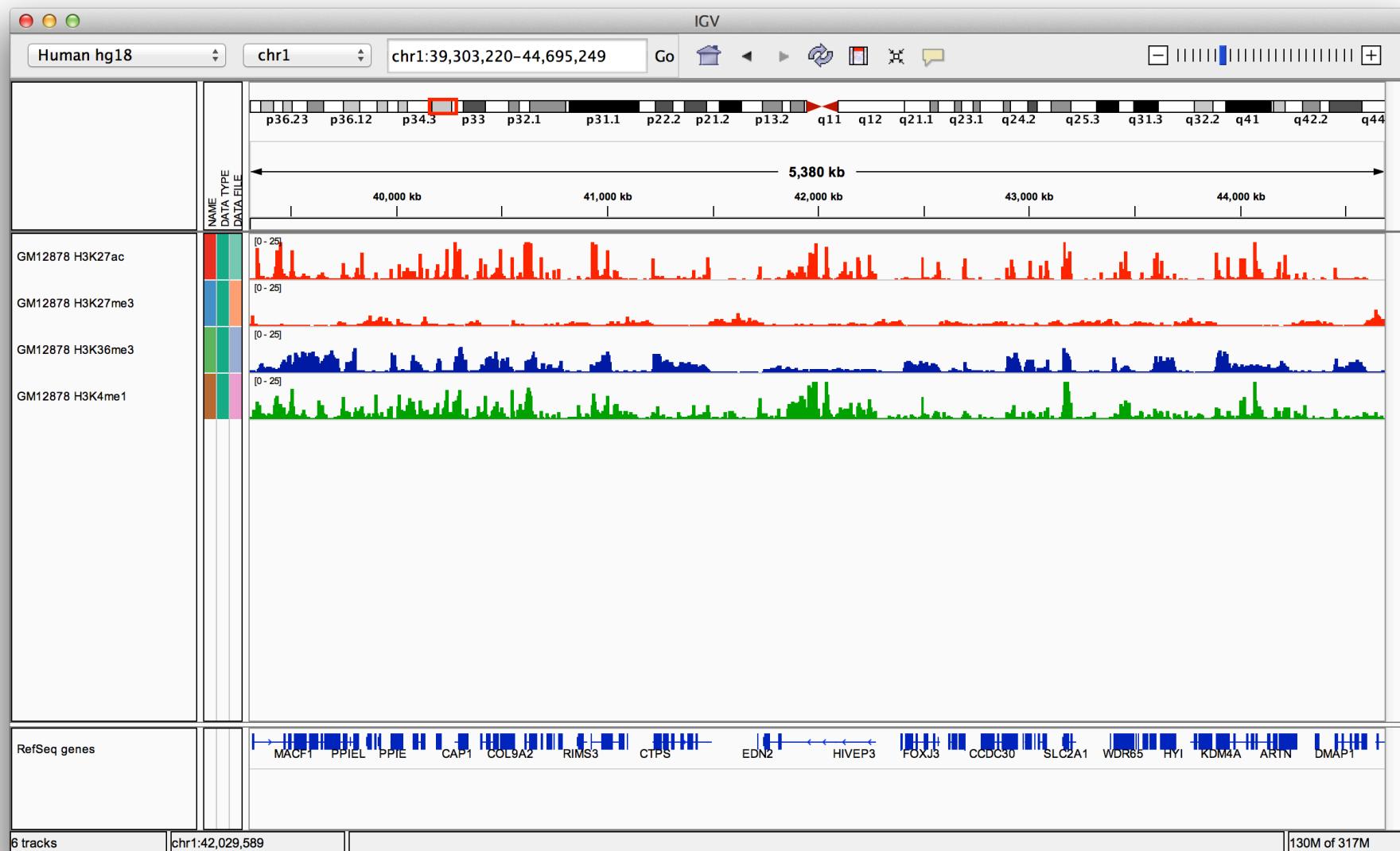
# Navigate



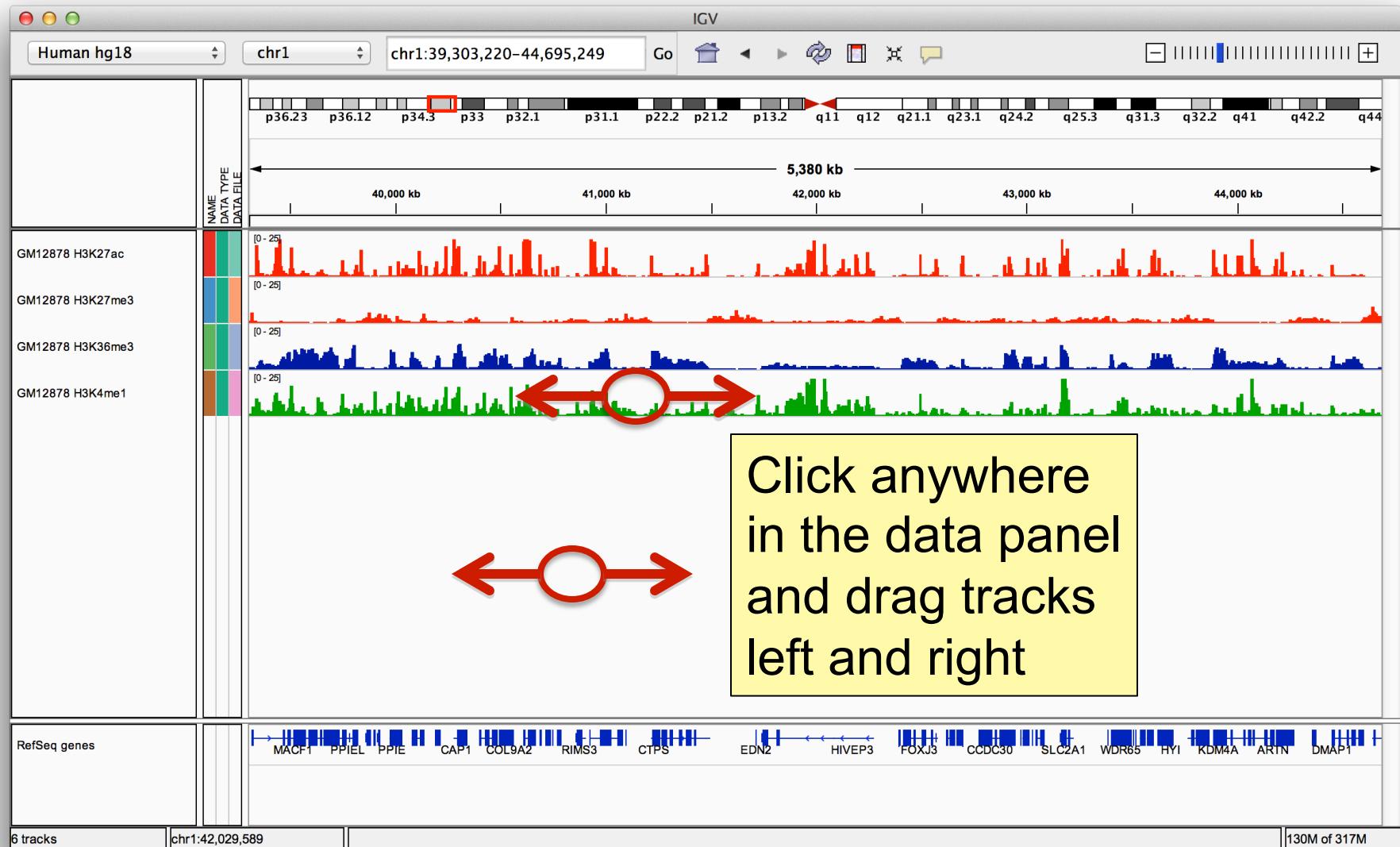
# Navigate



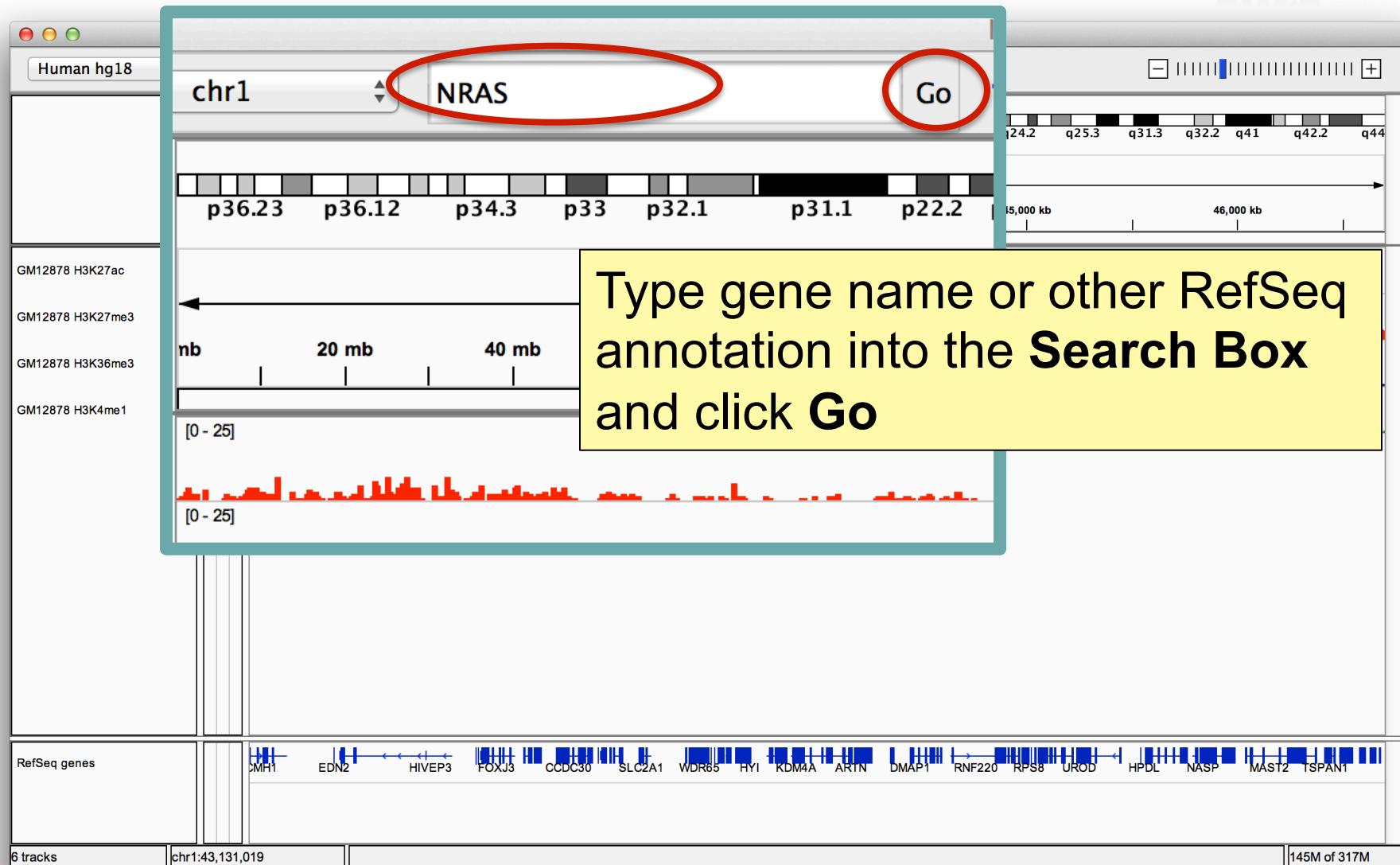
# Navigate



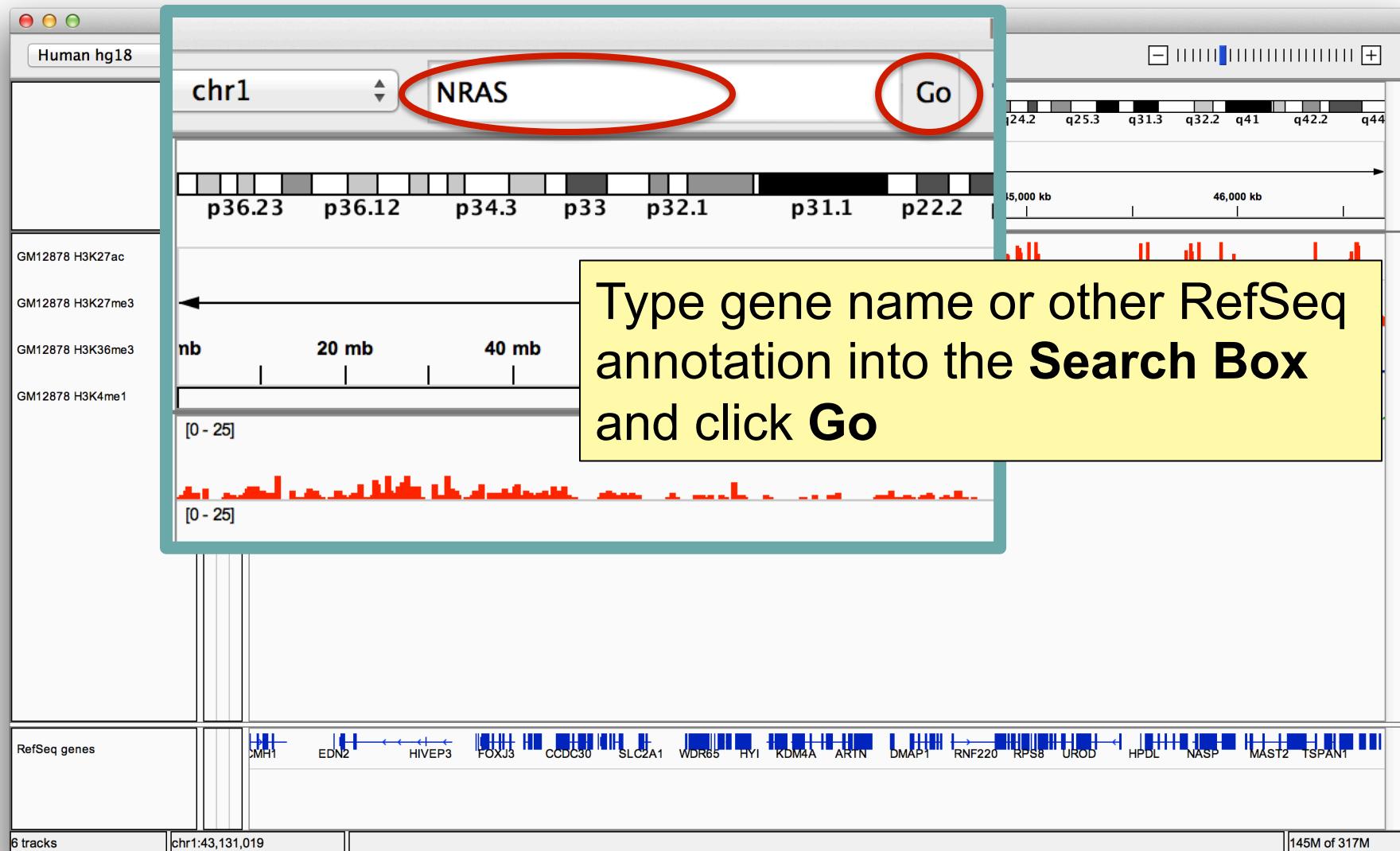
# Navigate



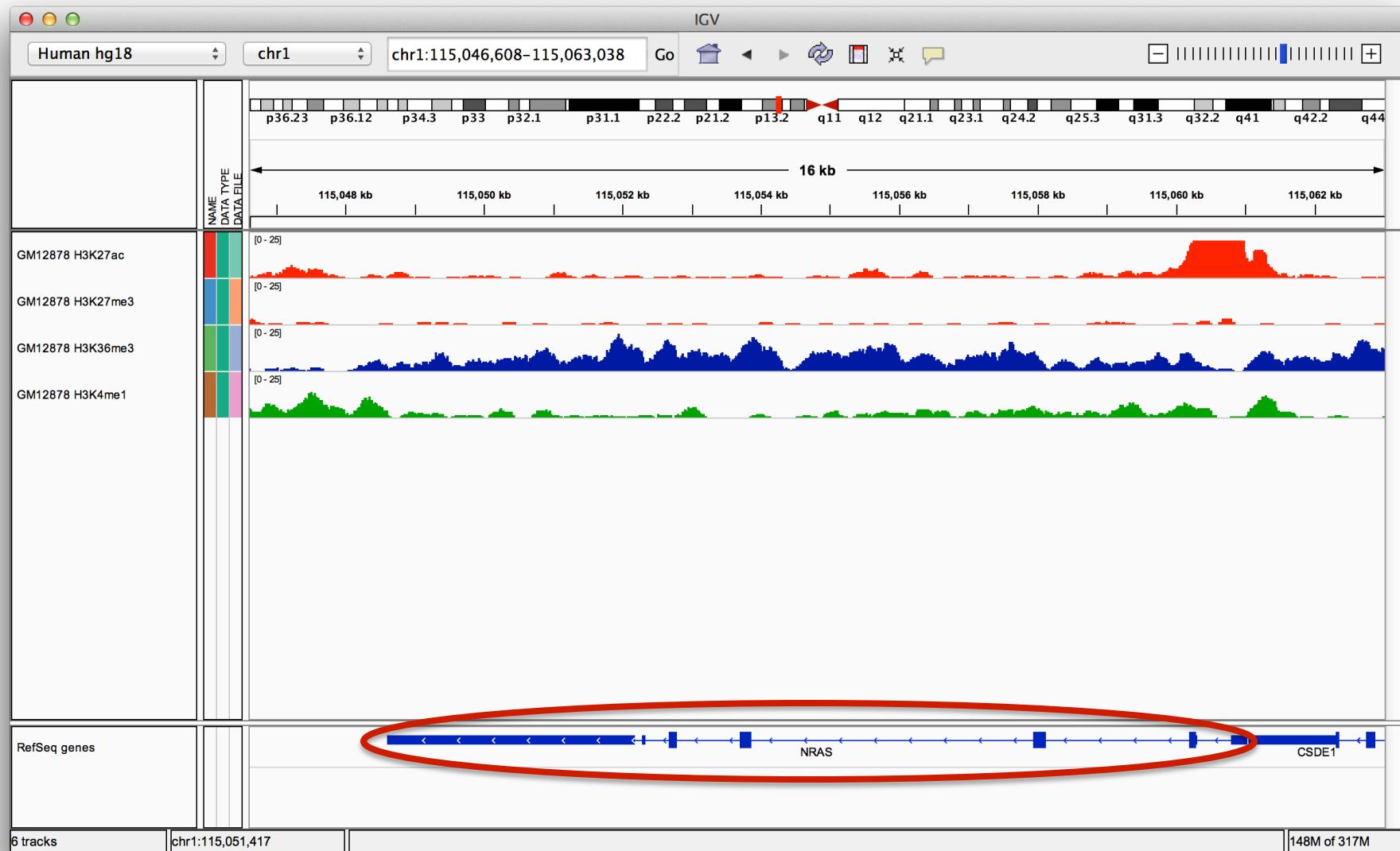
# Navigate



# Navigate



# Navigate

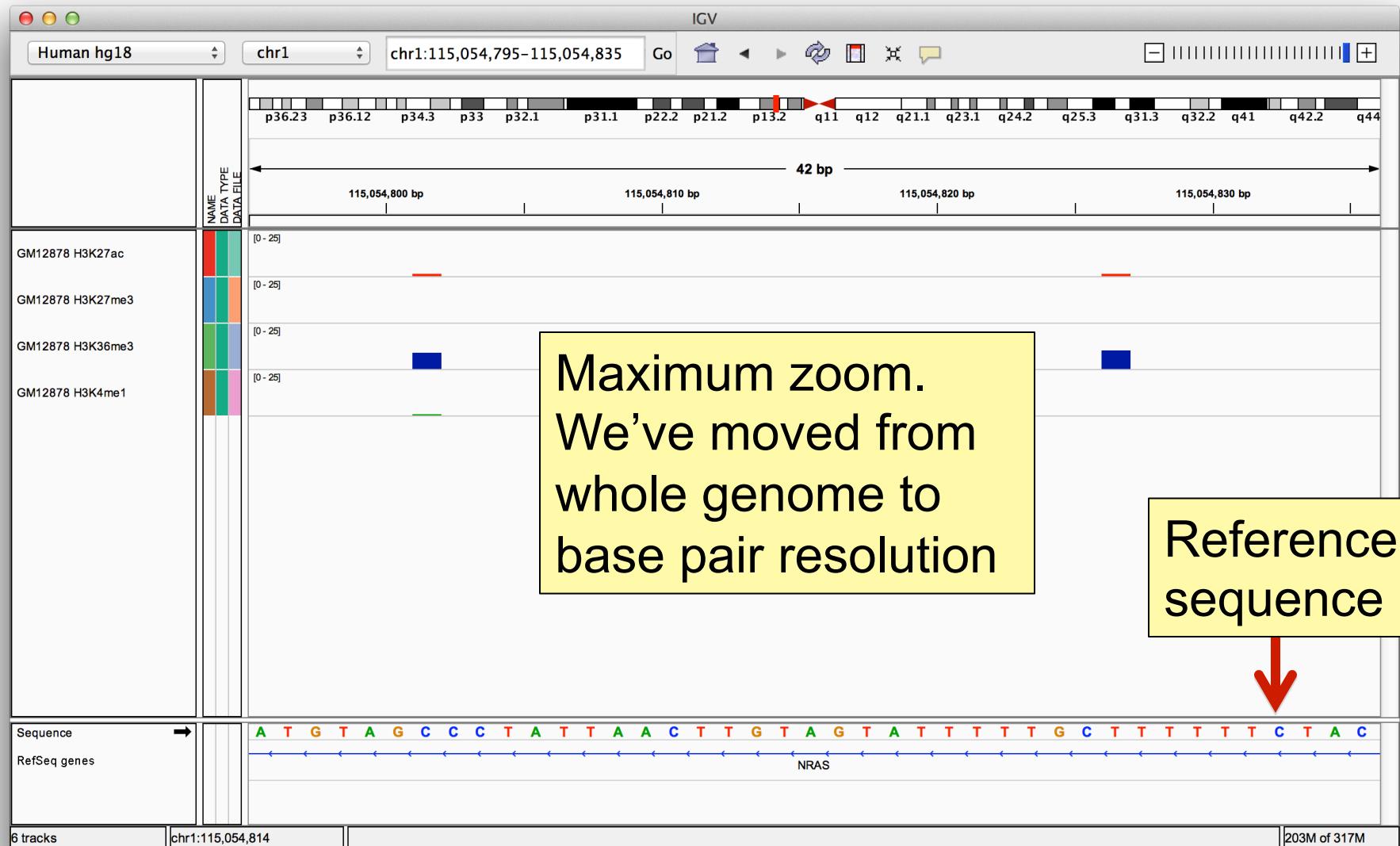


# Navigate



Click on the last tick on the “railroad track” to zoom in to maximum resolution

# Navigate

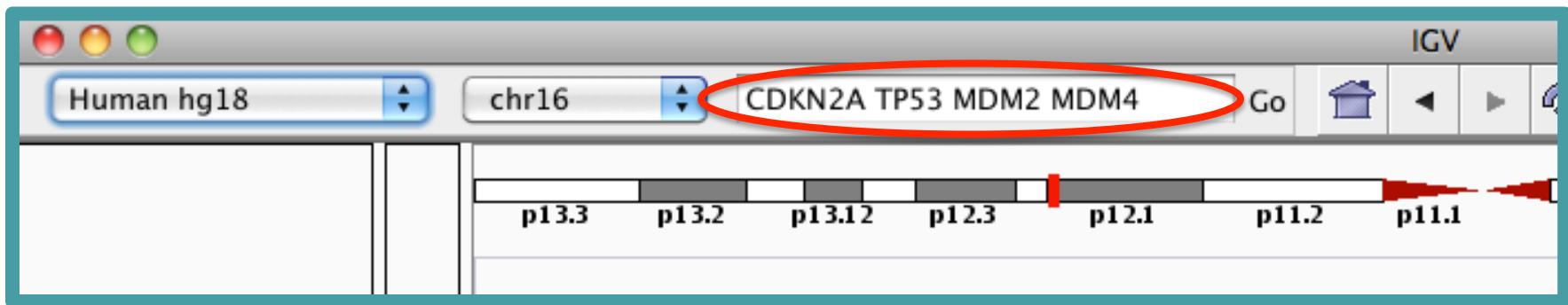


# Viewing multiple regions



- **Search box**

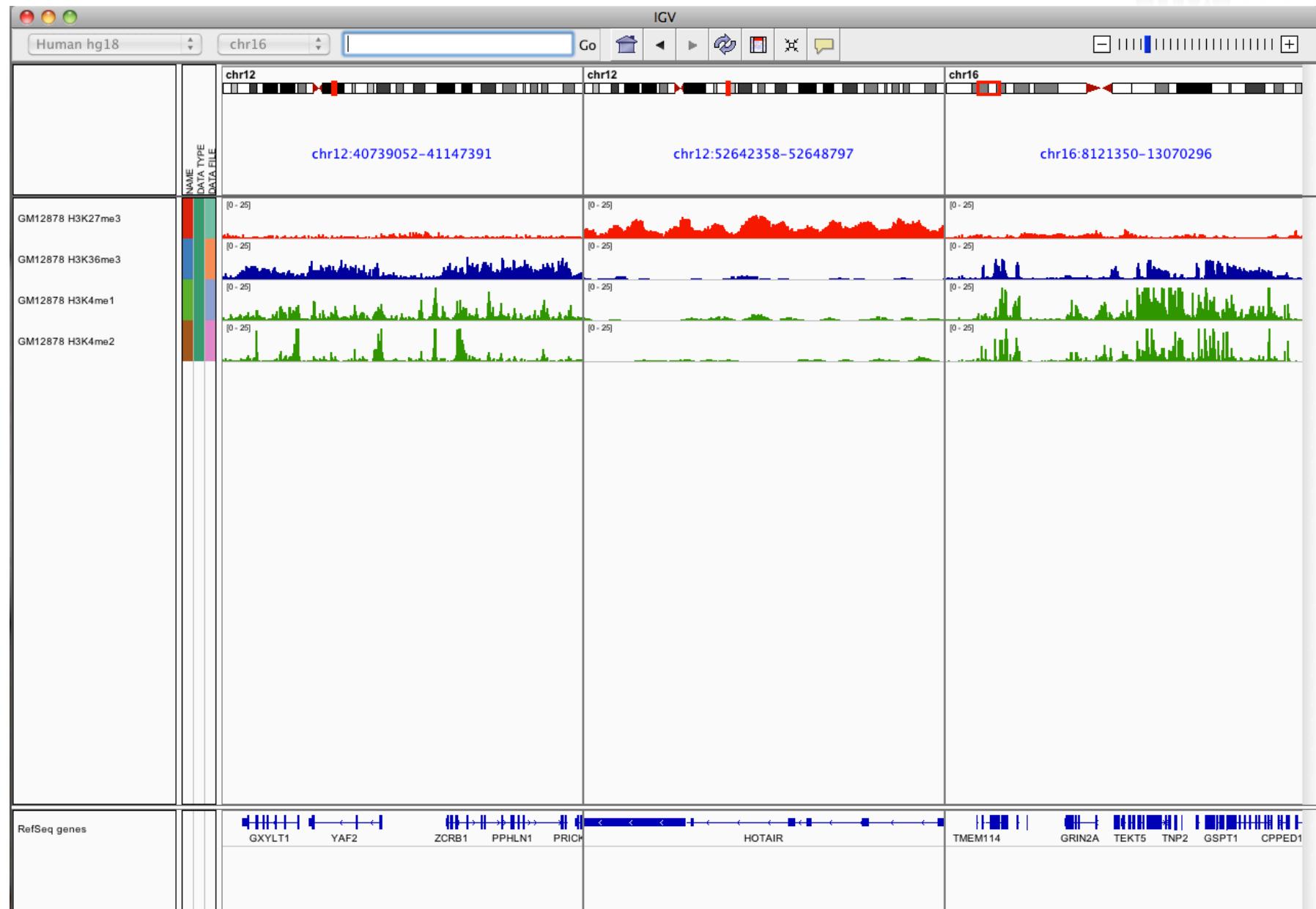
Enter multiple loci or features in the search box



- **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.

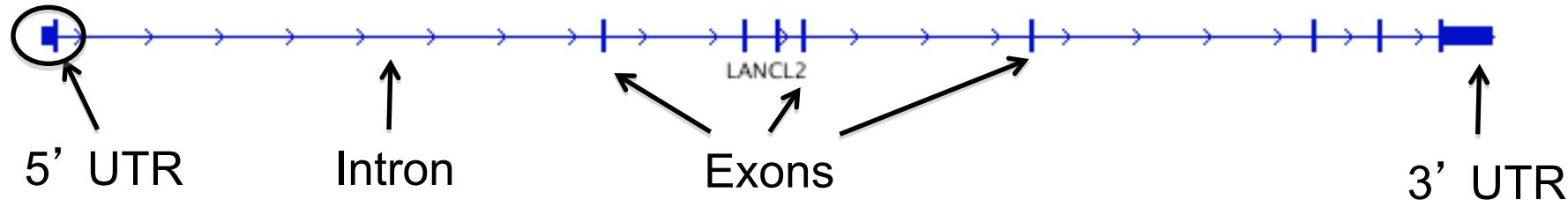
- [BAM](#)
- [BED](#)
- [BedGraph](#)
- [bigBed](#)
- [bigWig](#)
- [Birdsuite Files](#)
- [broadPeak](#)
- [CBS](#)
- [CN](#)
- [Cufflinks Files](#)
- [Custom File Formats](#)
- [Cytoband](#)
- [FASTA](#)
- [GCT](#)
- [genePred](#)
- [GFF](#)
- [GISTIC](#)
- [Goby](#)
- [GWAS](#)
- [IGV](#)
- [LOH](#)
- [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](http://www.broadinstitute.org/igv/FileFormats)

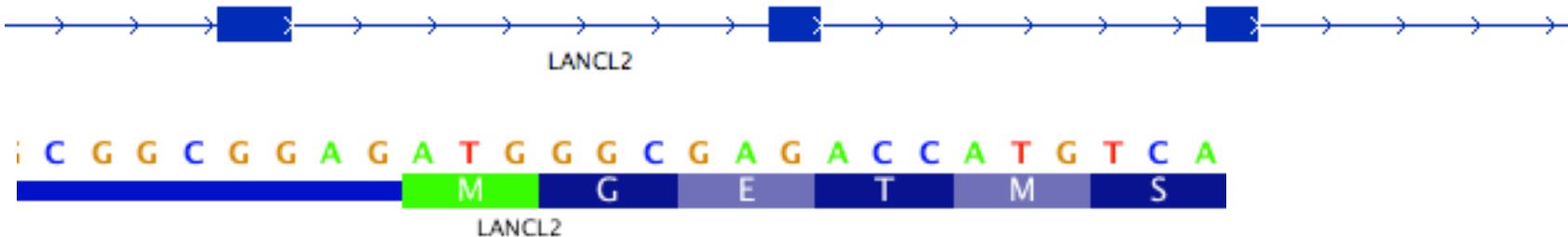
# Genome annotation track



## UCSC style gene representation



## Zoomed in views



## Zoomed out views

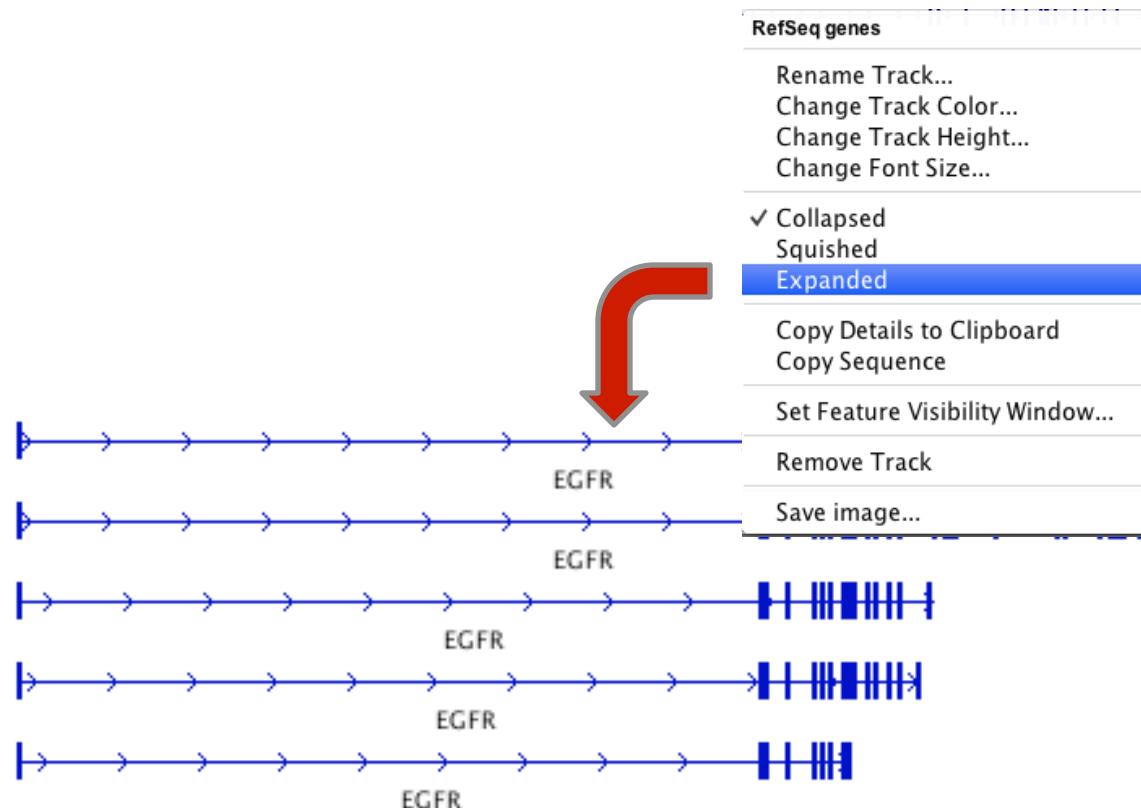


# Annotation display mode

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

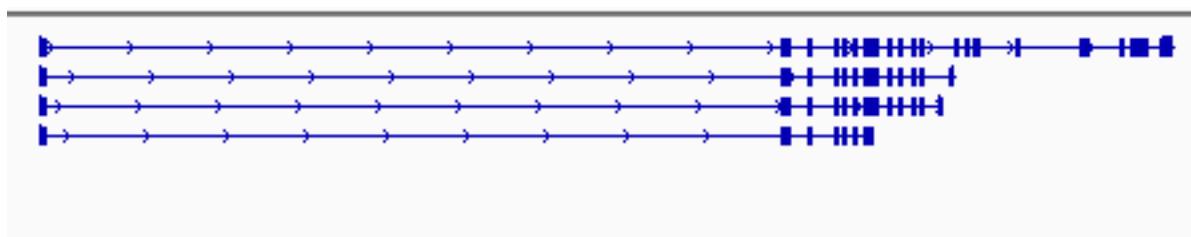
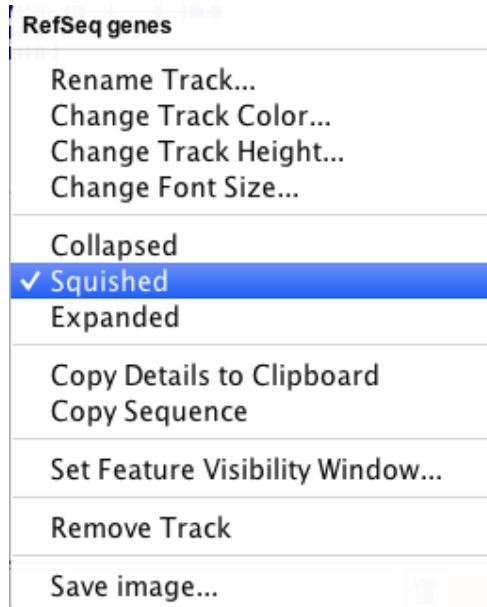


2. Expand the track using the popup menu



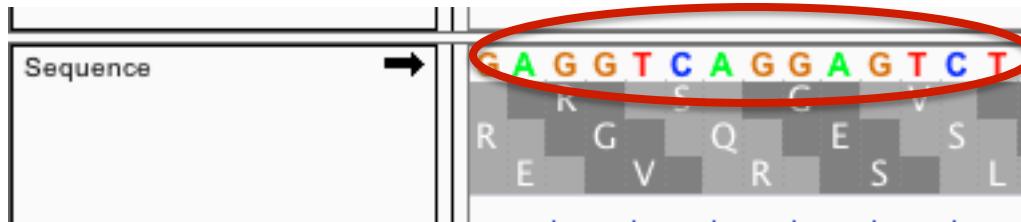
# Annotation display mode

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



# Reference sequence

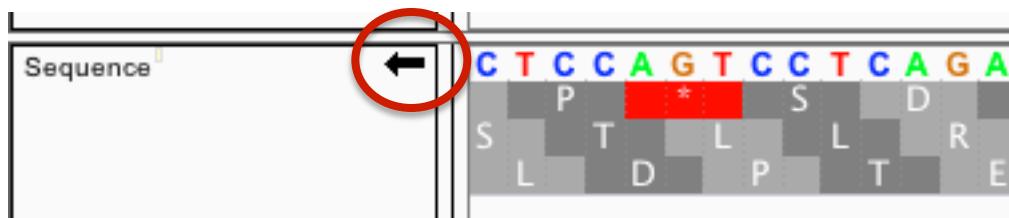
Click anywhere on the sequence to see a 3 frame translation.



By default the sequence for the forward strand is shown.



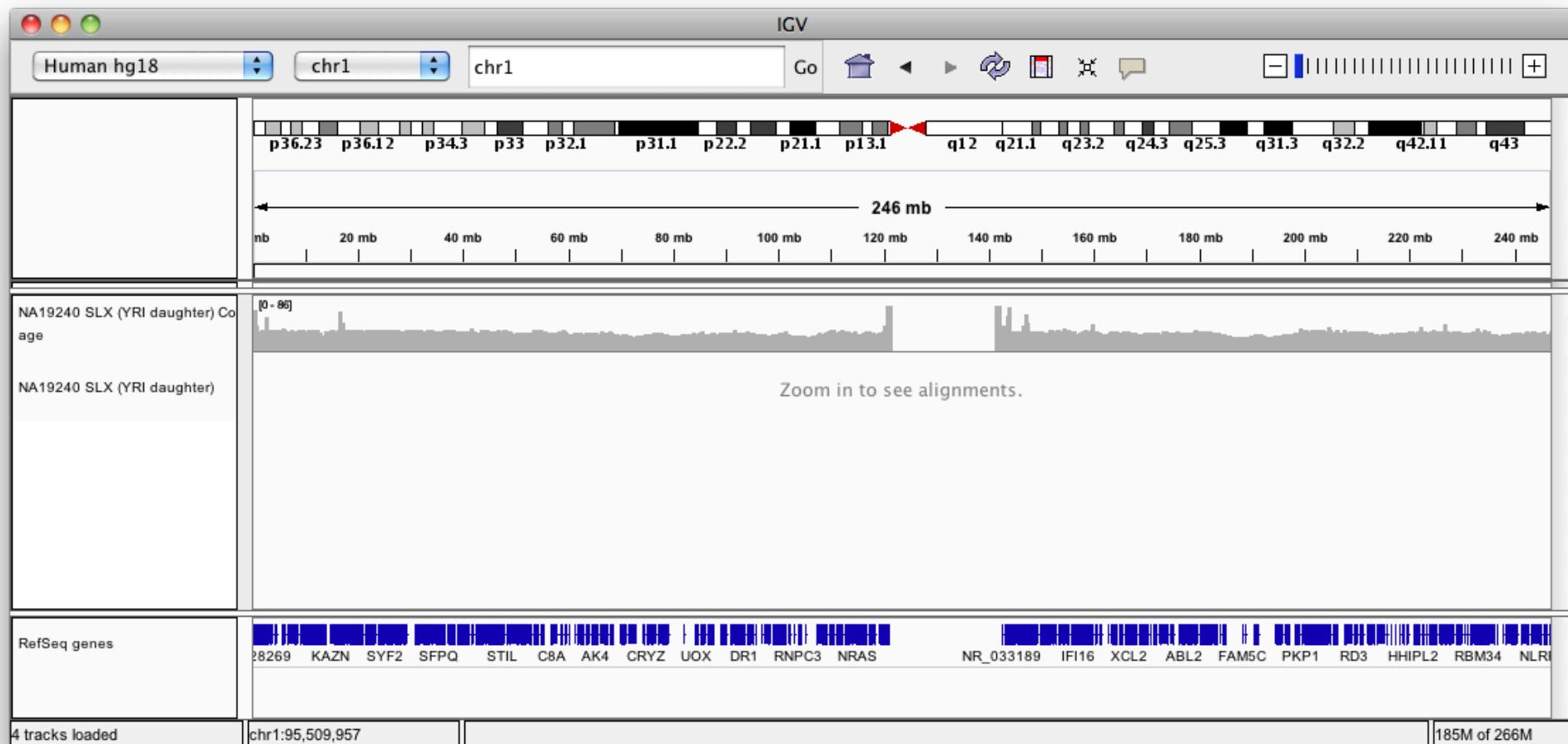
Click the arrow on the left to reverse the strand.



# Viewing NGS Data

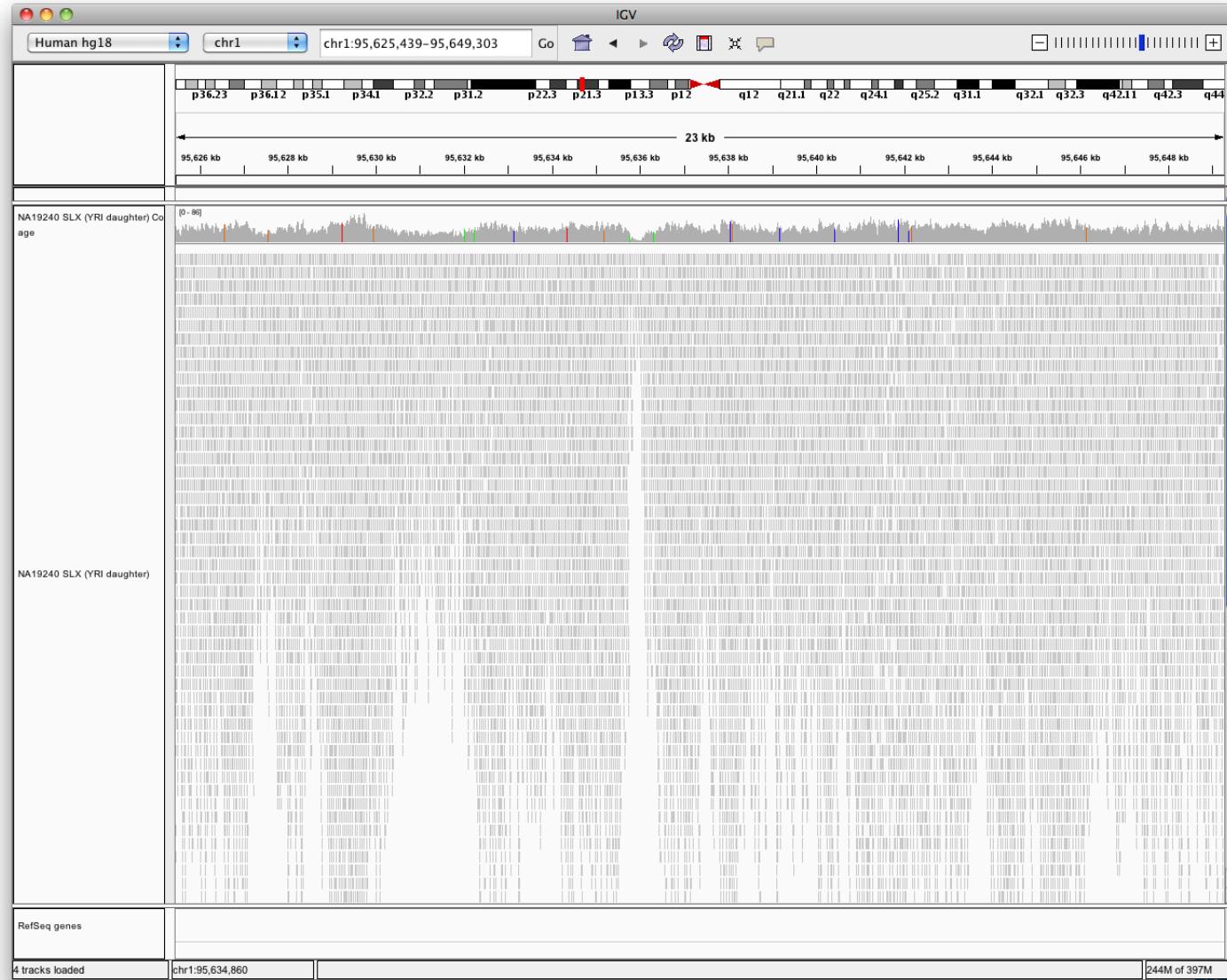
# Viewing alignments

## Whole chromosome view



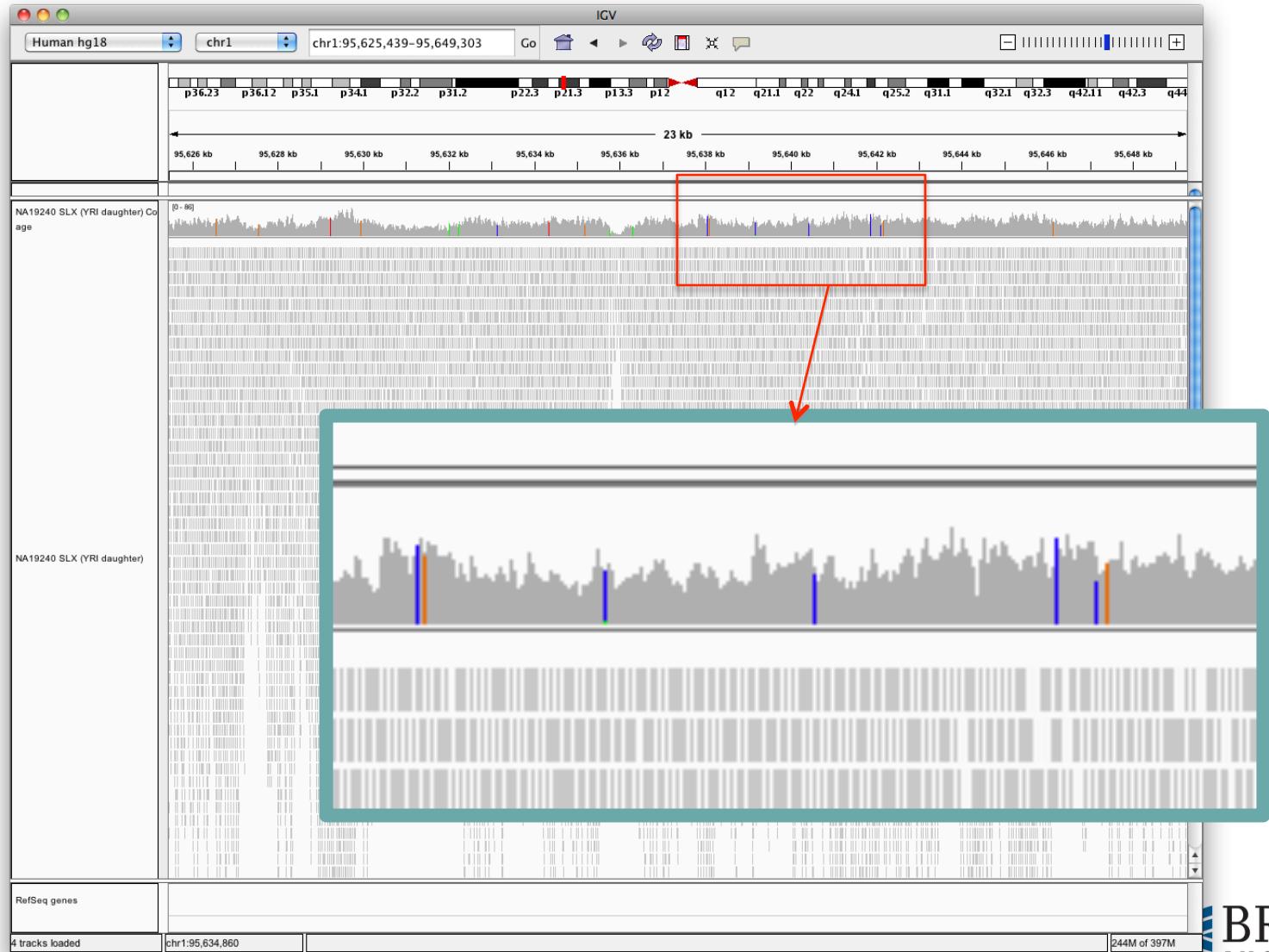
# Viewing alignments

Zoom in to view alignments



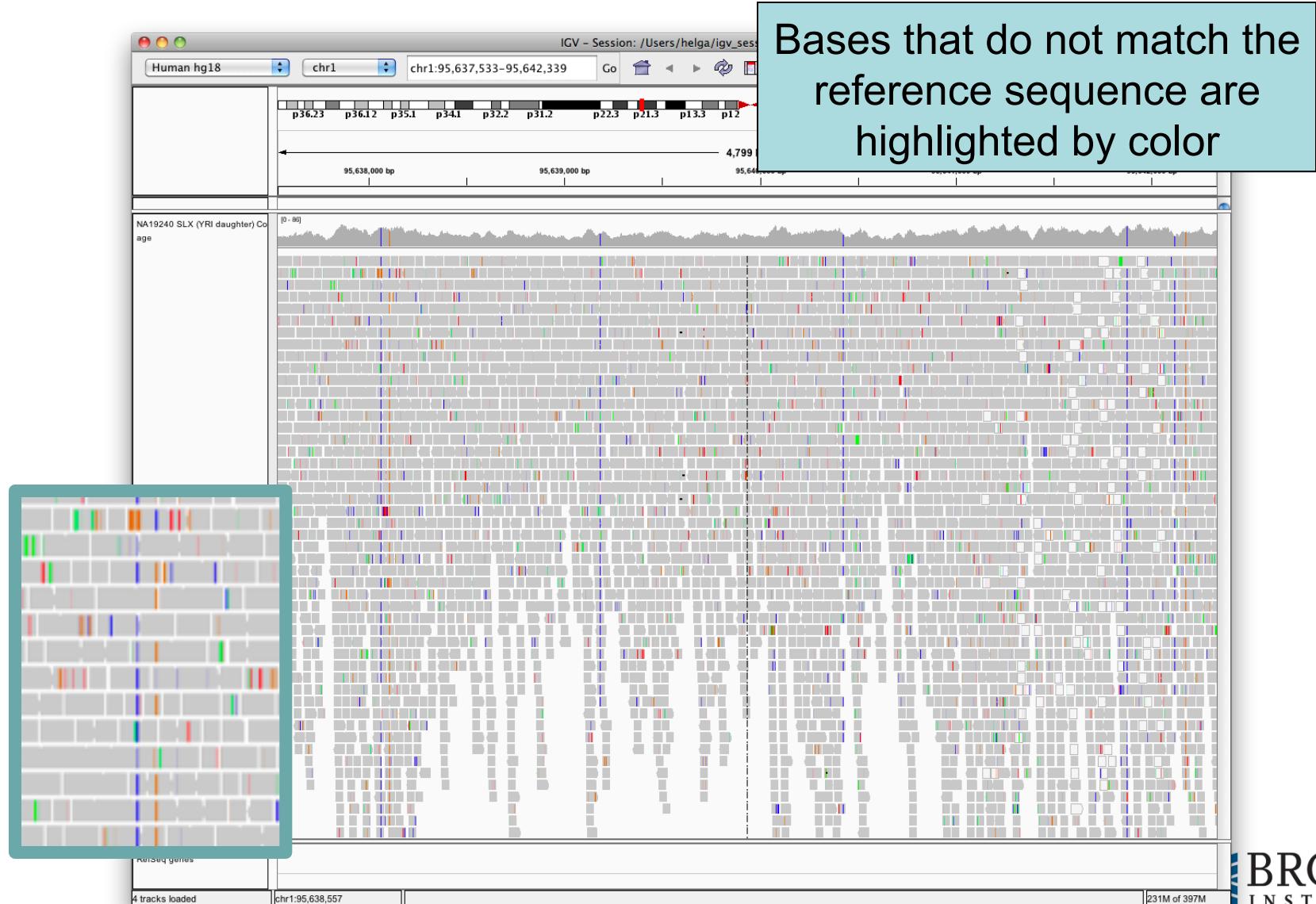
# Viewing alignments

Coverage track now has more detail



# Viewing alignments

Zoom in to see more detail



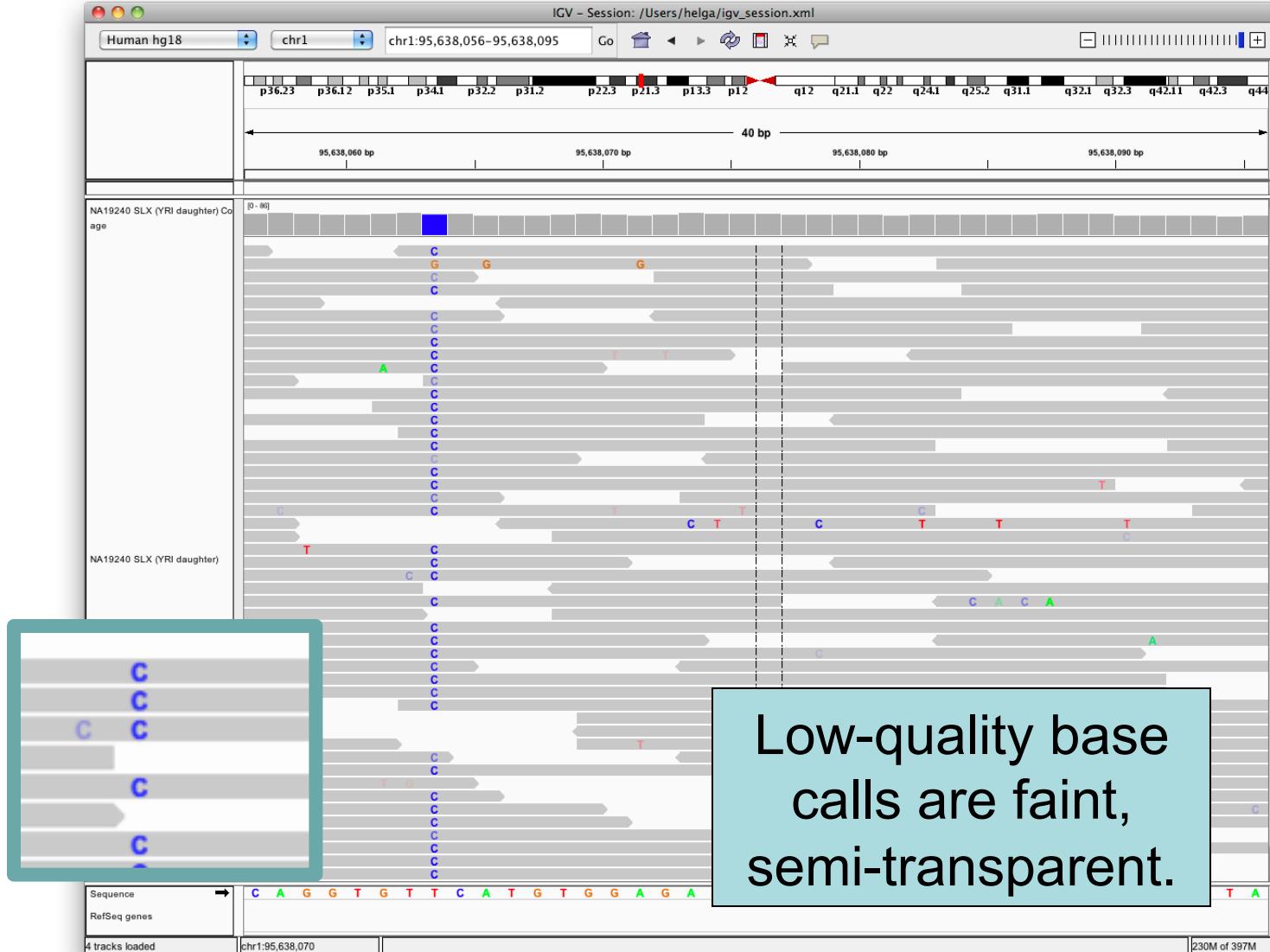
# Viewing alignments

Zoom in to see more detail

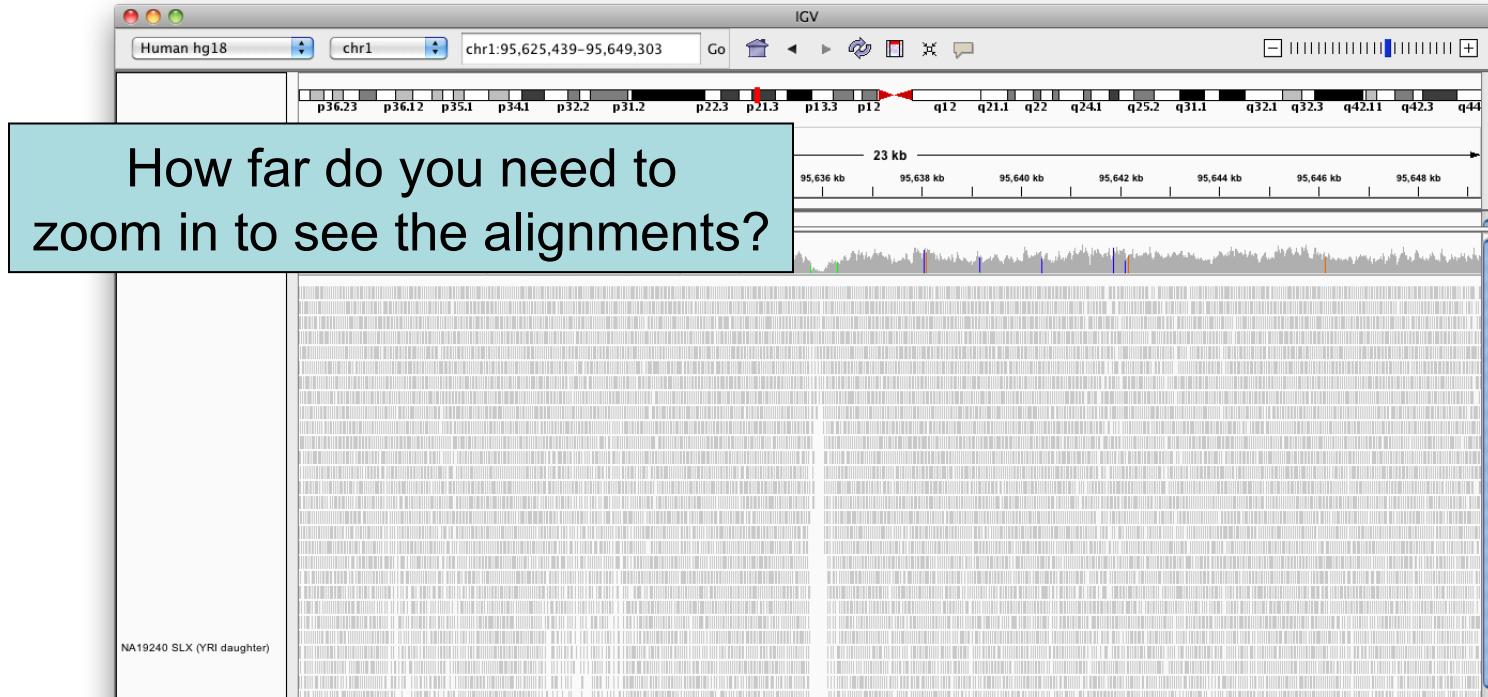


# Viewing alignments

Zoom in to see more detail

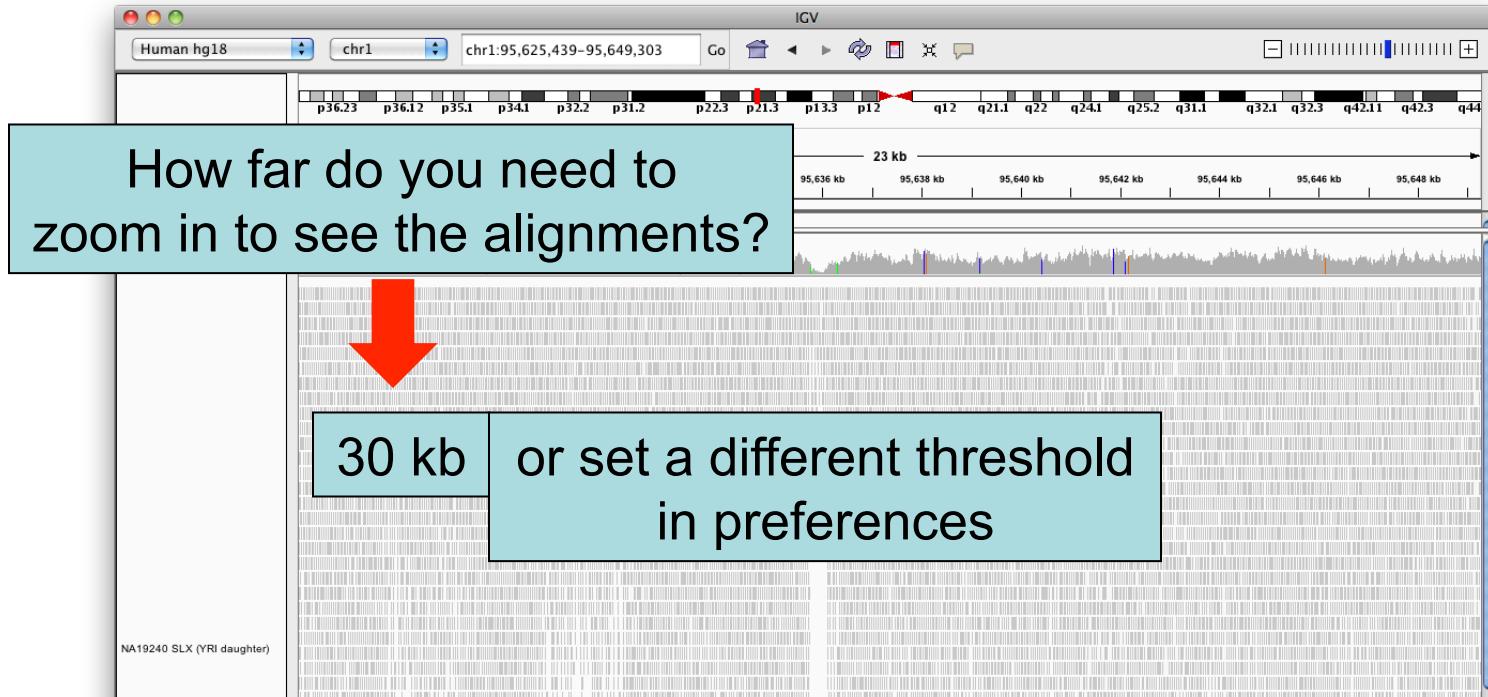


# Viewing alignments



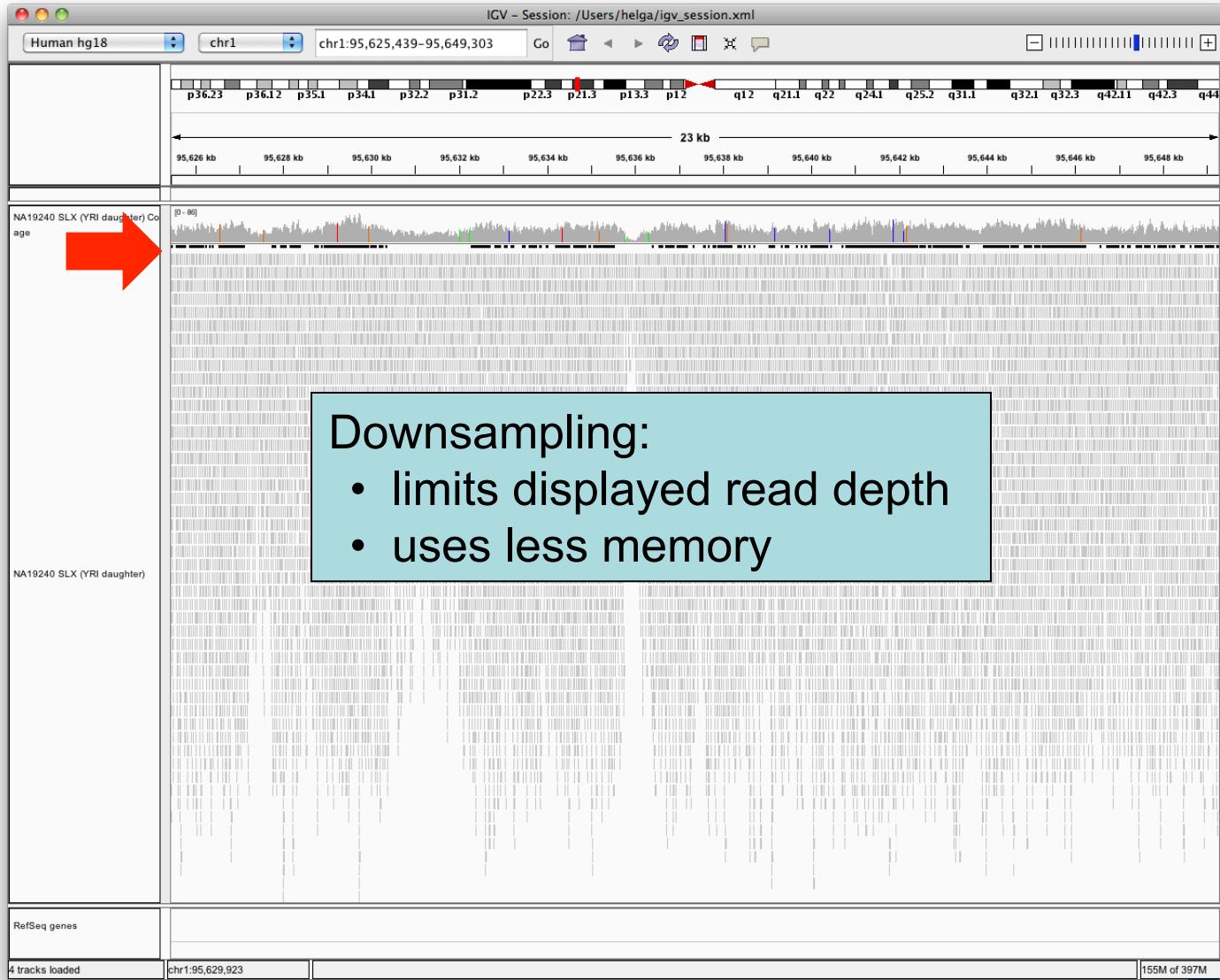
How far do you need to  
zoom in to see the alignments?

# 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



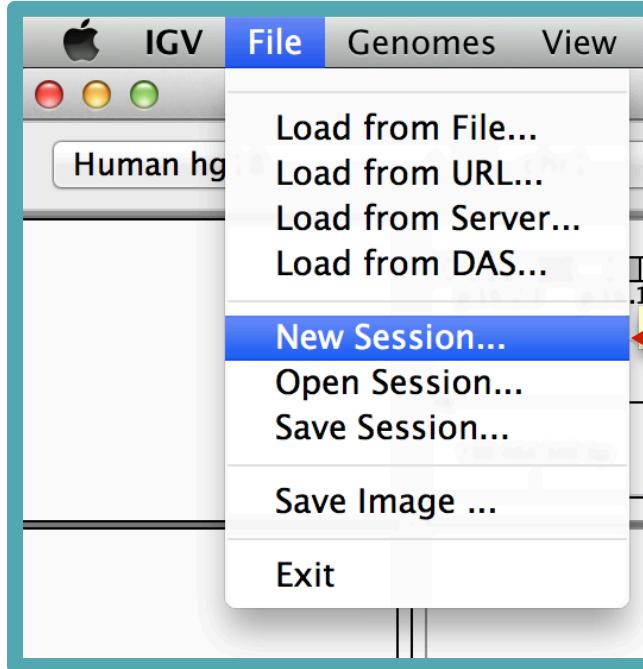
# Viewing SNPs



## Hands-on exercise

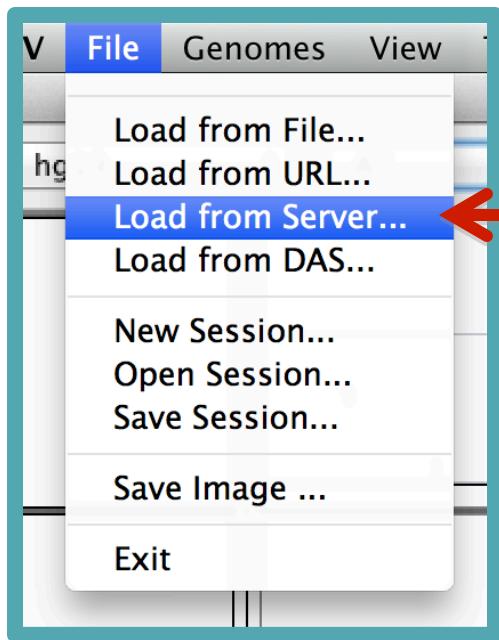
- Load alignments from whole genome sequencing
- View sites where SNPs were called
- Sort and color to highlight patterns

# Viewing SNPs



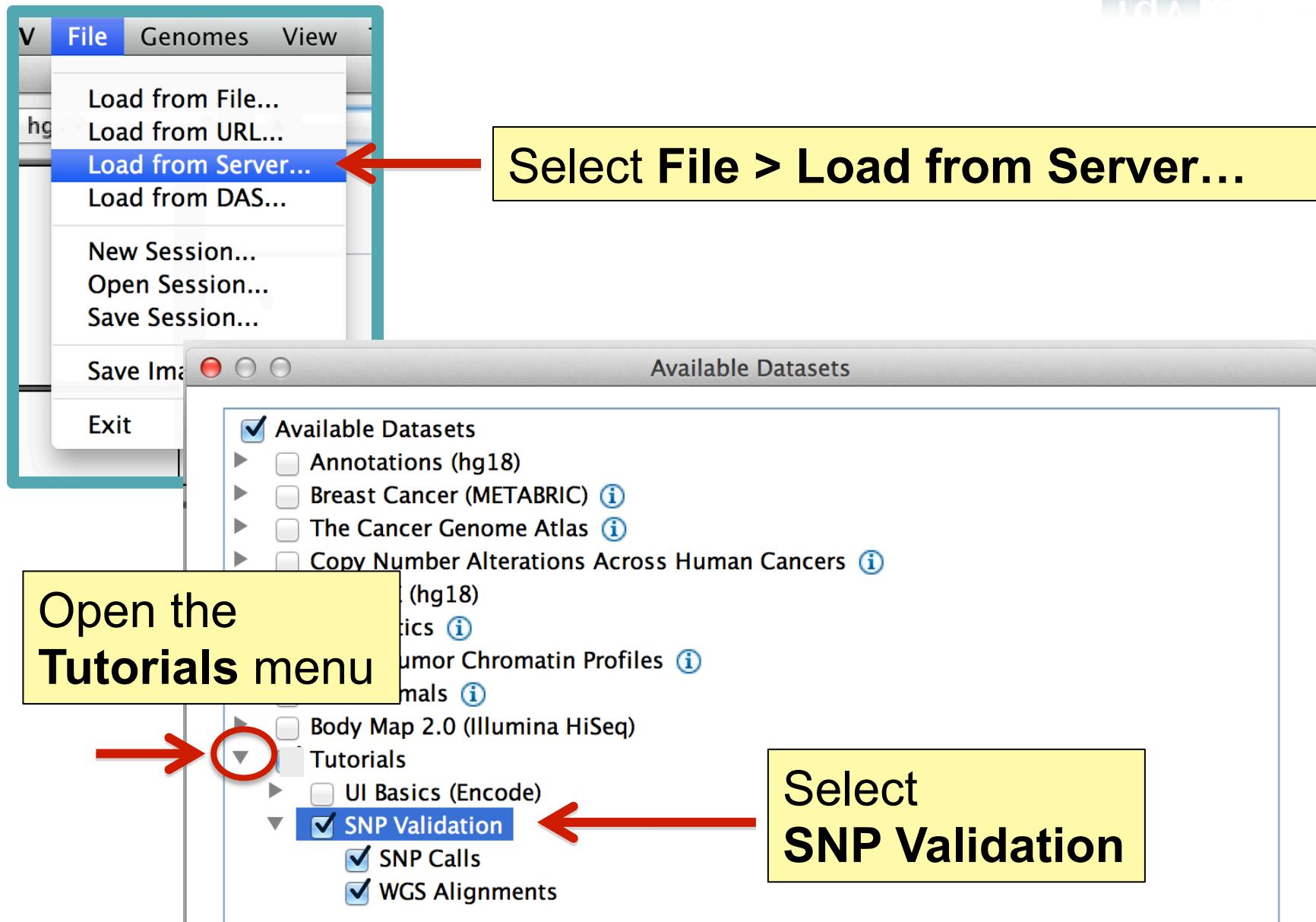
Before we start:  
**Select File > New Session**  
to clear IGV window

# Viewing SNPs

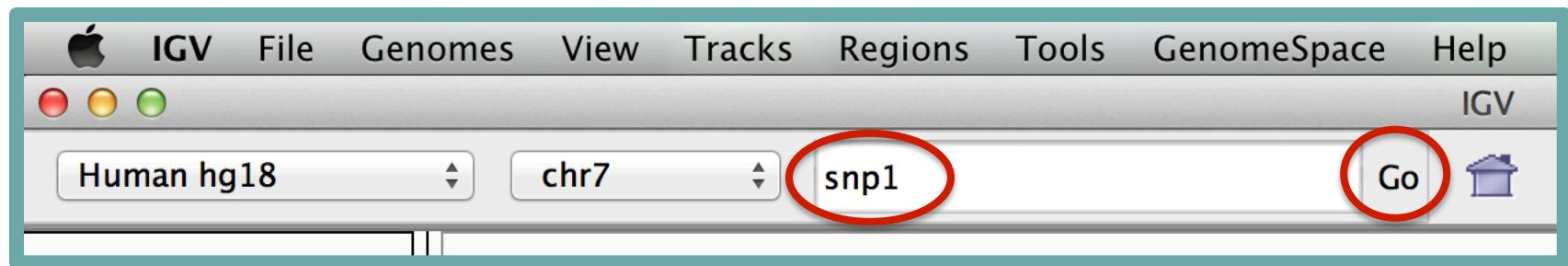


Select File > Load from Server...

# Viewing SNPs

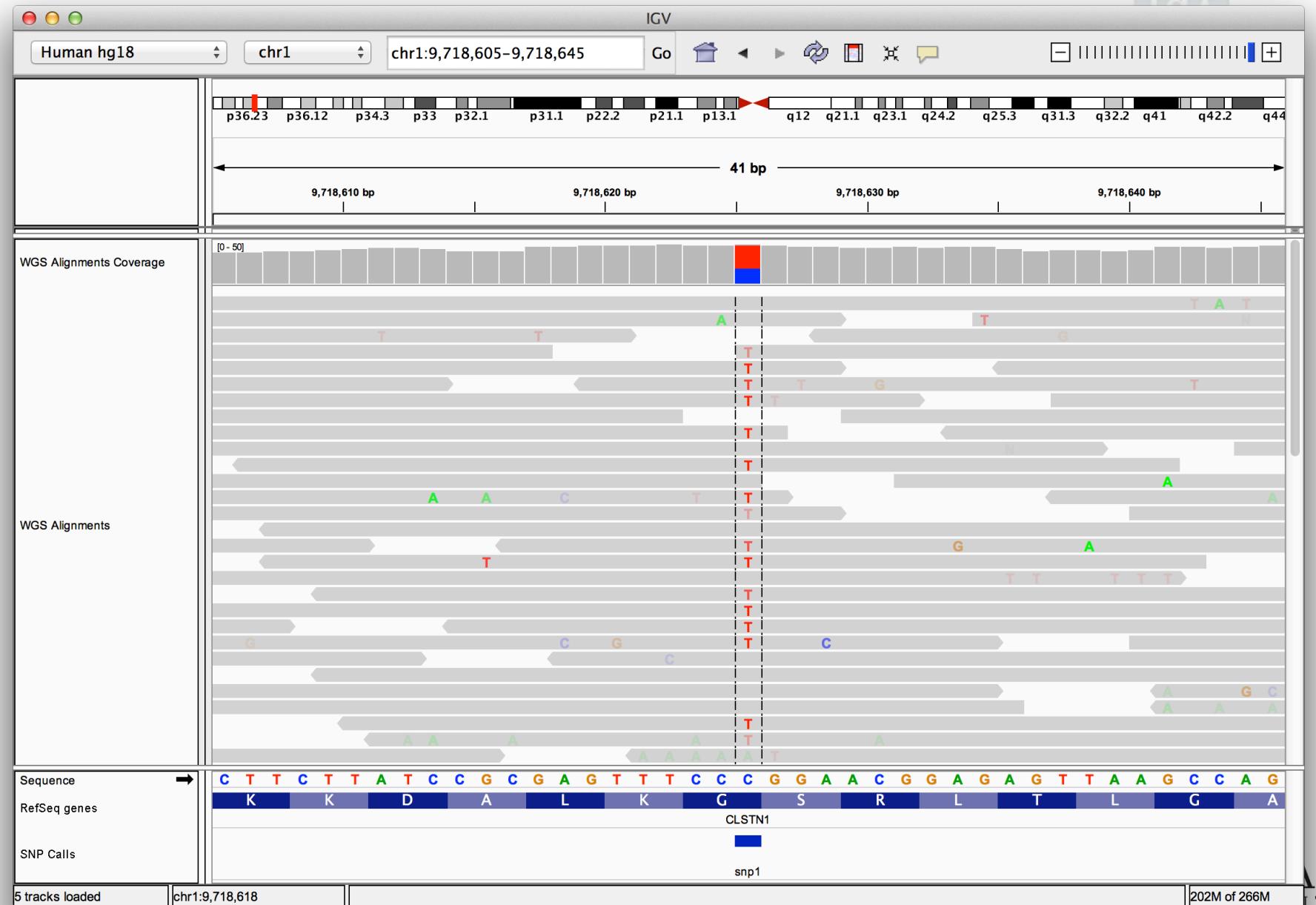


# Viewing SNPs

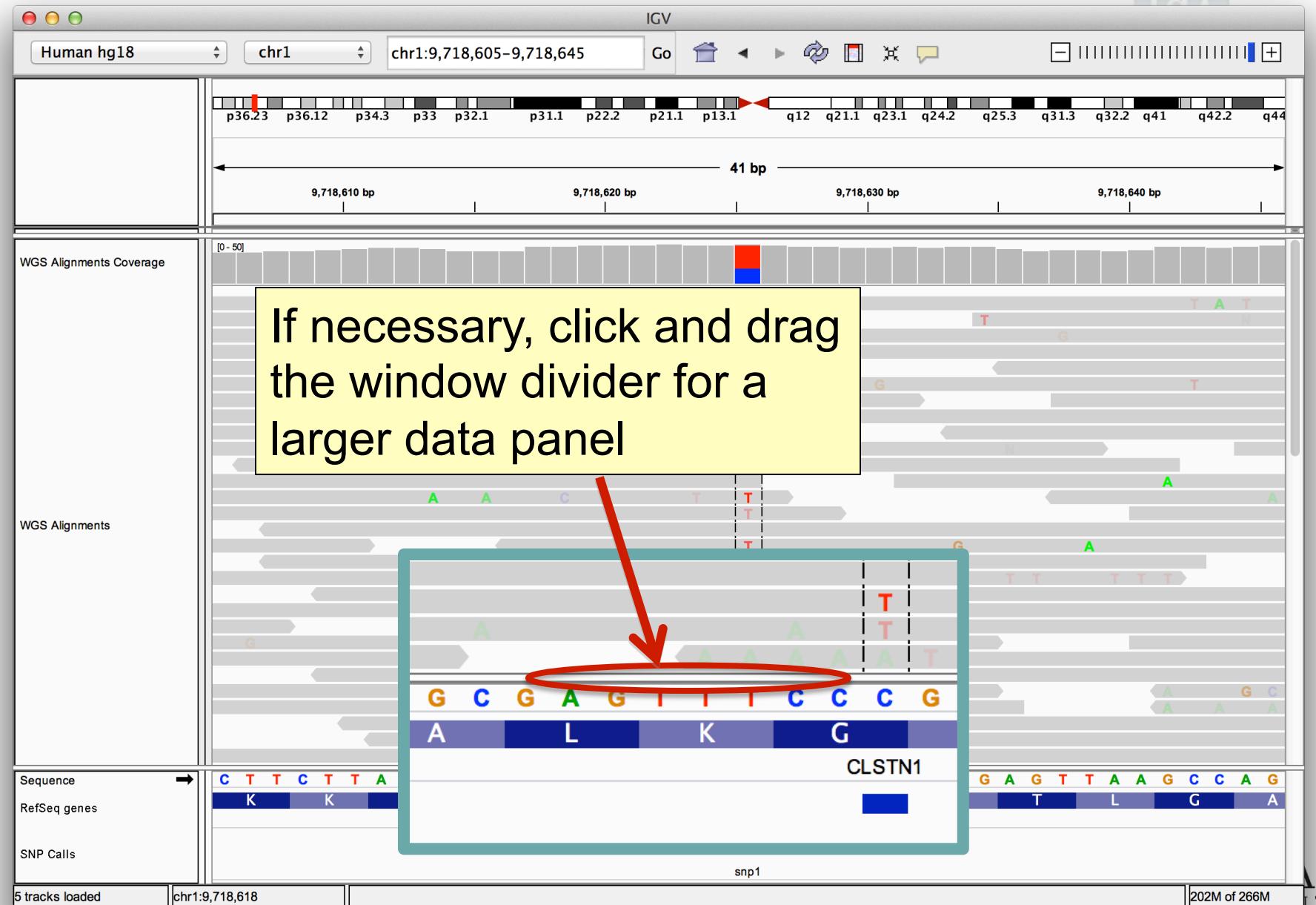


Type “snp1” in the **Search Box**  
and click **Go**

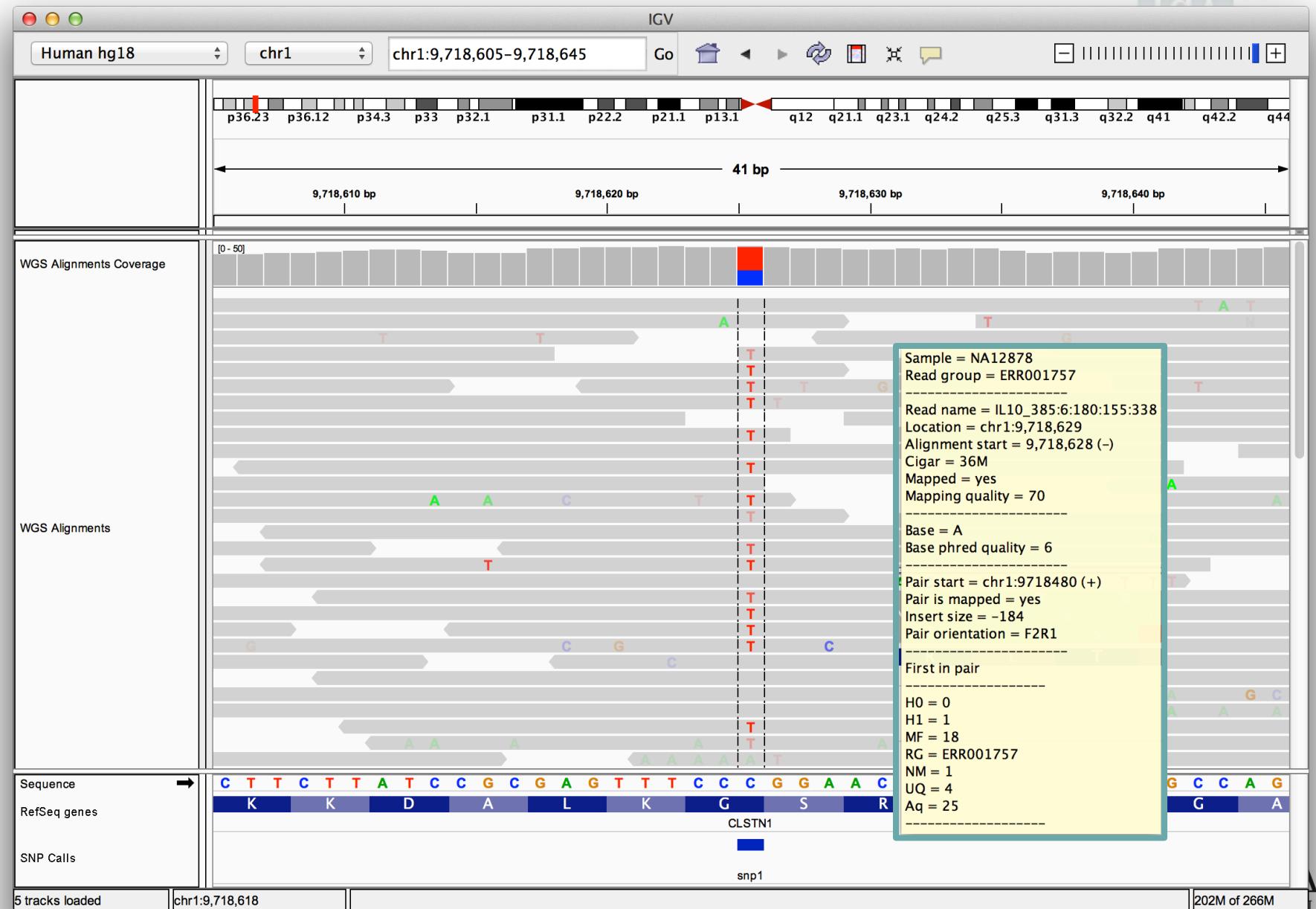
# Viewing SNPs



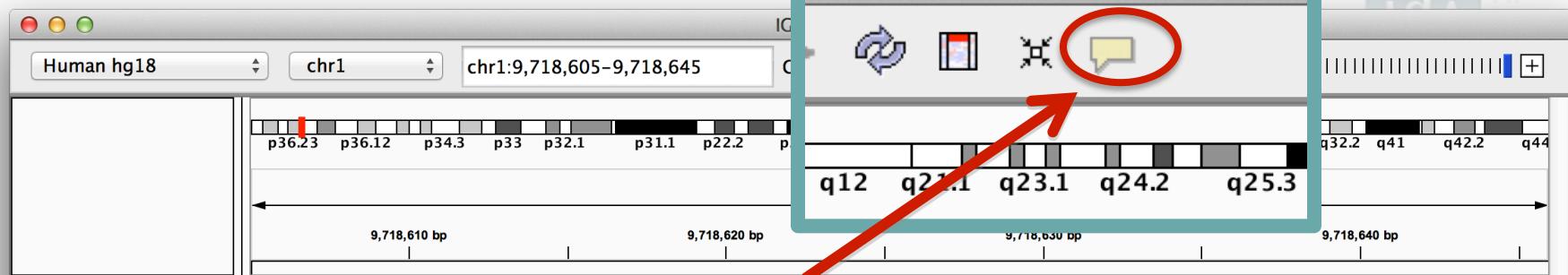
# Viewing SNPs



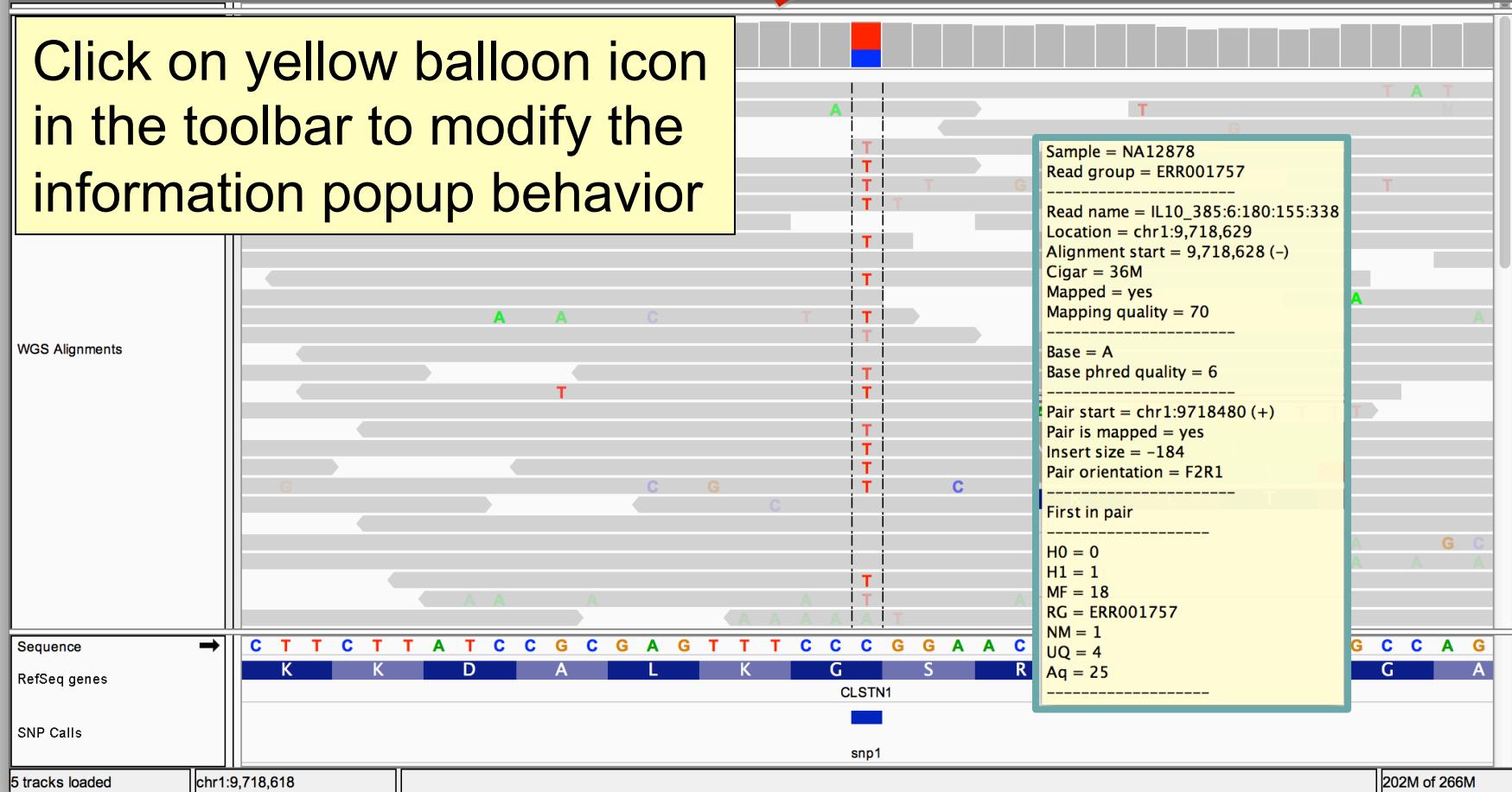
# Viewing SNPs



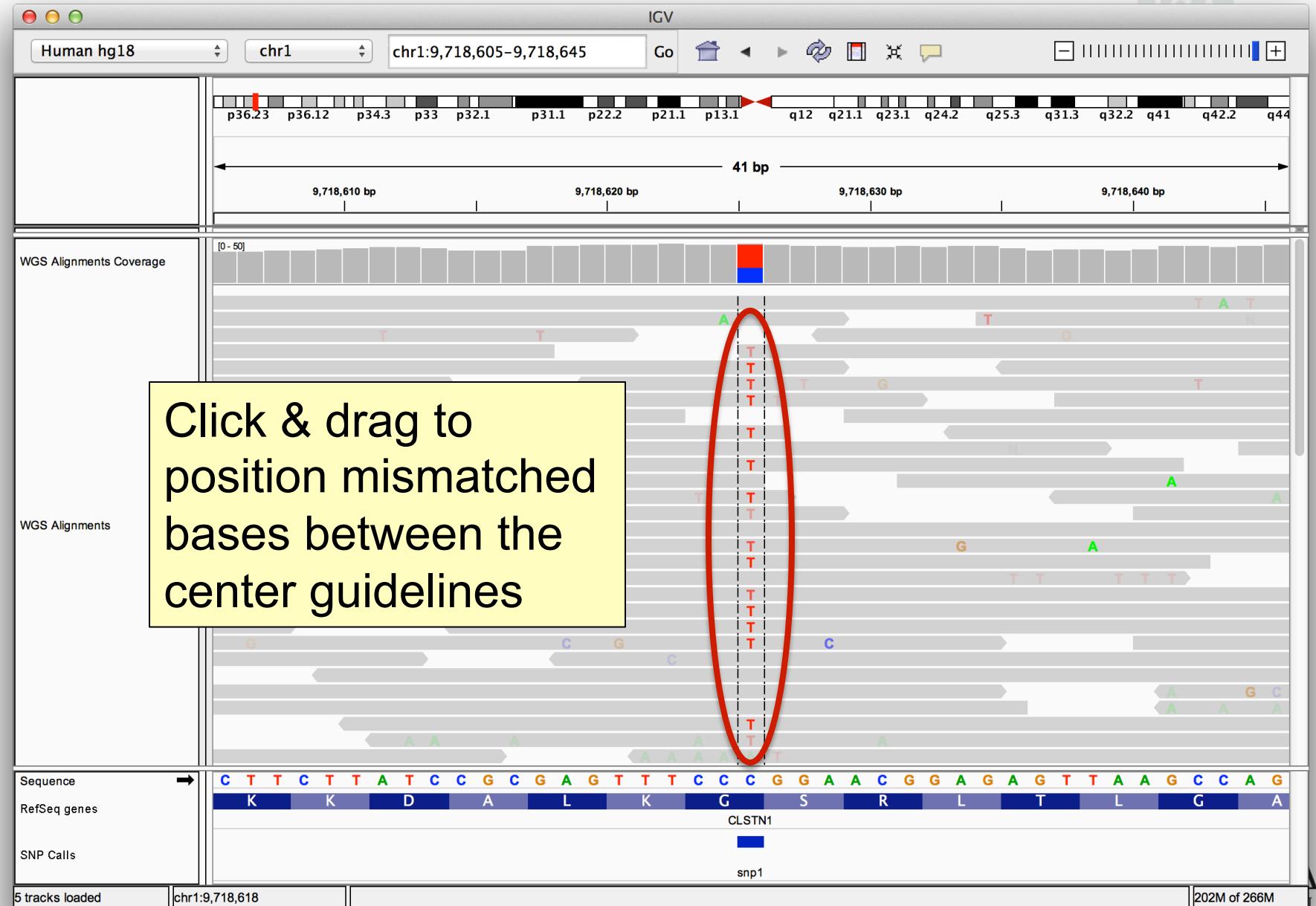
# Viewing SNPs



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 → C T T C T T A T C C G C G A G T T T C C C G G A A C G G A G T T T A A G C C C A G G  
K K D A L K G S R L T L G A SNP Calls CLSTN1  
snp1

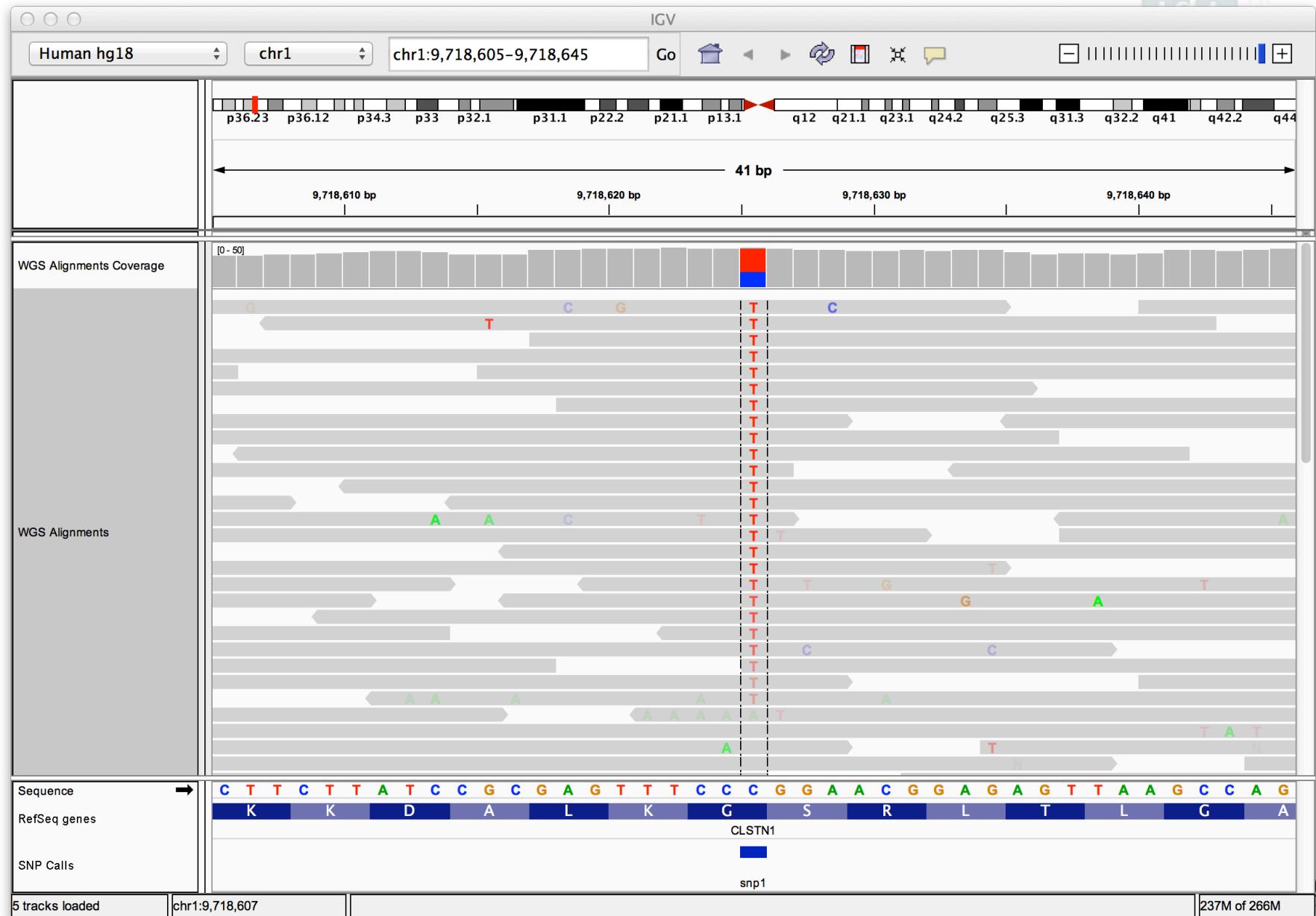
5 tracks loaded chr1:9,718,618 202M of 266M

**WGS Alignments**

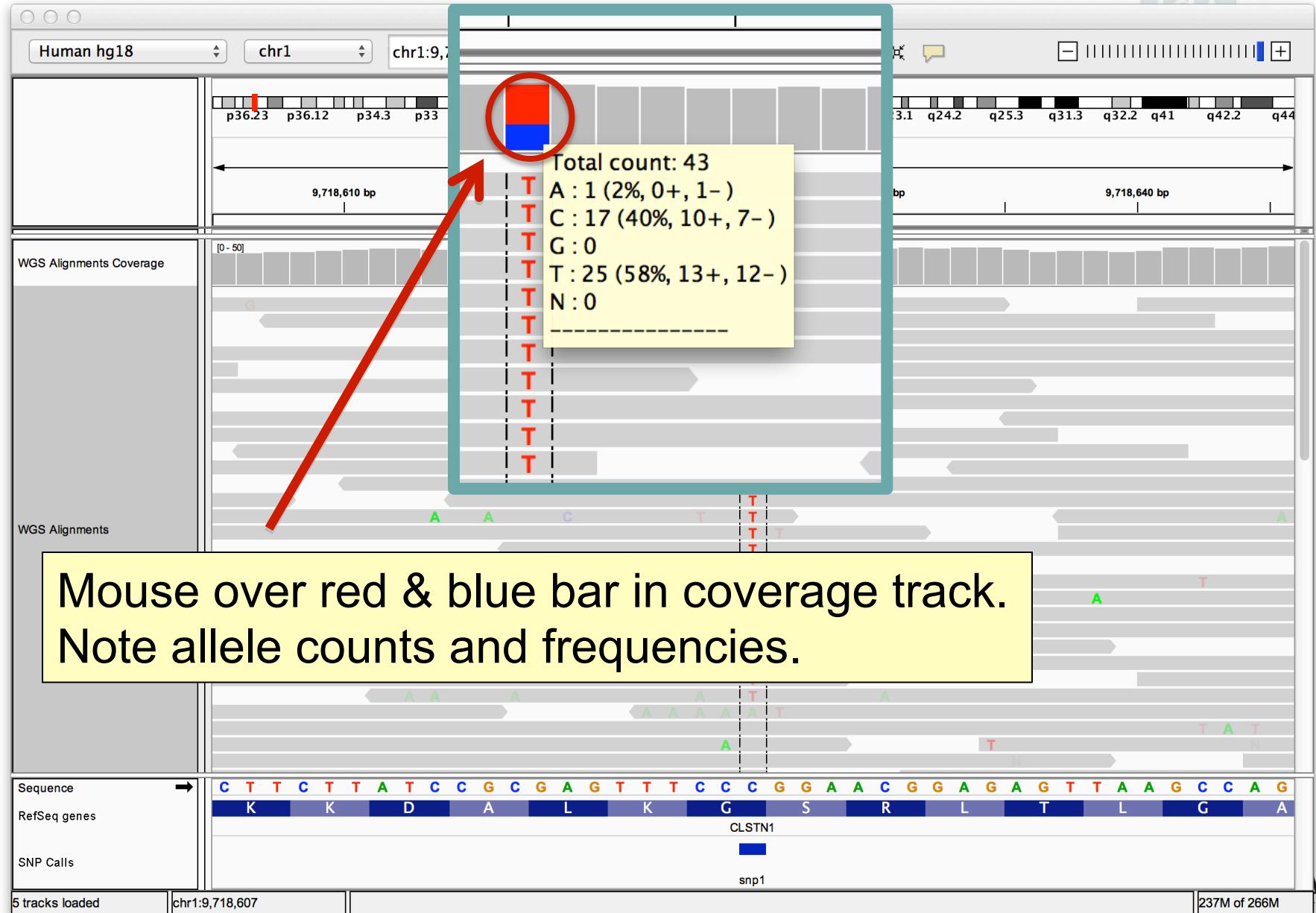
- Rename Track...
- Copy read details to clipboard
- Group alignments by ▶
- Sort alignments by ▶**
- Color alignments by ▶
- Shade base by quality
- Show mismatched bases
- Show all bases

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

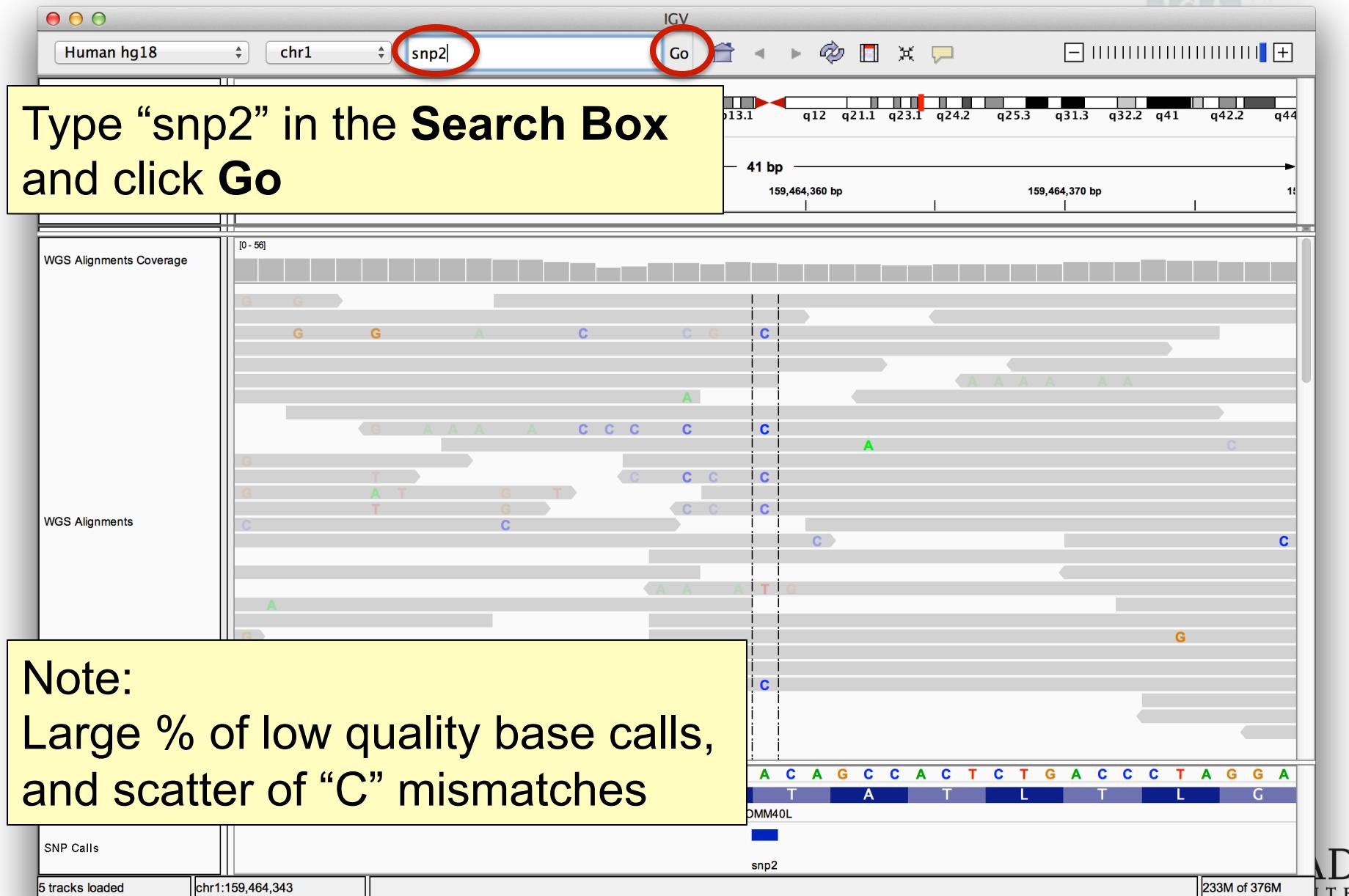
# Viewing SNPs



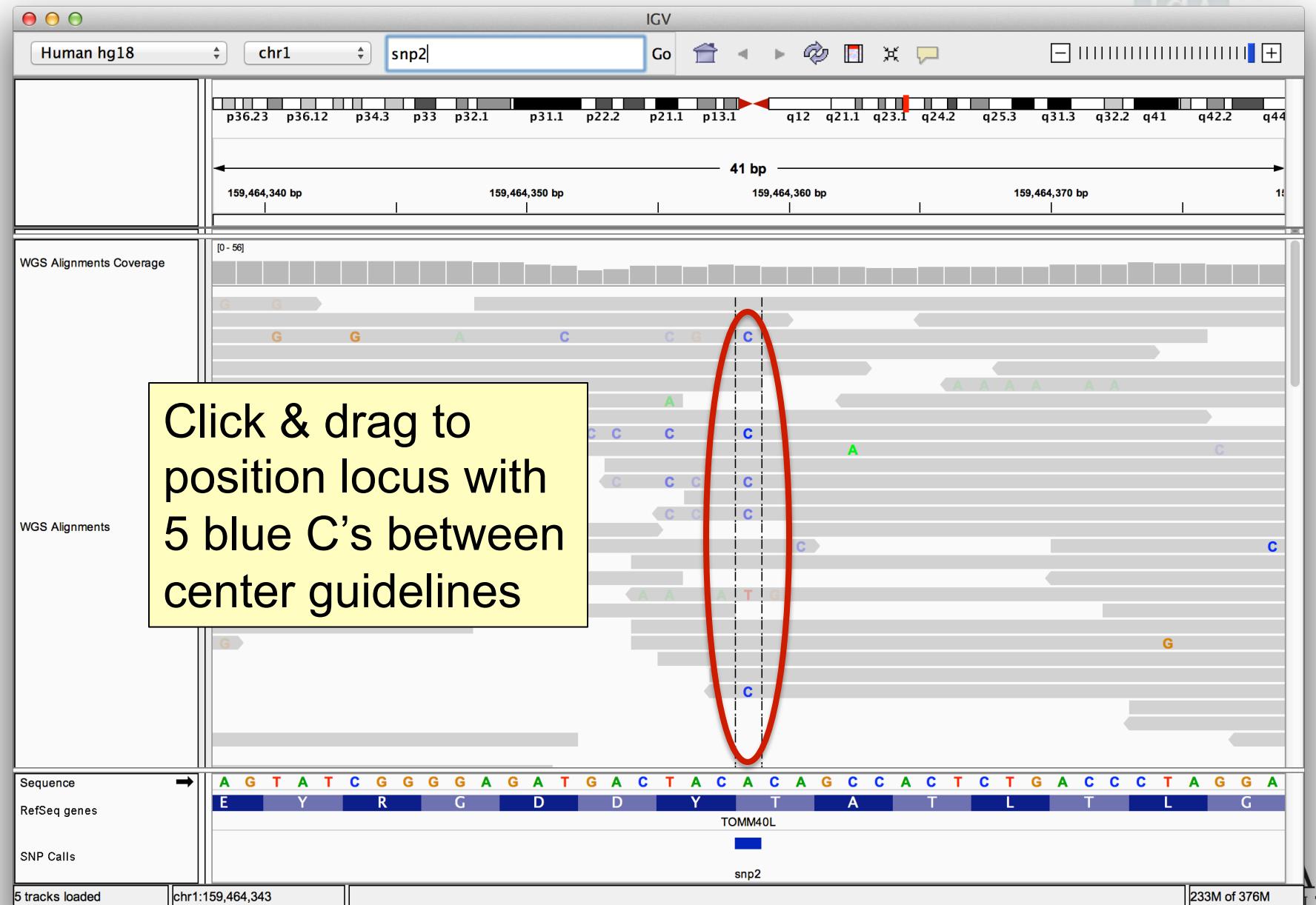
# Viewing SNPs



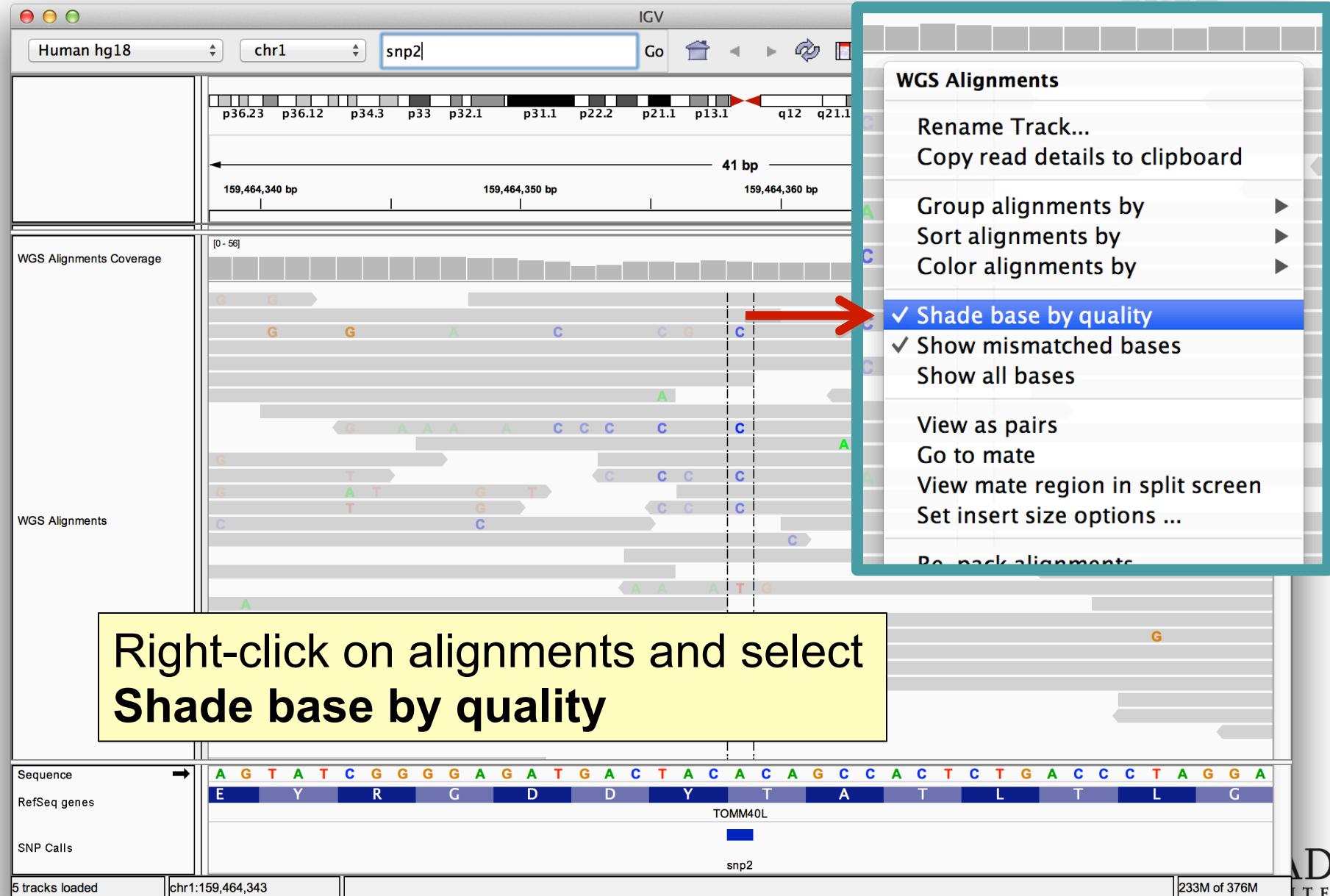
# Viewing SNPs



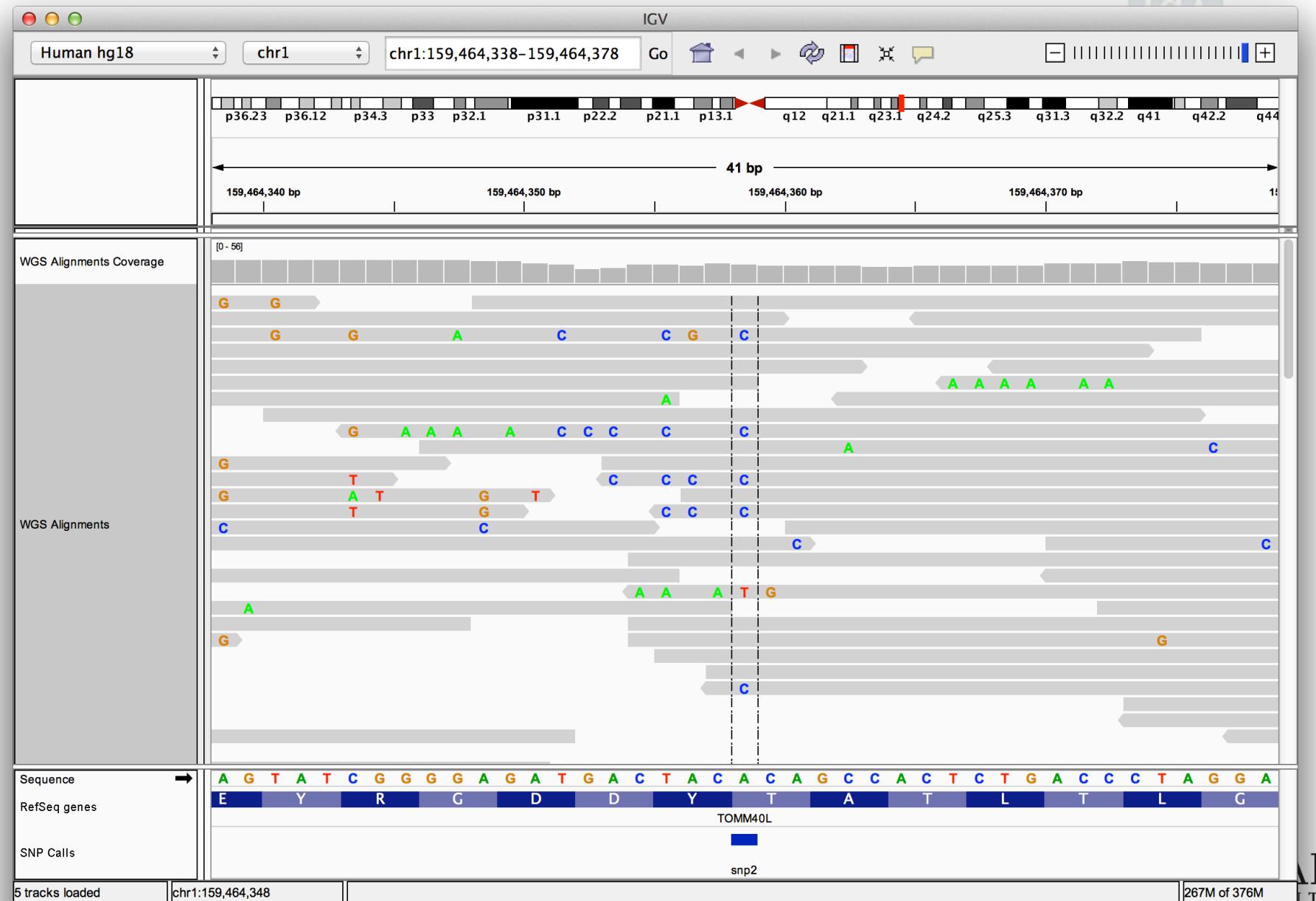
# Viewing SNPs



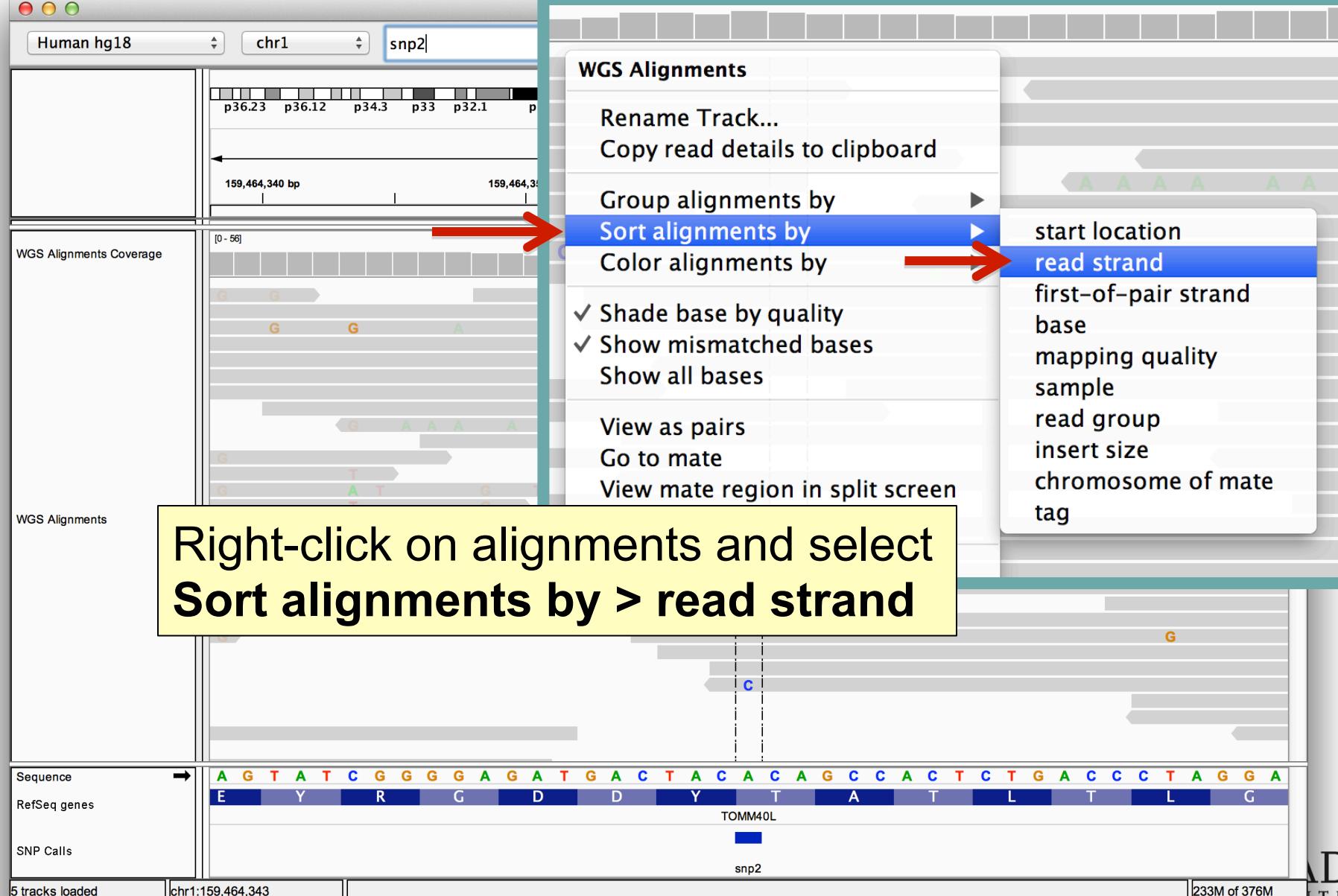
# Viewing SNPs



# 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 →

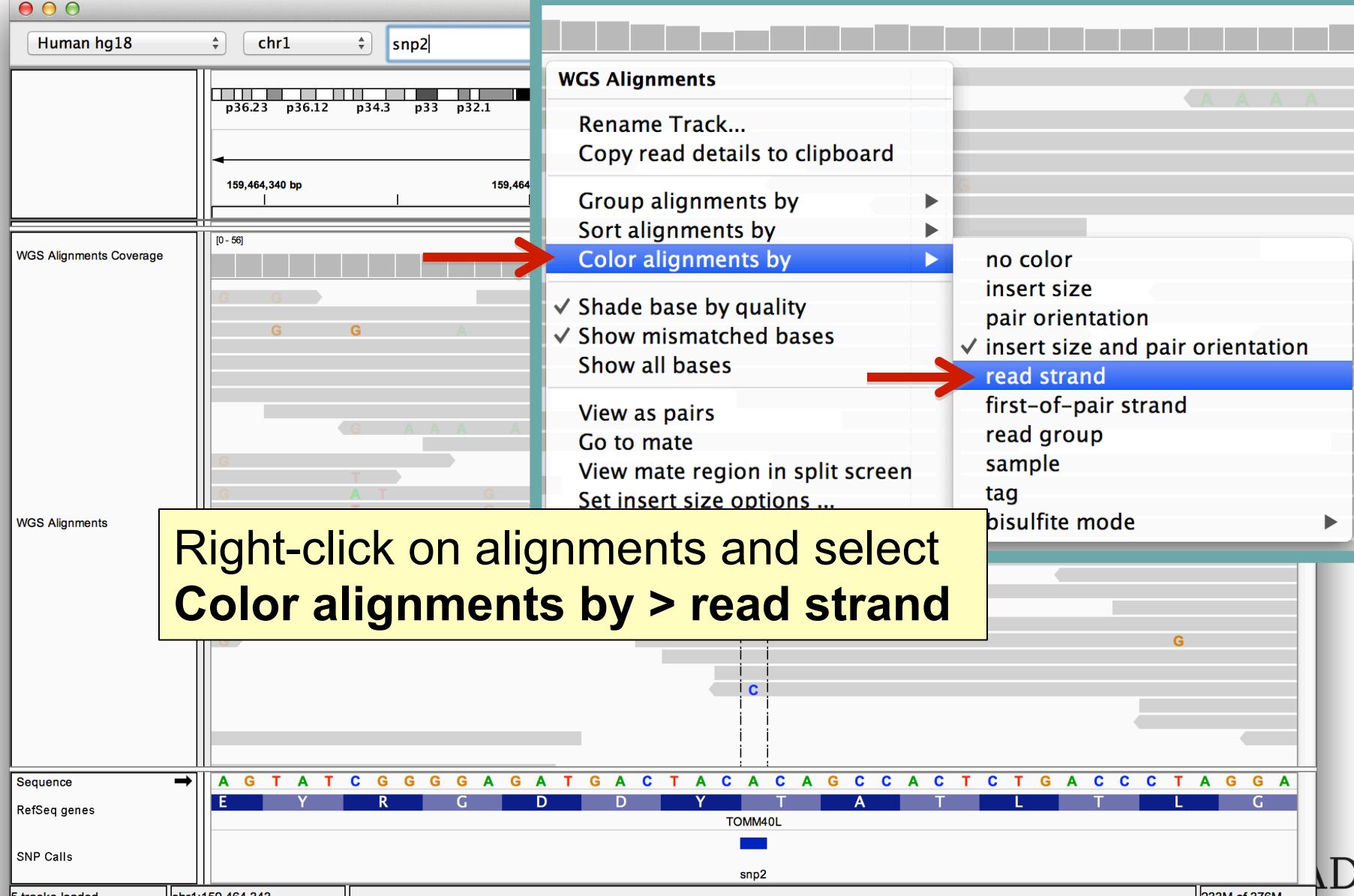
A	G	T	A	T	C	G	G	G	A	G	A	T	G	A	C	T	A	C	A	G	C	C	A	C	T	C	T	G	A	C	C	C	T	A	G	G
E	Y	R	G	D	D	D	Y	T	A	T	A	T	L	T	L	T	L	T	L	G																

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 Panel:** Human hg18 genome browser view. The chromosome is chr1 and the position is 159,464,340 bp, labeled "snp2".
- Left Panel:** WGS Alignments Coverage track for [0 - 56]. It displays multiple gray horizontal bars representing genomic regions, with colored arrows (red, green, blue) indicating the strand direction of each alignment. A red arrow points from the text instructions to this panel.
- Right Panel:** A context menu titled "WGS Alignments" is open over the coverage track. The "Color alignments by" option is selected (highlighted in blue). A second red arrow points from the menu to the "read strand" option in the list of choices.
- Bottom Panel:** Sequence track 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) and RefSeq genes (TOMM40L). A SNP call "snp2" is shown at position 159,464,340 bp.
- Bottom Status Bar:** Shows "5 tracks loaded" and "chr1:159,464,343" followed by a progress bar indicating "233M of 376M".

**Text Overlay:** A yellow callout box contains the instruction: "Right-click on alignments and select Color alignments by > read strand".

# Viewing SNPs



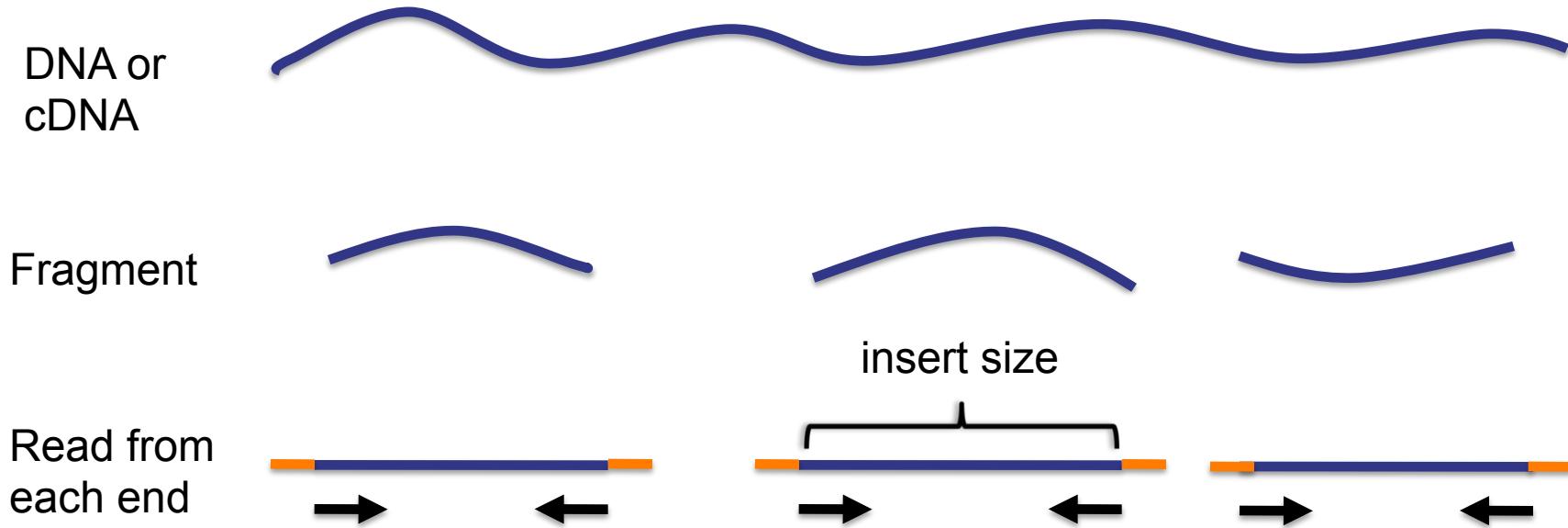
# Viewing Structural Events

# Structural events

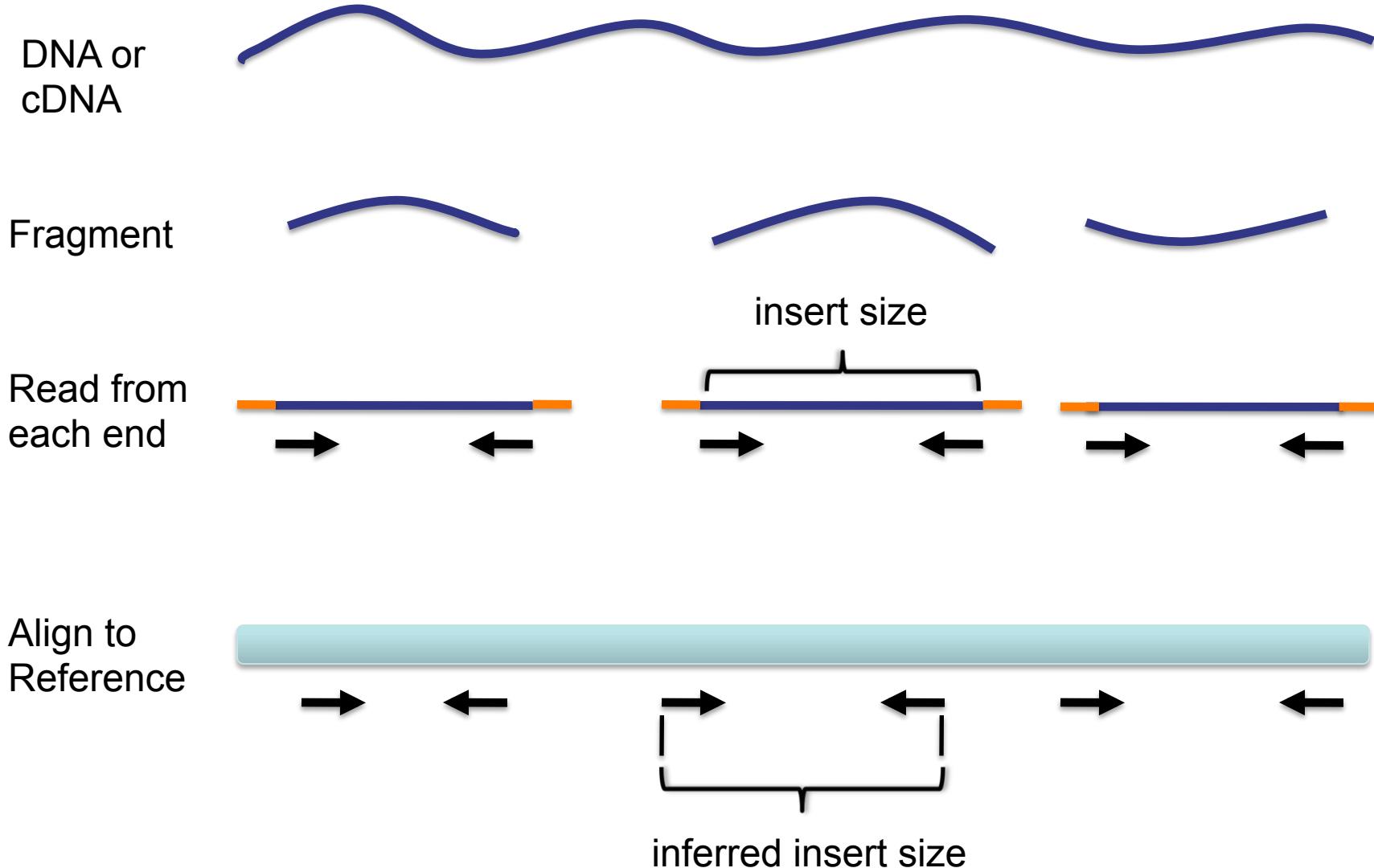


- Paired reads can yield evidence for genomic “structural events”, such as deletions, translocations, and inversions.
- Alignment coloring options help highlight these events based on:
  - Inferred insert size (template length)
  - Pair orientation (relative strand of pair)

# 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

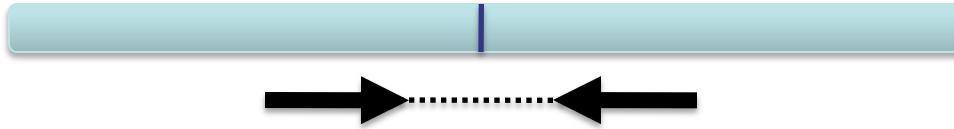


# Deletion

Reference  
Genome



Subject



# Deletion

Reference  
Genome

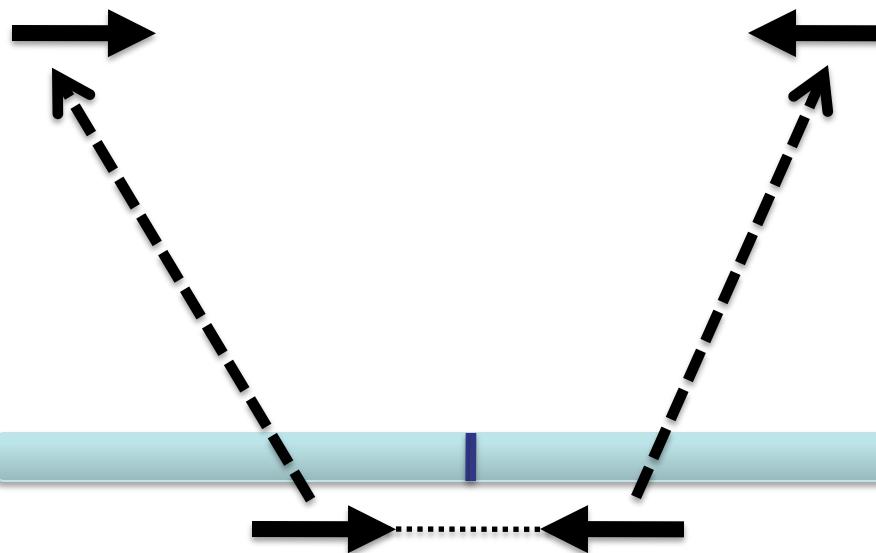


Subject



# Deletion

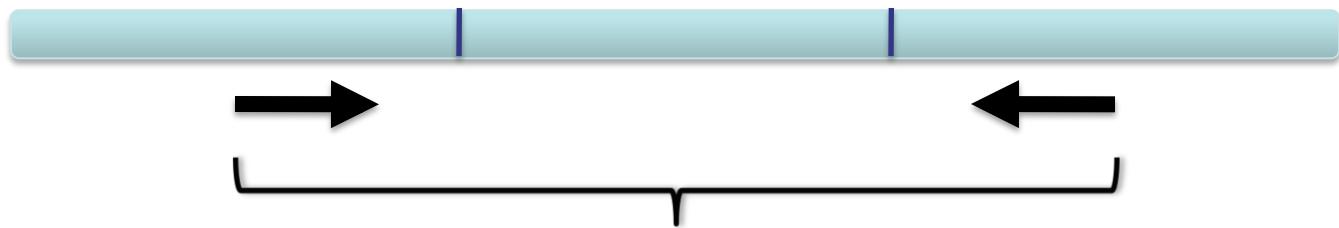
Reference  
Genome



Subject

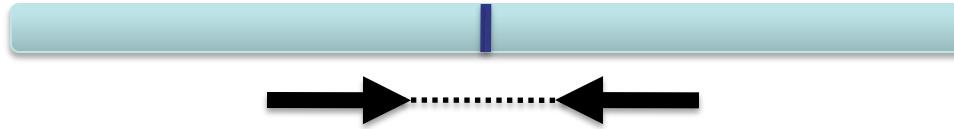
# Deletion

Reference  
Genome



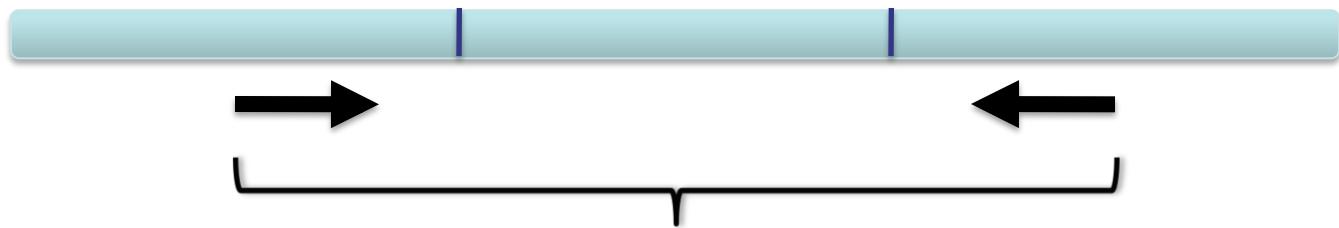
inferred insert size

Subject



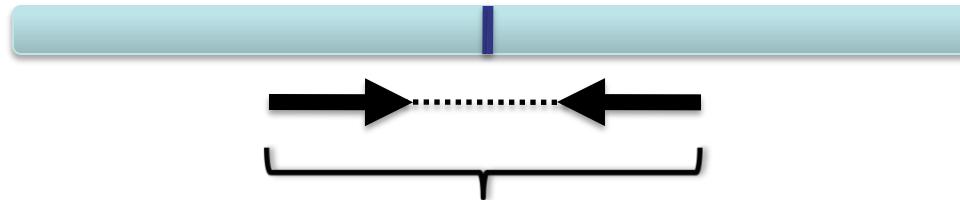
# Deletion

Reference  
Genome



inferred insert size

Subject

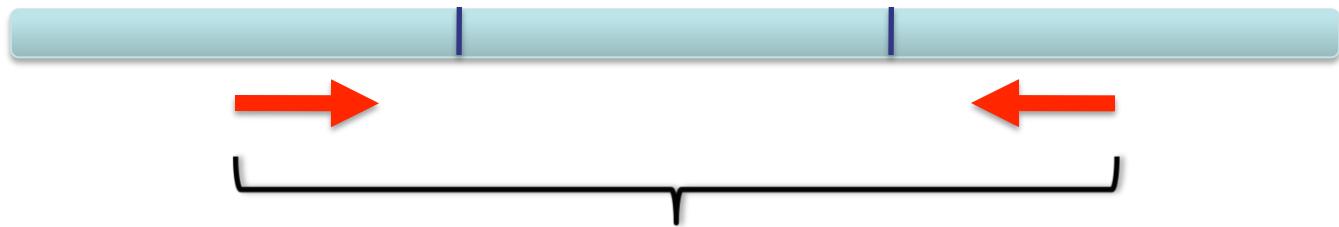


expected insert size

# Deletion

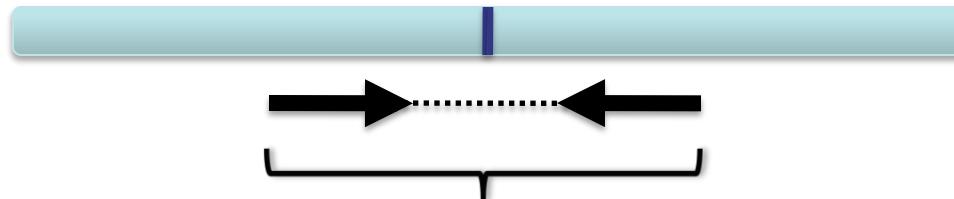
Inferred insert size is > expected value

Reference  
Genome



inferred insert size

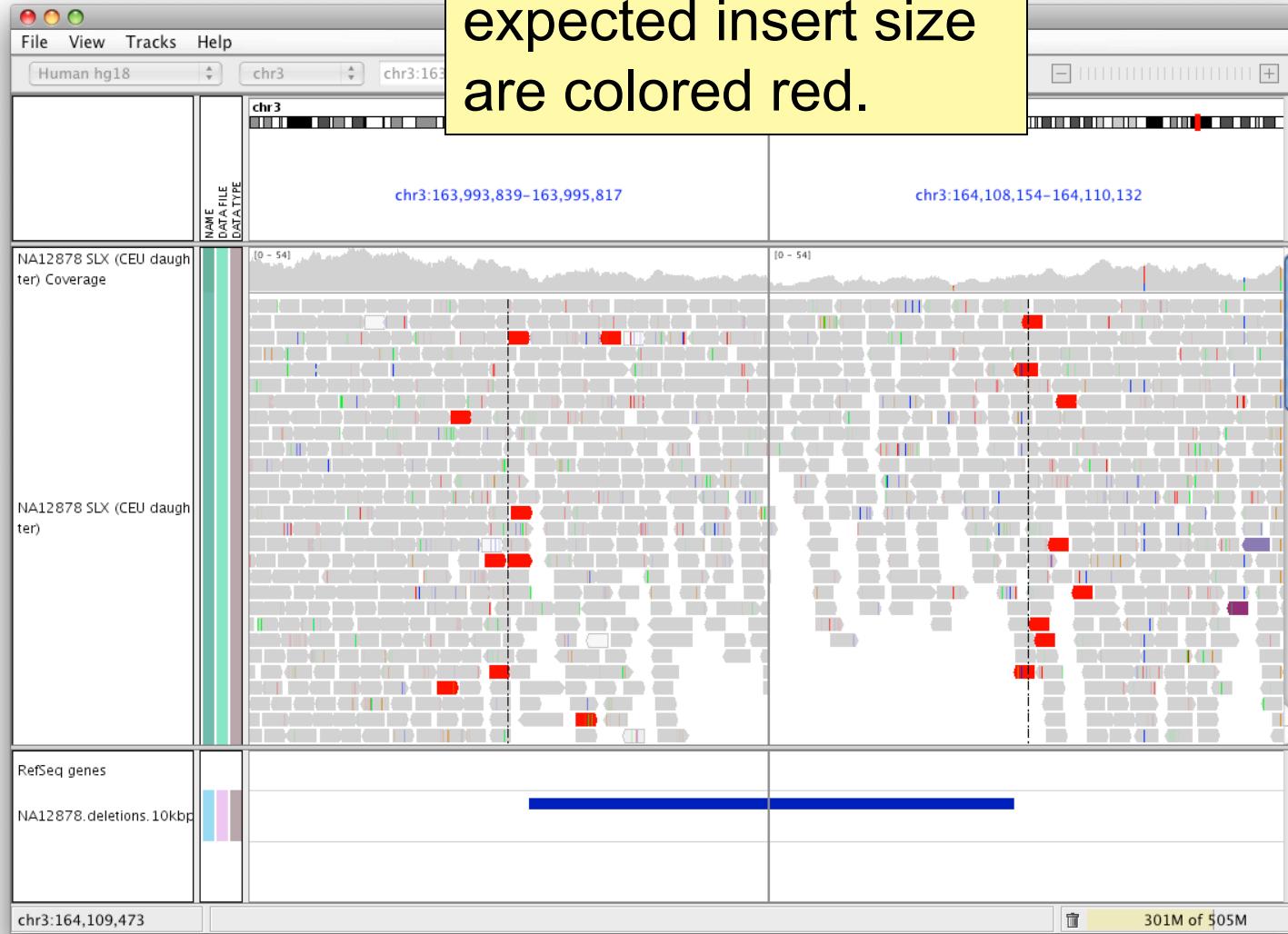
Subject



expected insert size

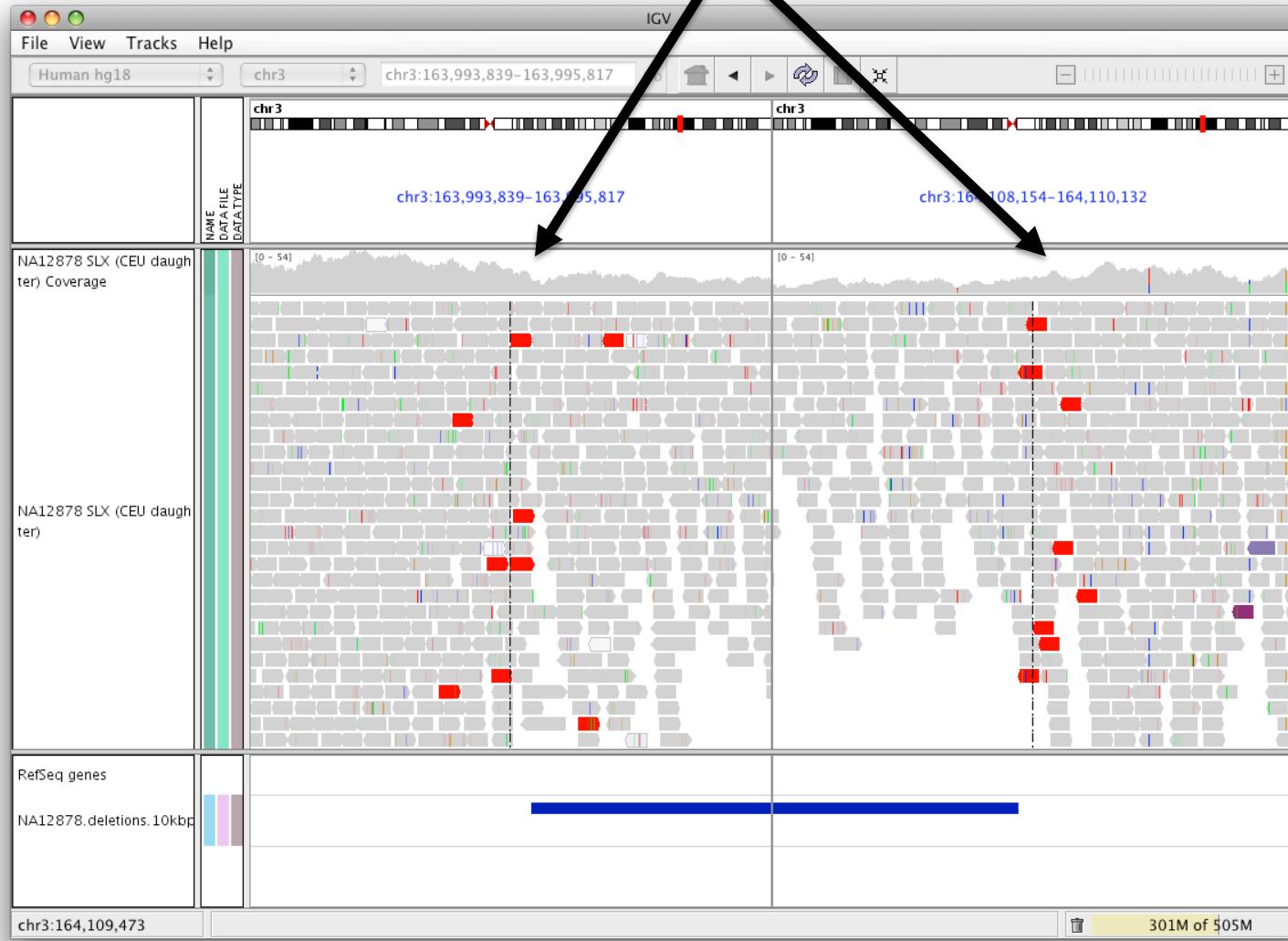
# Deletion

Pairs with larger than expected insert size are colored red.



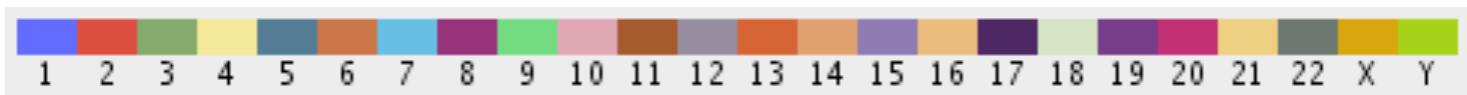
# Deletion

Note drop in coverage

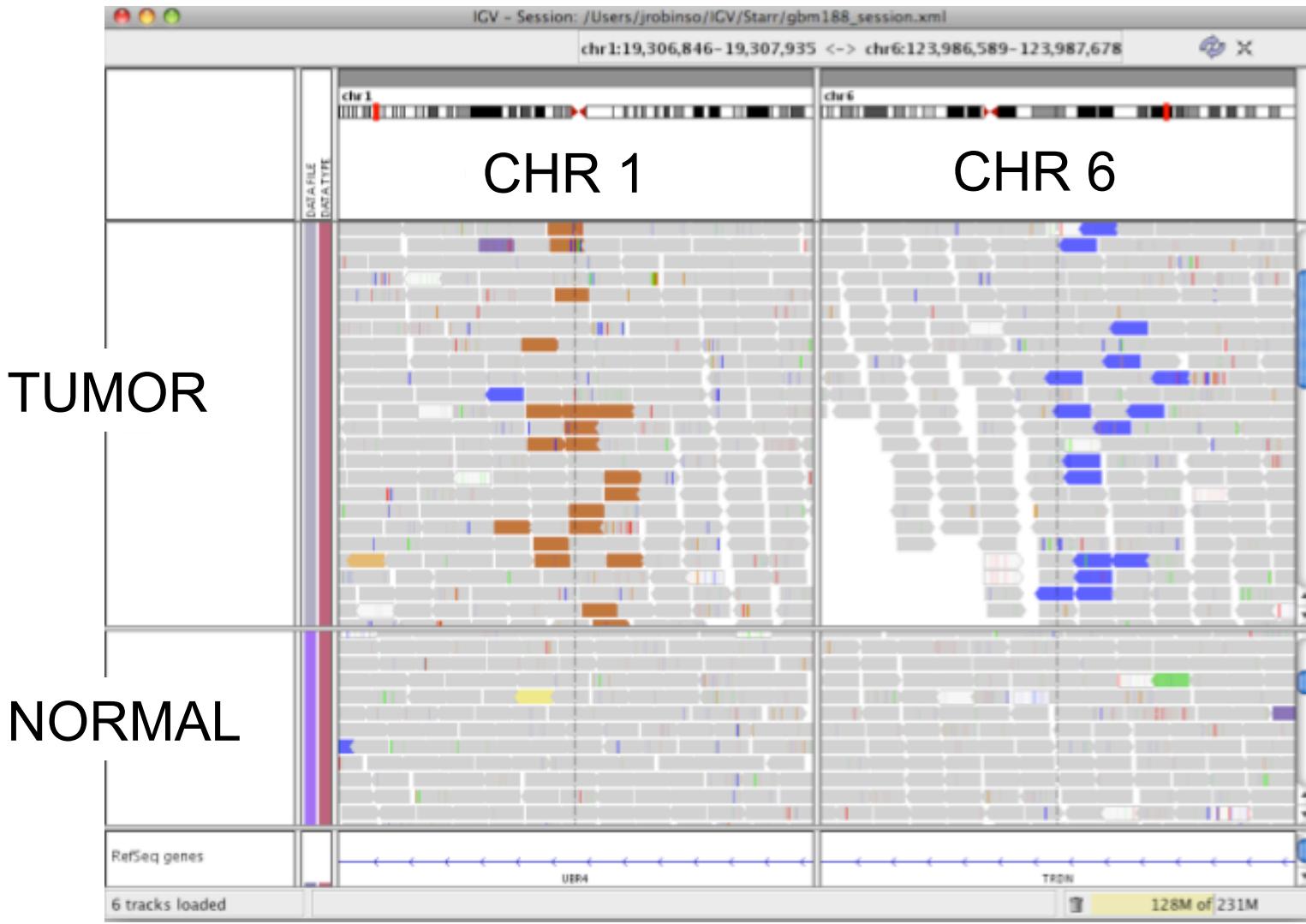


# Insert size color scheme

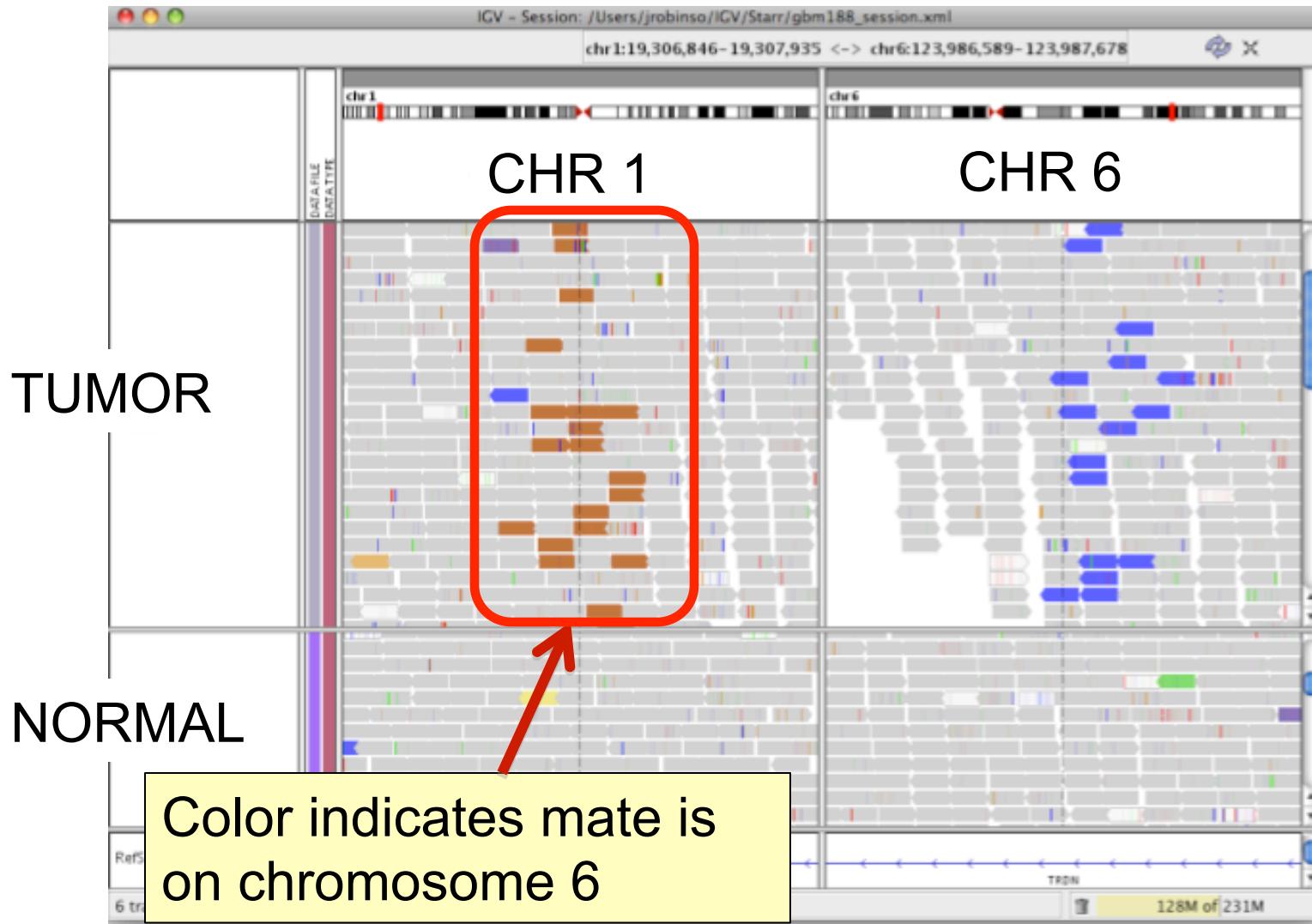
- Smaller than expected insert size: 
- Larger than expected insert size: 
- Pairs on different chromosomes  
*Each end colored by chromosome of its mate*



# 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

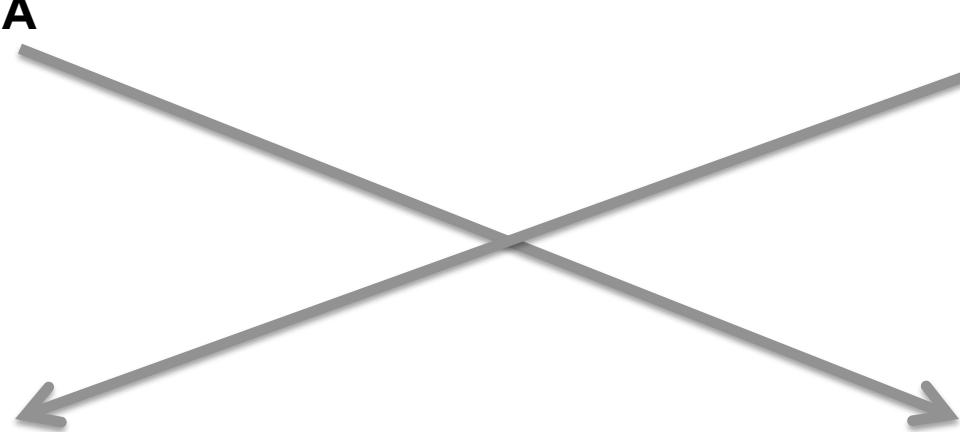
B

Subject



B

A



# Inversion

Reference  
Genome



Subject



# Inversion

Reference  
Genome



A

B

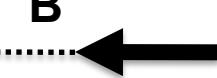


Subject

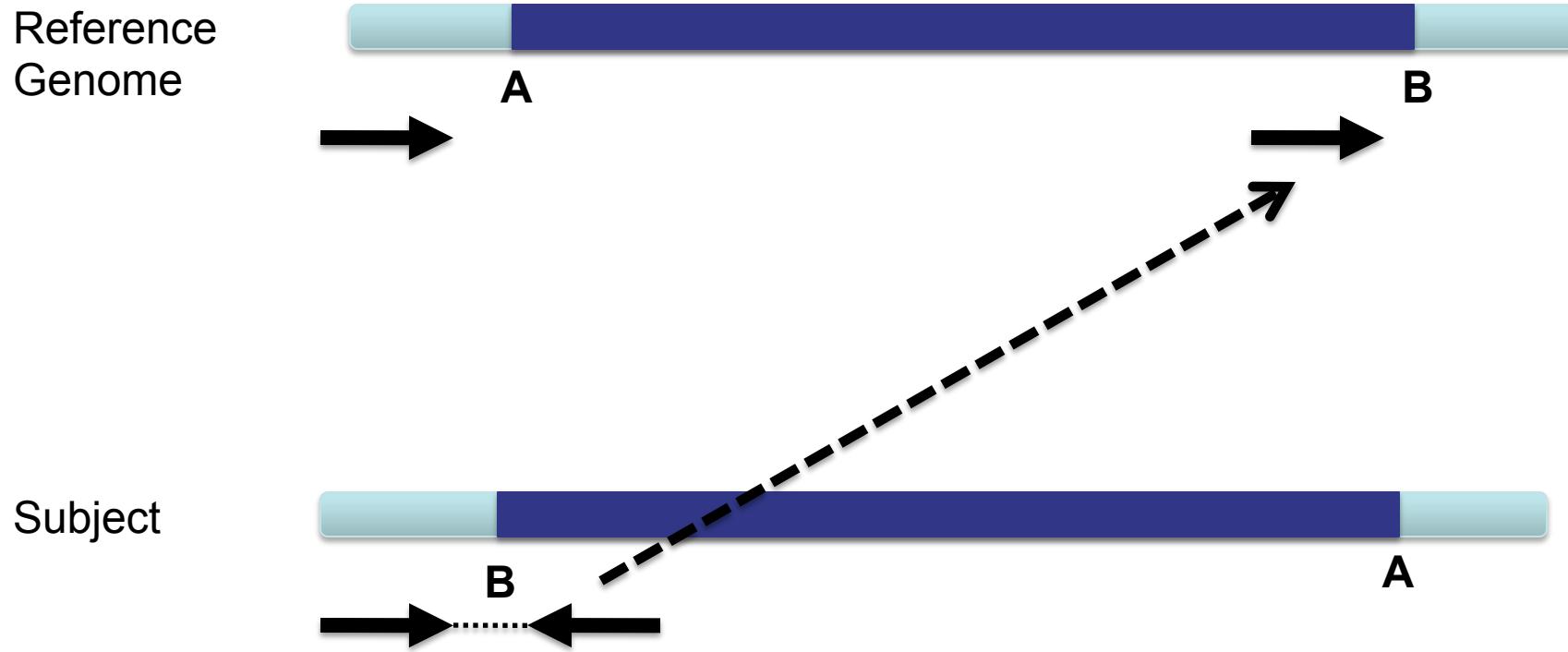


B

A



# Inversion



# Inversion

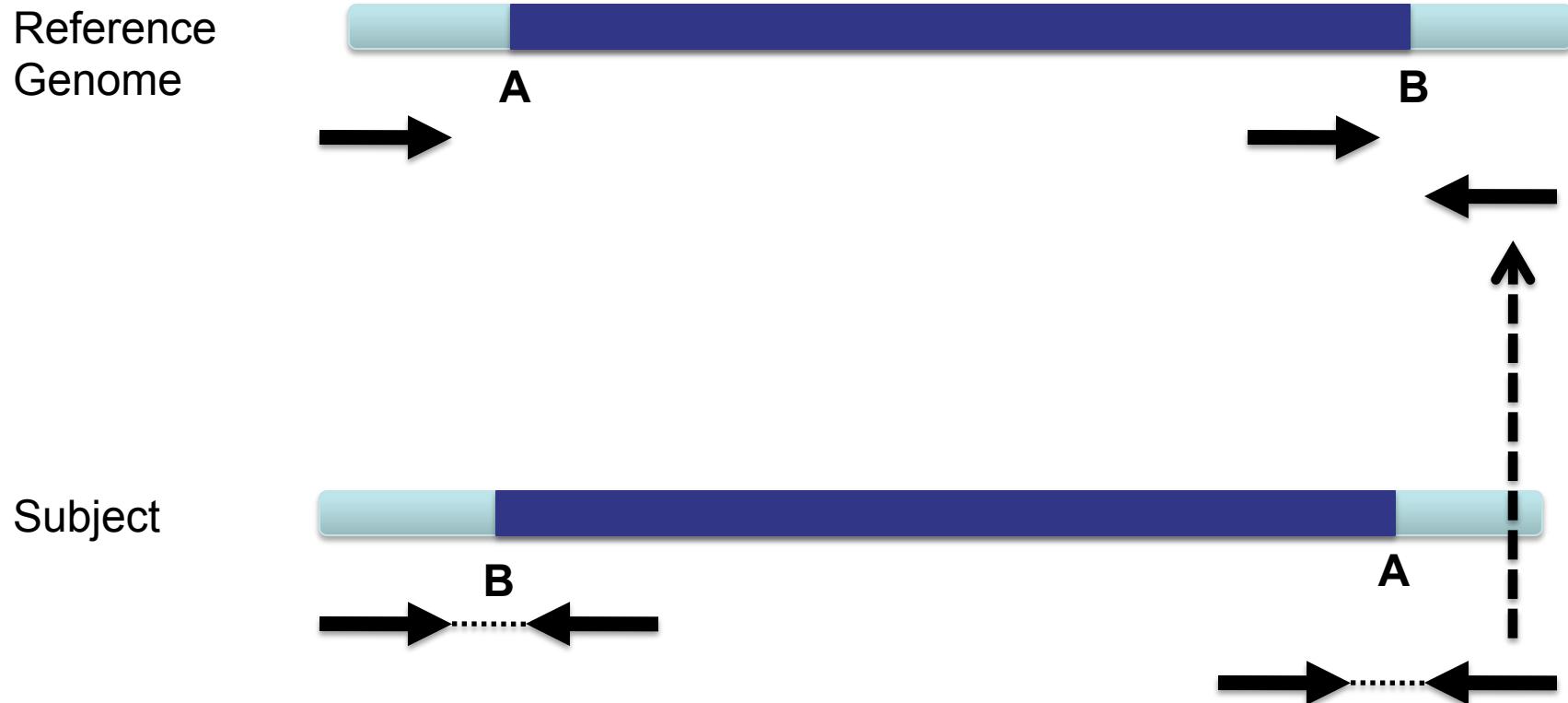
Reference  
Genome



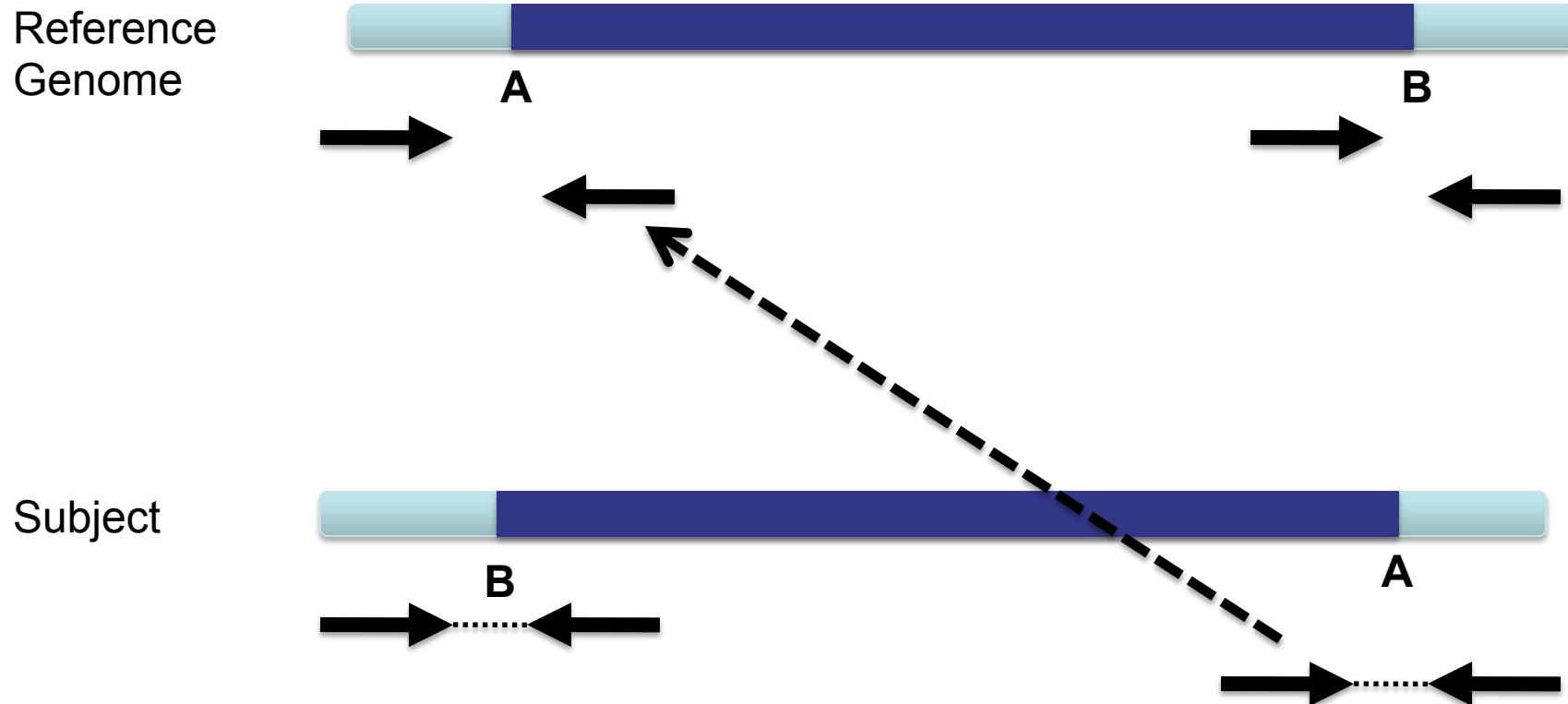
Subject



# Inversion

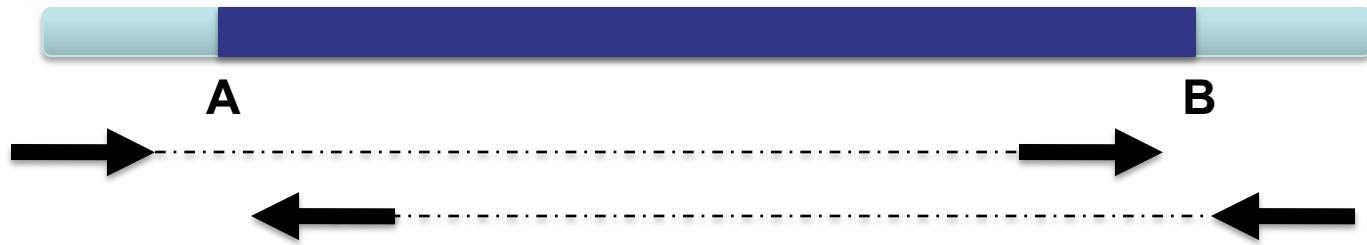


# Inversion



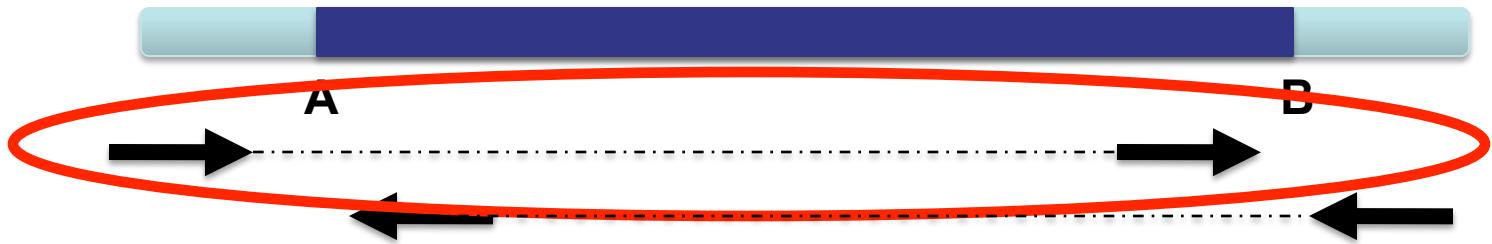
# Inversion

Reference  
Genome



# Inversion

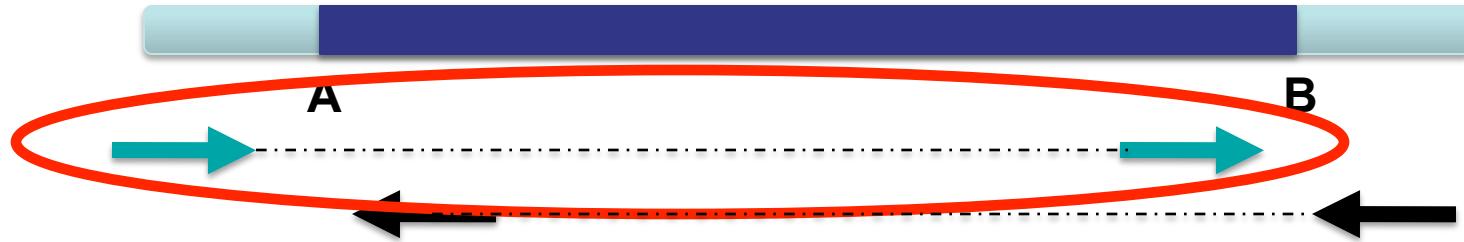
Reference  
Genome



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 ►**

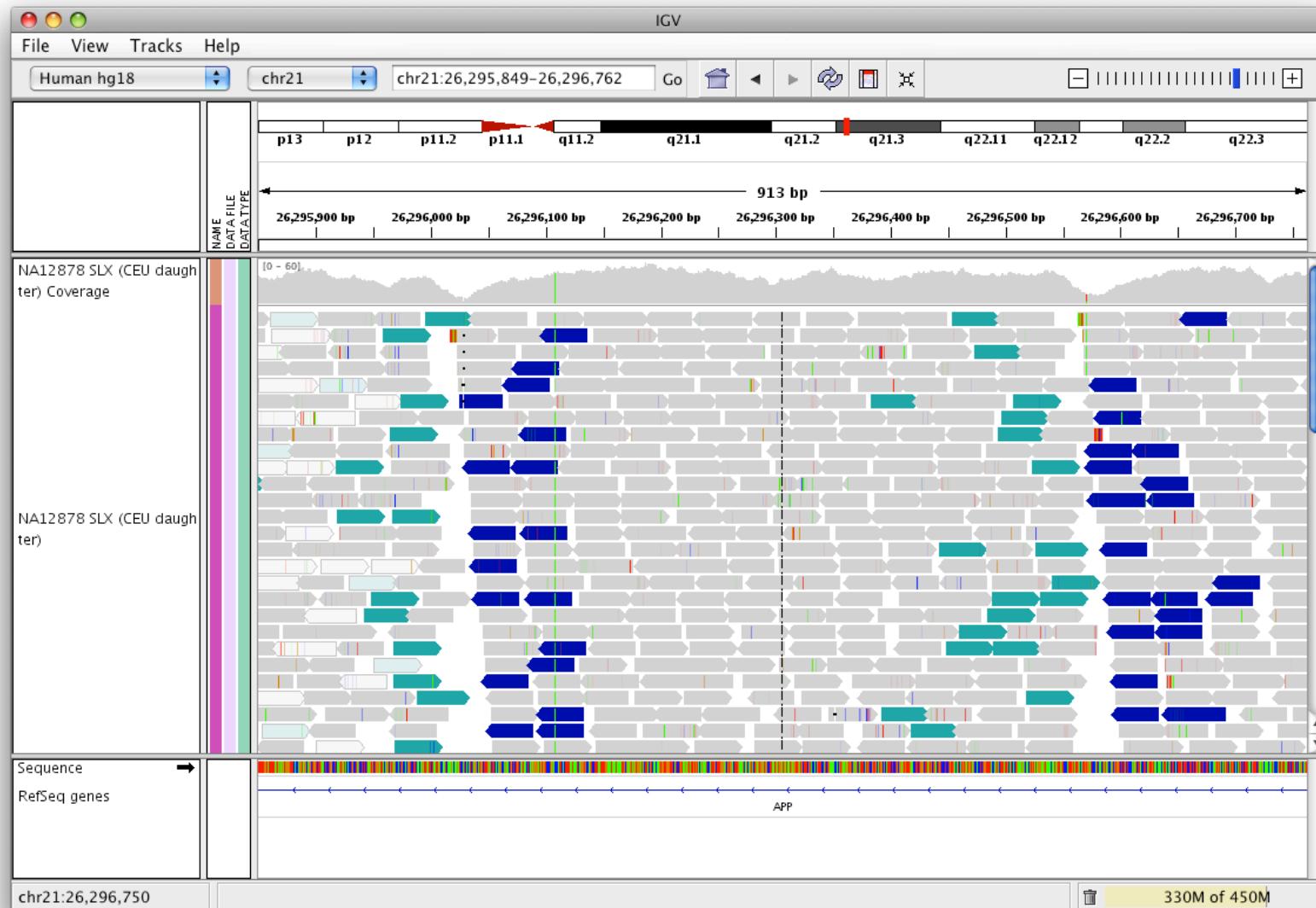
- ✓ Shade base by quality
- ✓ Show mismatched bases
- Show all bases

- 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

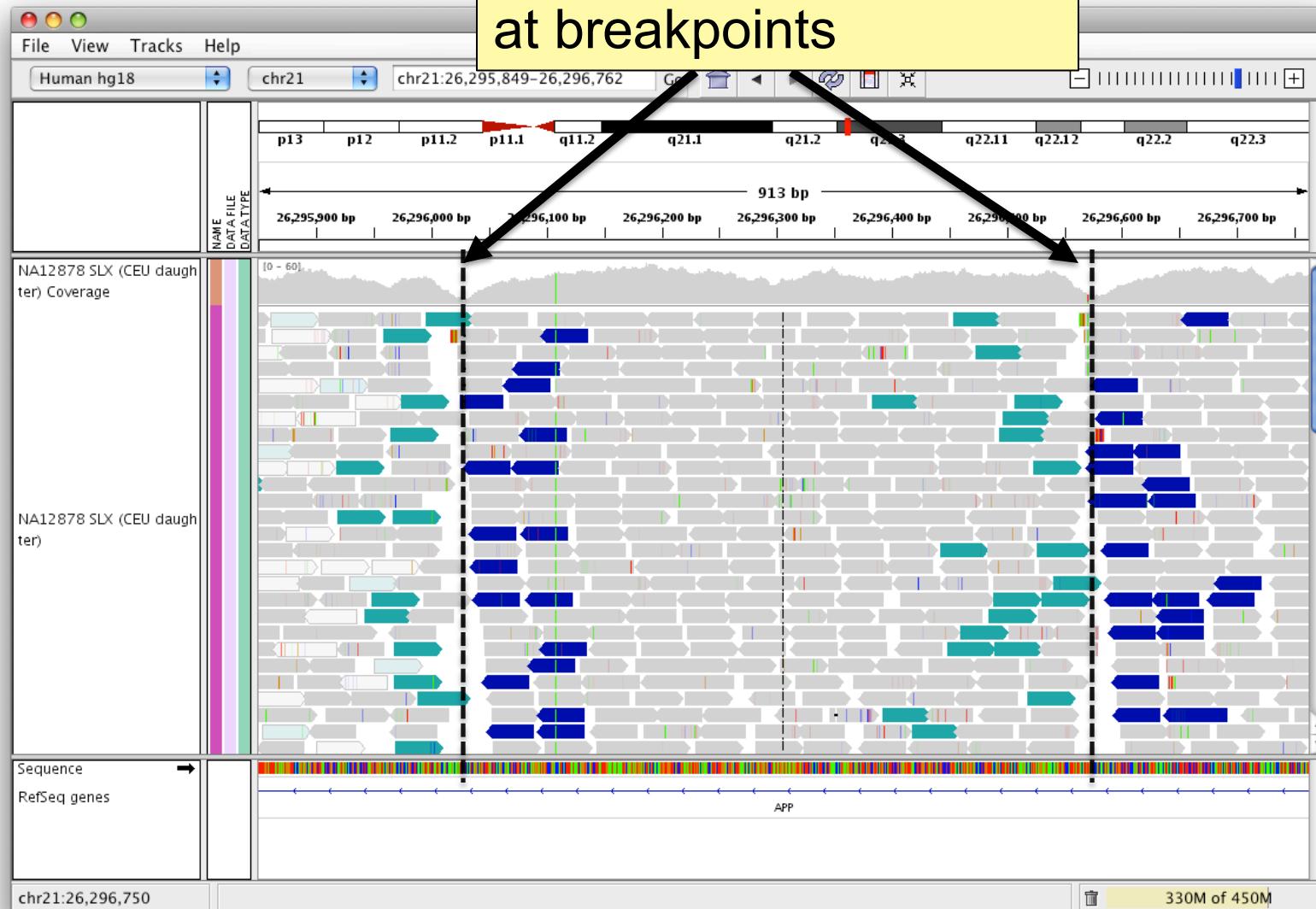


# Inversion

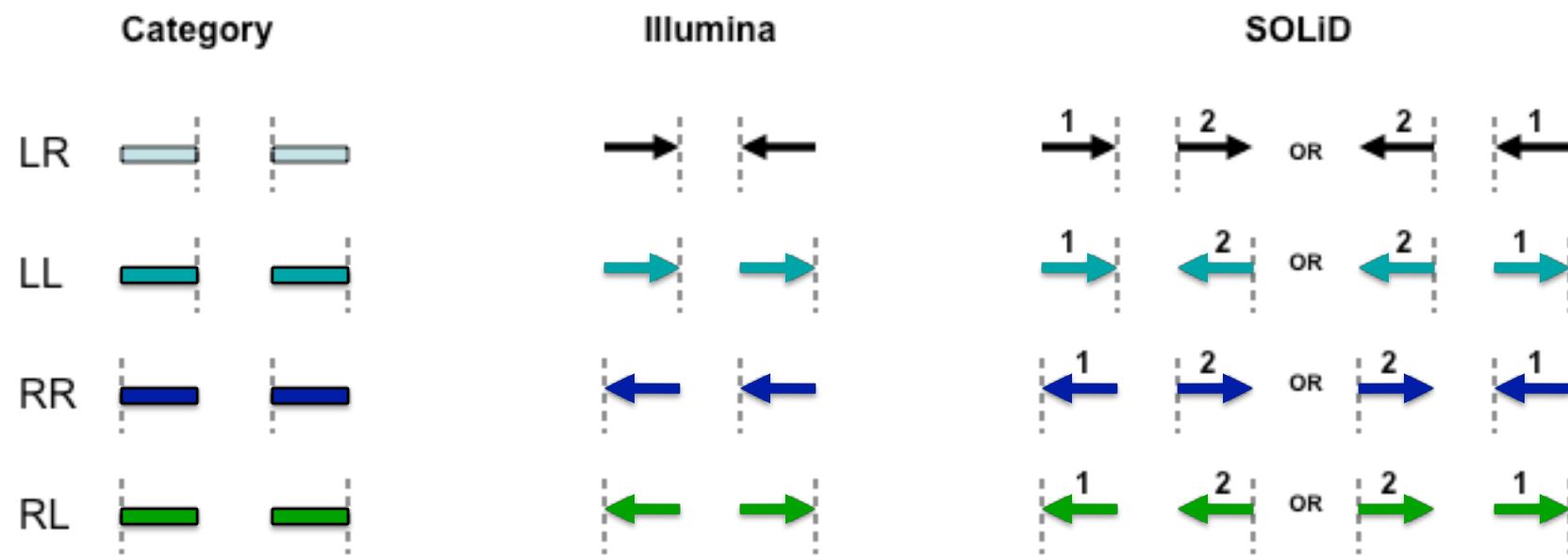


Integrative  
Genomics  
Viewer

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.

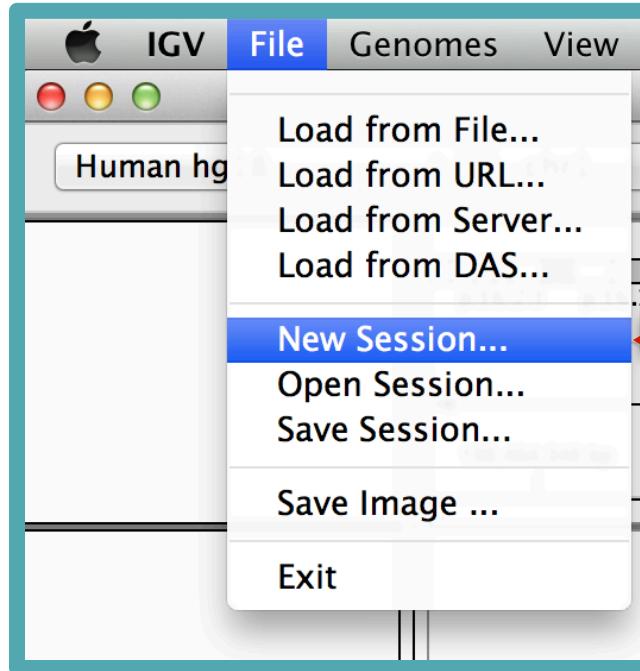
These categories only apply to reads where both mates map to the same chromosome.

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](http://www.illumina.com/science/data_library.ilmn)



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

# RNA-Seq Setup

---

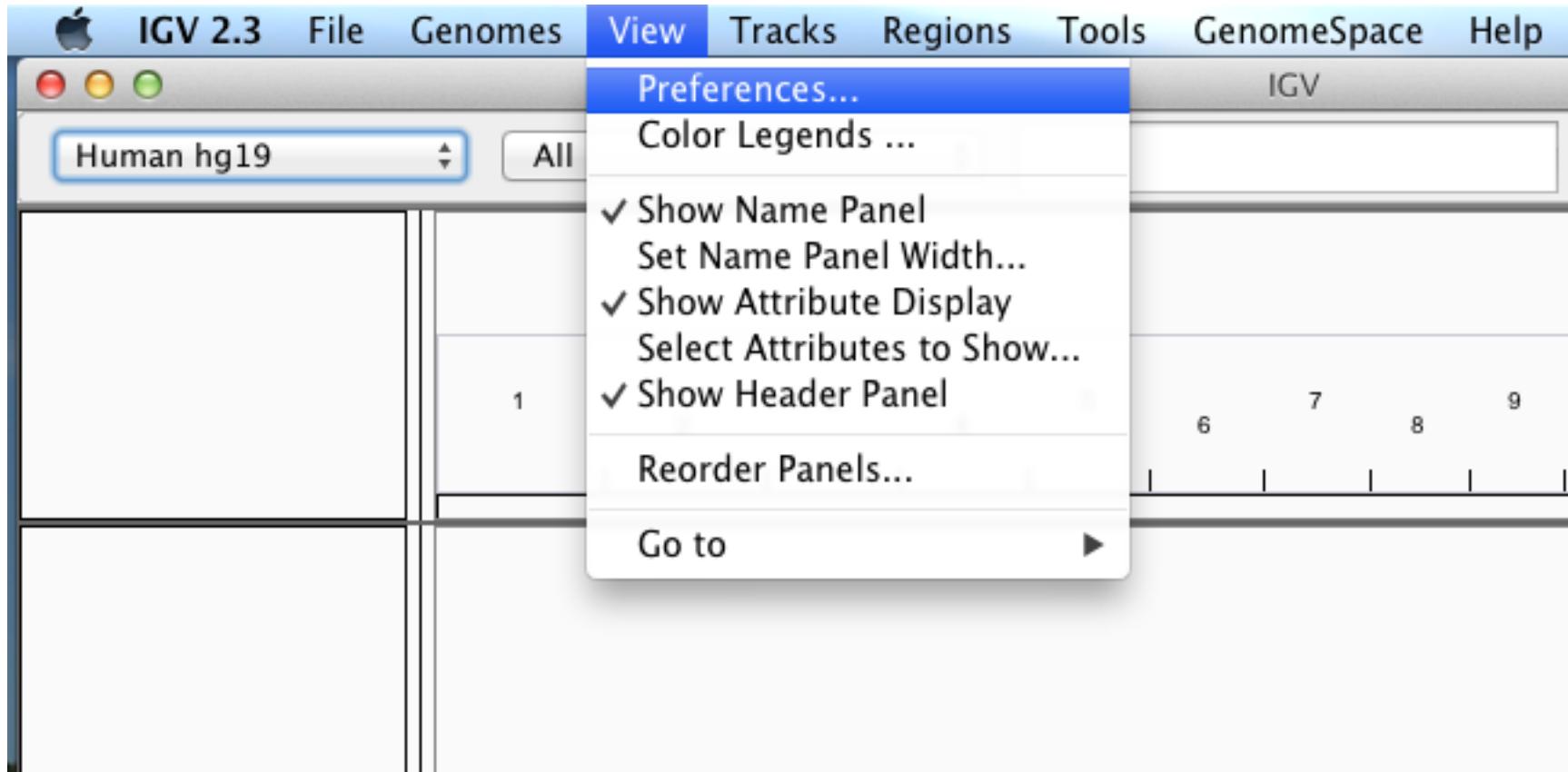
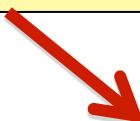


- Step 1: Tune settings for RNA.

# 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 menu bar includes General, Tracks, Mutations, Chars, Alignments (which is highlighted with a red box and has a red arrow pointing to it), Probes, Proxy, Advanced, and IonTorrent. The main panel contains several configuration sections:

- Visibility range threshold (kb):** 30 (Nominal window size at which alignments become visible)
- Filter and shading options:**
  - Coverage allele-freq threshold: 0.2
  - Mapping quality threshold: 0
  - Filter duplicate reads
  - Filter vendor failed reads
  - Filter secondary alignments
  - Flag unmapped pairs
  - 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	<input checked="" type="checkbox"/> 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". There are OK and Cancel buttons at the bottom right.

# 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 yellow callout)
  - 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 | 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

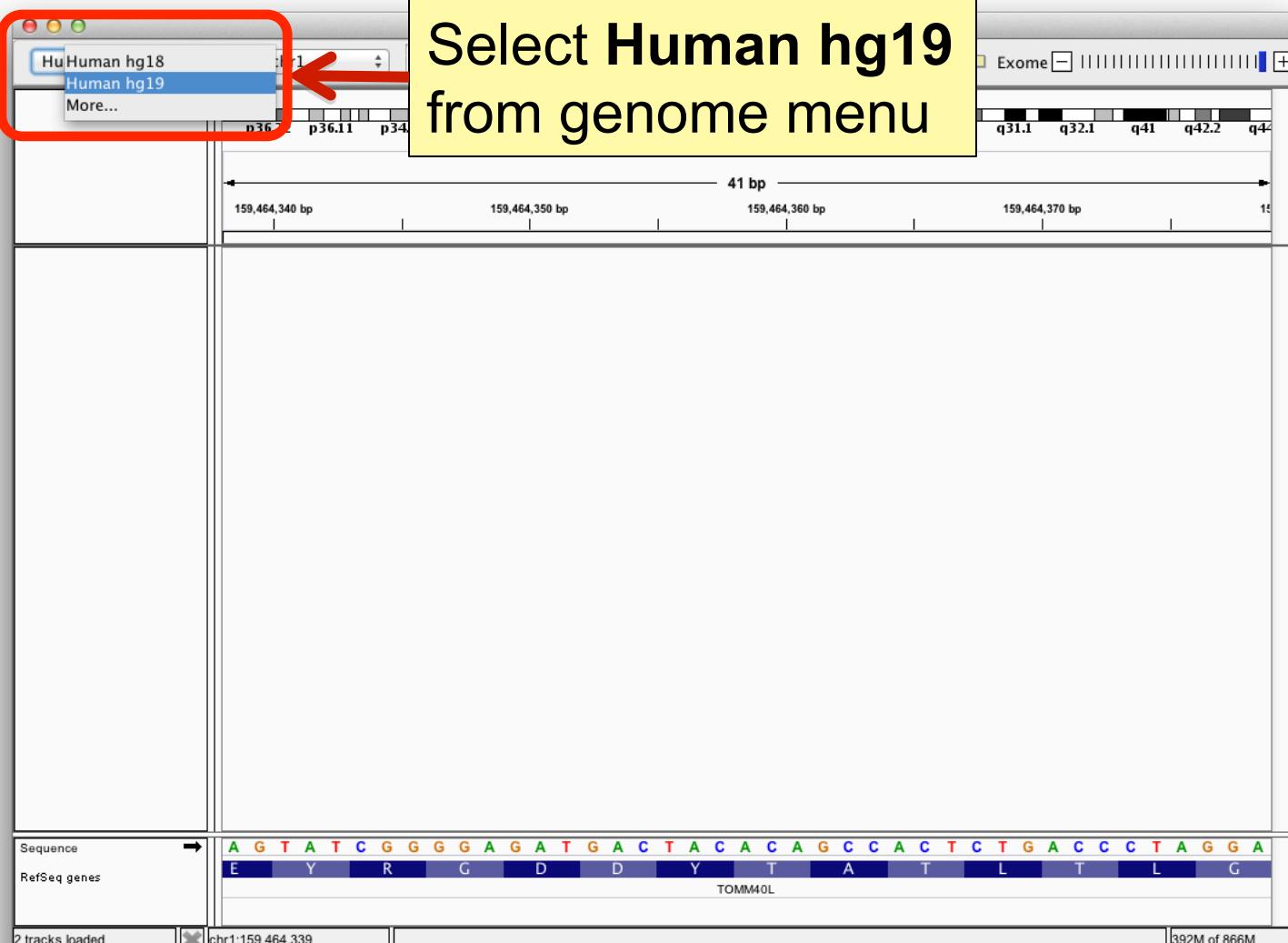
Maximum (bp): 1000 Maximum (percentile): 99.5

5 tracks loaded 386M of 866M

Click OK to save changes

OK Cancel

# RNA-seq alignments

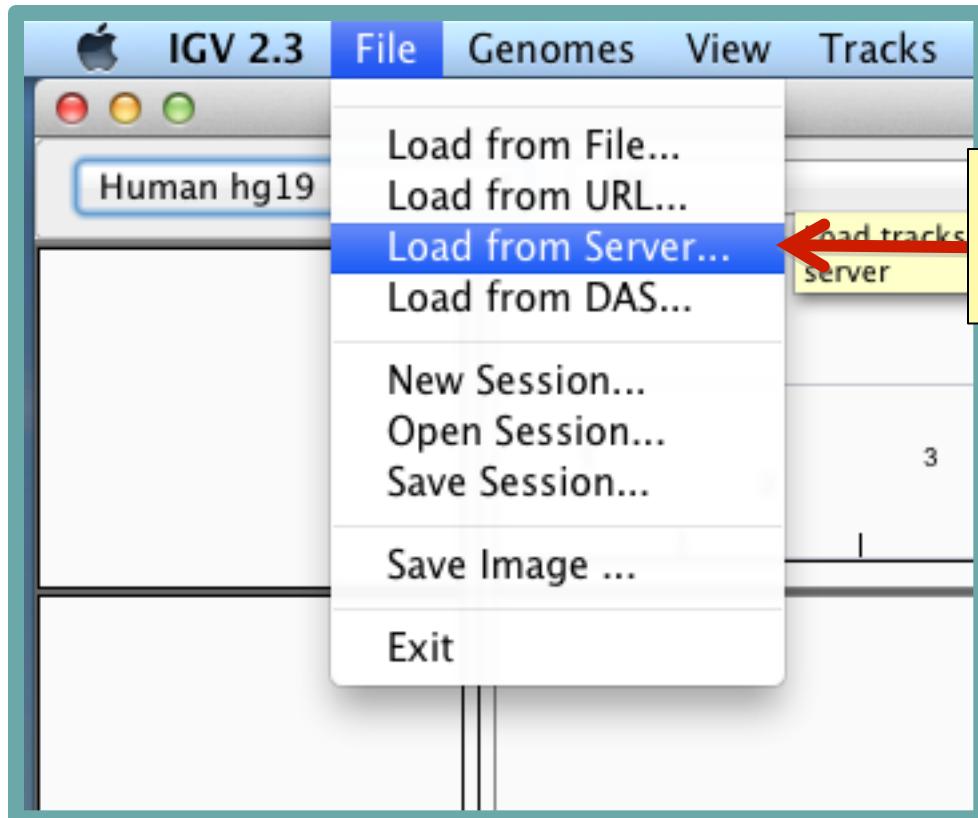


Select Human hg19 from genome menu

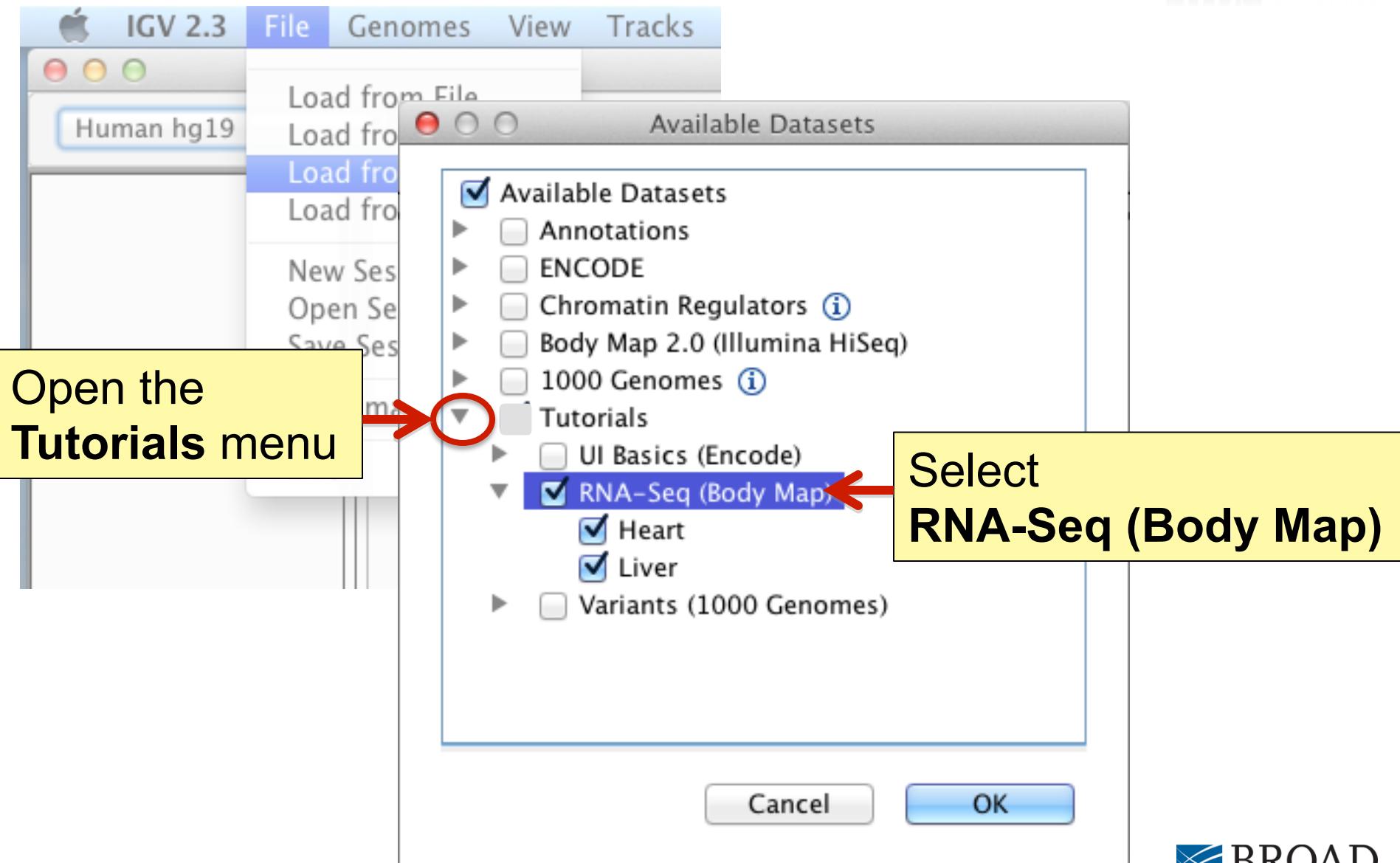
The screenshot shows the IGV interface with the following details:

- Genome Menu:** A red box highlights the "HuHuman hg18" dropdown menu. A yellow box with a red arrow points to the "Human hg19" option, which is currently selected.
- Chromosome View:** The main panel displays chromosome 1 with a zoomed-in view of a 41 bp region between 159,464,340 bp and 159,464,380 bp. The p36 band is visible on the left, and q31.1, q32.1, q41, q42.2, and q42 bands are visible on the right.
- Sequence View:** At the bottom, a sequence track shows the DNA sequence: 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 it, a RefSeq genes track shows the TOMM40L gene, with exons in various colors (red, green, blue, purple) and introns in grey. The gene spans from approximately 159,464,339 bp to 159,464,380 bp.
- Status Bar:** The bottom status bar indicates "2 tracks loaded" and "chr1:159,464,339". It also shows memory usage: "392M of 866M".

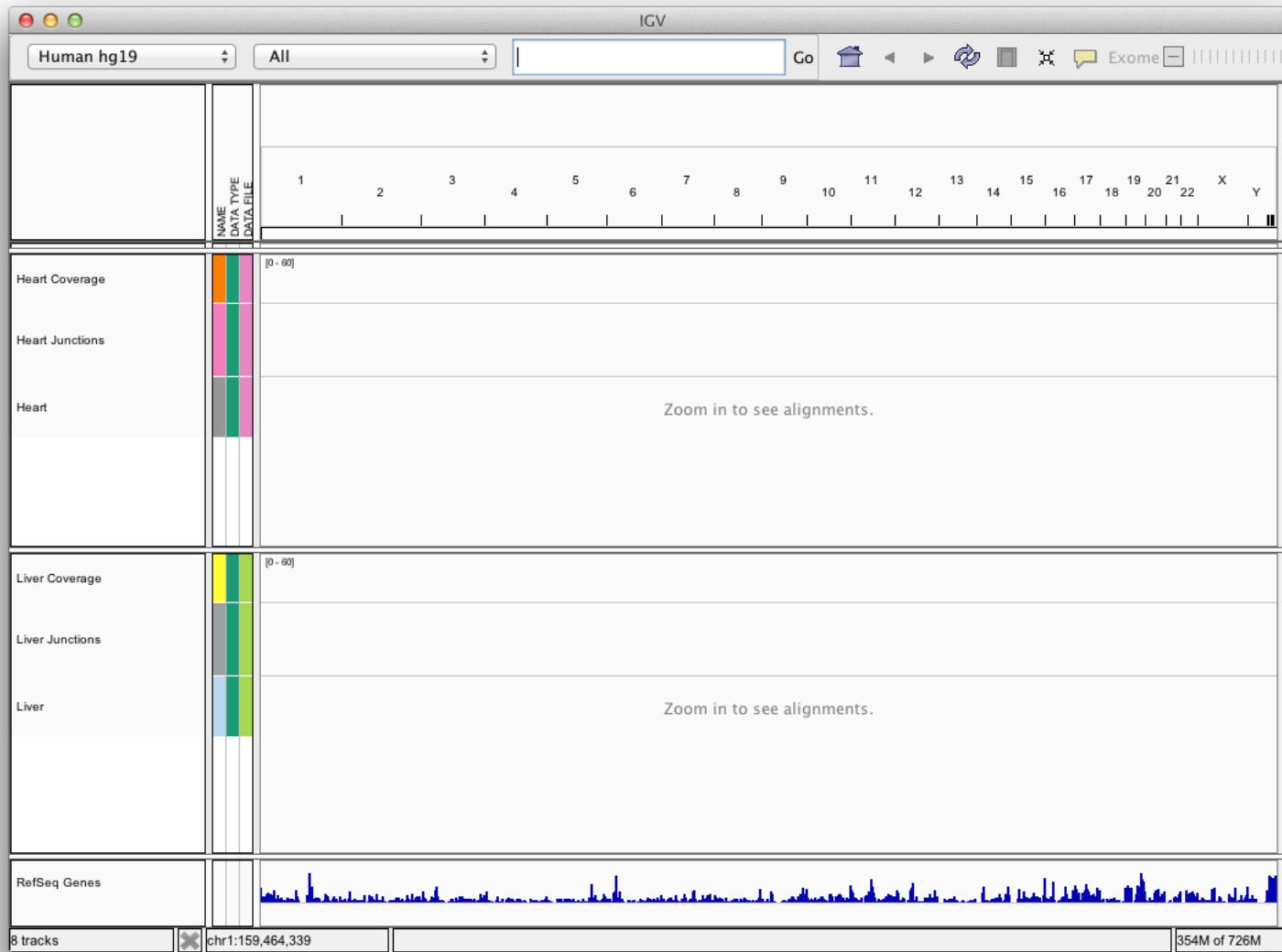
# RNA-seq alignments



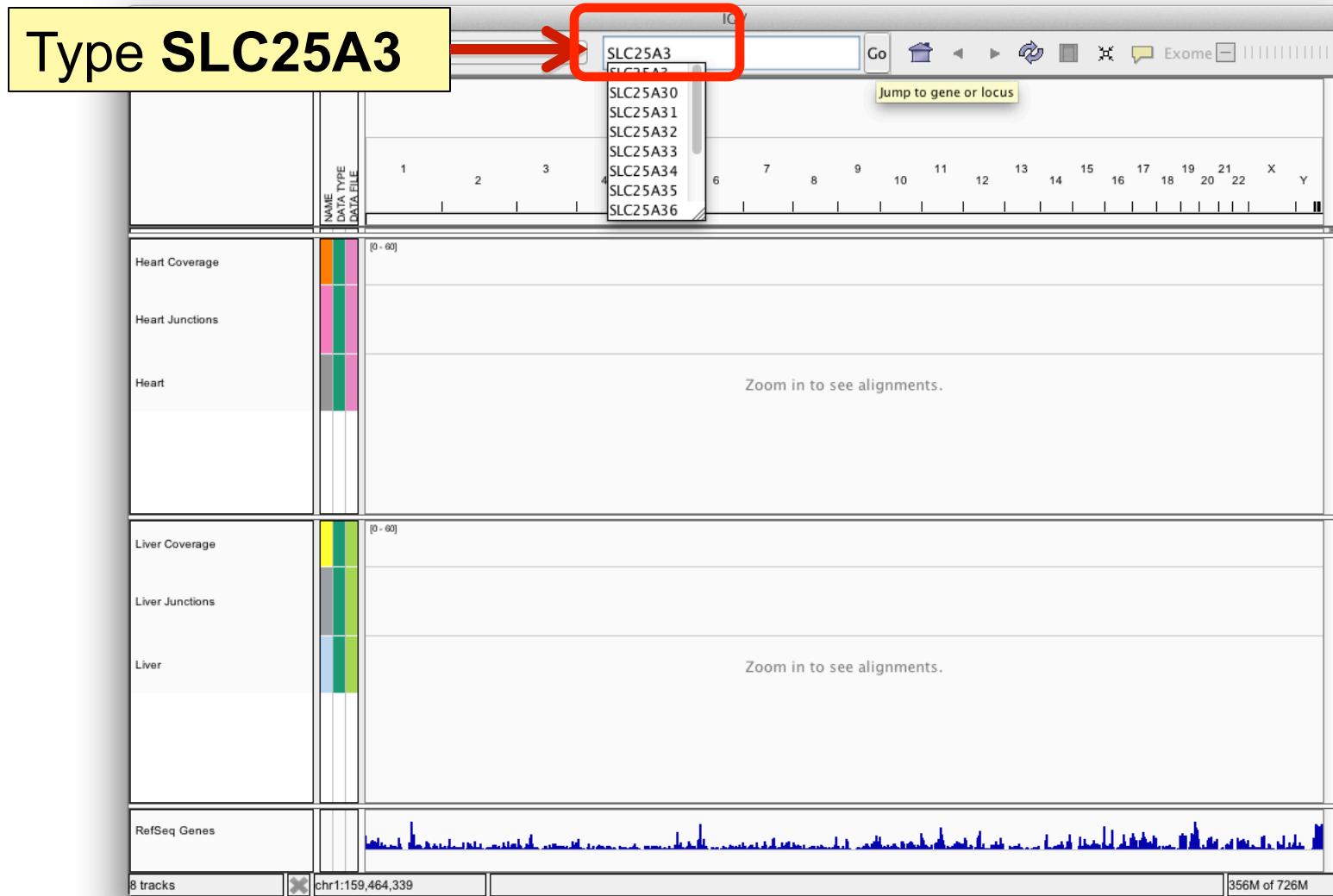
# RNA-seq alignments



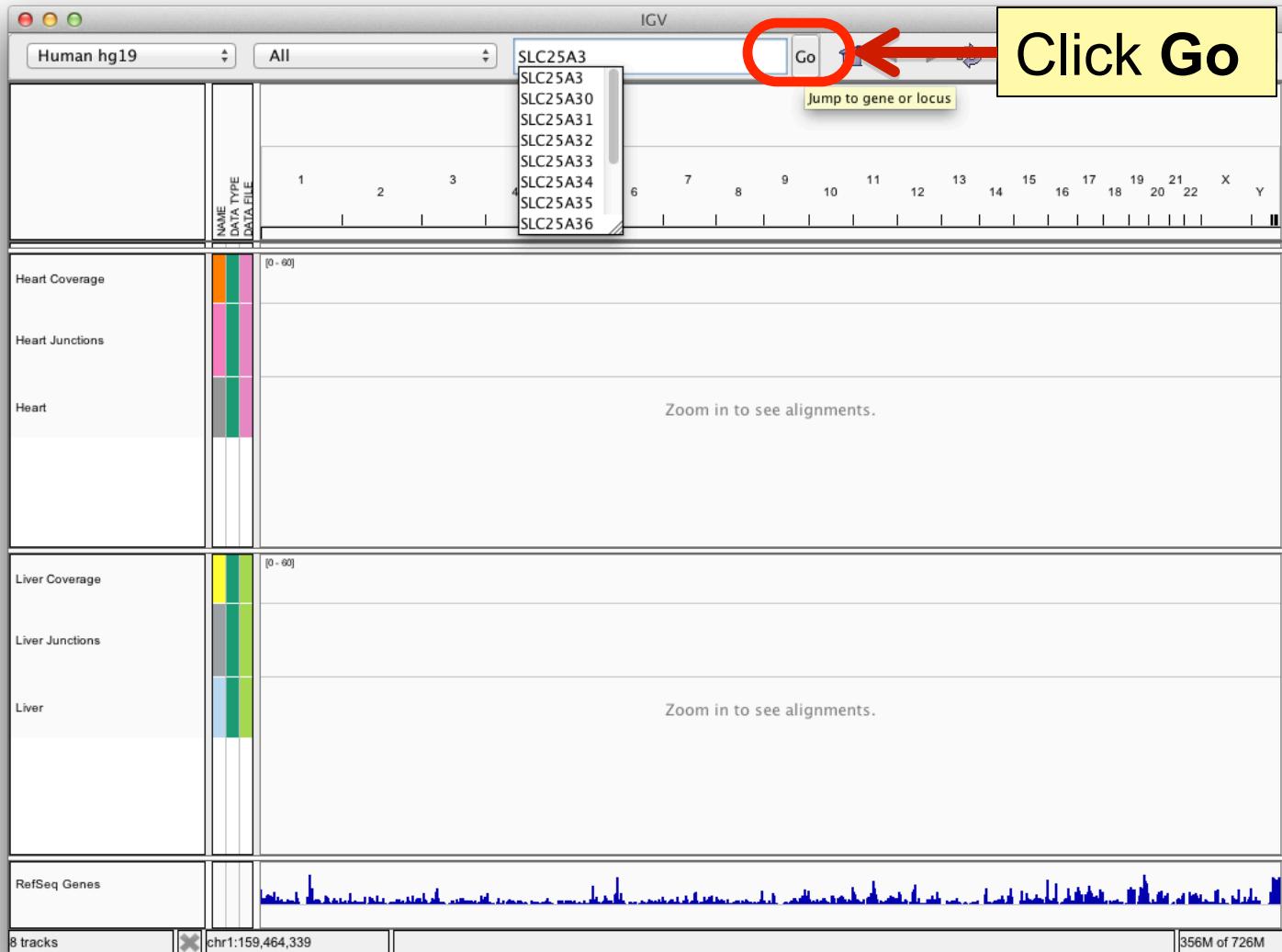
# RNA-seq alignments



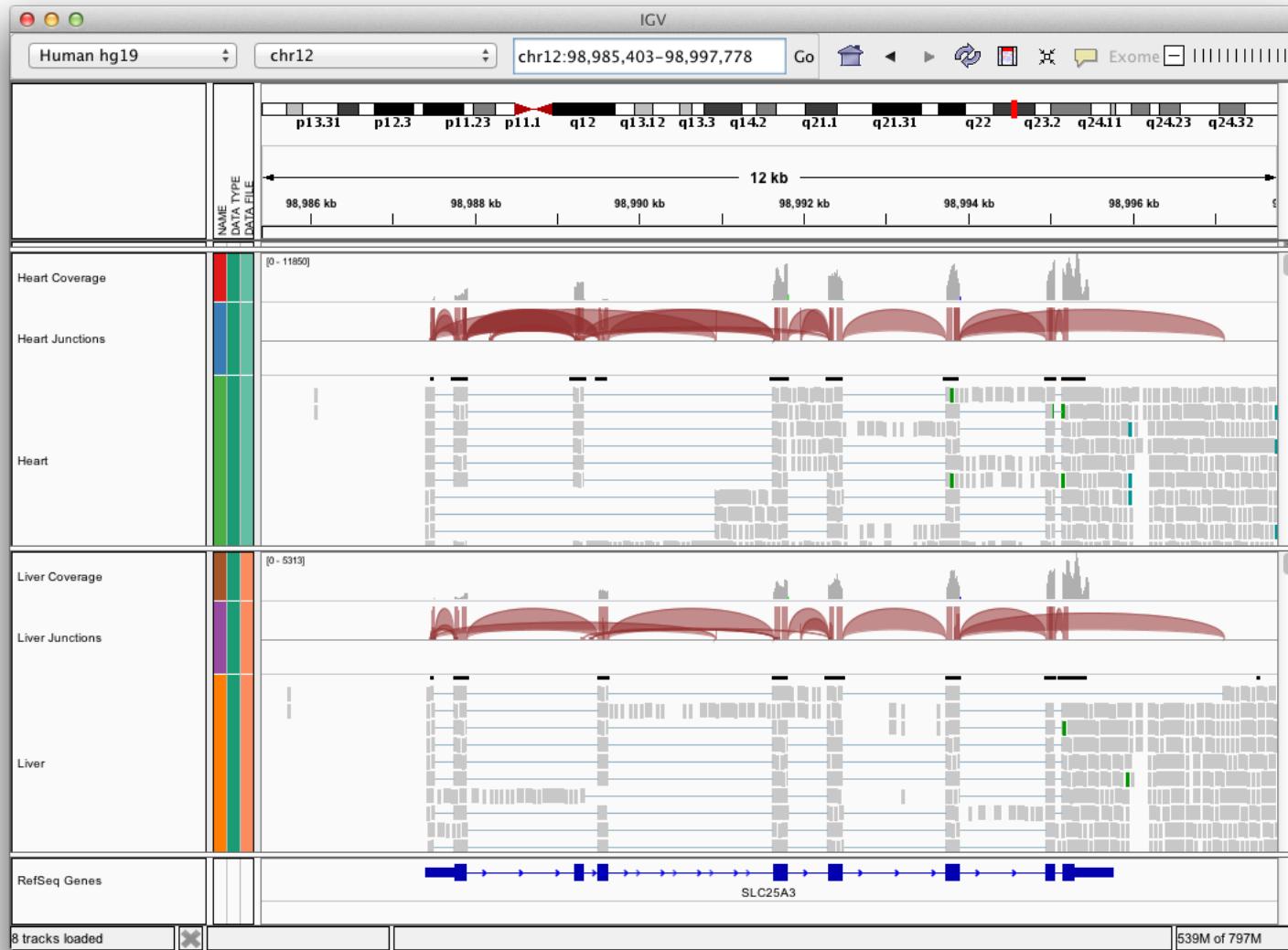
# RNA-seq alignments



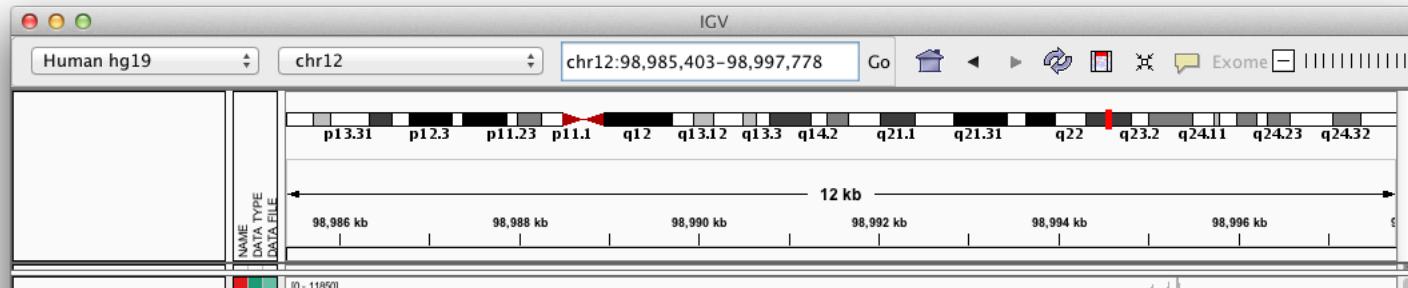
# RNA-seq alignments



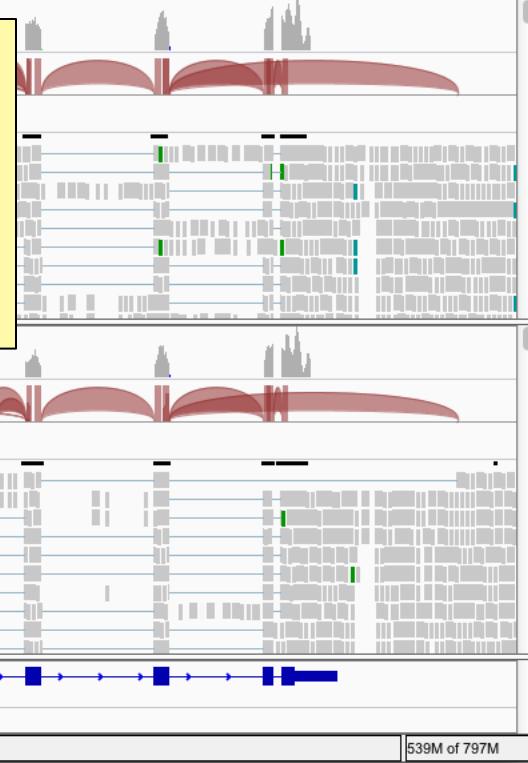
# RNA-seq alignments



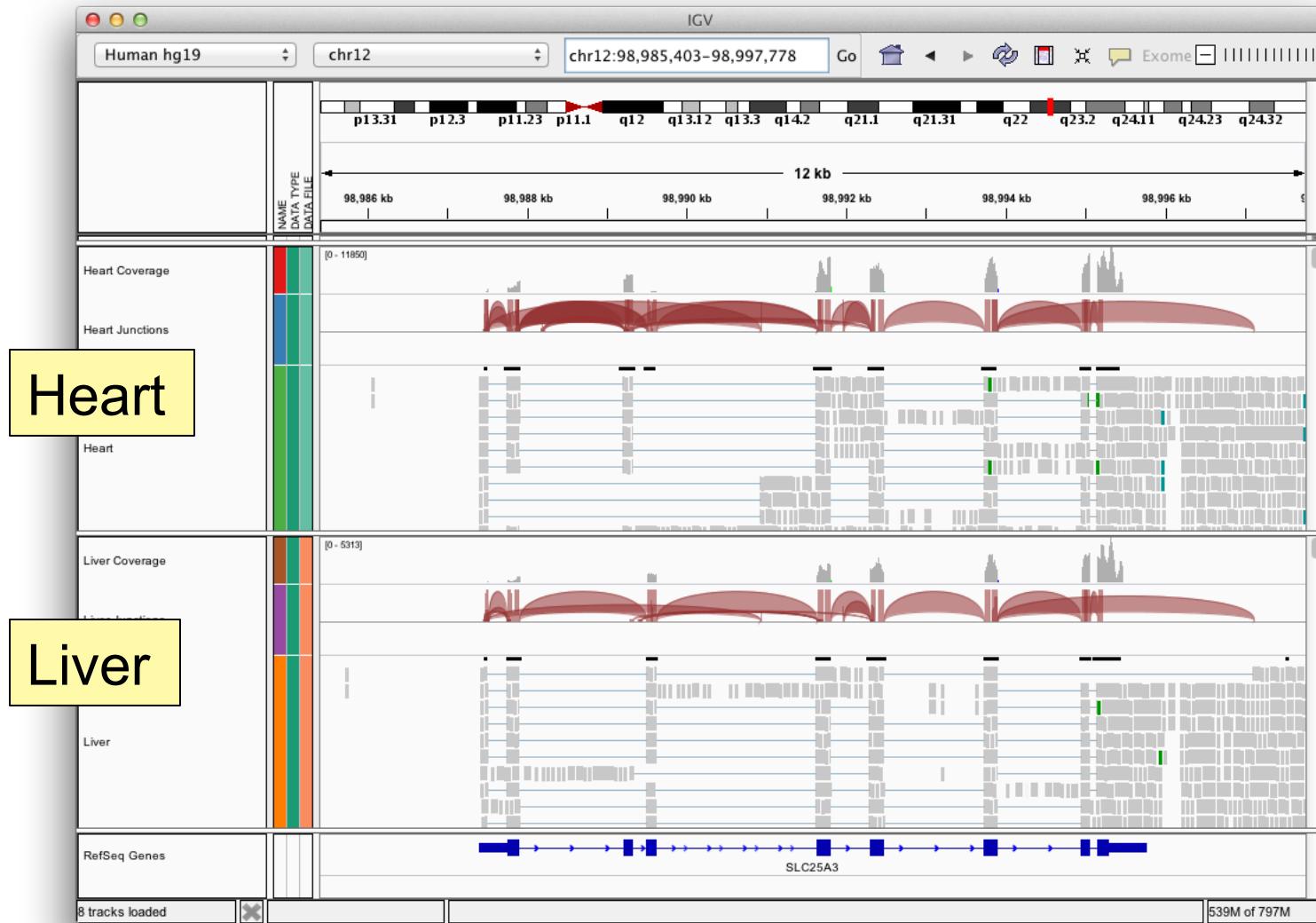
# RNA-seq alignments



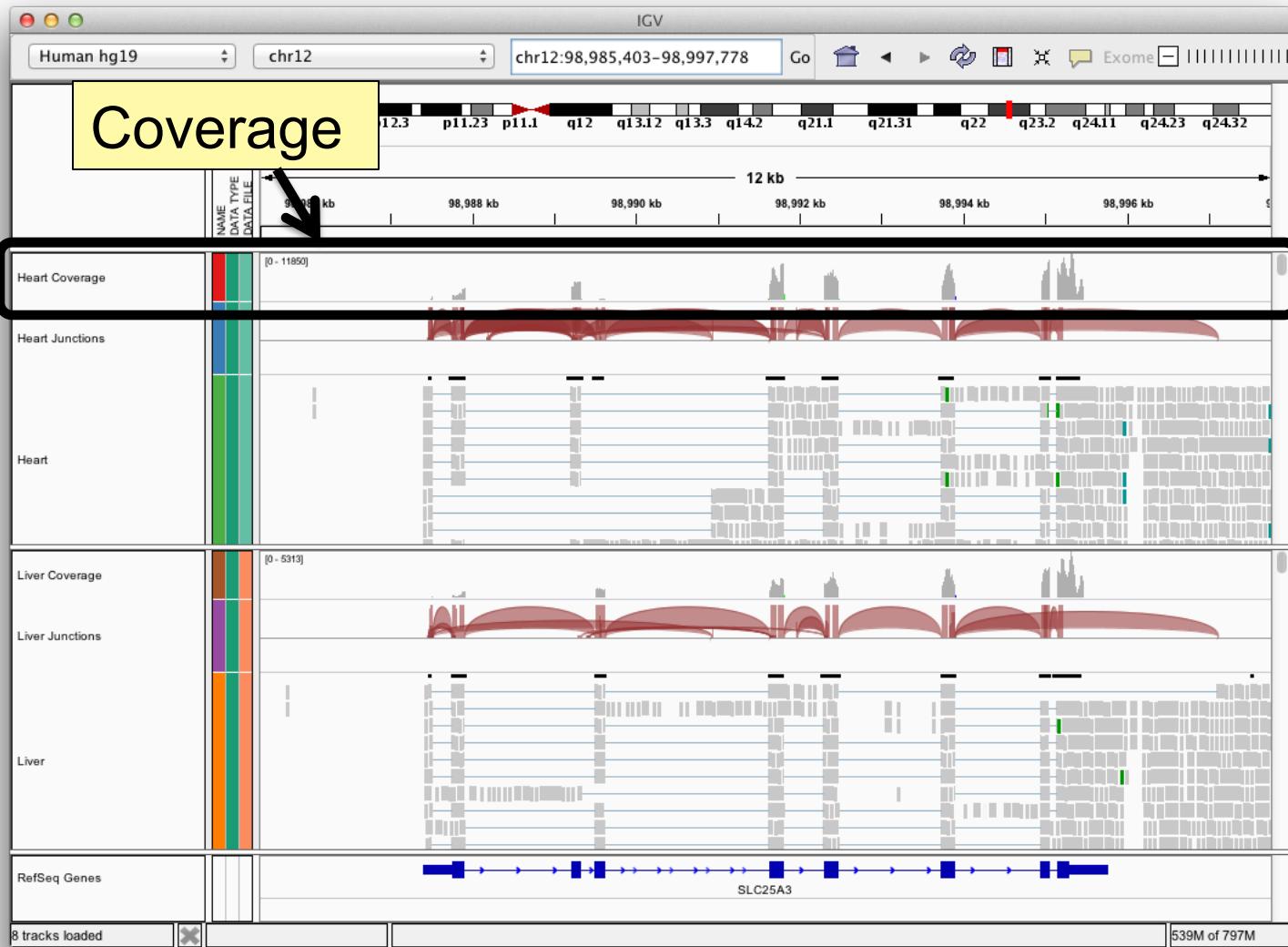
If reads are still blue & red from the settings for the last exercise, then right-click and select  
**Color alignments by > no color**



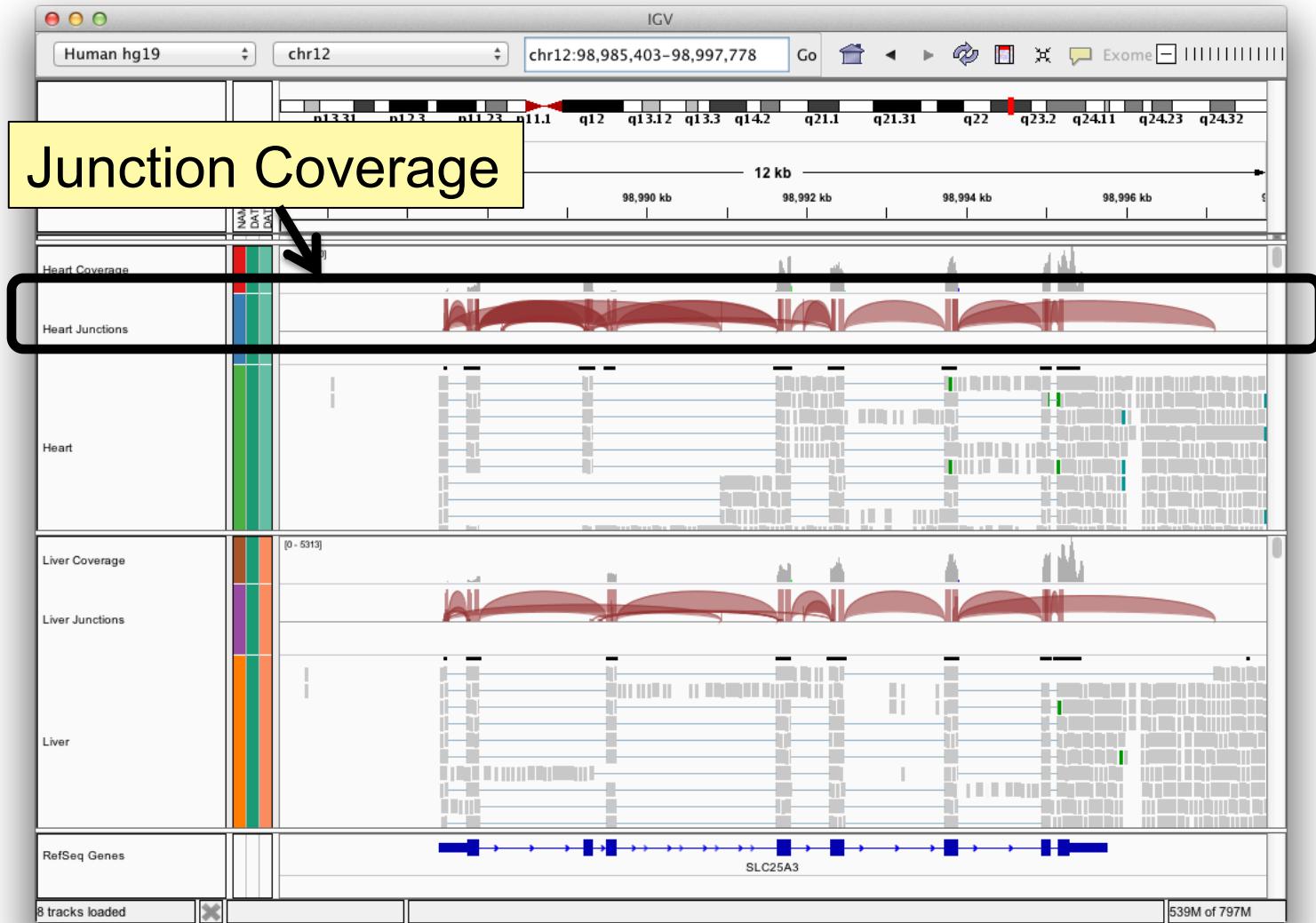
# RNA-seq alignments



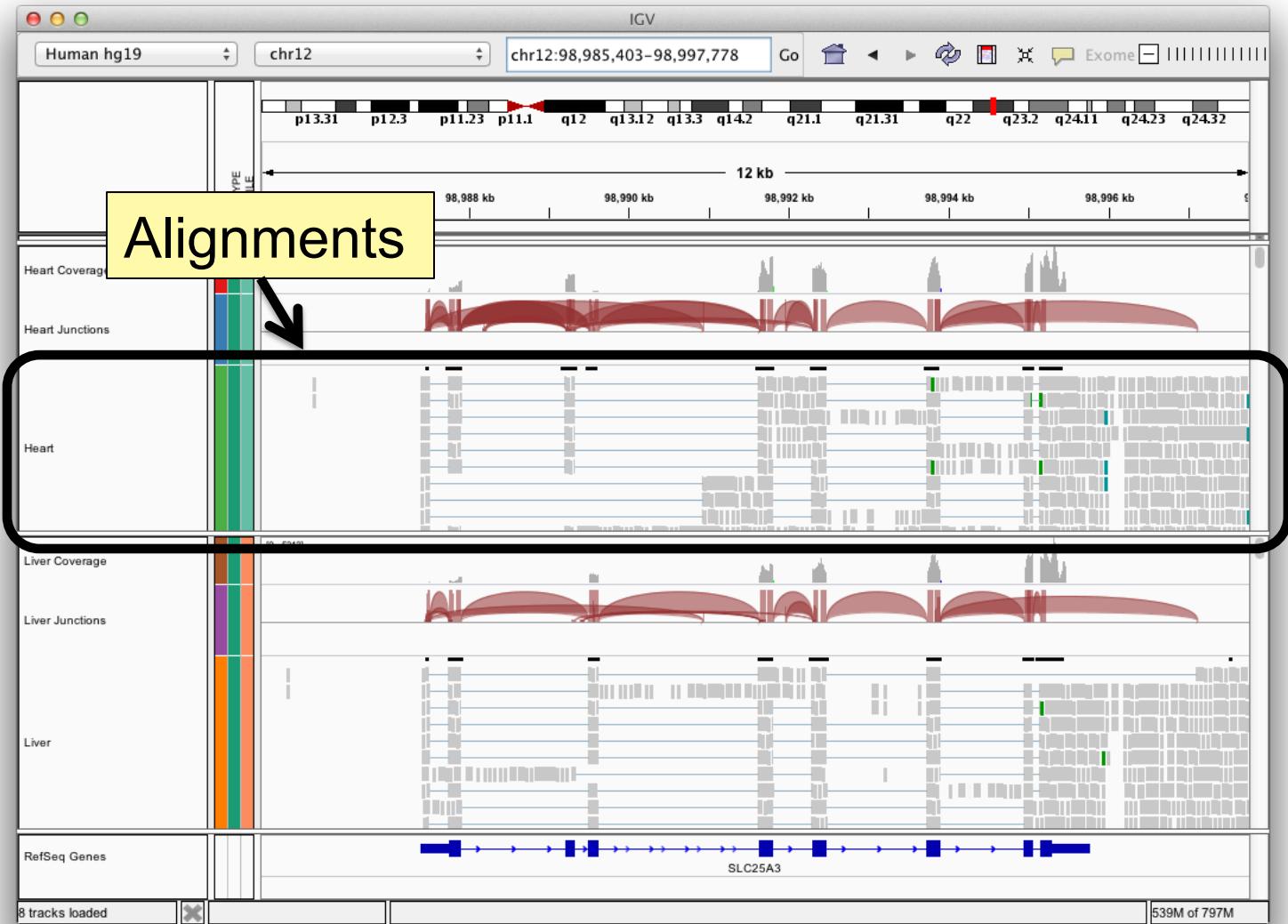
# RNA-seq alignments



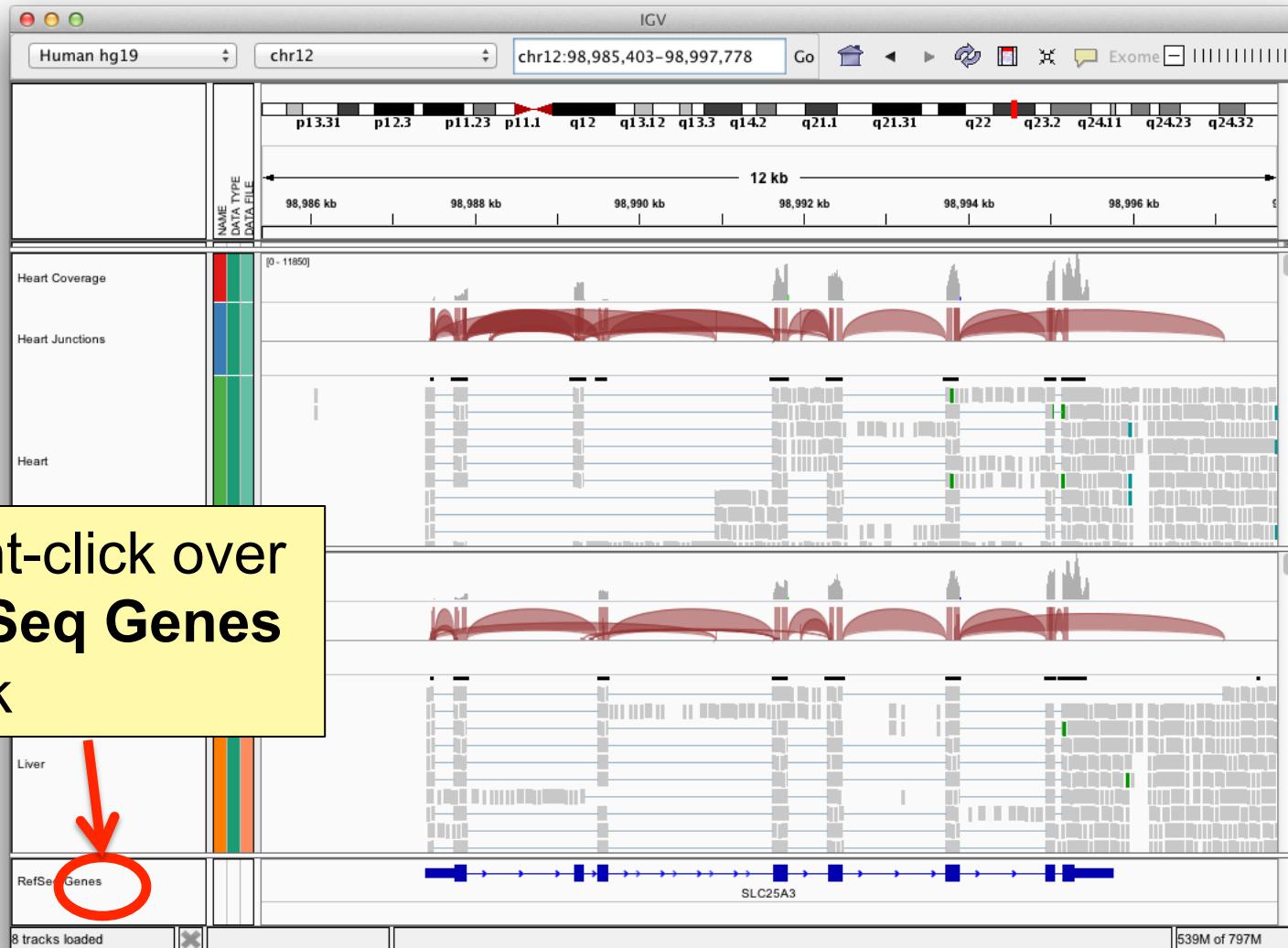
# RNA-seq alignments



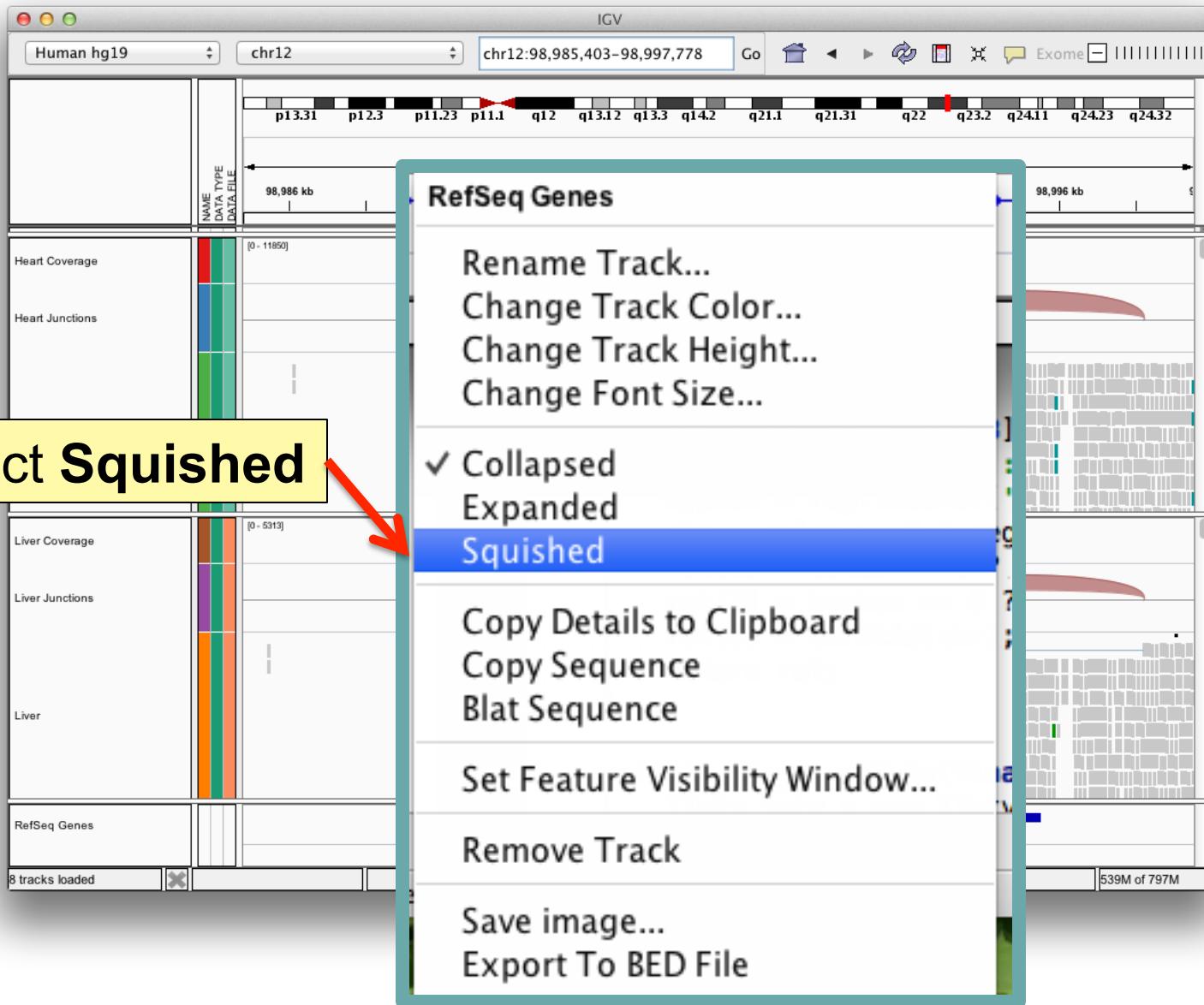
# RNA-seq alignments



# RNA-seq alignments



# RNA-seq alignments

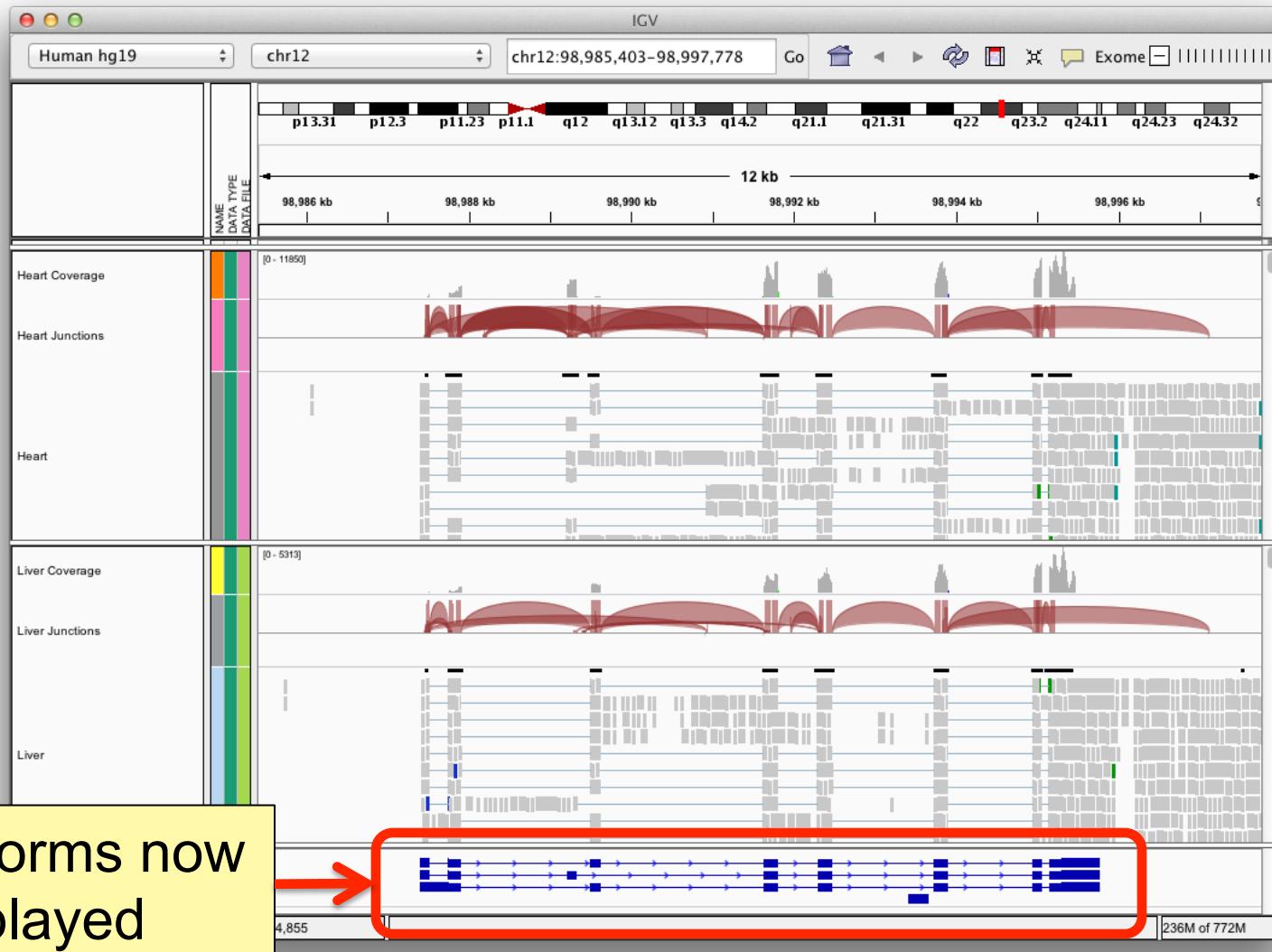


Select Squished

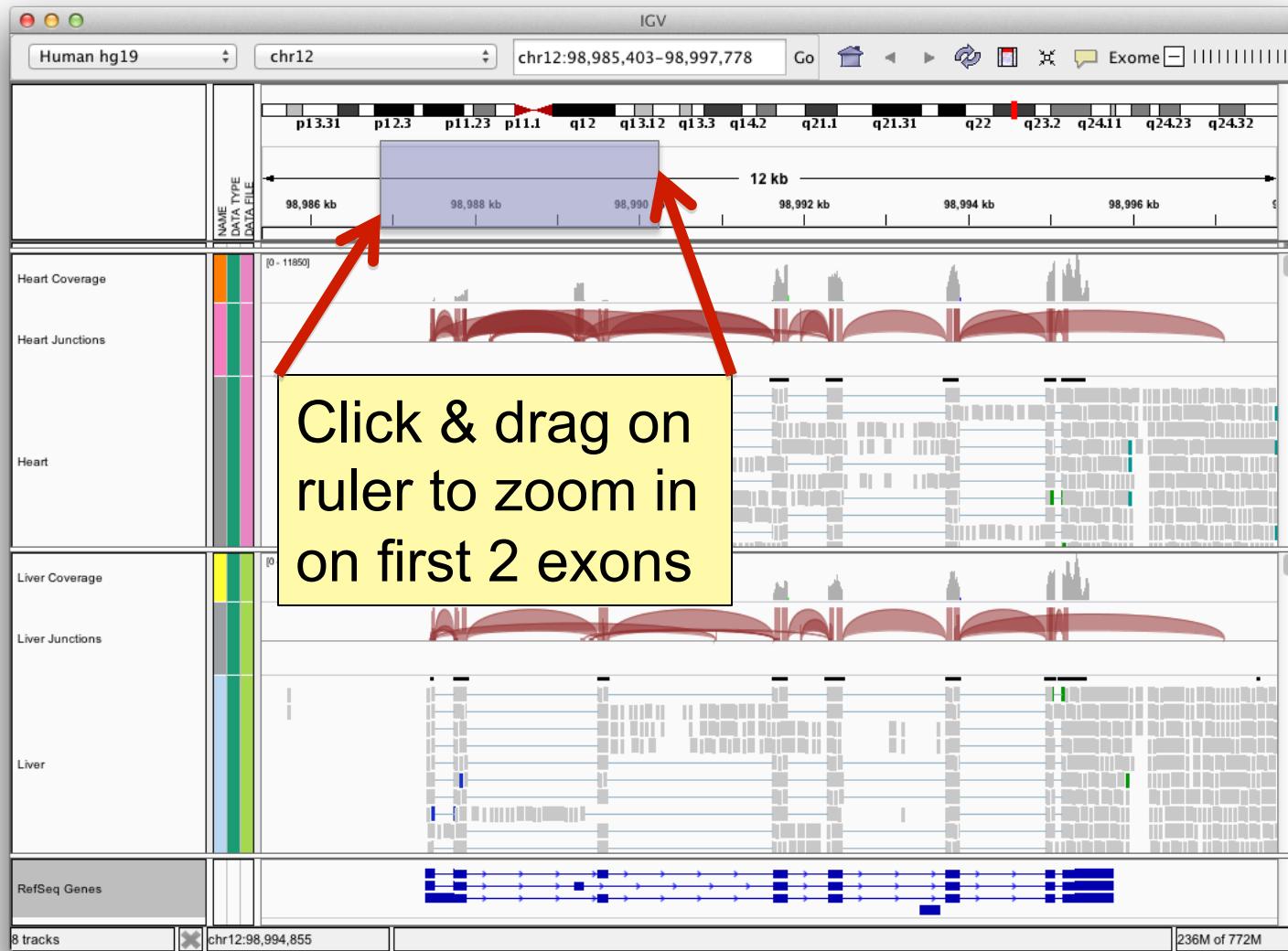
**RefSeq Genes**

- 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

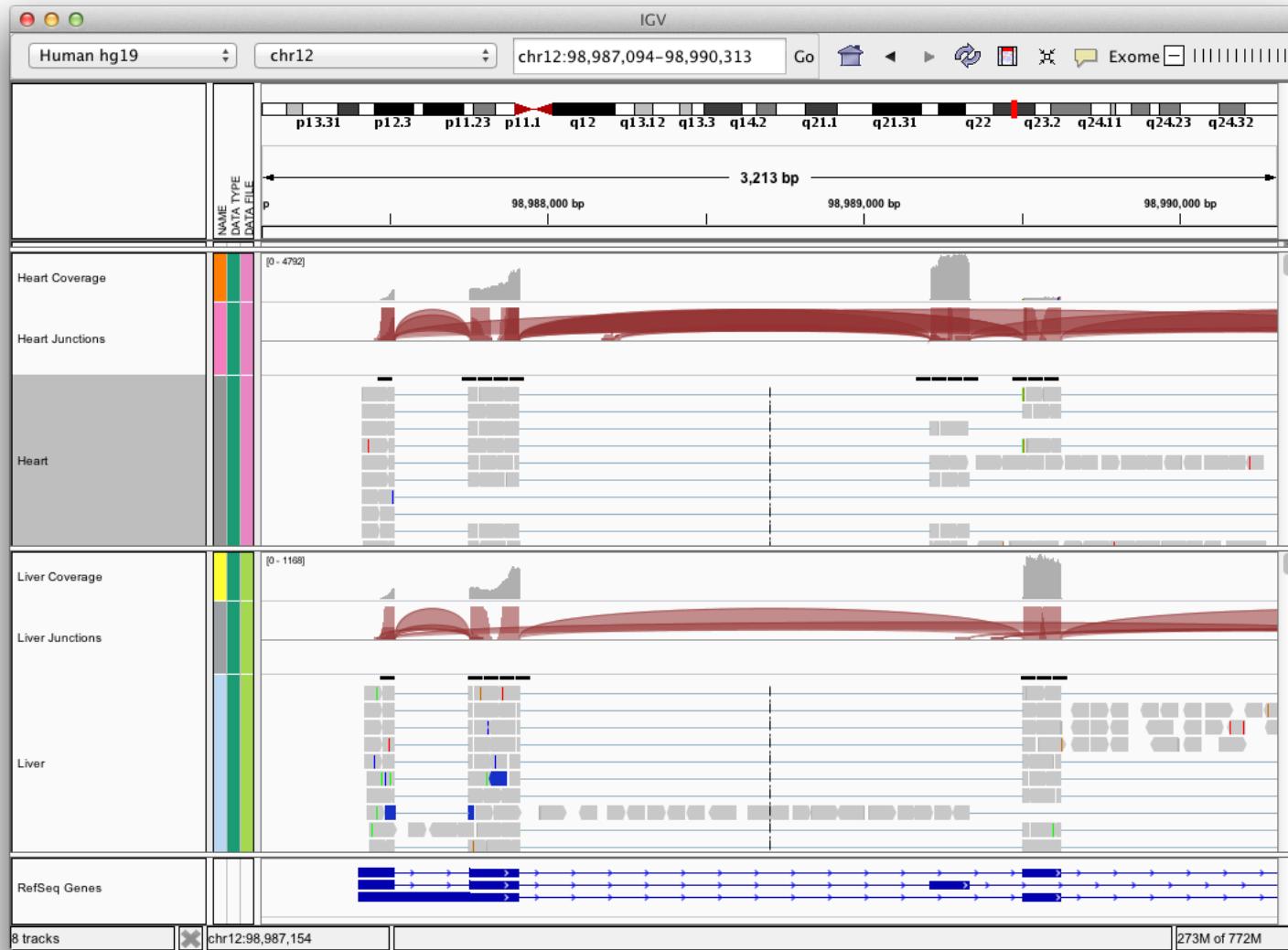
# RNA-seq alignments



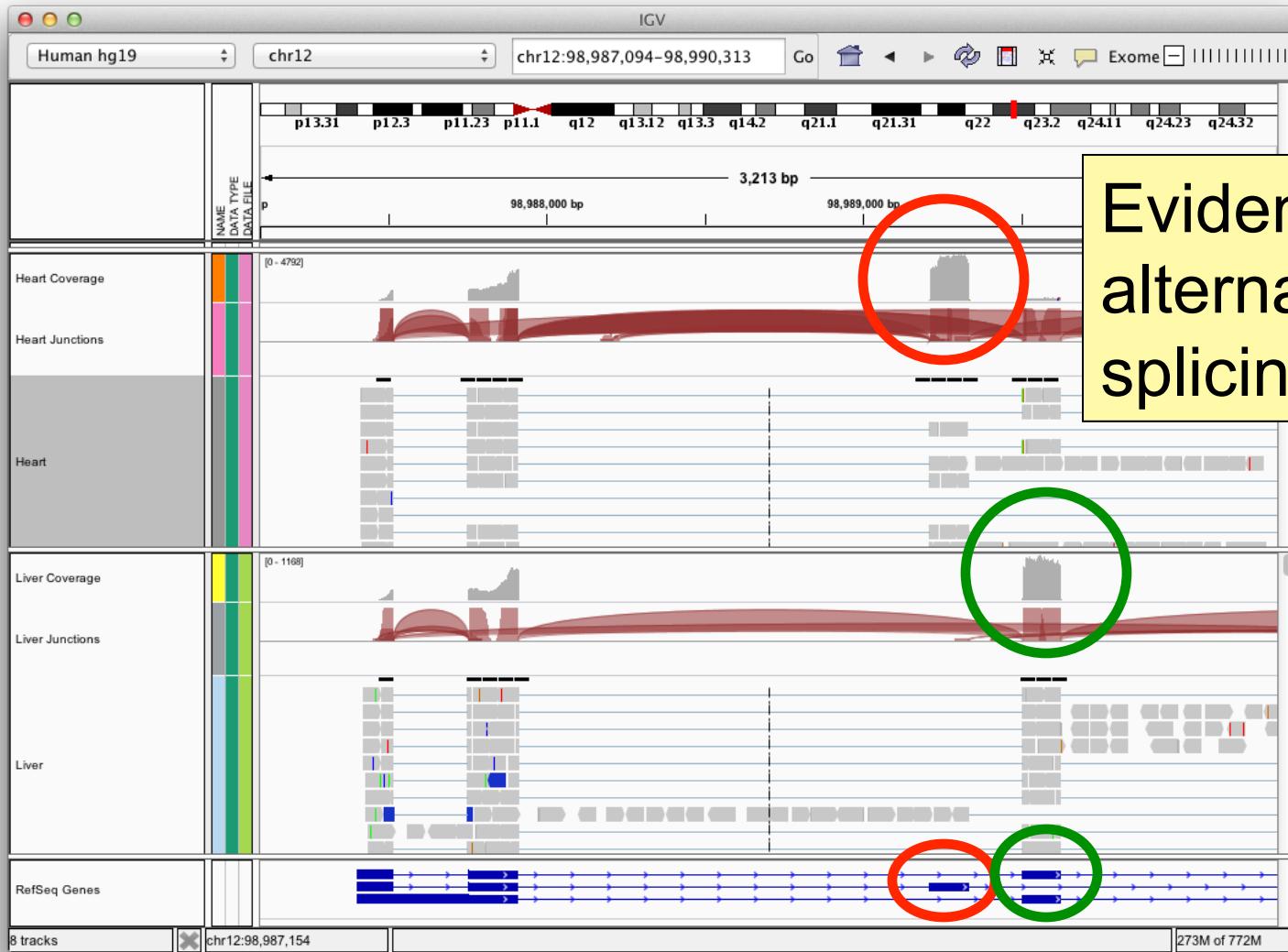
# RNA-seq alignments



# RNA-seq alignments



# RNA-seq alignments



Evidence of alternative splicing

# Sashimi plot

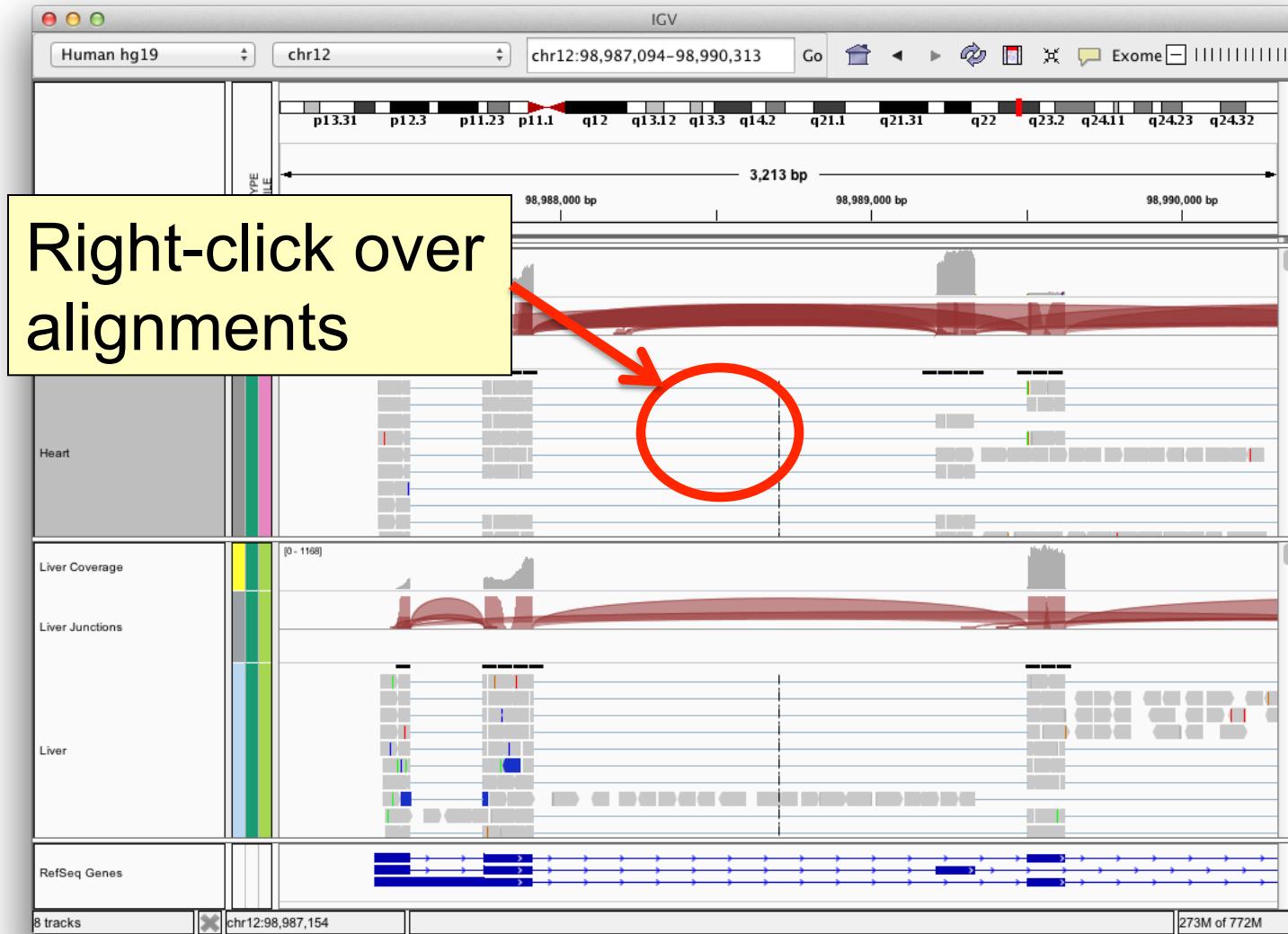
---

## Viewing RNA splicing with Sashimi Plots

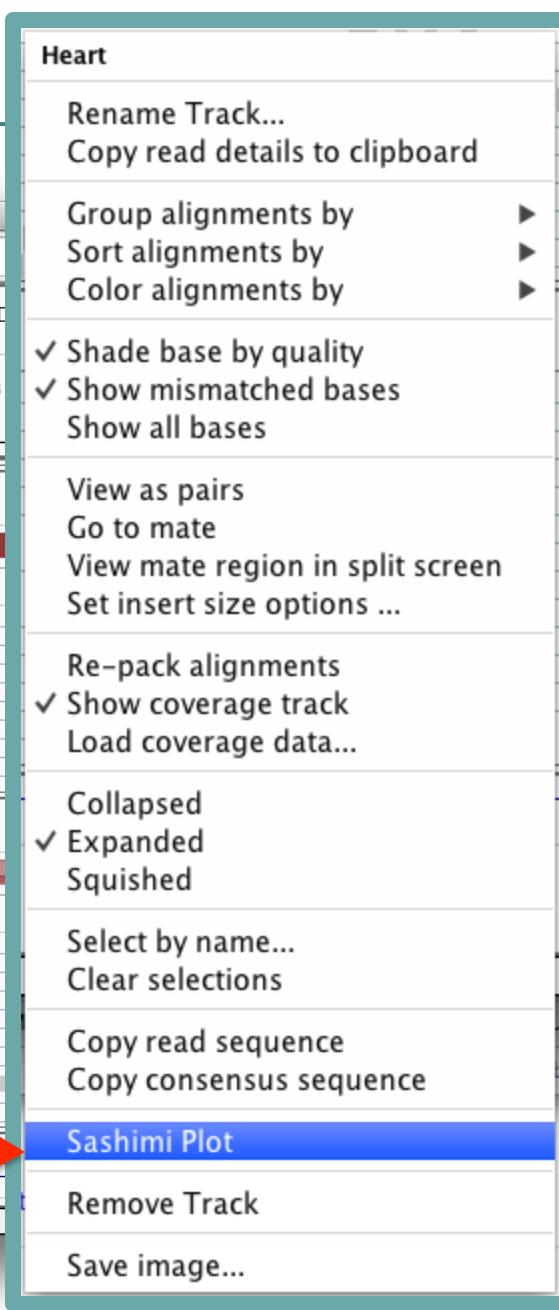
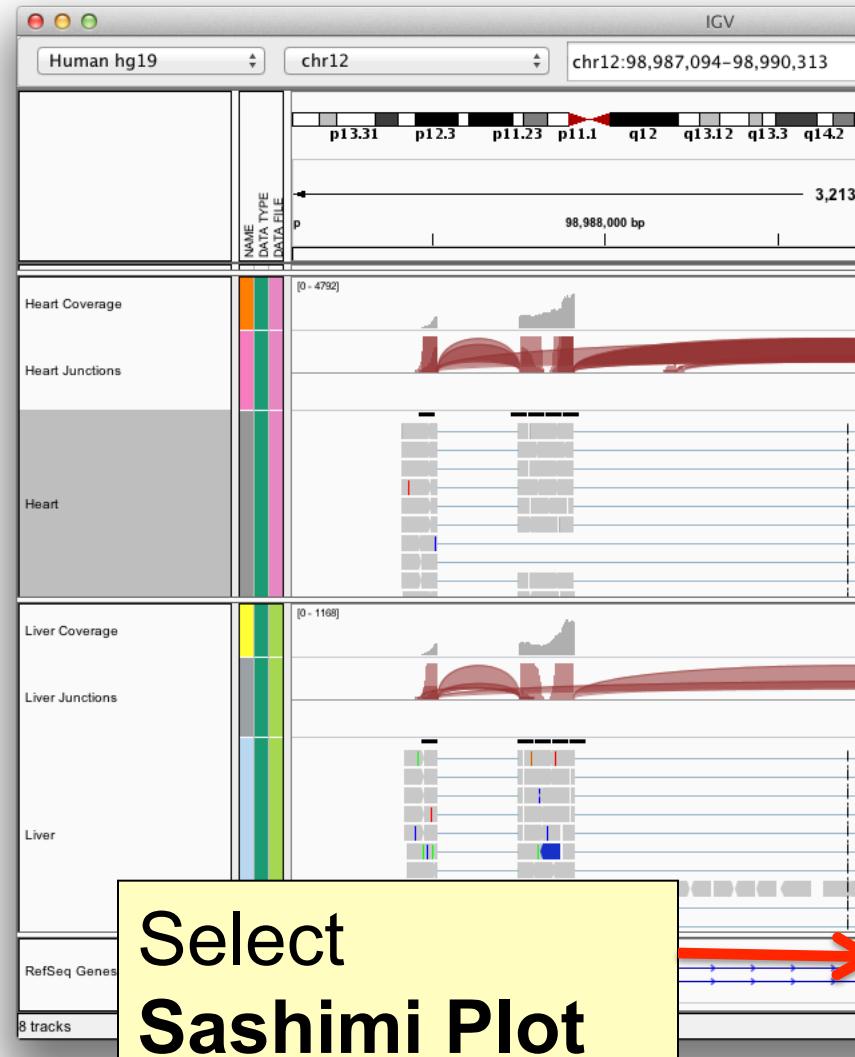
Reference: Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airoldi EM, Burge, CB.

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

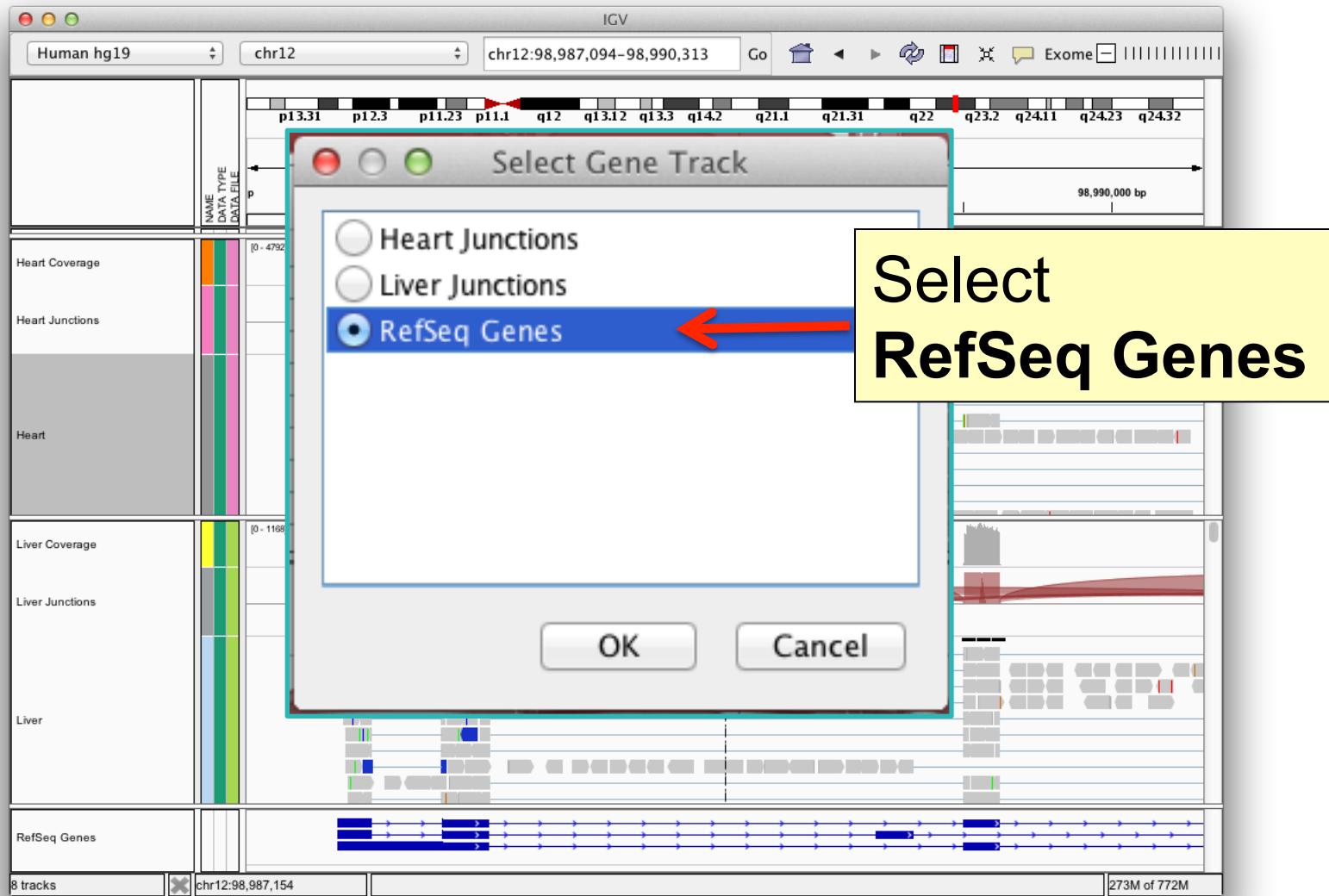
# RNA-seq alignments



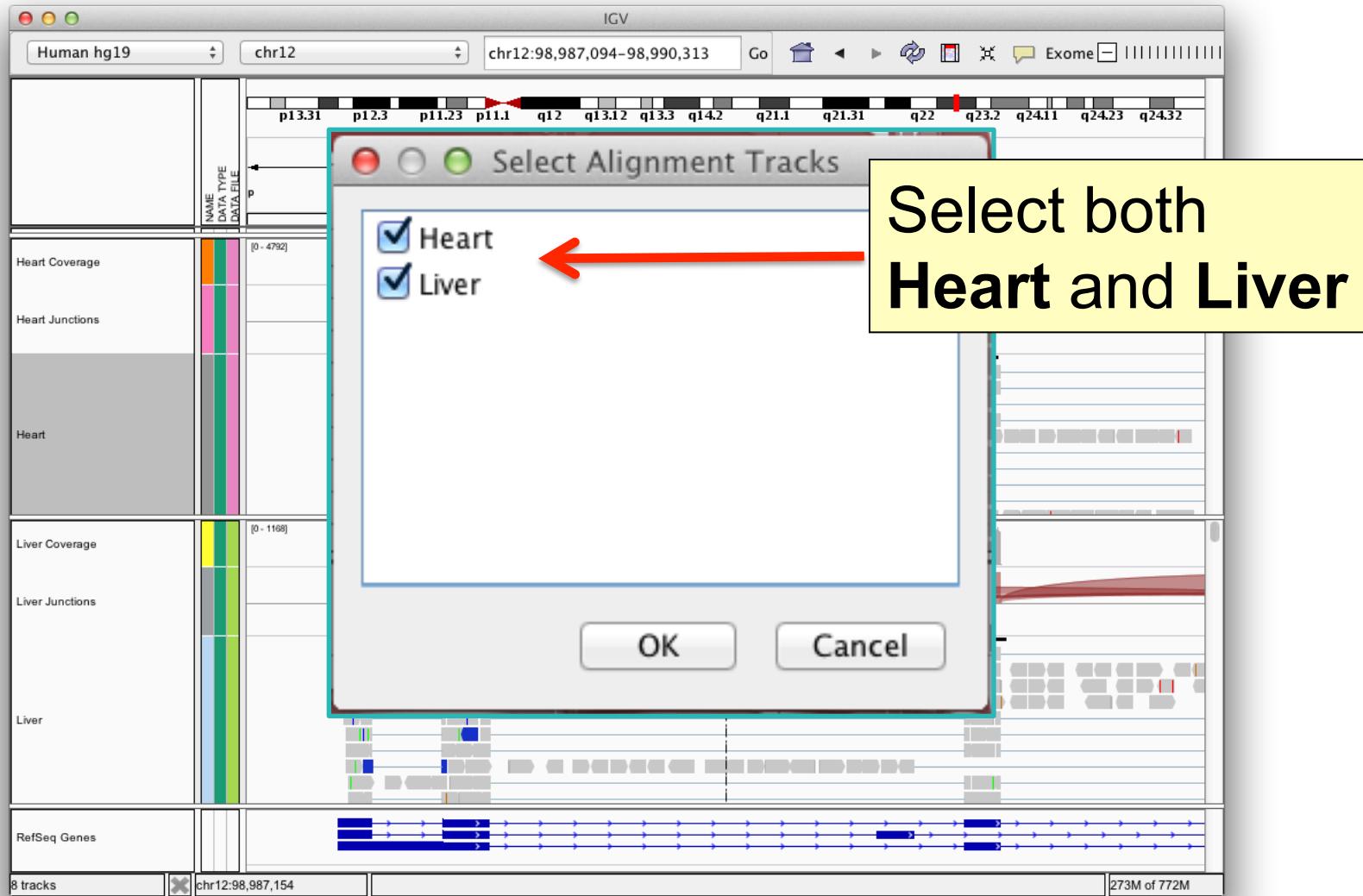
# RNA-seq alignments



# RNA-seq alignments



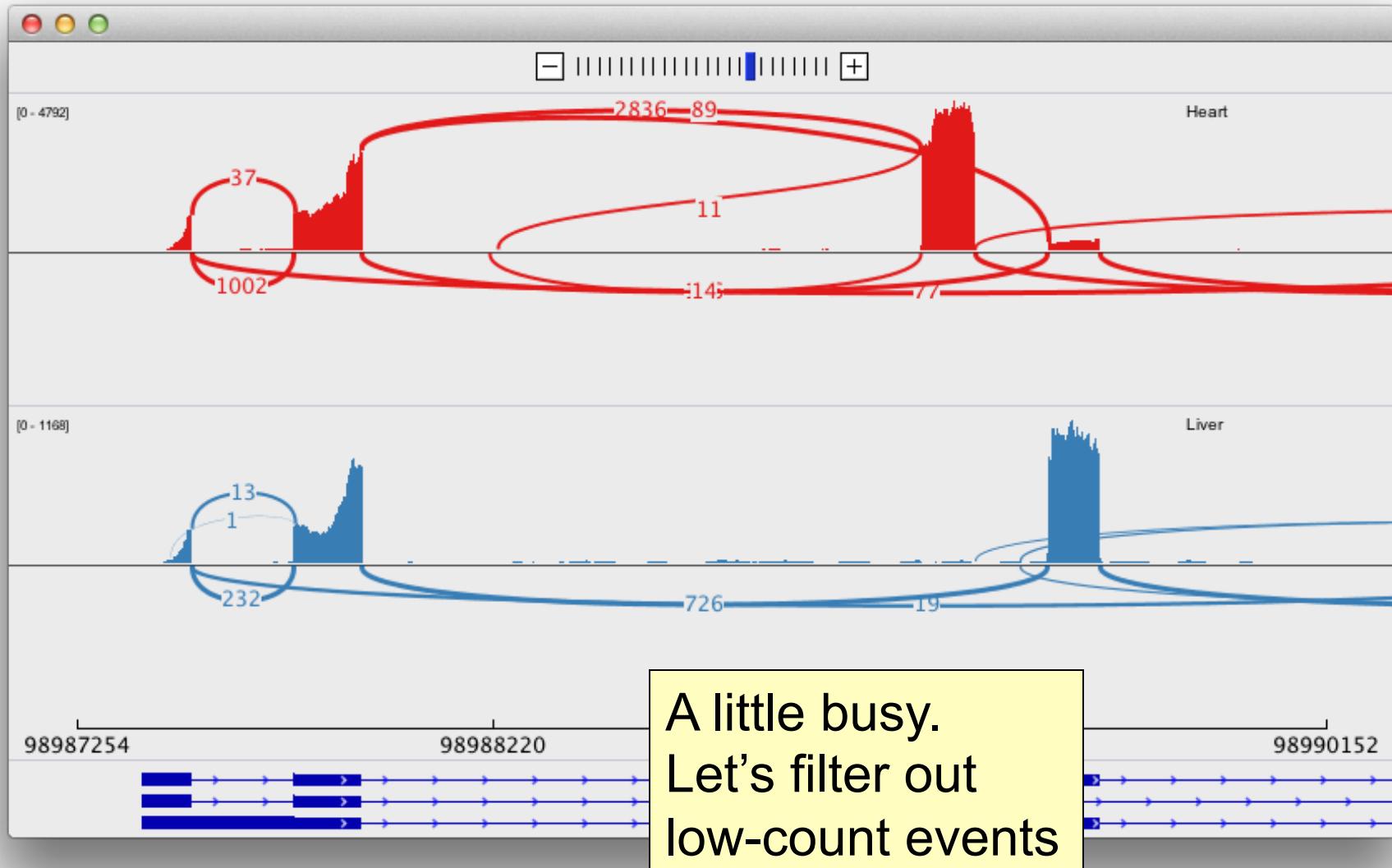
# RNA-seq alignments



# RNA-seq alignments



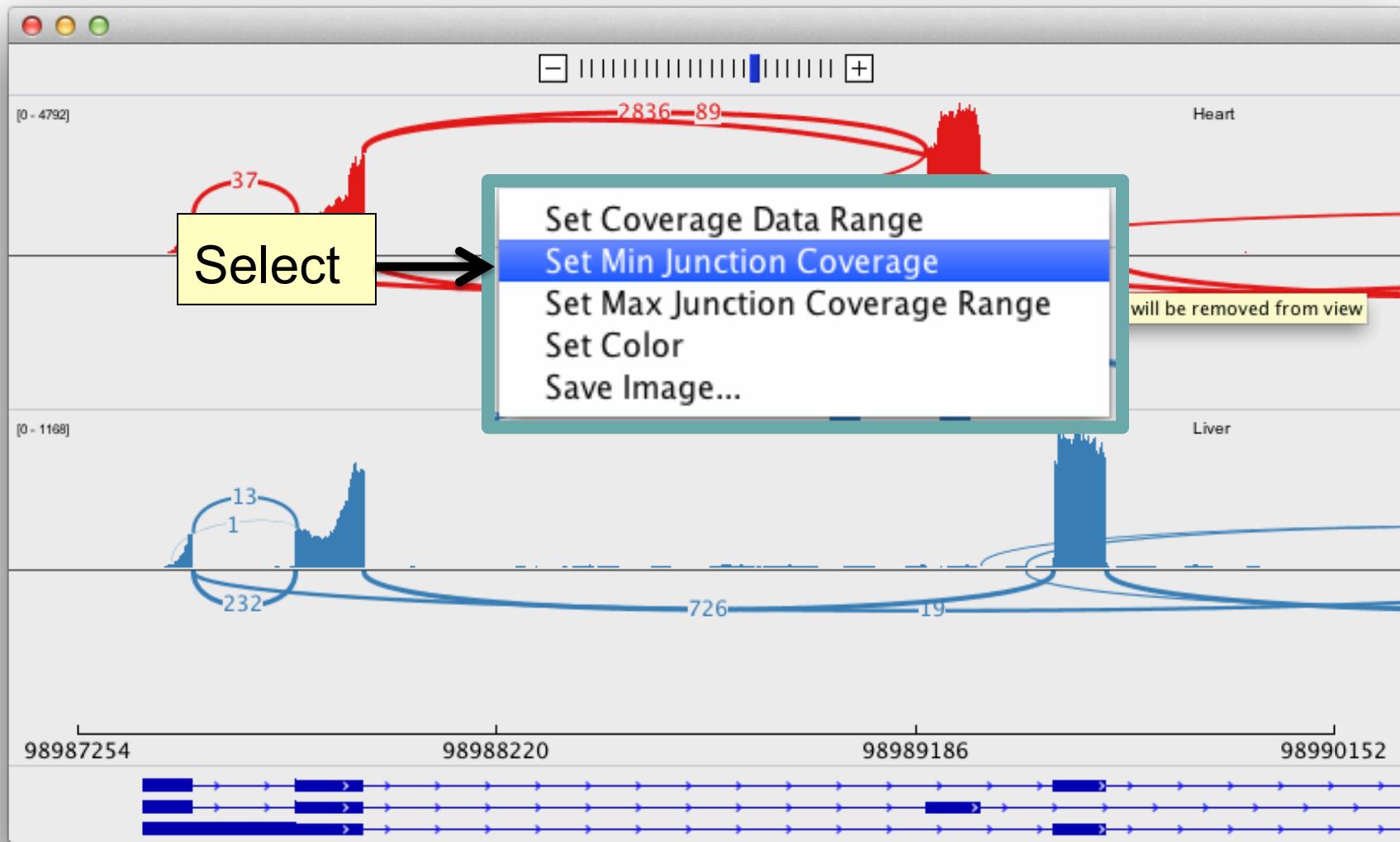
# RNA-seq alignments



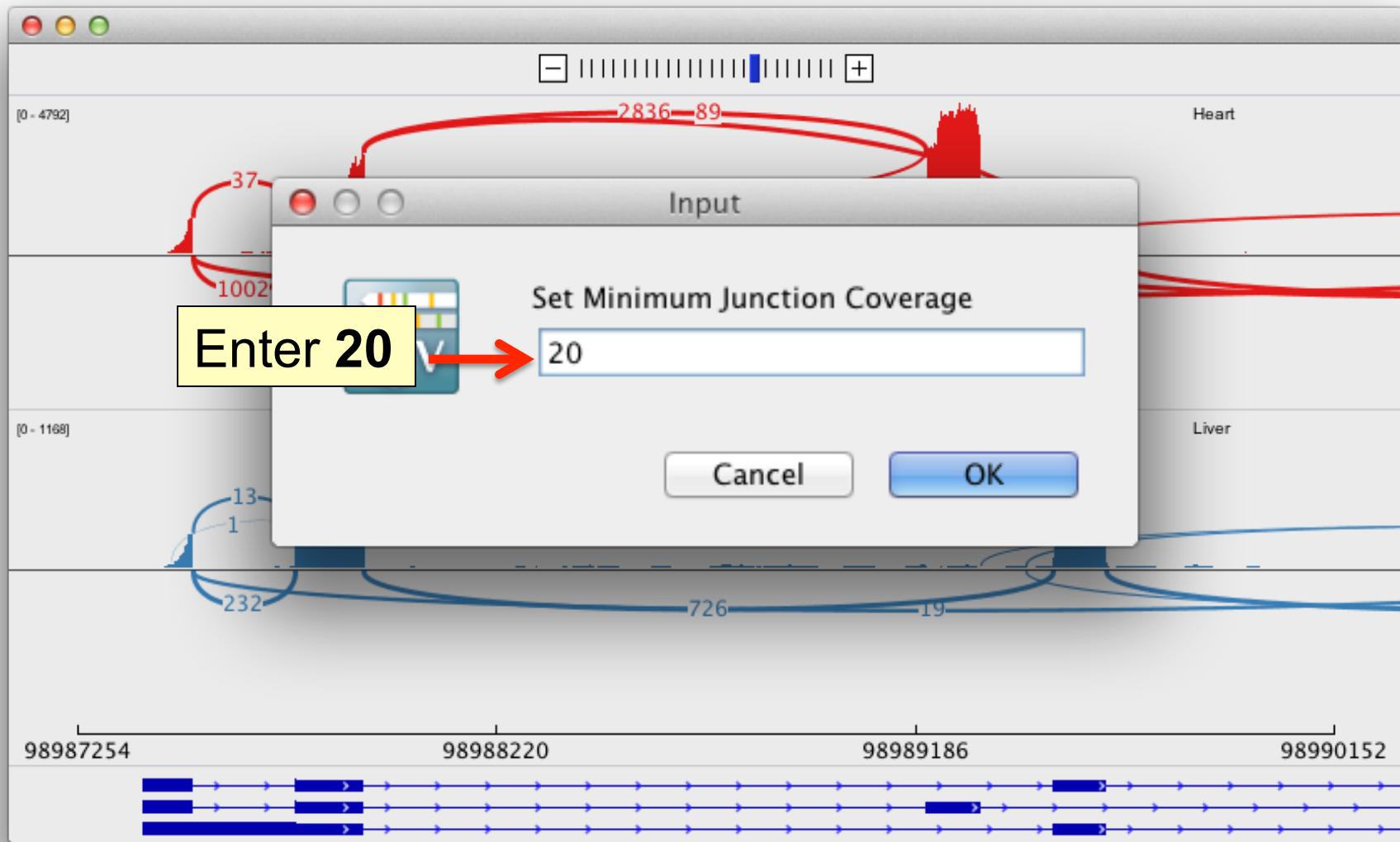
# RNA-seq alignments



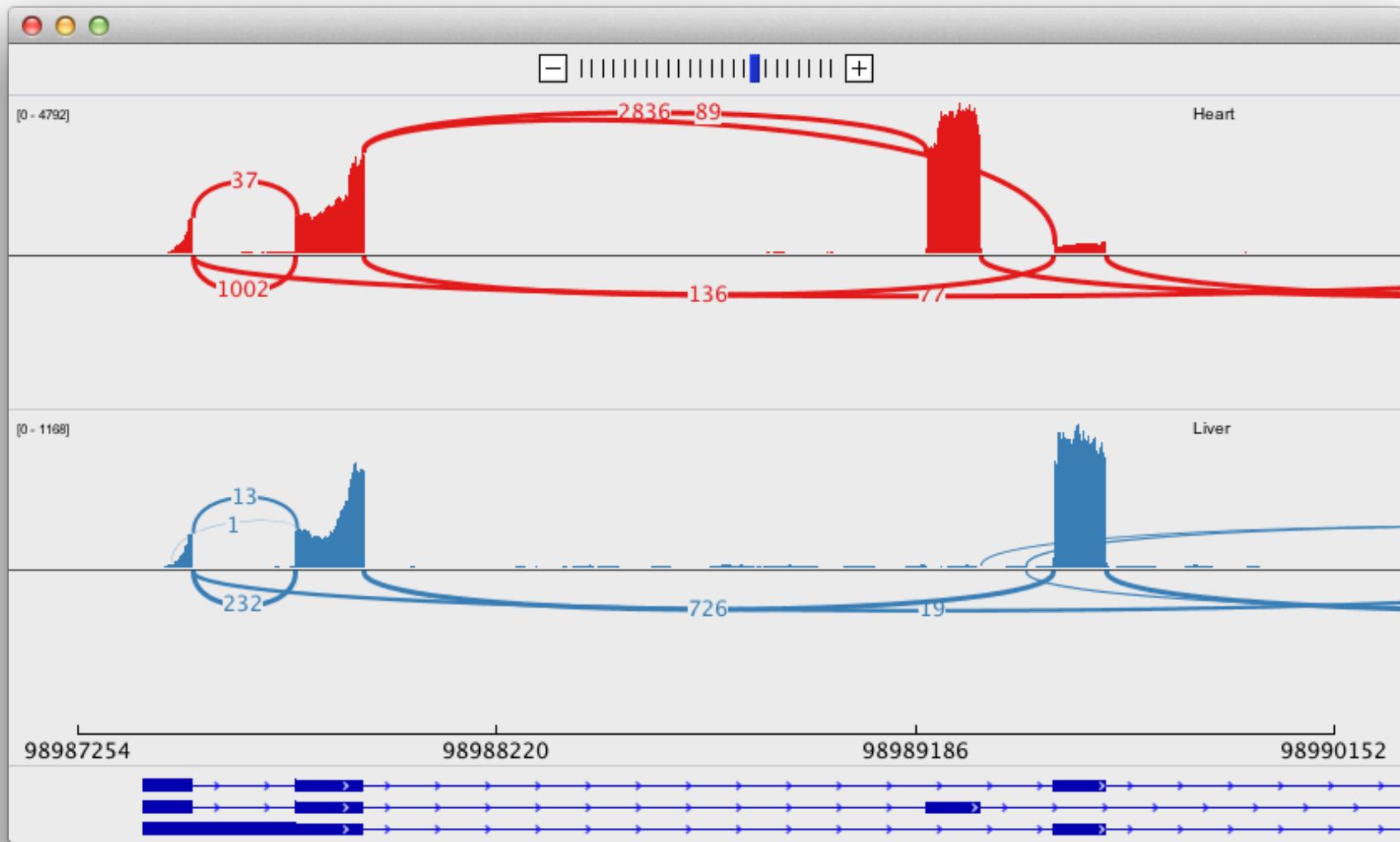
# RNA-seq alignments



# 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

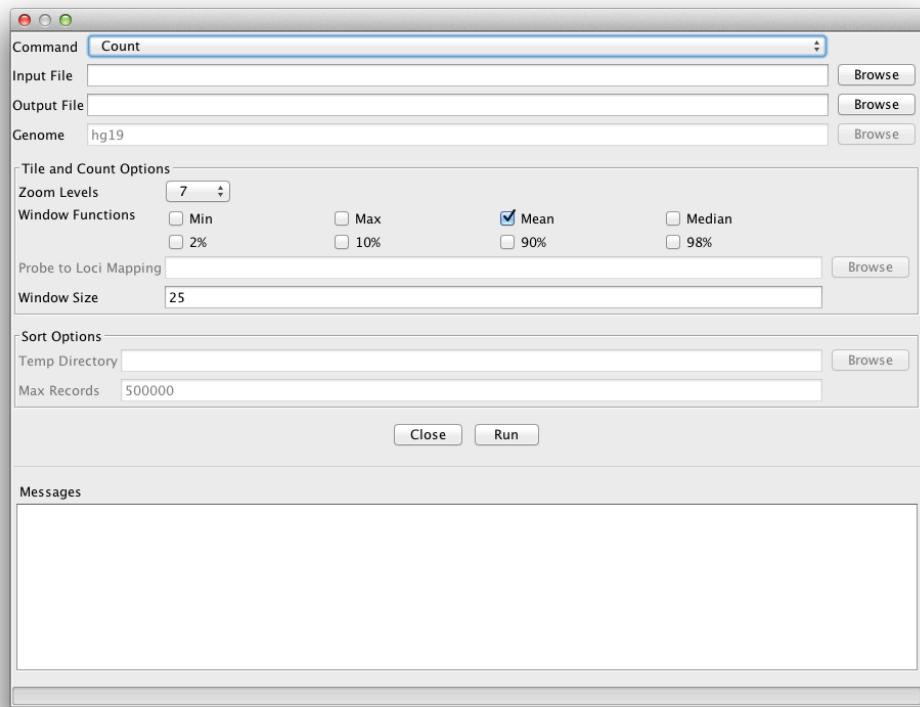
A set of utilities for preparing files for efficient display.

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

# 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

# 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).