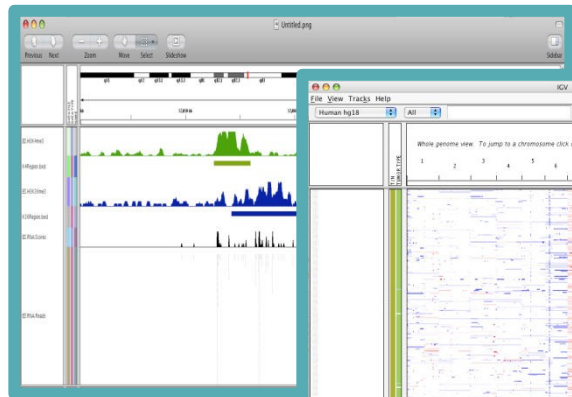


# Introduction to NGS Visualization with the Integrative Genomics Viewer (IGV)

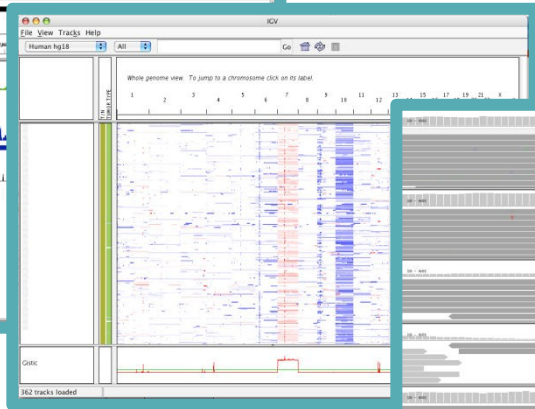
# Integrative Genomics Viewer (IGV)



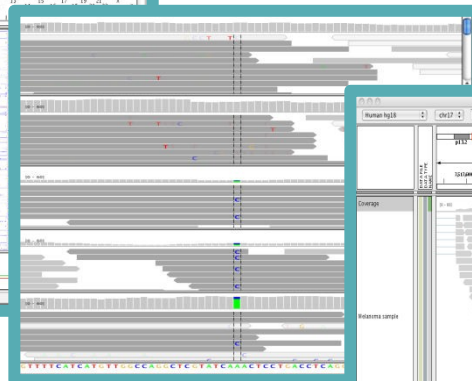
*Desktop application for the interactive visual exploration of integrated genomic datasets*



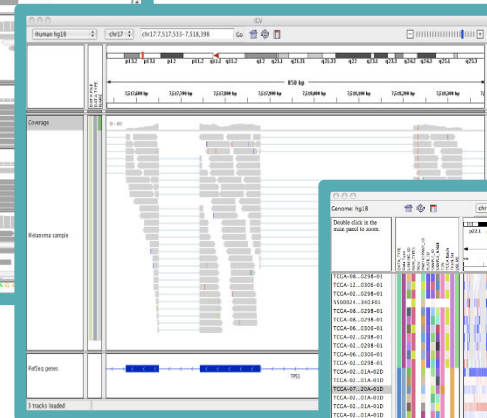
**Epigenomics**



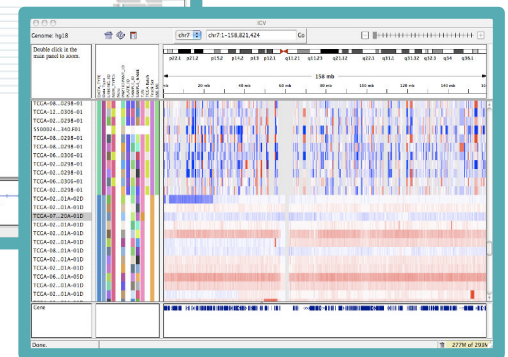
**Microarrays**



**NGS alignments**



**RNA-Seq**



**mRNA, CNV, Seq**

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

65,000 registrations



# Features

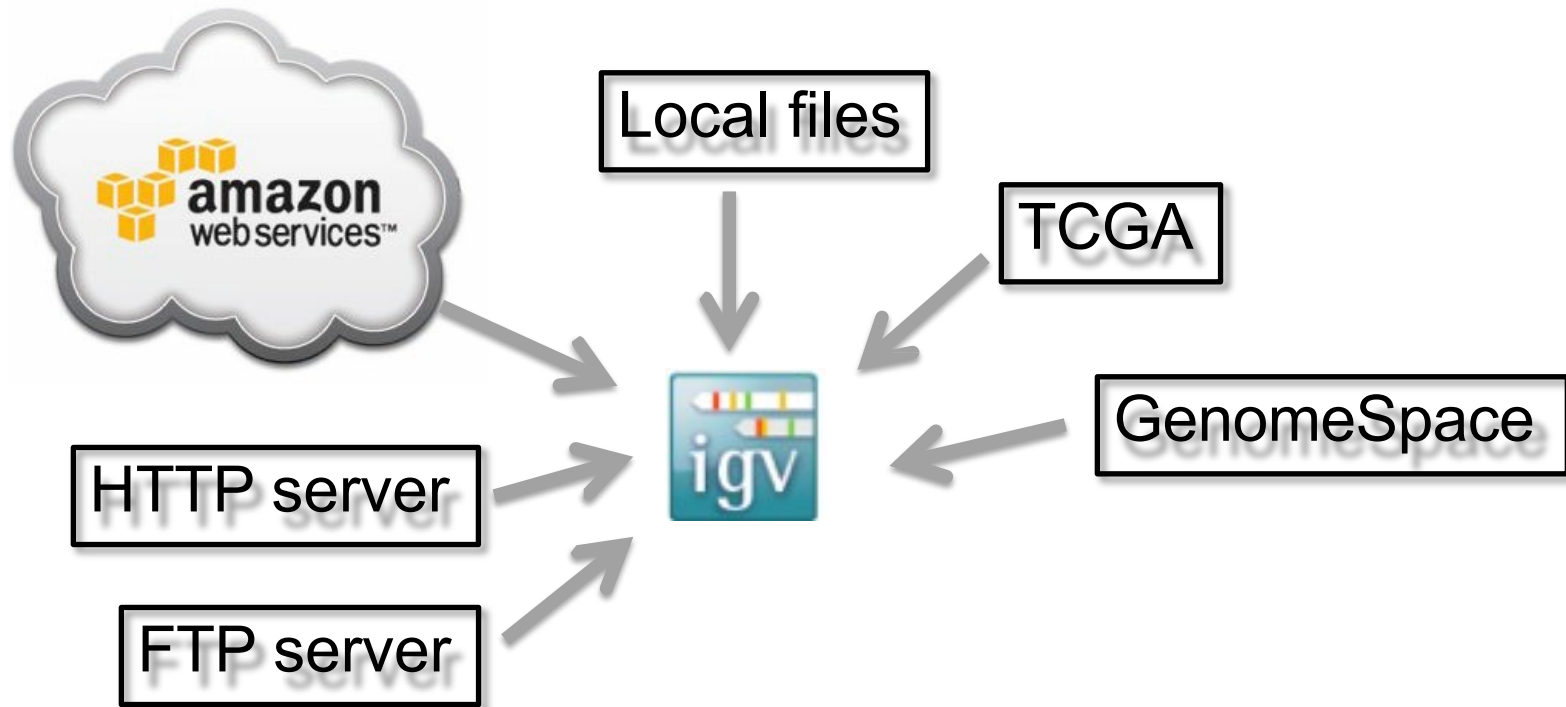
---



With IGV you can...

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

# IGV data sources



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

# Using IGV: The Basics

# Using IGV: the basics

---



Hands-on exercise

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

# Launch IGV



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

The screenshot shows the homepage of the Integrative Genomics Viewer (IGV) website. The browser address bar displays <https://www.broadinstitute.org/igv/>. On the left side, there is a navigation menu with the following items: Home, Downloads (highlighted with a red circle and a red arrow), Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below the menu is a search box and the Broad Institute logo. The main content area features a large banner with the text "Integrative Genomics Viewer" and a background image of the IGV interface. Below the banner, there are sections for "What's New" (with news items from July 3, 2012, and April 20, 2012), "Citing IGV" (with citation information for James T. Robinson et al. and Helga Thorvaldsdottir et al.), "Overview" (describing IGV as a high-performance visualization tool), "Downloads" (with a registration requirement), and "Funding" (listing funding from the National Cancer Institute, National Institute of General Medical Sciences, and Starr Cancer Consortium). The Broad Institute logo is also present at the bottom right of the page.

# Launch IGV



Registration | Integrative Genomics Viewer

www.broadinstitute.org/software/igv/?q=registration

Home » Registration

## Registration

### IGV Registration

IGV is an open-source application, released under the terms of the [GNU Lesser General Public License \(LGPL\)](#). To download IGV fill in the form below and click "Agree" to indicate you have reviewed and agreed to the licensing terms. This information is only used to help us track usage for reports to our funding agencies and will not be used for other purposes.

Name

Email

Organization

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

Search website  
  
search  
[Broad Home](#)  
[Cancer Program](#)

BROAD INSTITUTE  
© 2013 Broad Institute



# Launch IGV



Downloads | Integrative Genomics Viewer

www.broadinstitute.org/igv/download

Home > Downloads

### Downloads

#### Integrative Genomics Viewer (Version 2.3)

**Mac Users:** Apple has pushed out an update that blocks all but the latest versions of Java. See [this article](#) for details. To run IGV from the web launch buttons below, you need the [latest version of Java](#). Another option which avoids Mac security issues is to use the "zip" distribution below. After unzipping double-click the "igv.command" file to launch IGV.

**Java:** IGV 2.3 requires Java 6 or greater. To use the launch buttons below on MacOS Java 7 is required.

**Chrome:** Chrome does not launch 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 you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error "could not create virtual machine".

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

[Nightly Build](#) Latest development build.

[Archived Versions](#)

#### igvtools

Utilities for preprocessing data files.

- [igvtools 2.3.20.zip](#)

#### Download

A downloadable version that does not require launching from the web. For Windows, Mac OS X, and Linux.

- [IGV 2.3.20.zip](#)

#### Source Code

Source distribution archive:

- [v2.3.20.zip](#)

Source code repository is hosted at github:

- <https://github.com/broadinstitute/IGV/>



# Launch IGV



Downloads | Integrative Genomics Viewer

www.broadinstitute.org/igv/download

Home > Downloads

### Downloads





#### Integrative Genomics Viewer (Version 2.3)

**Mac Users:** Apple has pushed out an update that blocks all but the latest versions of Java. See [this article](#) for details. To run IGV from the web launch buttons below, you need the [latest version of Java](#). Another option which avoids Mac security issues is to use the "zip" distribution below. After unzipping double-click the "igv.command" file to launch IGV.

**Java:** IGV 2.3 requires Java 6 or greater. To use the launch buttons below on MacOS Java 7 is required.

**Chrome:** Chrome does not launch 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 you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error "could not create virtual machine".

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

[Nightly Build](#) Latest development build.

[Archived Versions](#)

#### igvtools

Utilities for preprocessing data files.

- [igvtools 2.3.20.zip](#)

#### Download

A downloadable version that does not require launching from the web. For Windows, Mac OS X, and Linux.

- [IGV 2.3.20.zip](#)

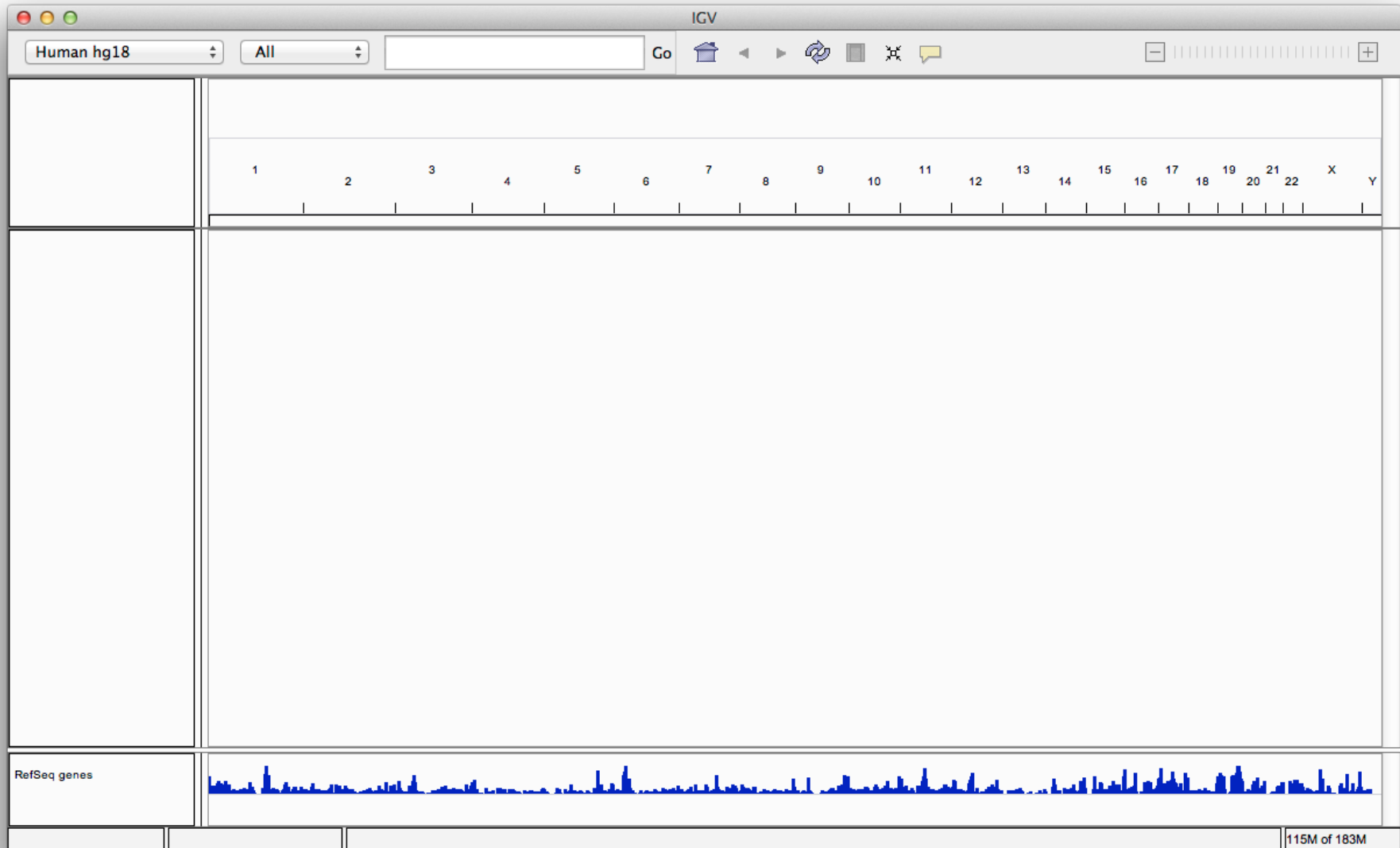
#### Source Code

Source distribution archive:

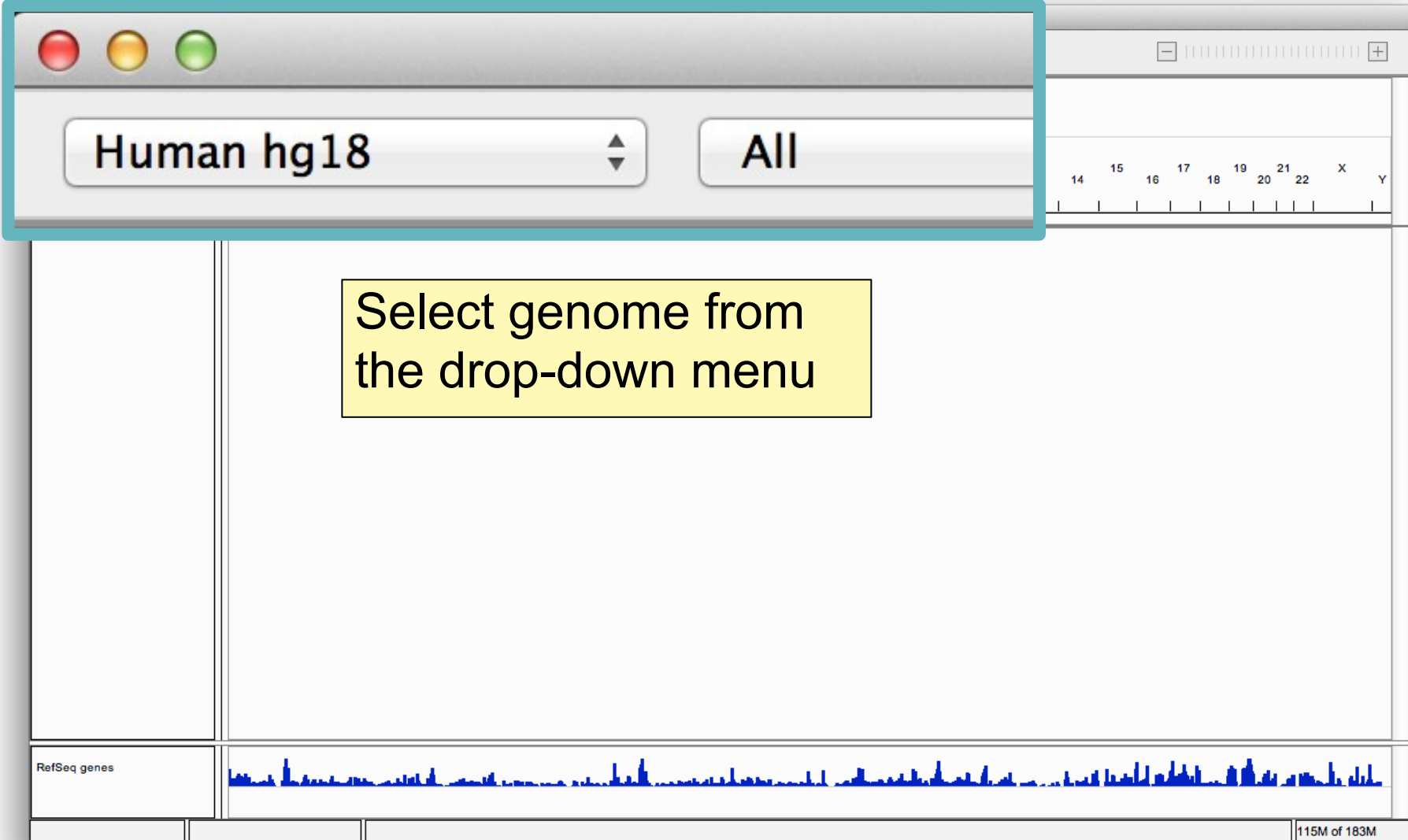
- [v2.3.20.zip](#)

Source code repository is hosted at github:

- <https://github.com/broadinstitute/IGV/>



# Select the reference genome



Human hg18 All

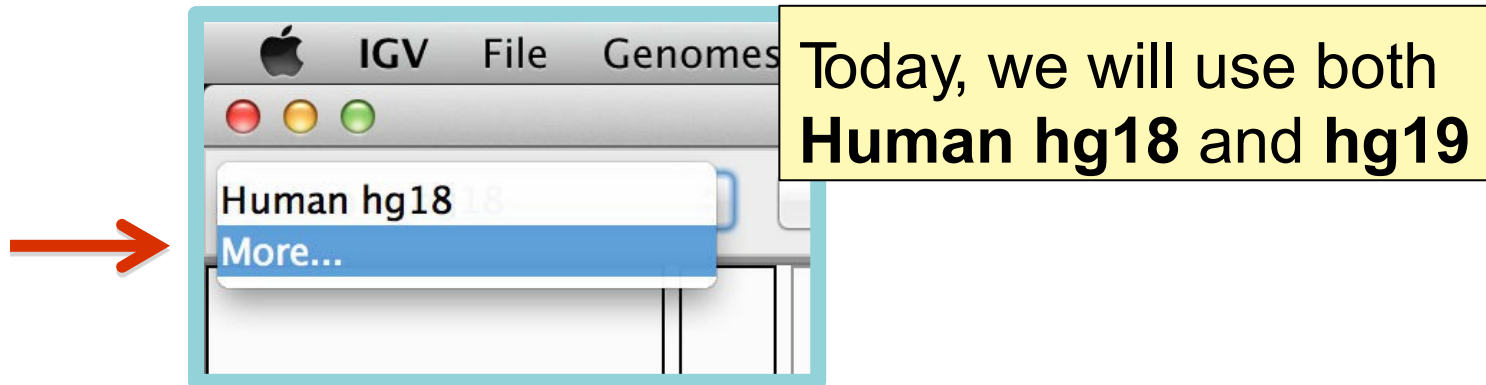
Select genome from the drop-down menu

14 15 16 17 18 19 20 21 22 X Y

RefSeq genes

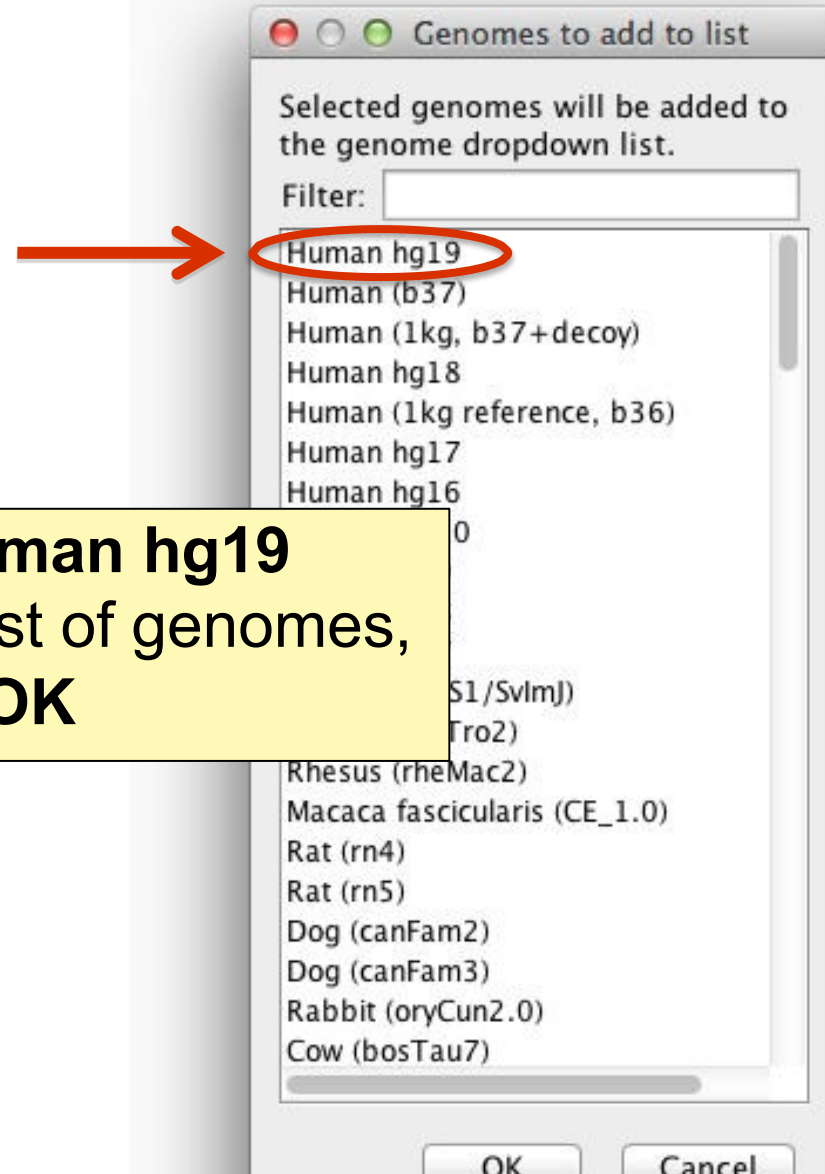
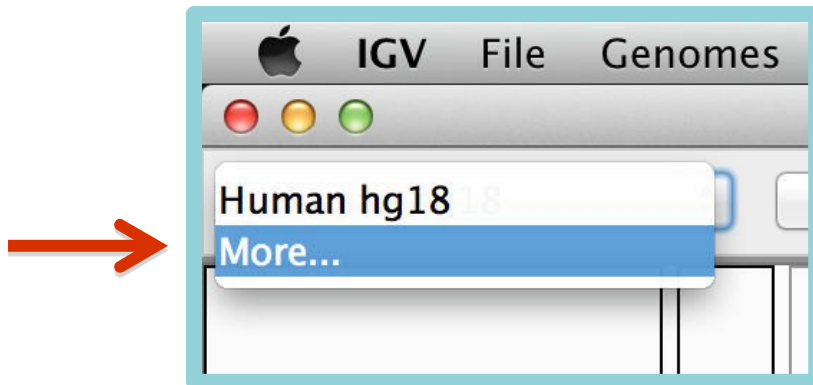
115M of 183M

# Select the reference genome



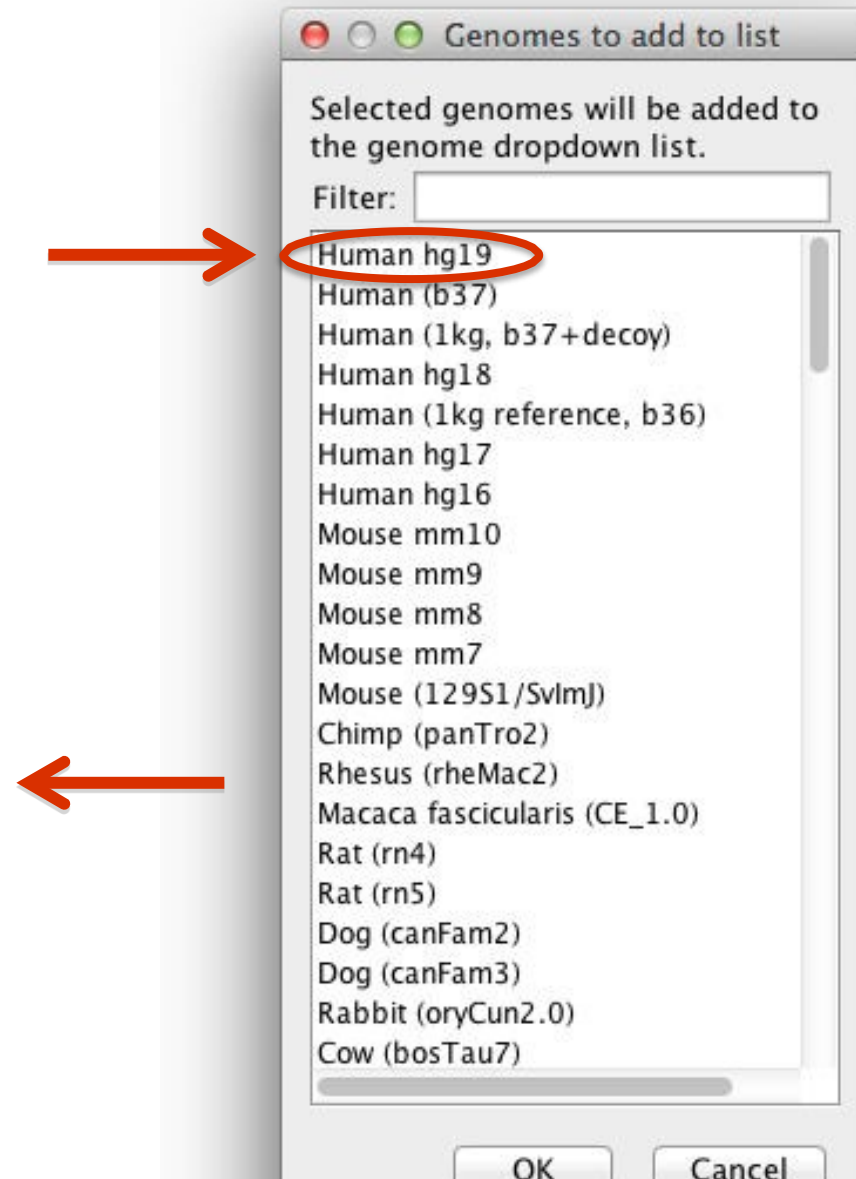
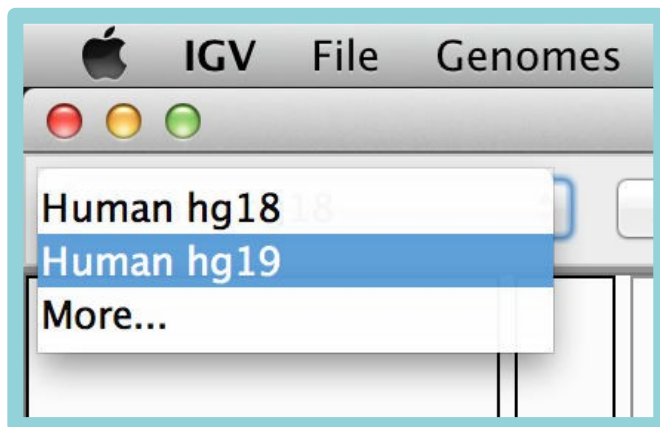
If **Human hg19** is not in the menu,  
then click on ***More...***

# Select the reference genome



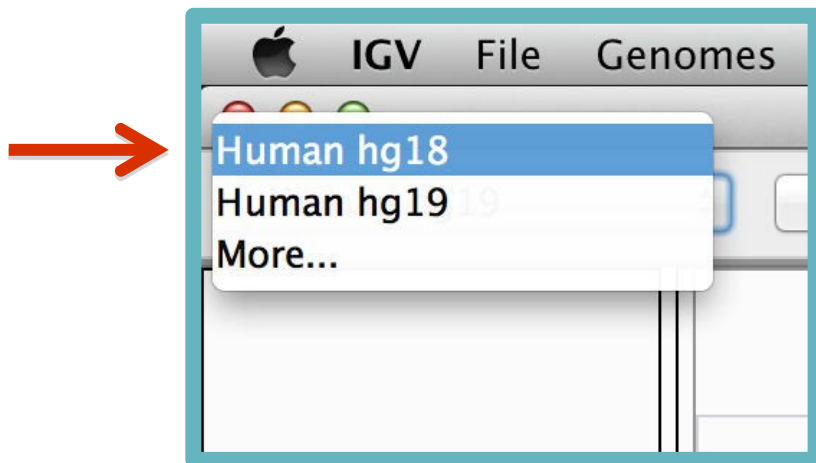
Select **Human hg19**  
from the list of genomes,  
and click **OK**

# Select the reference genome



# Select the reference genome

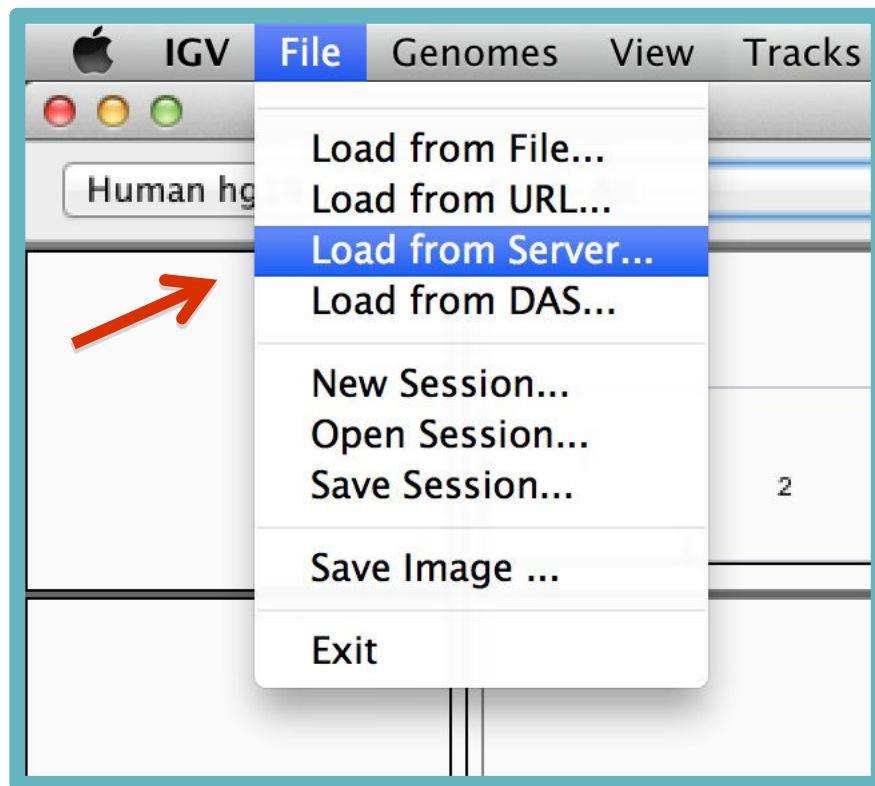
Select **Human hg18**





# Load data

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

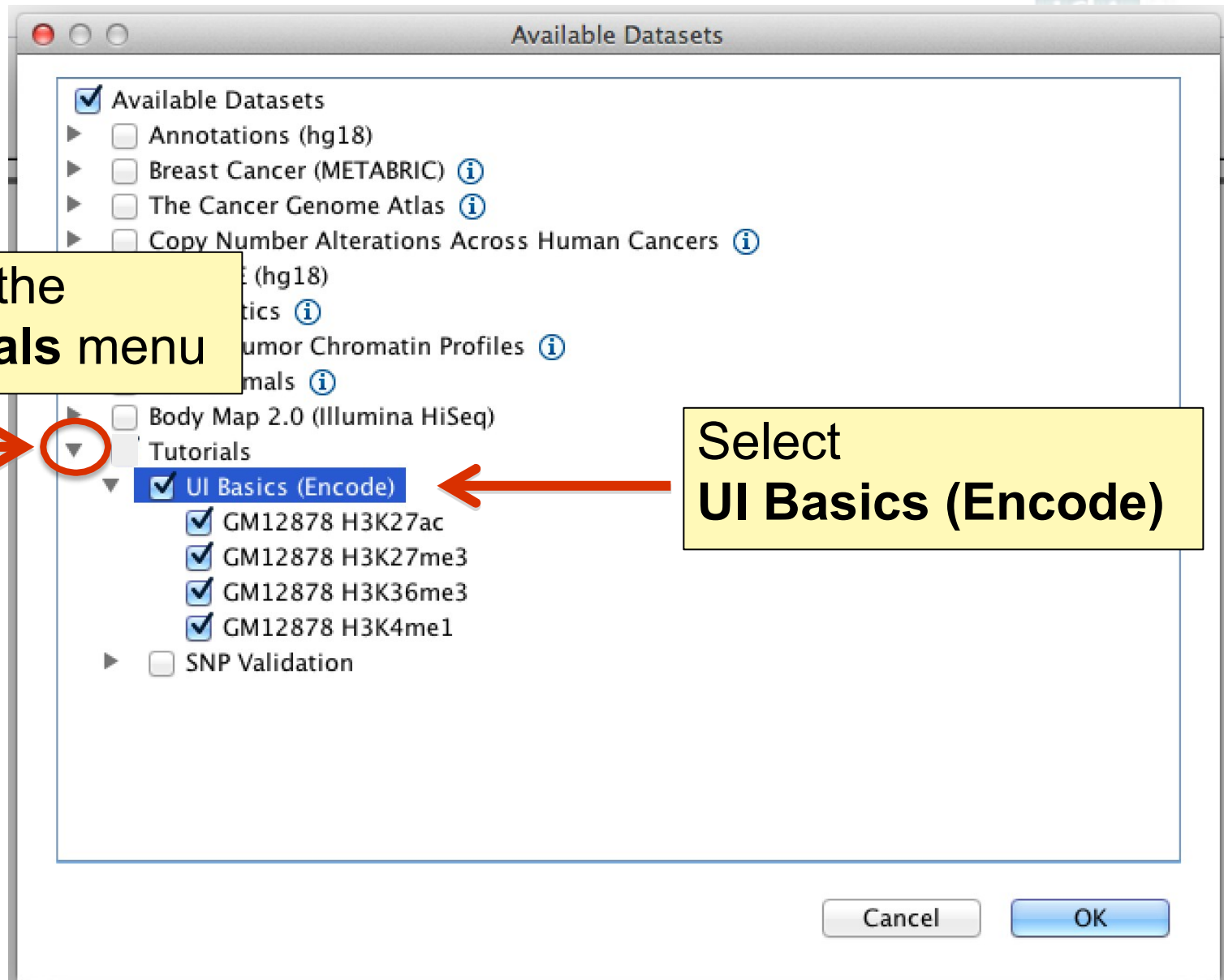


# Load data

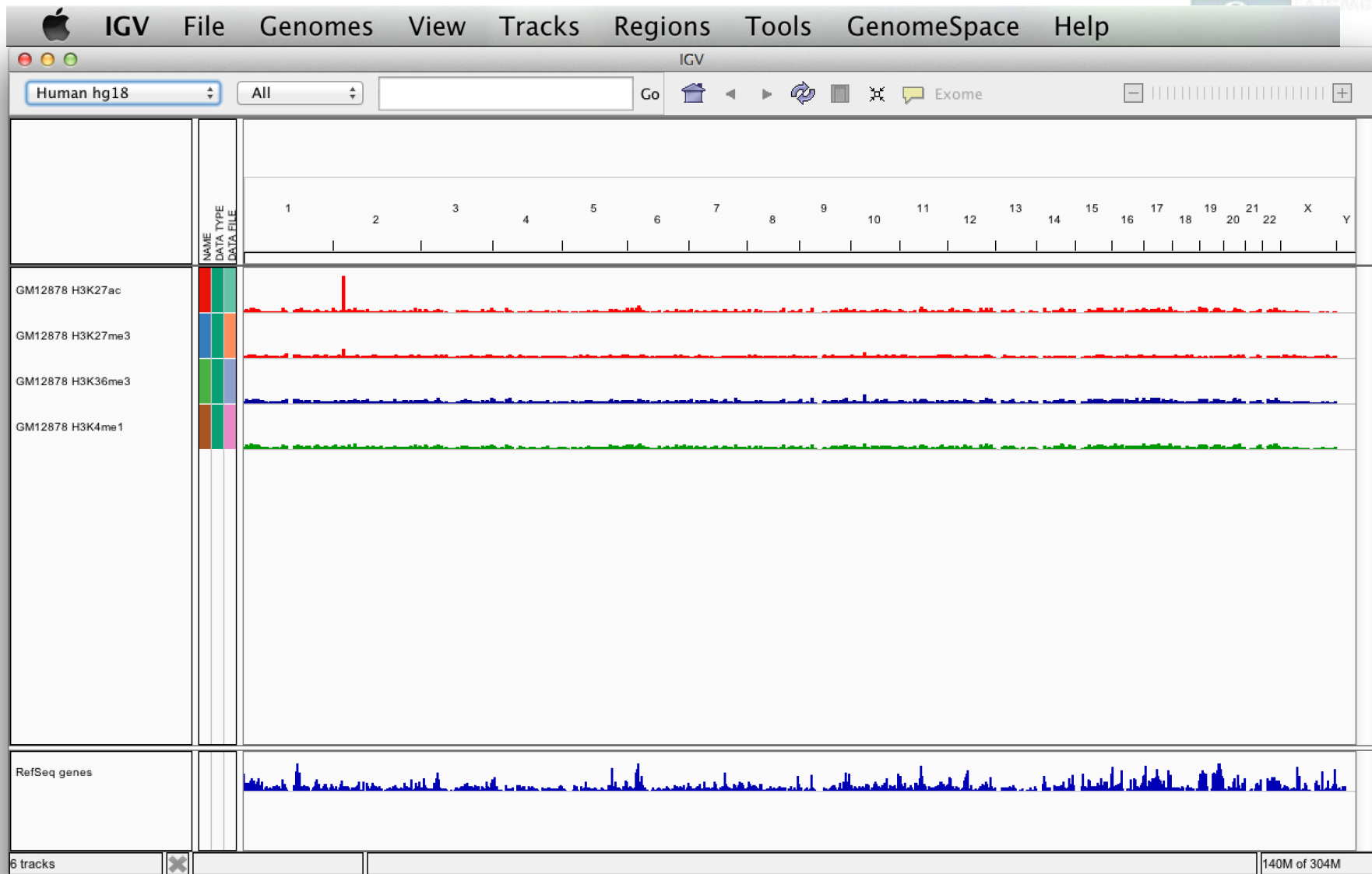
Open the  
**Tutorials** menu



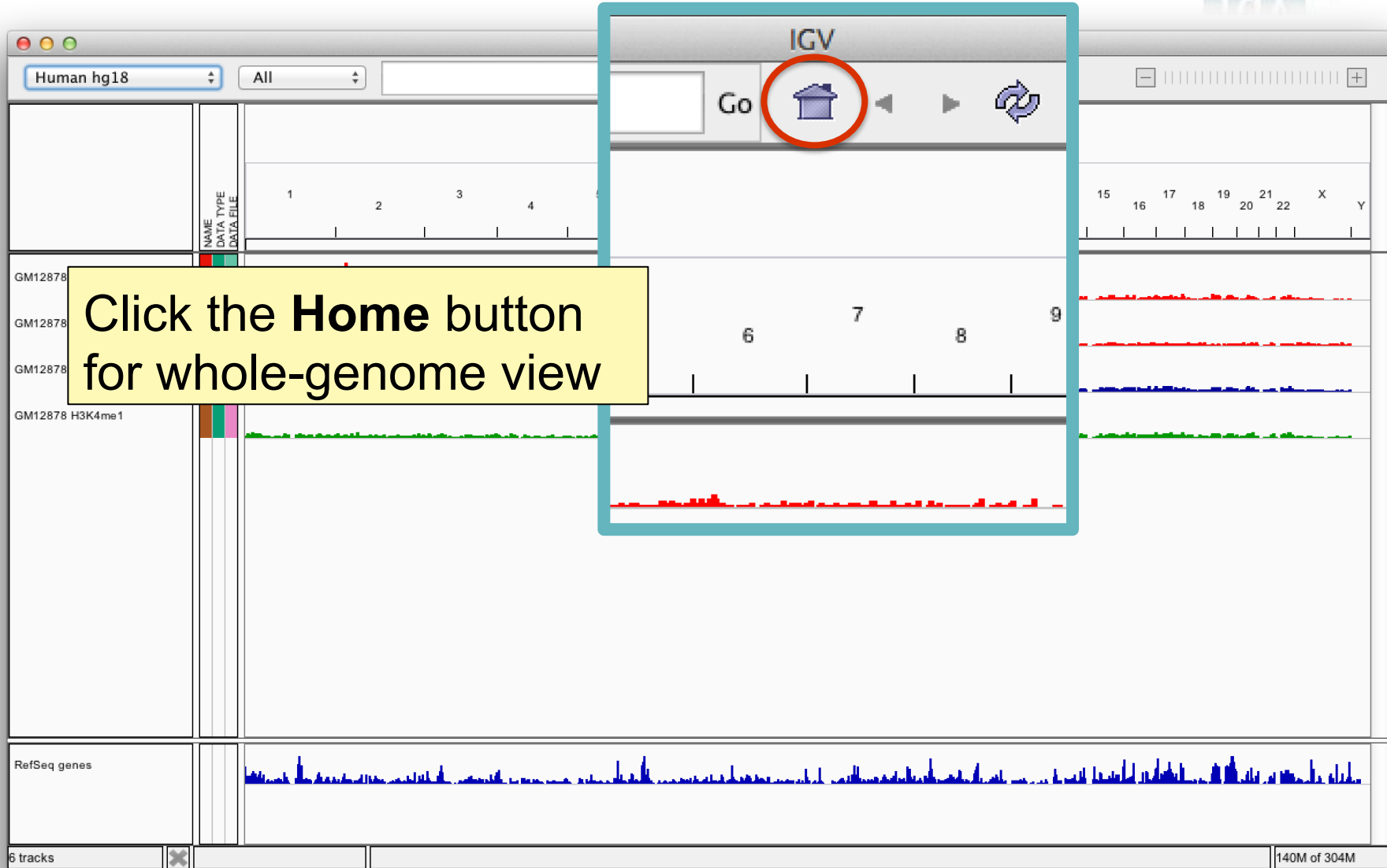
Select  
**UI Basics (Encode)**



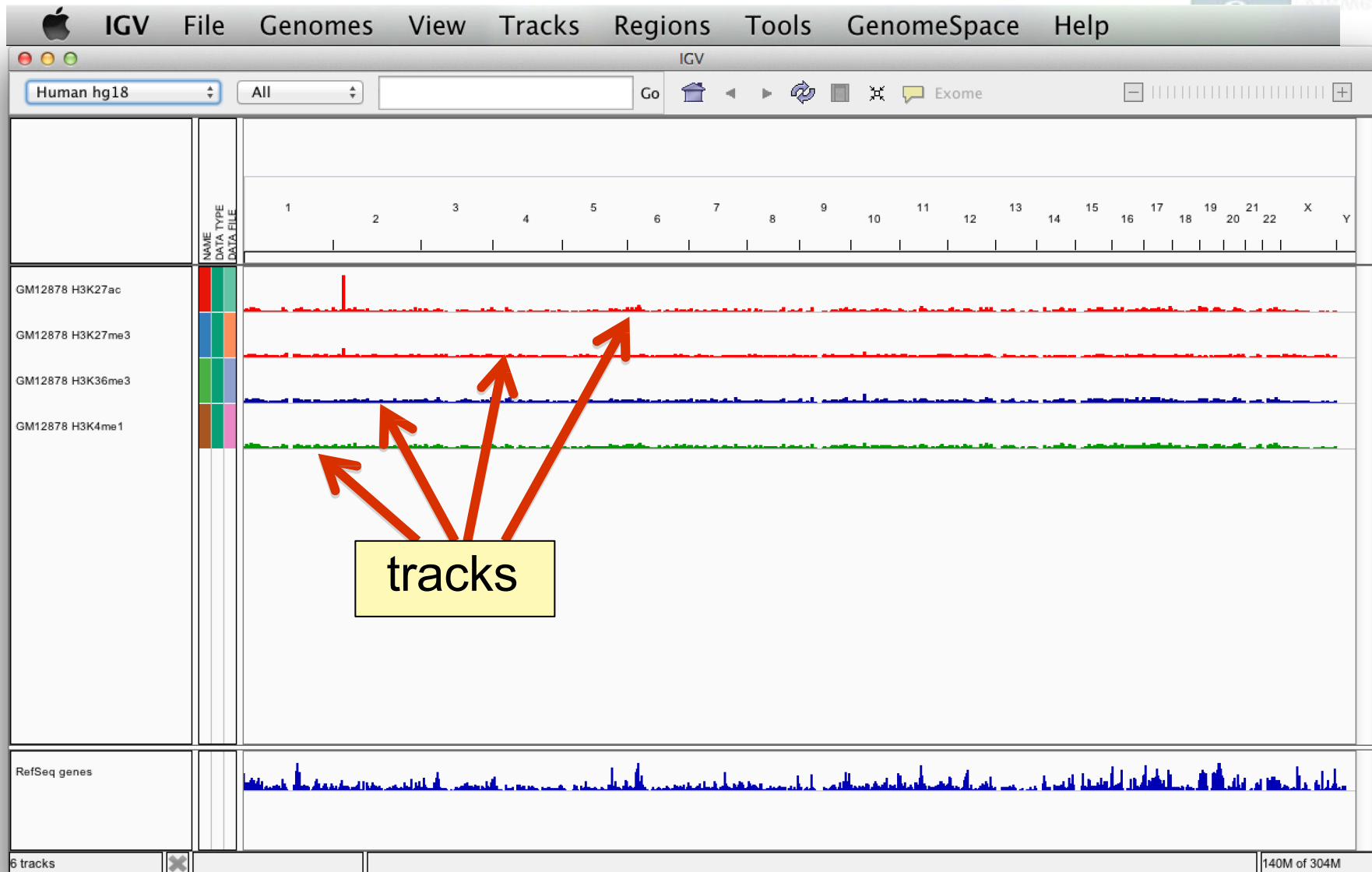
# Screen layout



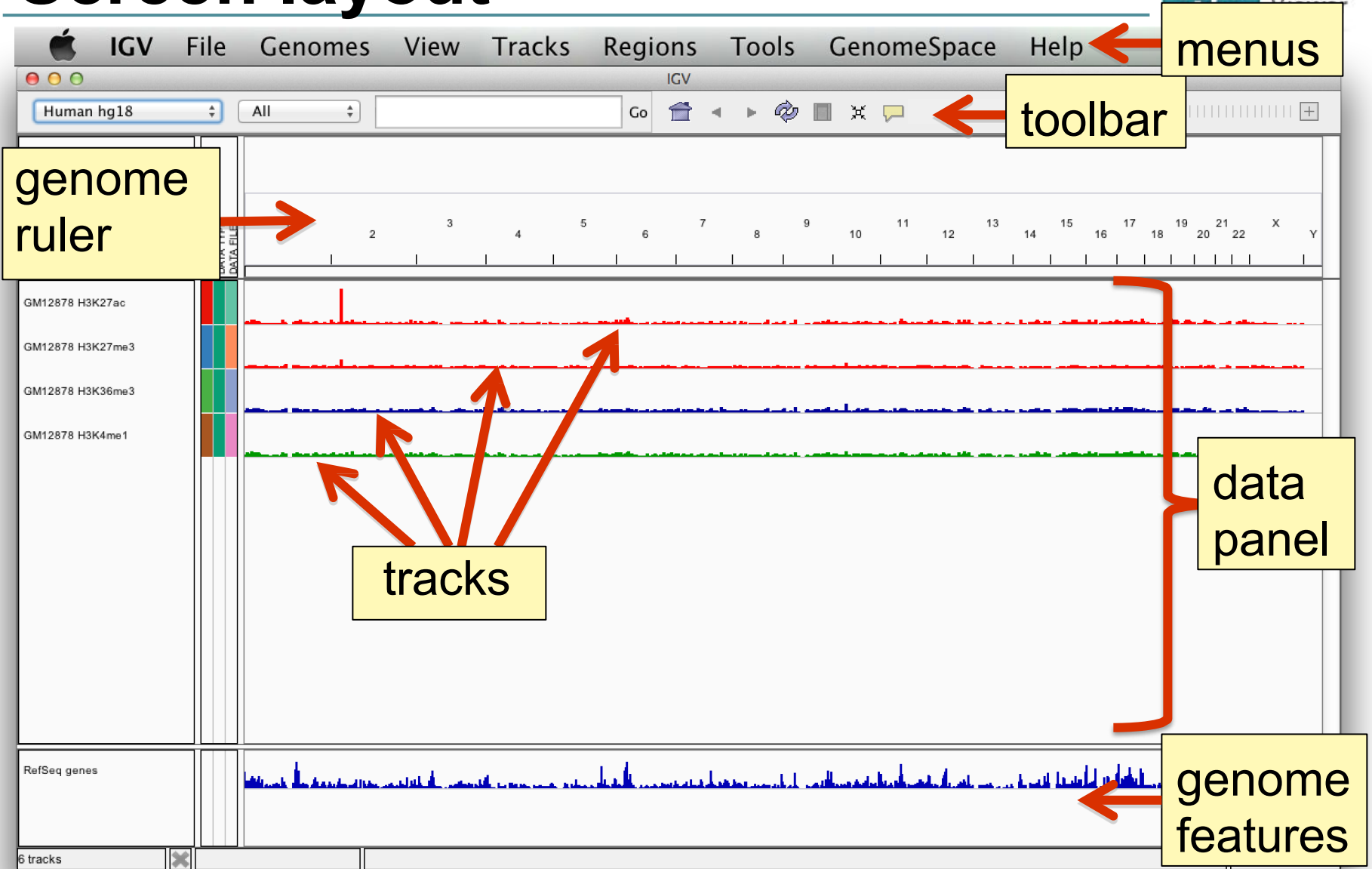
# Screen layout



# Screen layout



# Screen layout



# File formats and track types

---



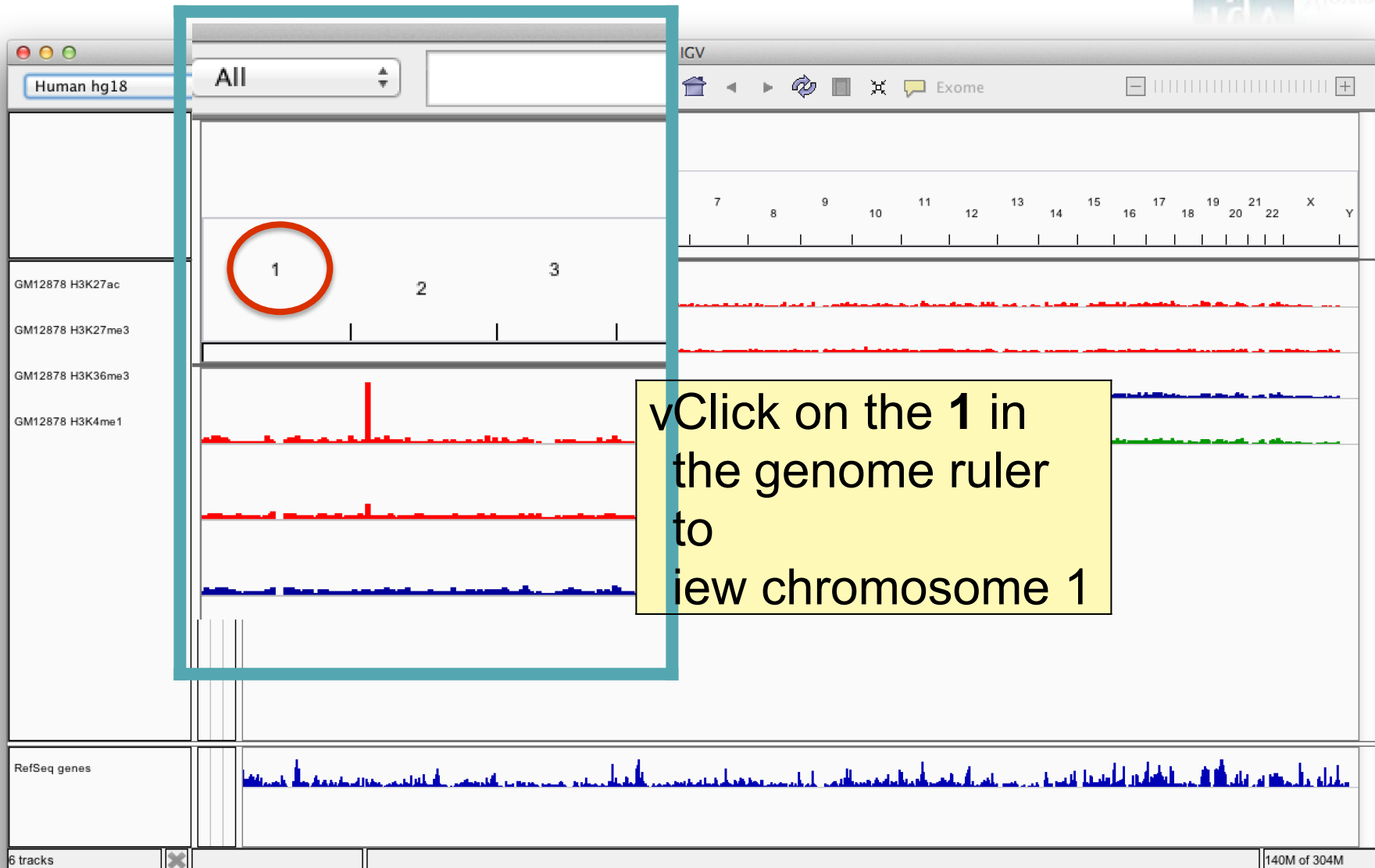
- The **file format** defines the track type.
- The **track type** determines the display options

# File formats and track types

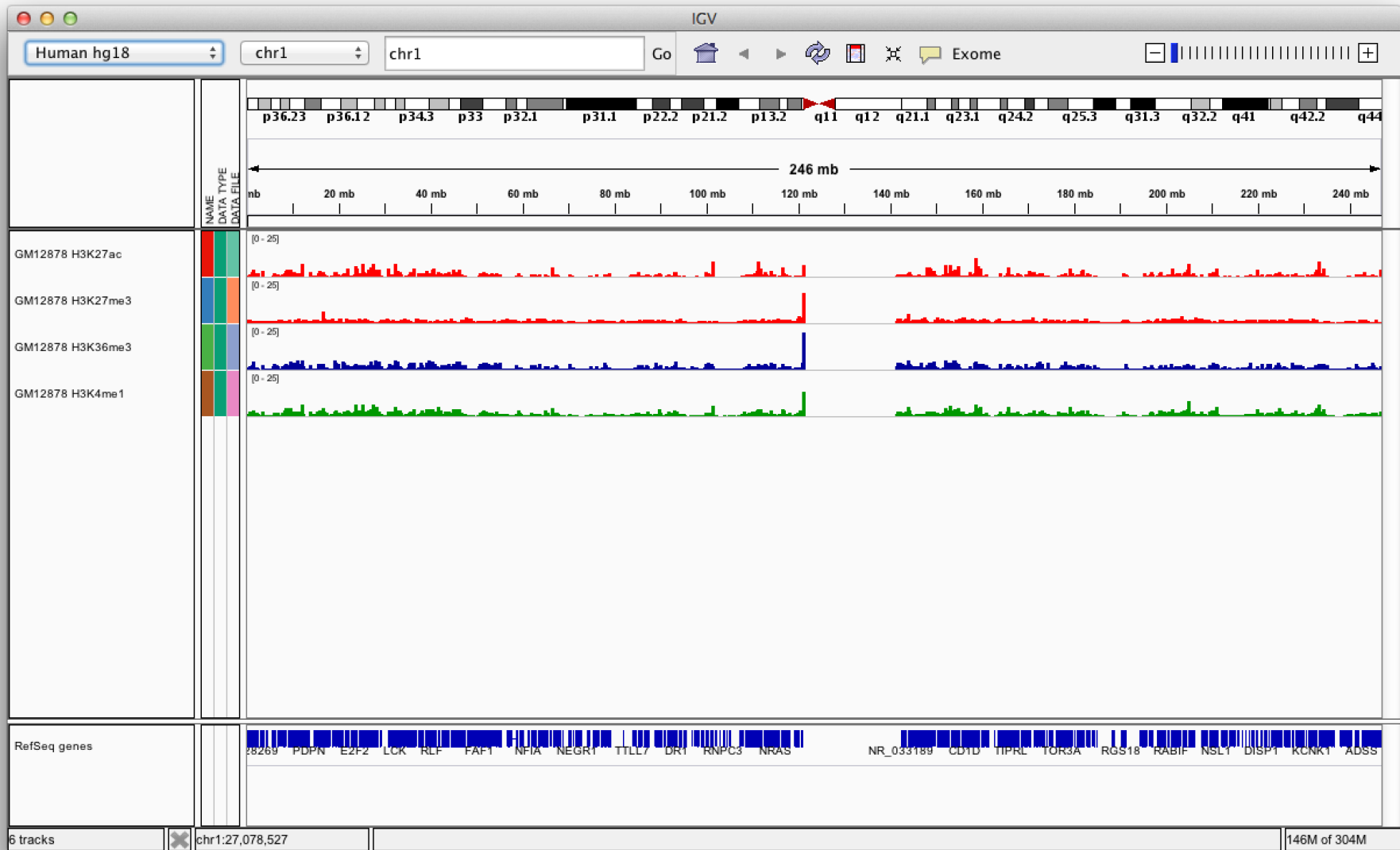
- The **file format** defines the track type.
- The **track type** determines the display options
- IGV supports many different file formats.
  - [BAM](#)
  - [BED](#)
  - [BedGraph](#)
  - [bigBed](#)
  - [bigWig](#)
  - [Birdsuite Files](#)
  - [broadPeak](#)
  - [CBS](#)
  - [CN](#)
  - [Cufflinks 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)



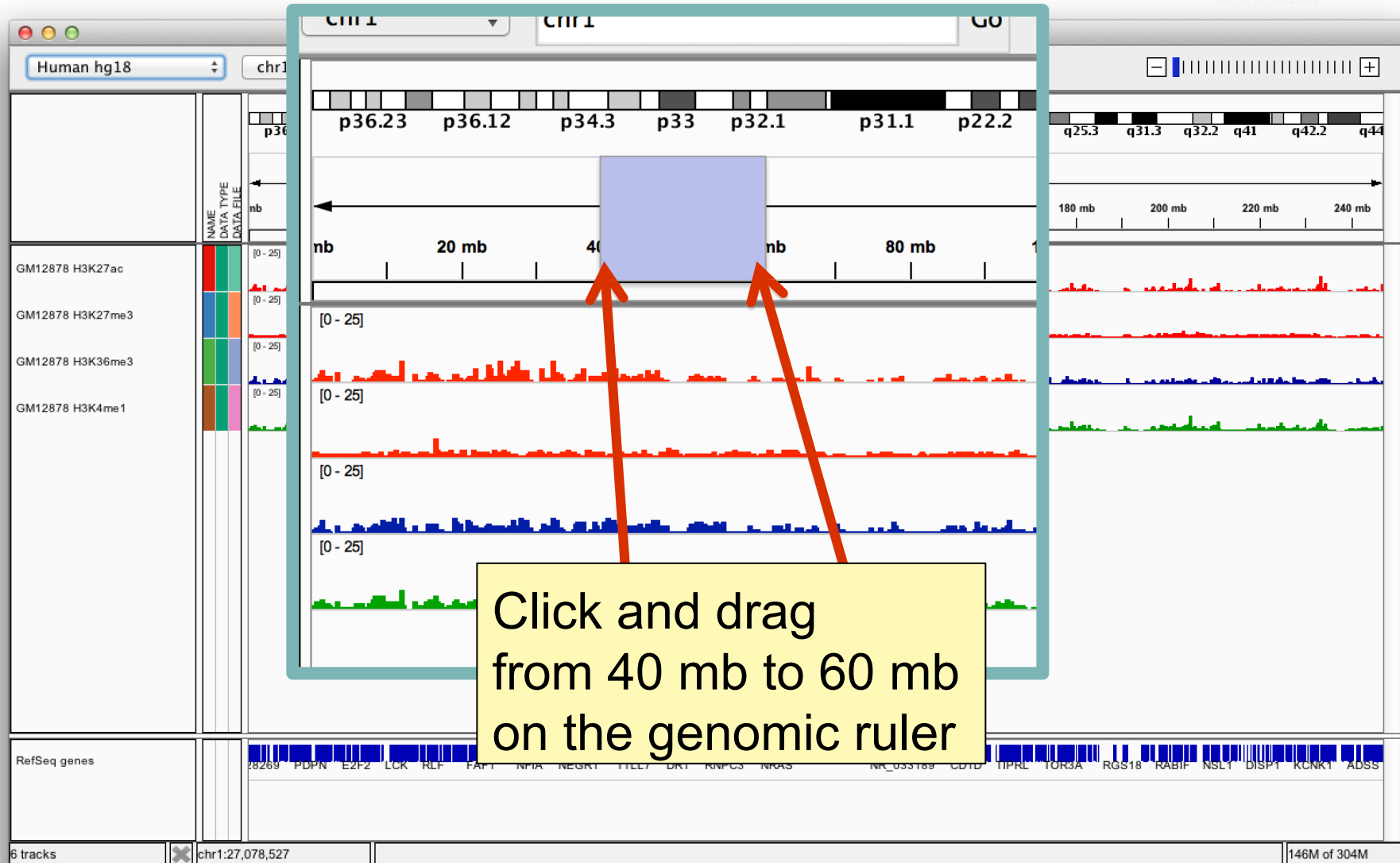
# Navigate



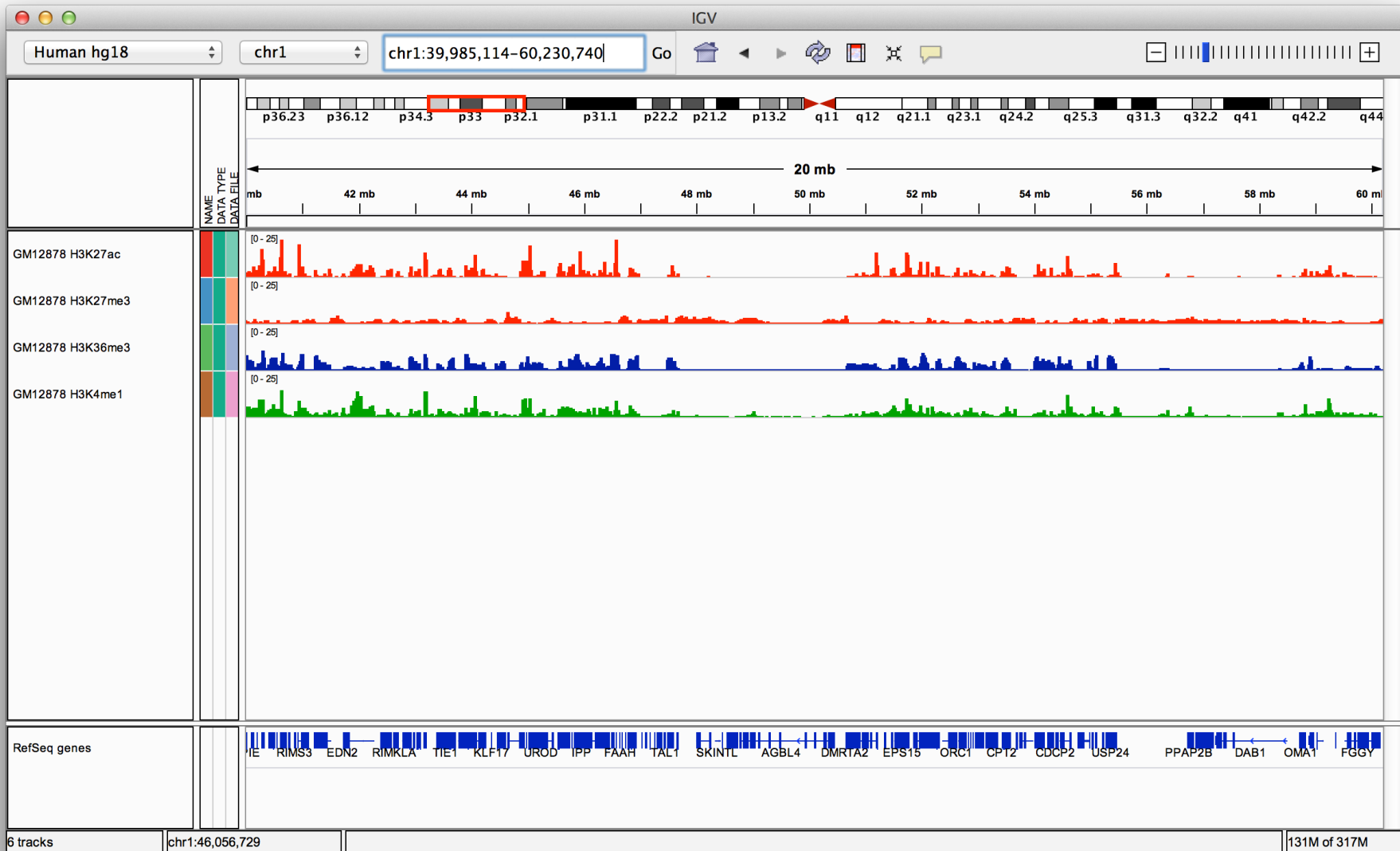
# Navigate



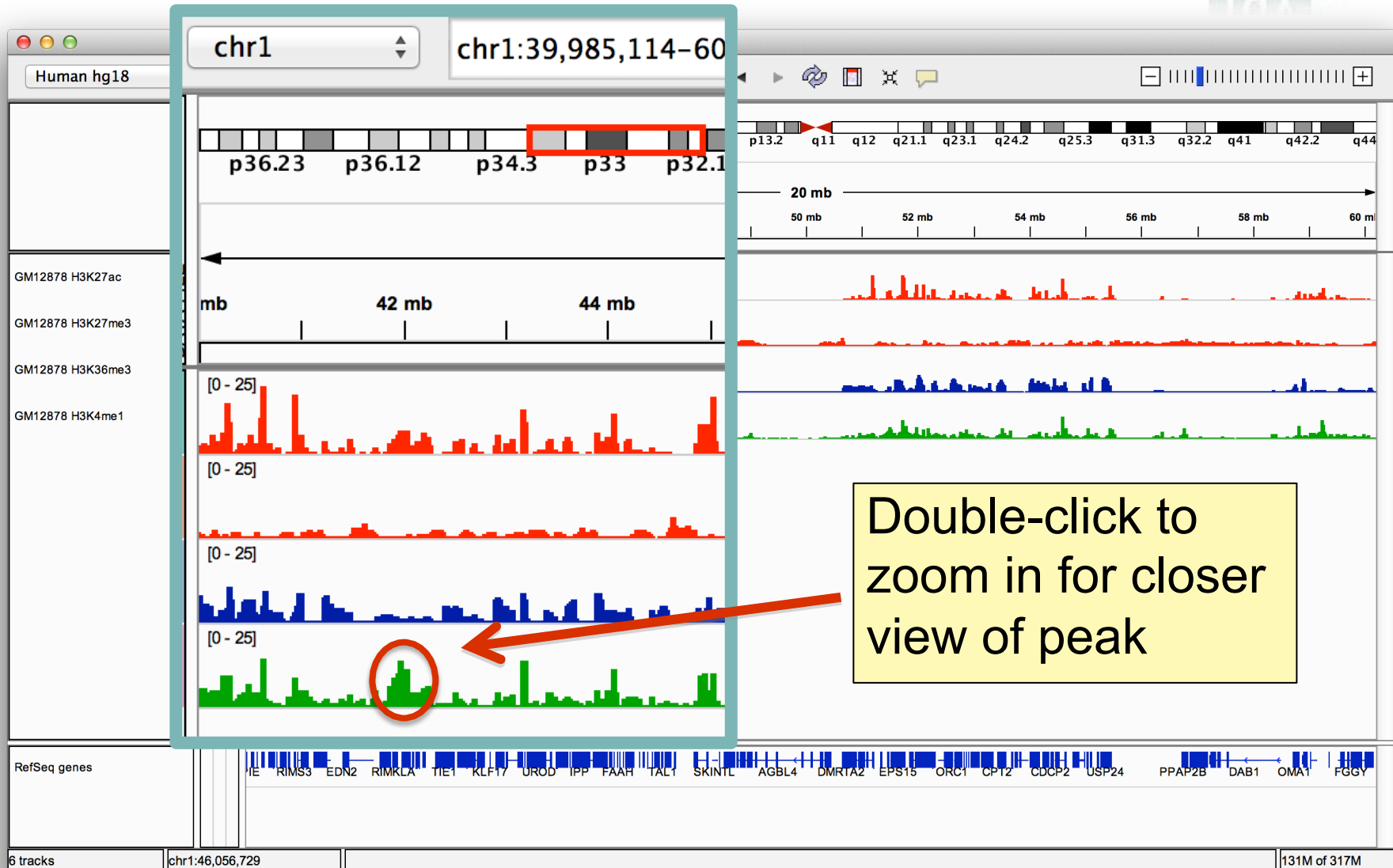
# Navigate



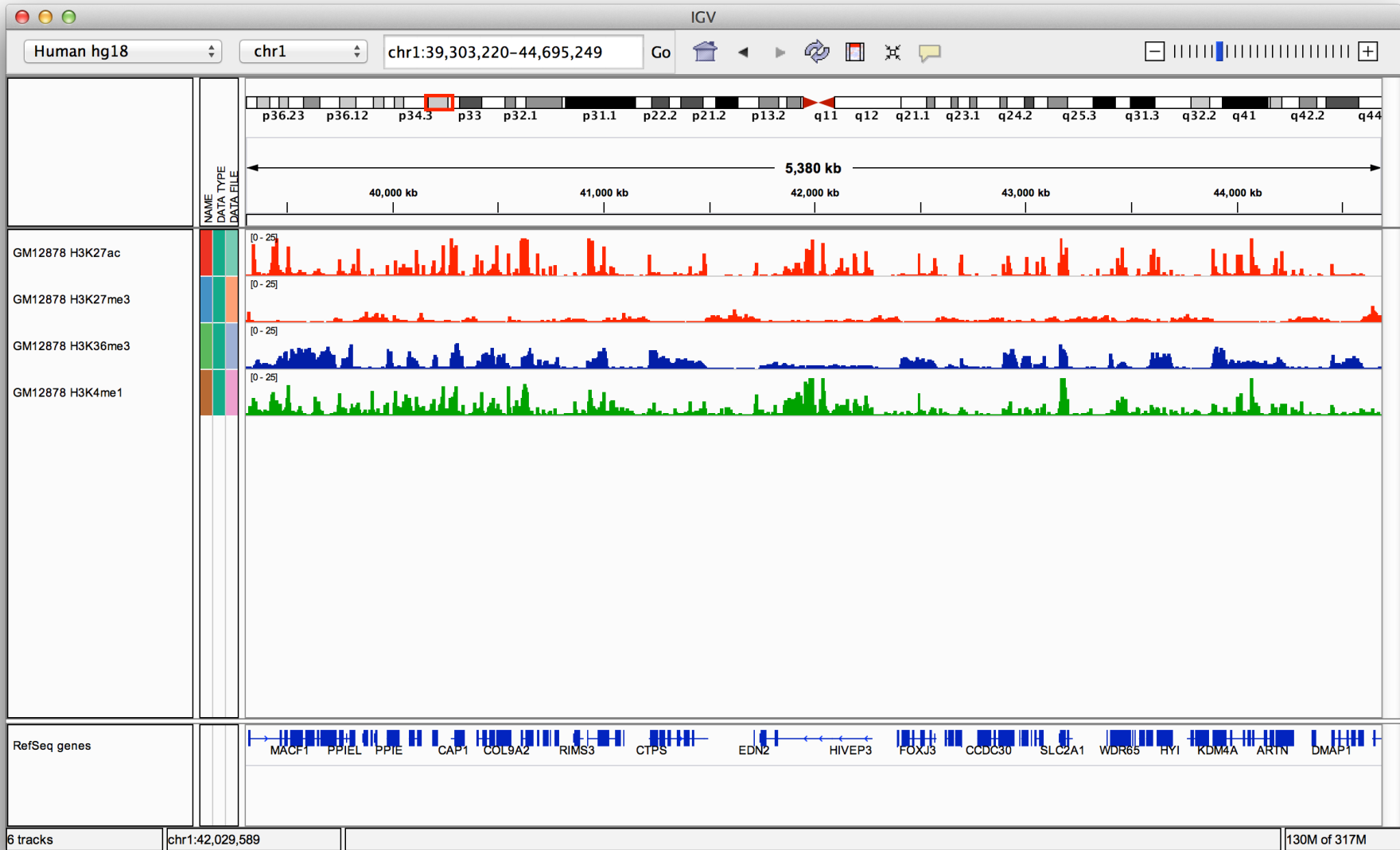
# Navigate



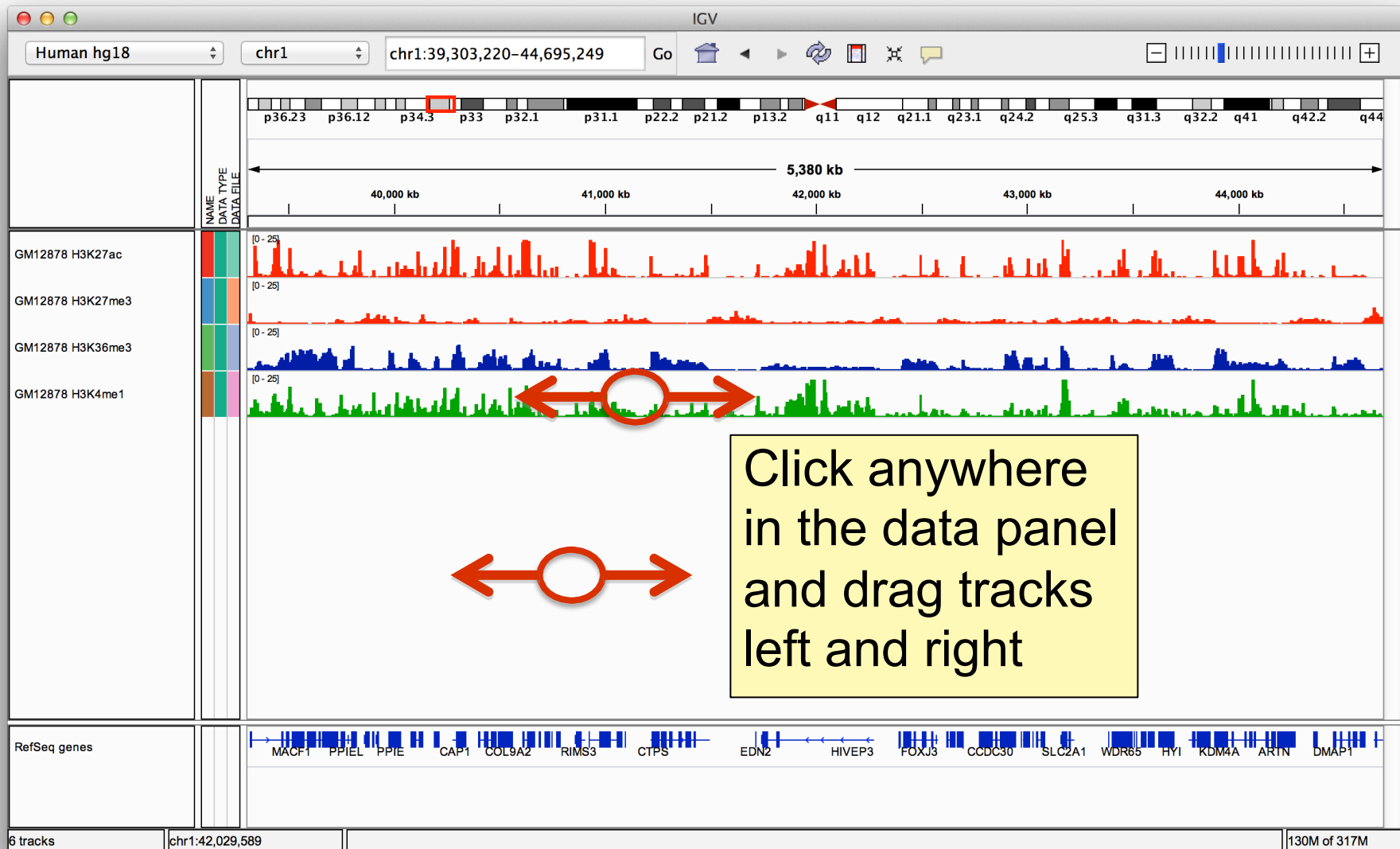
# Navigate



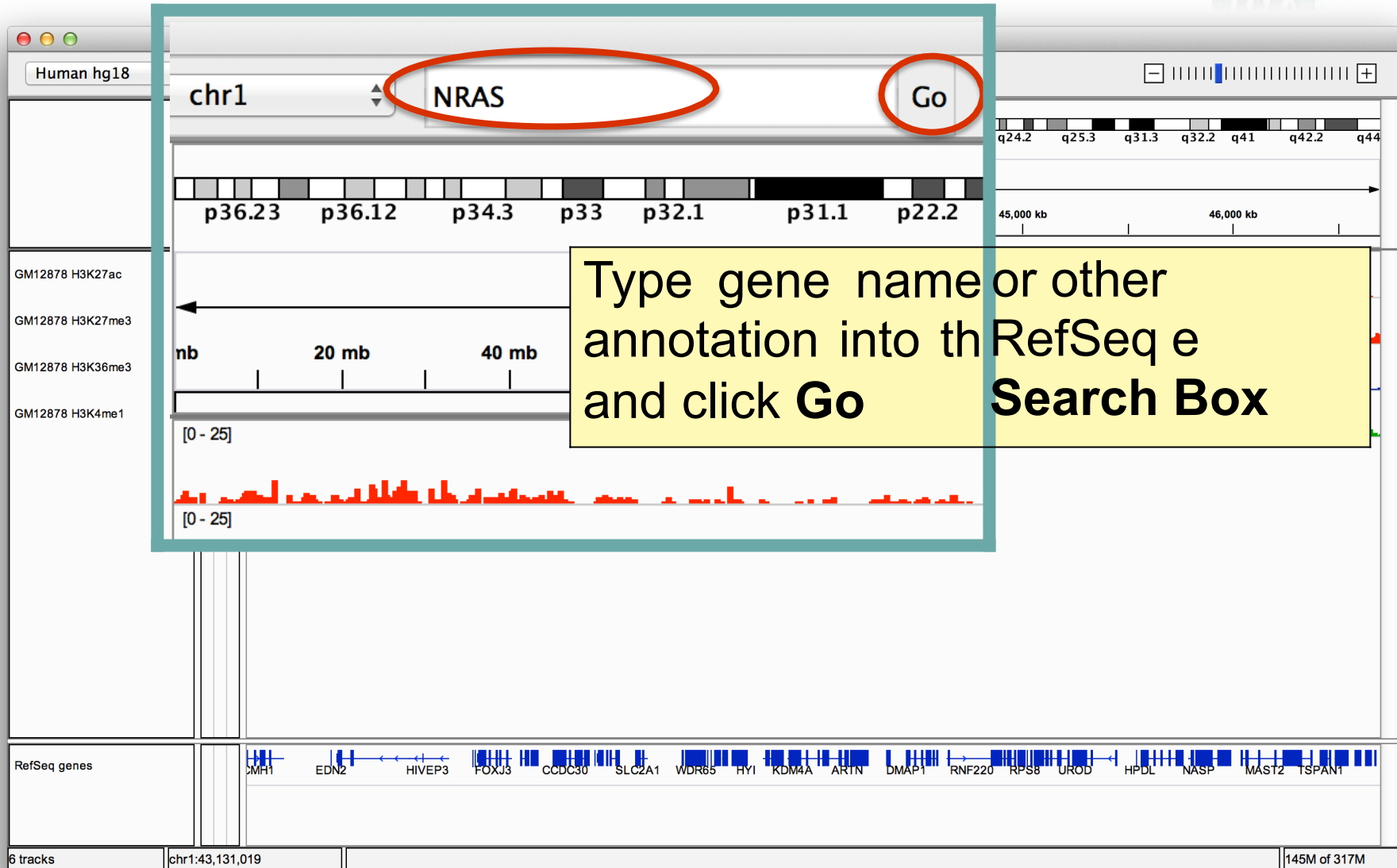
# Navigate



# Navigate



# Navigate



Human hg18

chr1

q24.2 q25.3 q31.3 q32.2 q41 q42.2 q44

p36.23 p36.12 p34.3 p33 p32.1 p31.1 p22.2

45,000 kb 46,000 kb

GM12878 H3K27ac

GM12878 H3K27me3

GM12878 H3K36me3

GM12878 H3K4me1

[0 - 25]

[0 - 25]

nb 20 mb 40 mb

Type gene name or other annotation into the RefSeq search box and click **Go**

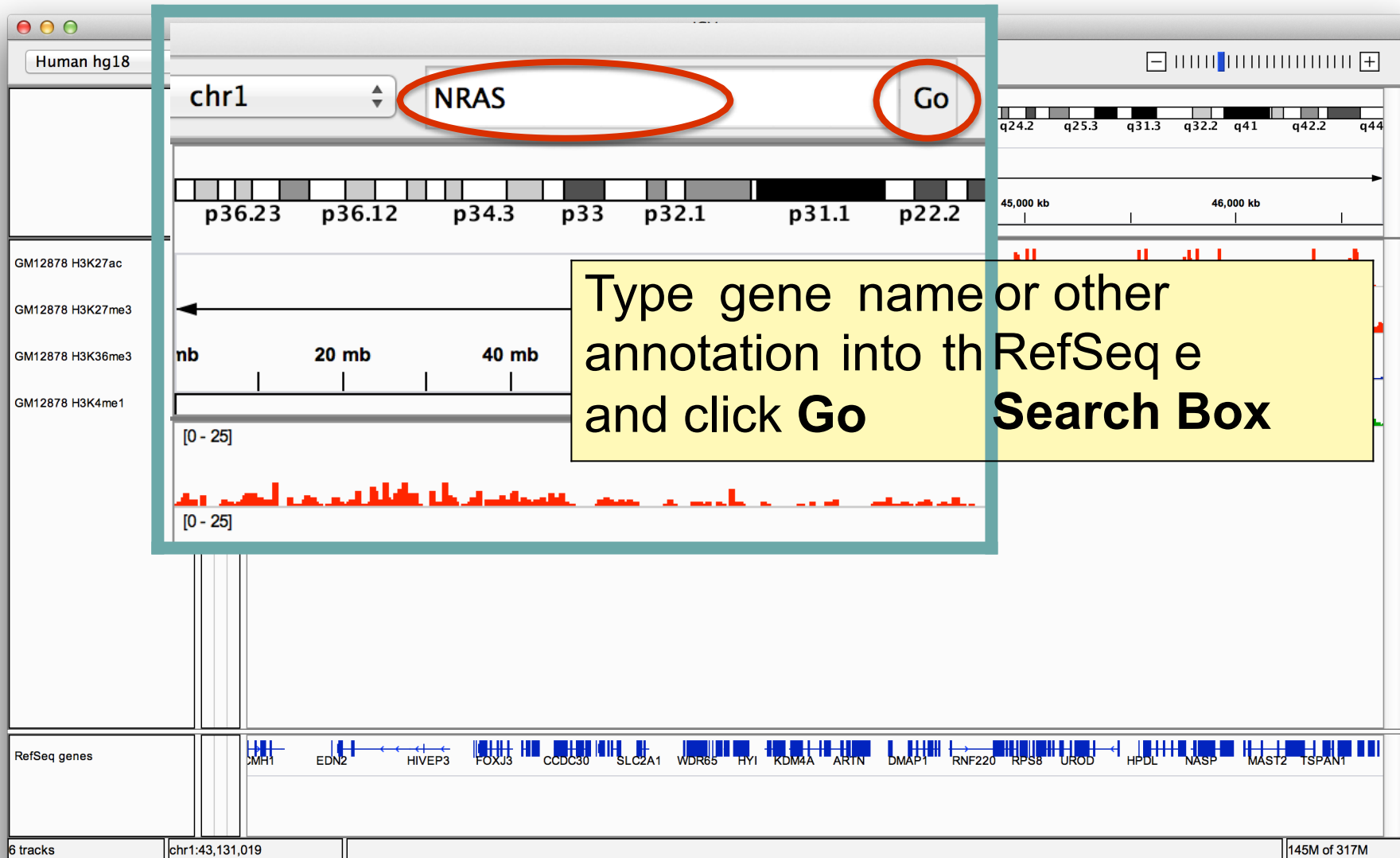
RefSeq genes

DMH1 EDN2 HIVEP3 FOXJ3 CCDC30 SLC2A1 WDR65 HY1 KDM4A ARTN DMAP1 RNF220 RPS8 UROD HPDL NASP MAST2 TSPAN1

6 tracks chr1:43,131,019 145M of 317M



# Navigate



Human hg18

chr1

q24.2 q25.3 q31.3 q32.2 q41 q42.2 q44

p36.23 p36.12 p34.3 p33 p32.1 p31.1 p22.2

45,000 kb 46,000 kb

GM12878 H3K27ac

GM12878 H3K27me3

GM12878 H3K36me3

GM12878 H3K4me1

nb 20 mb 40 mb

[0 - 25]

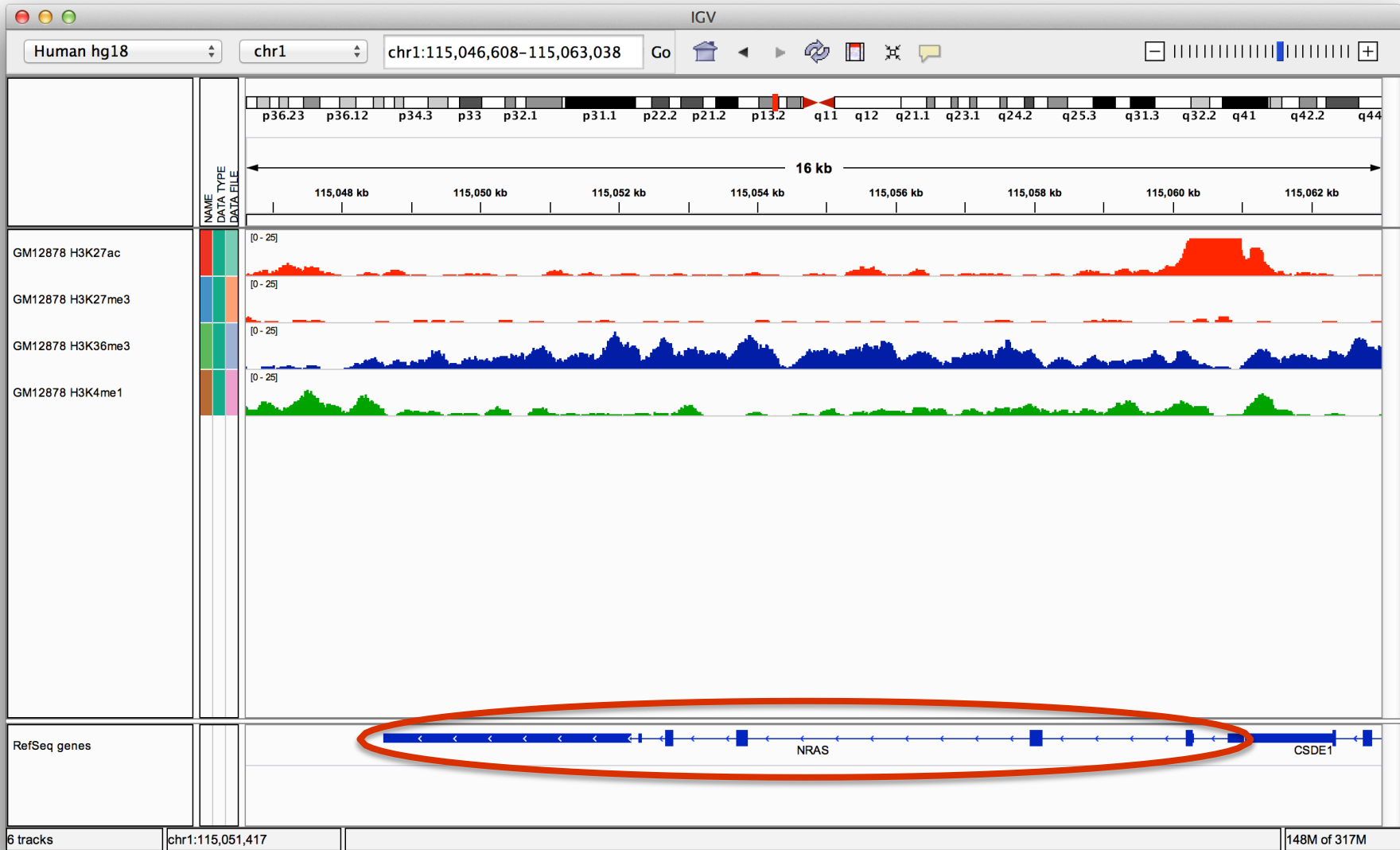
[0 - 25]

RefSeq genes

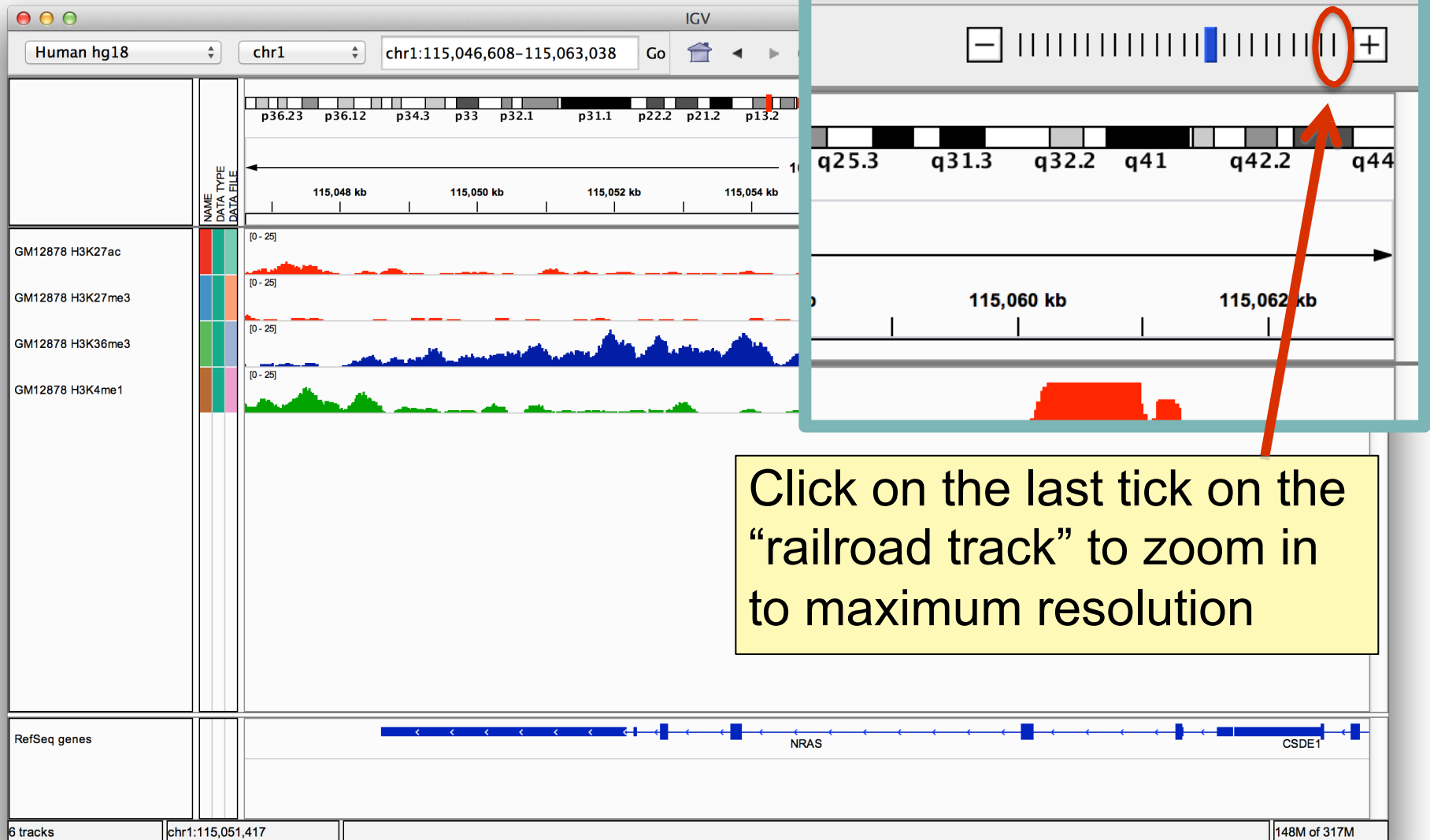
DMH1 EDN2 HIVEP3 FOXJ3 CCDC30 SLC2A1 WDR65 HY1 KDM4A ARTN DMAP1 RNF220 RPS8 UROD HPDL NASP MAST2 TSPAN1

6 tracks chr1:43,131,019 145M of 317M

# Navigate

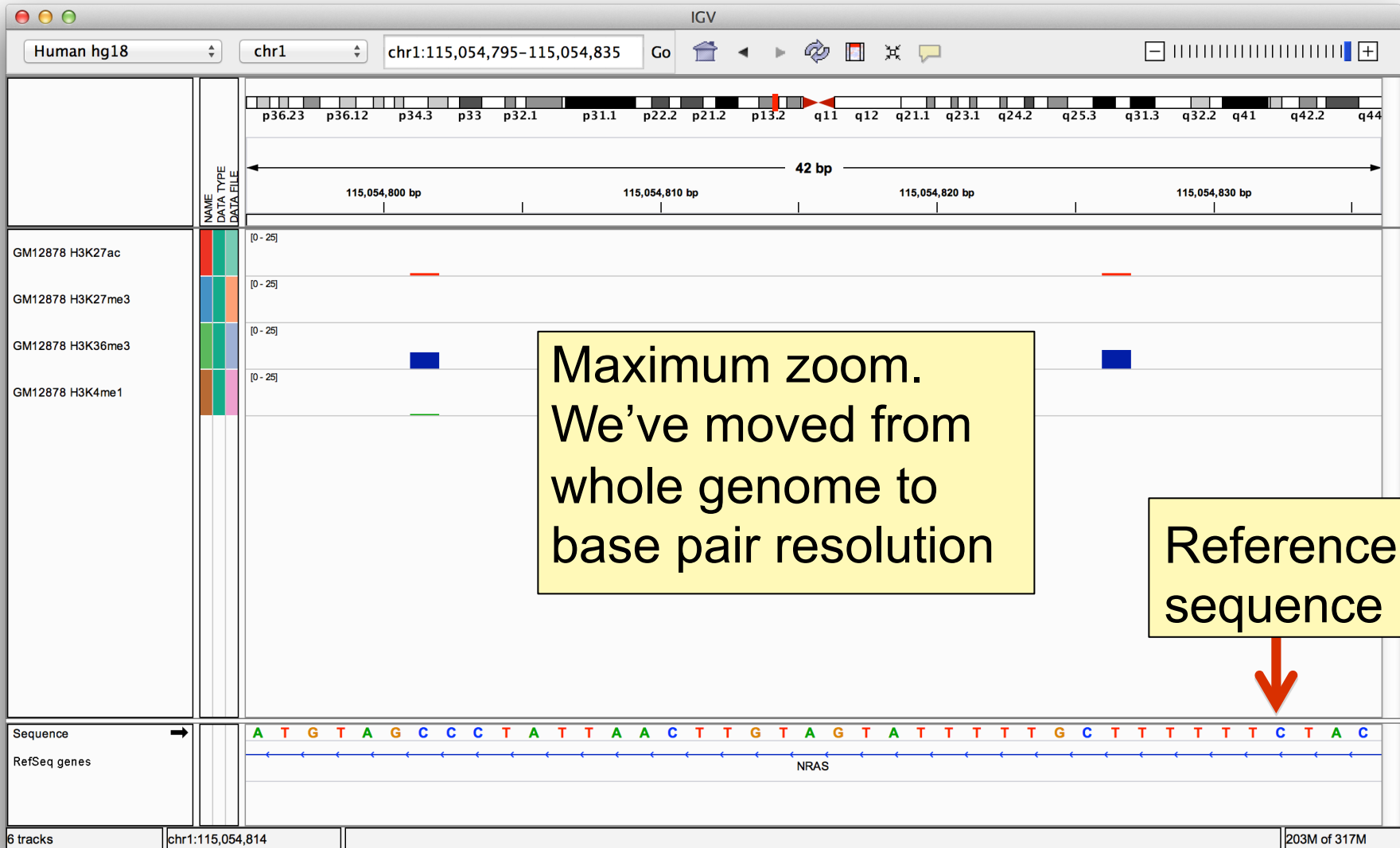


# Navigate



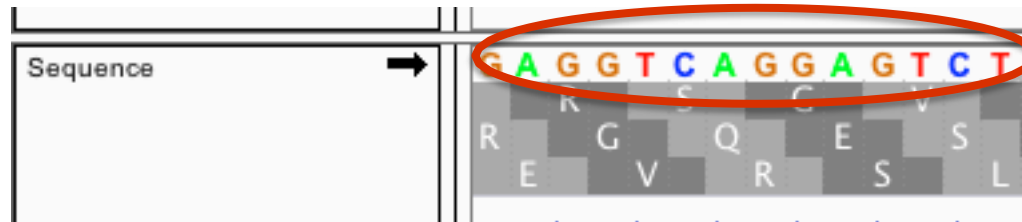
Click on the last tick on the “railroad track” to zoom in to maximum resolution

# Navigate



# Reference sequence

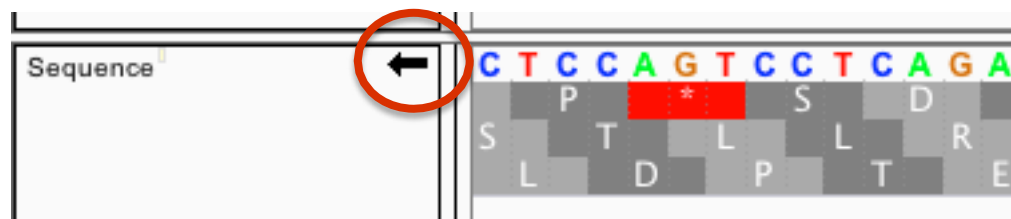
Click anywhere on the sequence to see a 3 frame translation.



By default the sequence for the forward strand is shown.



Click the arrow on the left to reverse the strand.



# Genome annotation track



## UCSC style gene representation

5' UTR

Intron

Exons

3' UTR

Zoomed in views

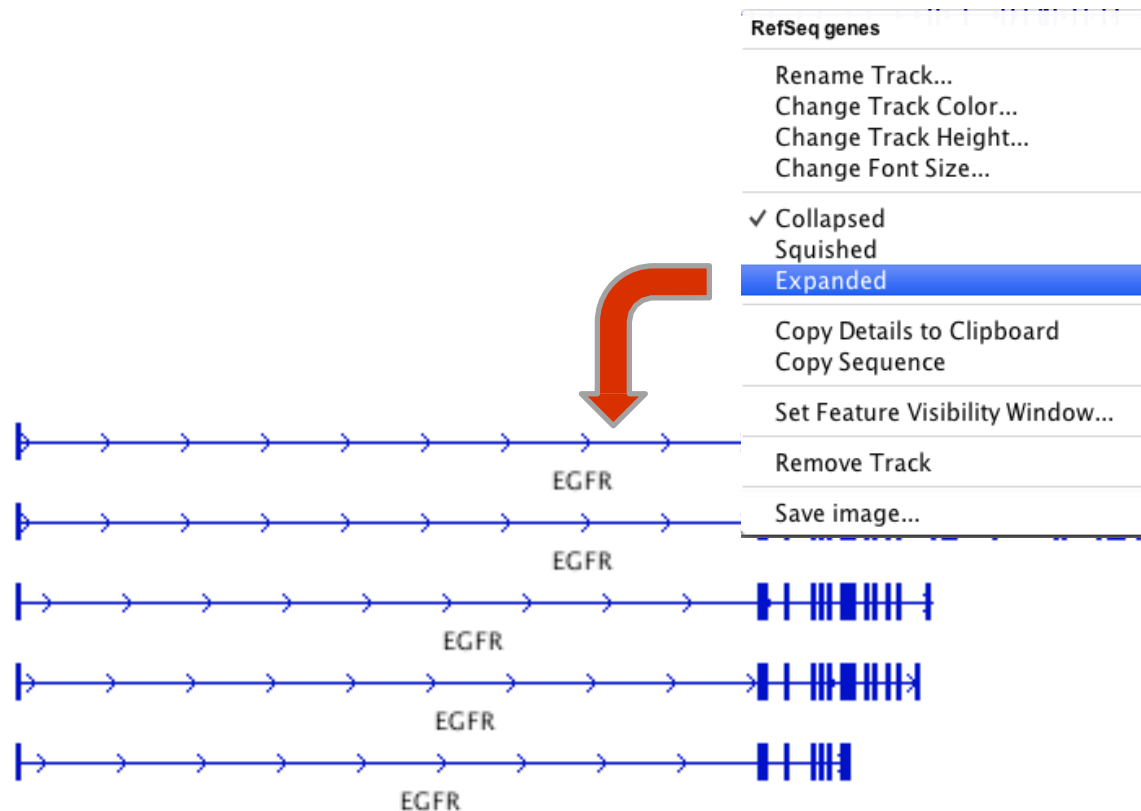
Zoomed out views

# Annotation display mode

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

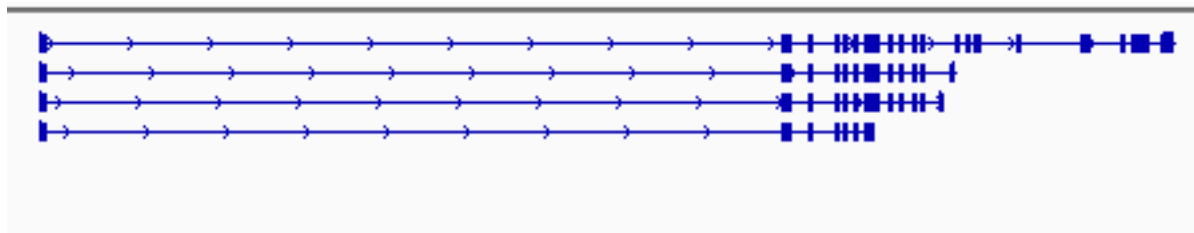
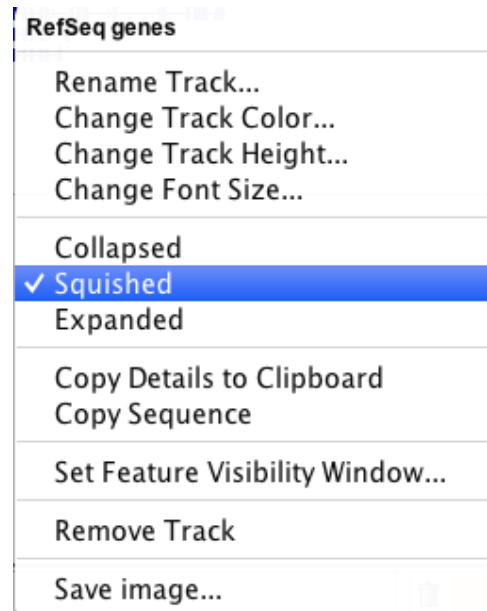


2. Expand the track using the popup menu



# Annotation display mode

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



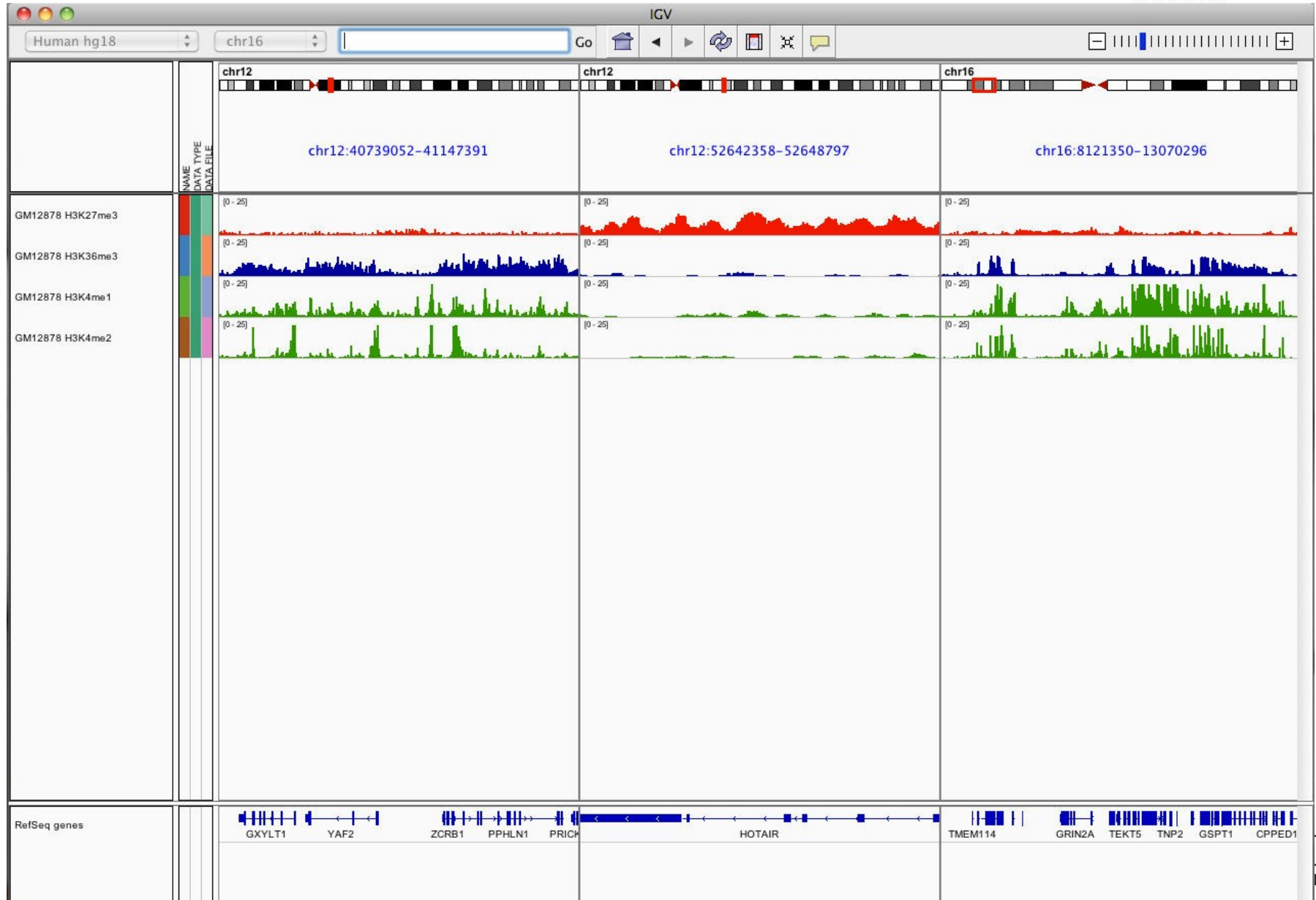


# Viewing multiple regions

---



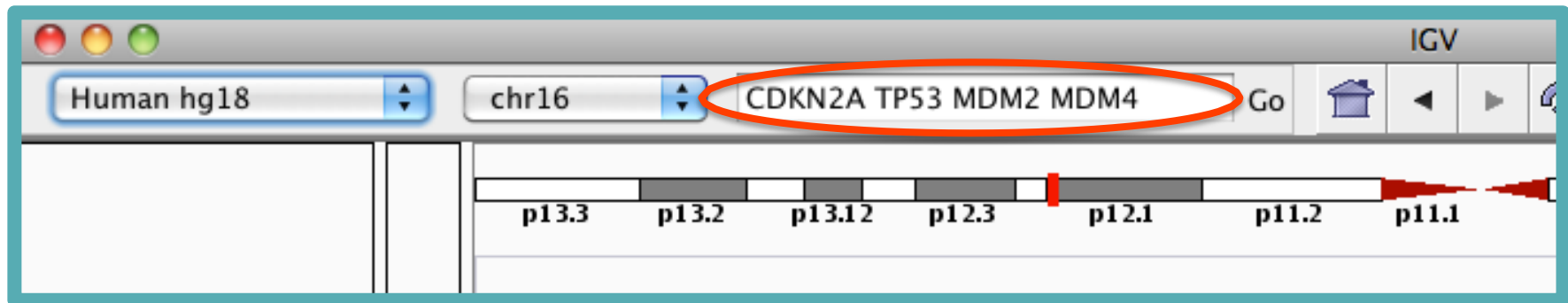
# Viewing multiple regions



# Viewing multiple regions

- **Search box**

Enter multiple loci or features in the search box



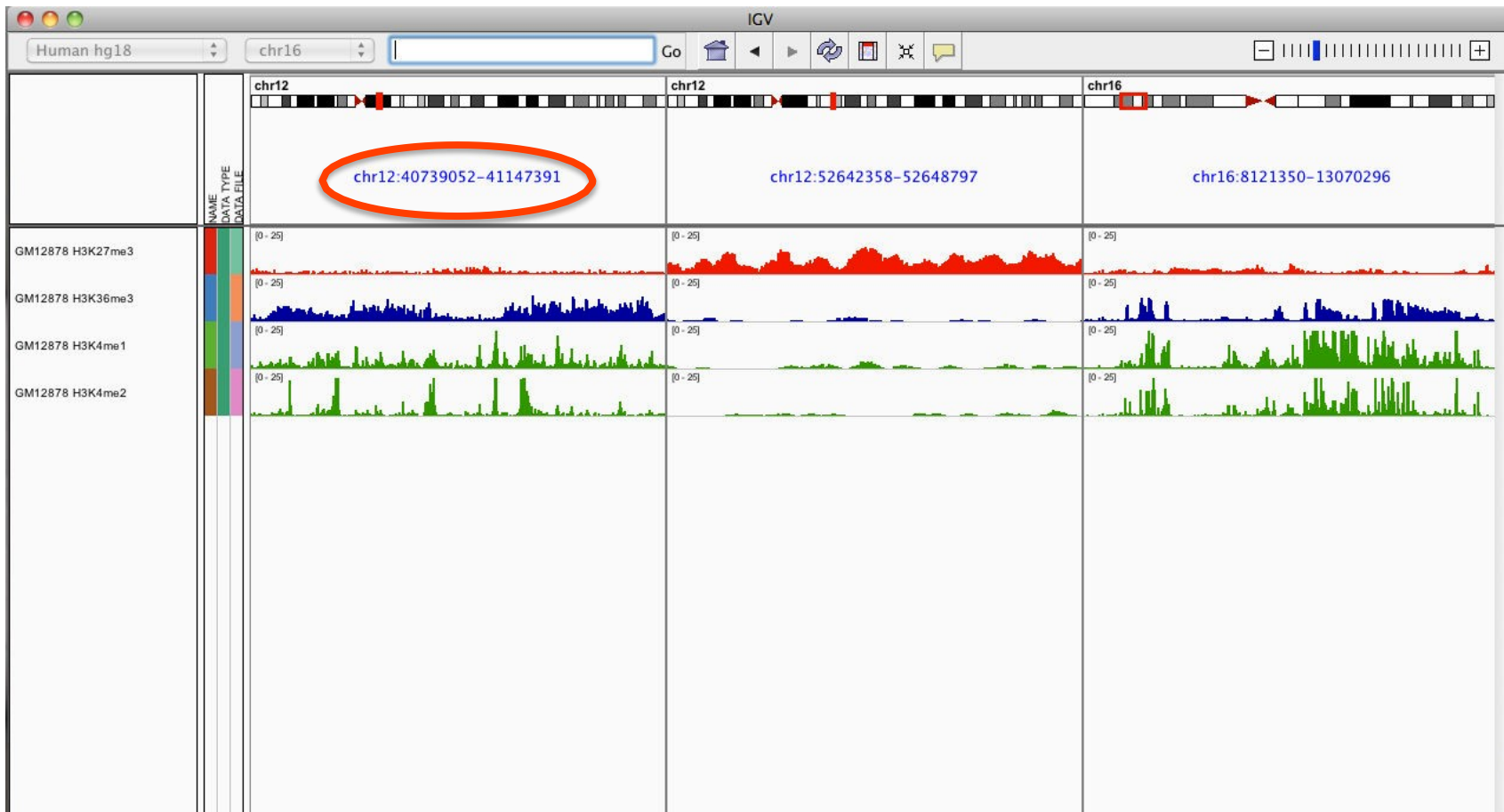
- **Regions > Gene Lists...**

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

# Viewing multiple regions

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

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

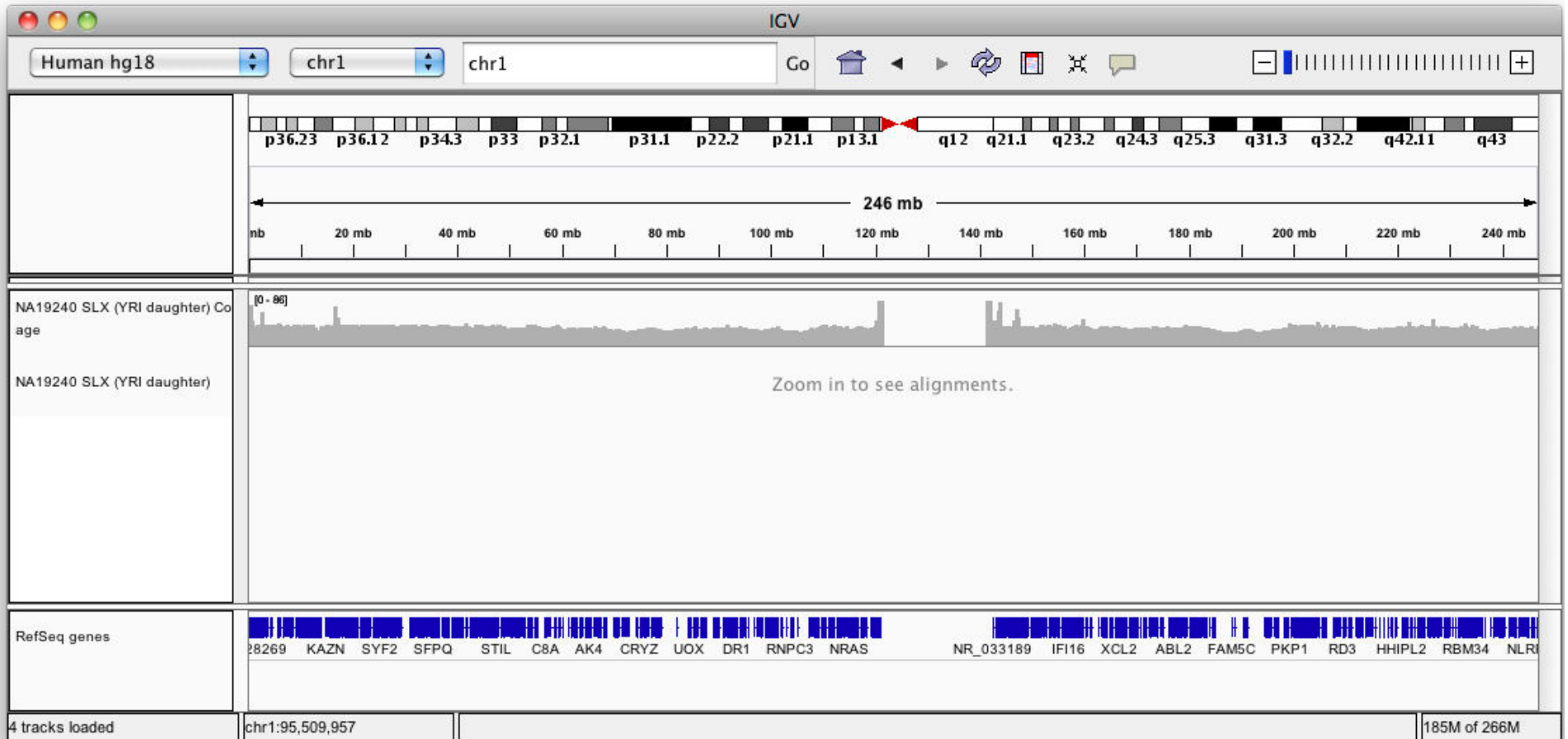


# Viewing NGS Data

# Viewing alignments



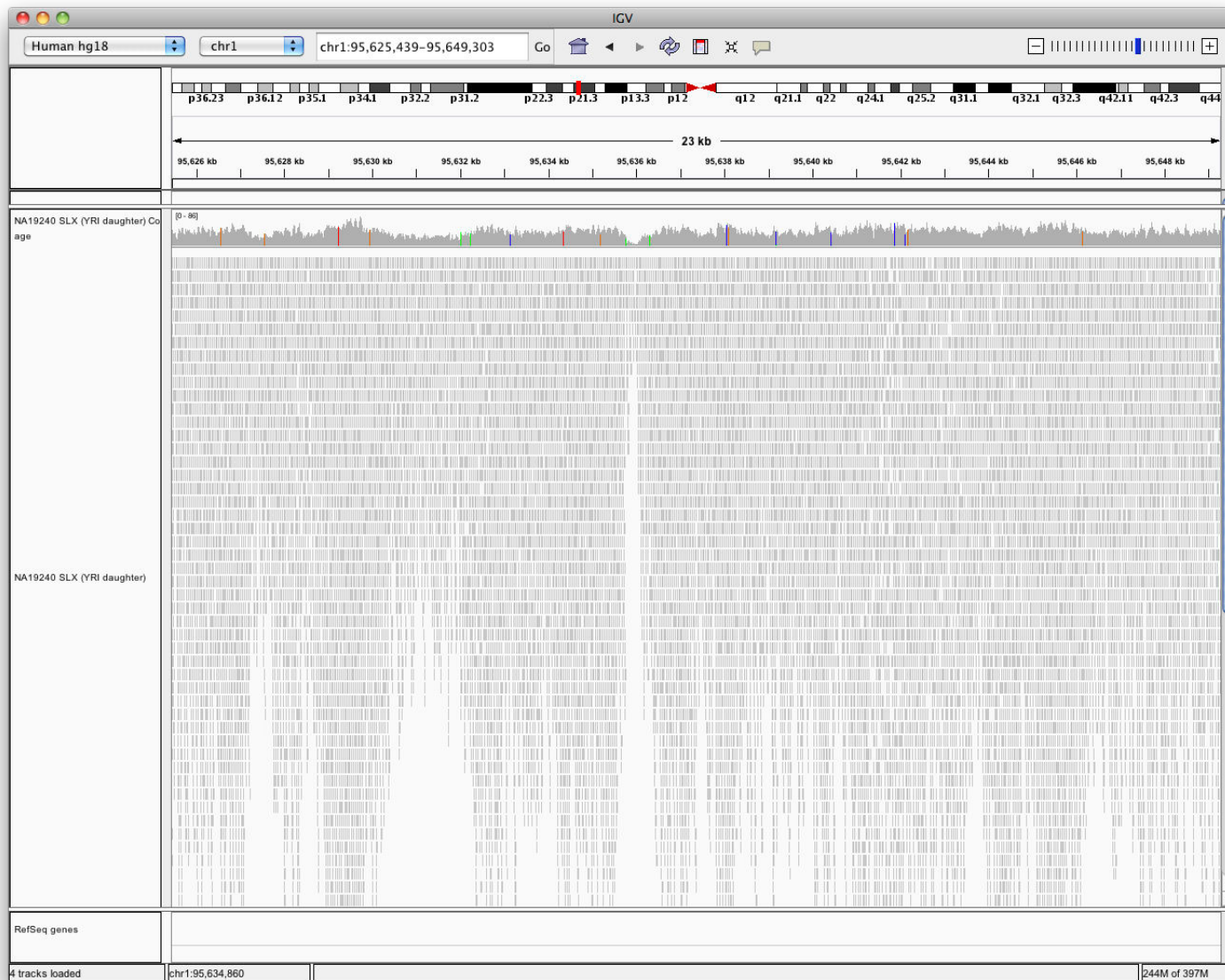
## Whole chromosome view



# Viewing alignments

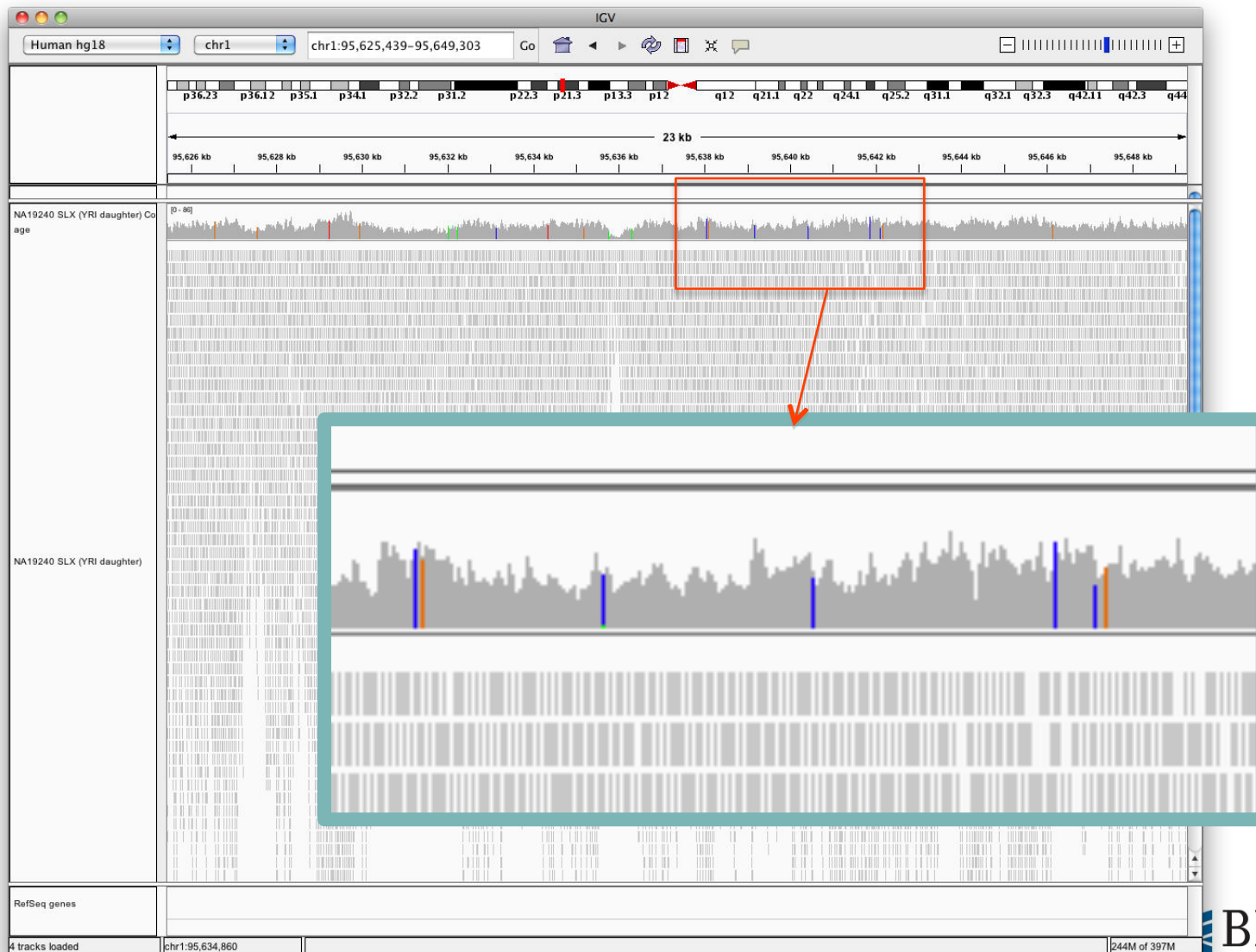


## Zoom in to view alignments



# Viewing alignments

Coverage track now has more detail





# Viewing alignments

Zoom in to see more detail

Bases that do not match the reference sequence are highlighted by color



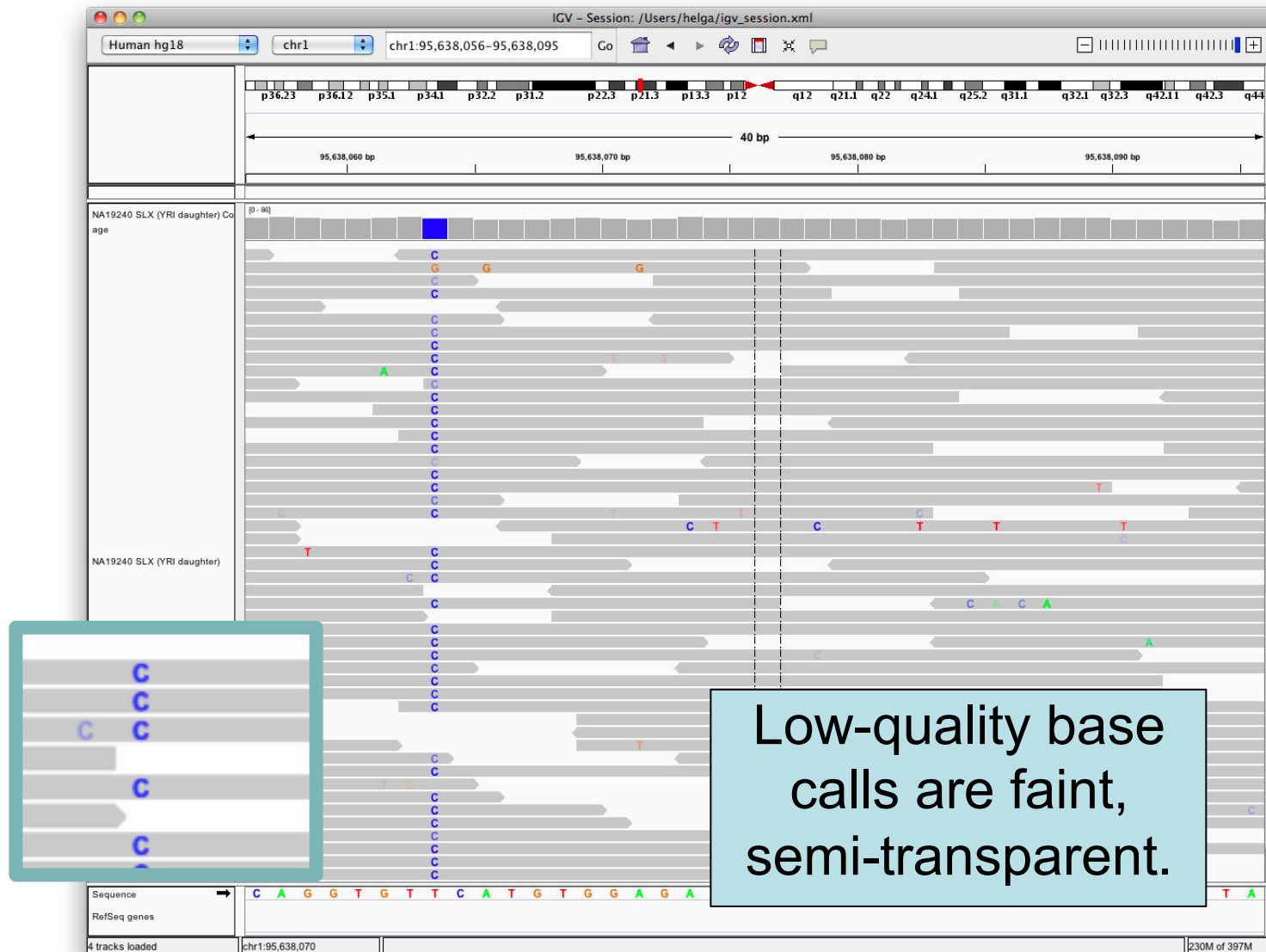
# Viewing alignments

Zoom in to see more detail

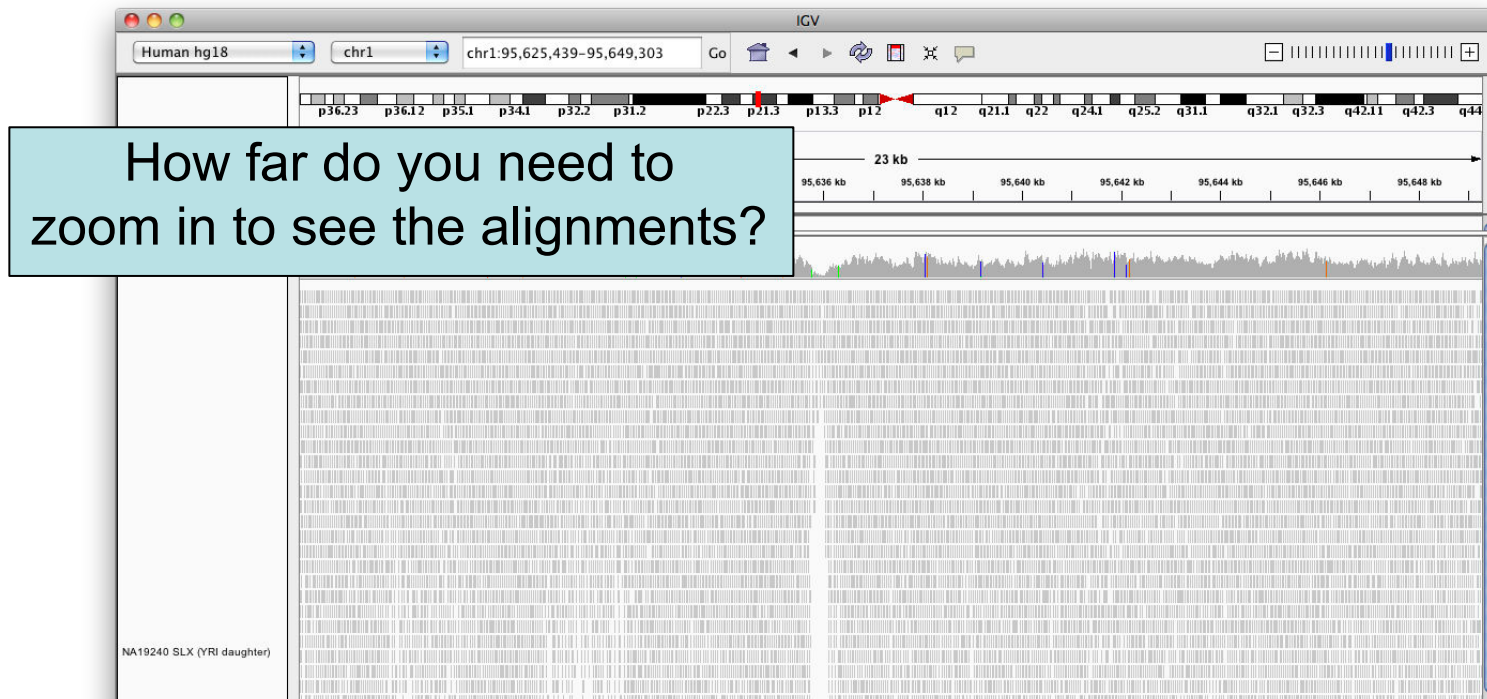


# Viewing alignments

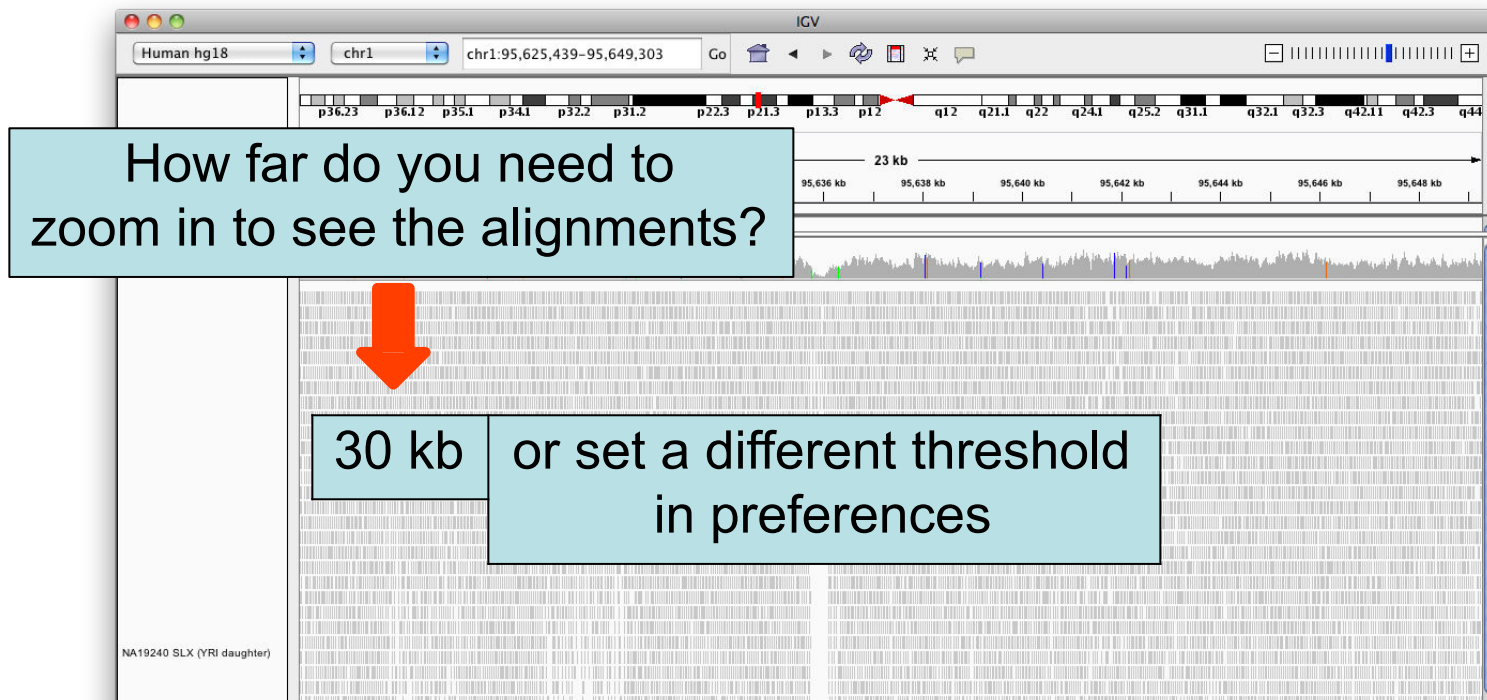
Zoom in to see more detail



# Viewing alignments

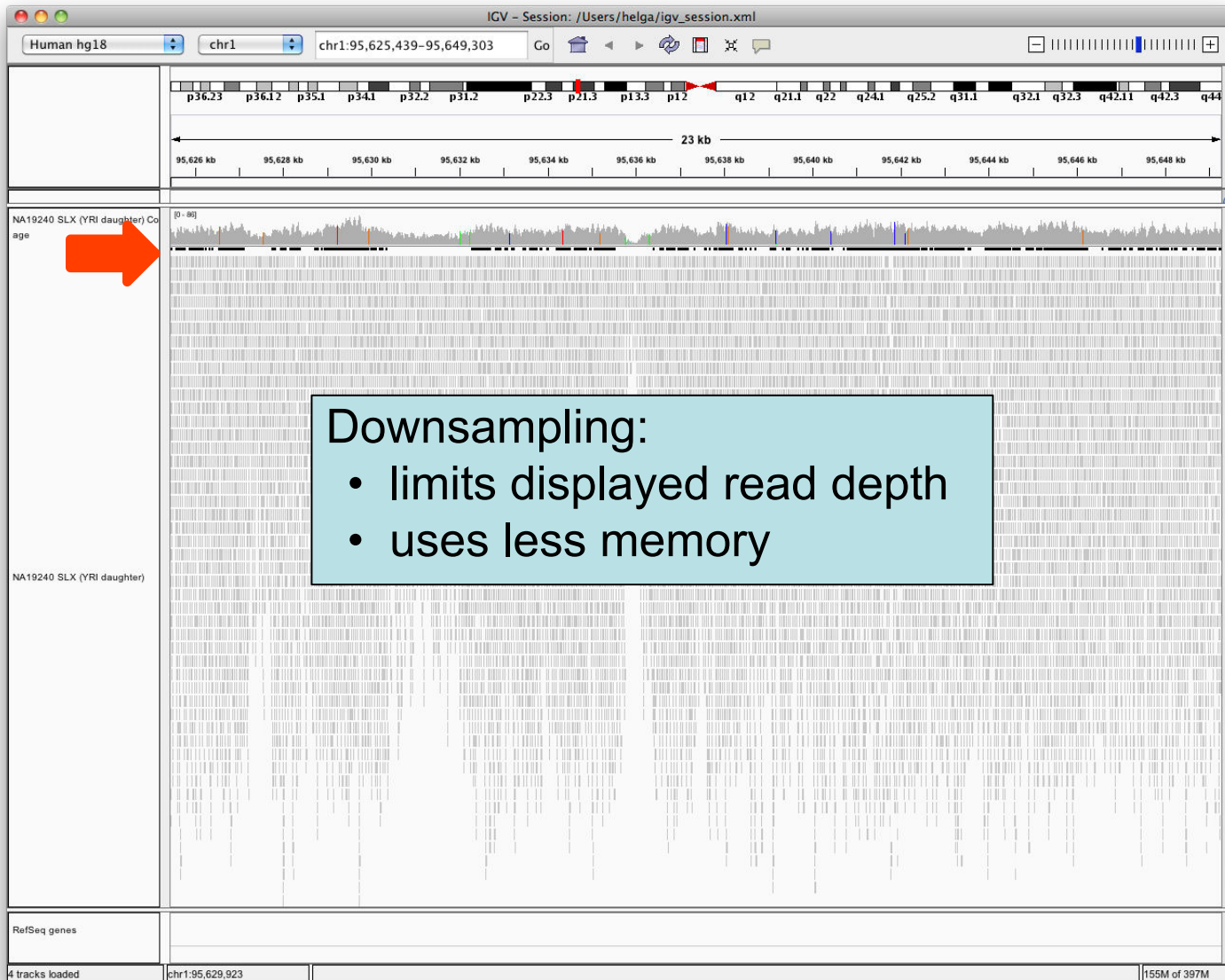


# Viewing alignments



- 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

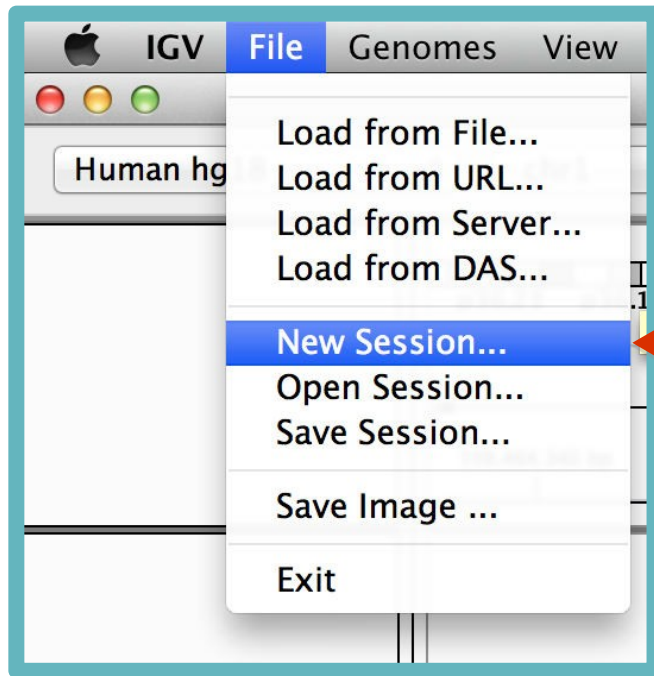
---



Hands-on exercise

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

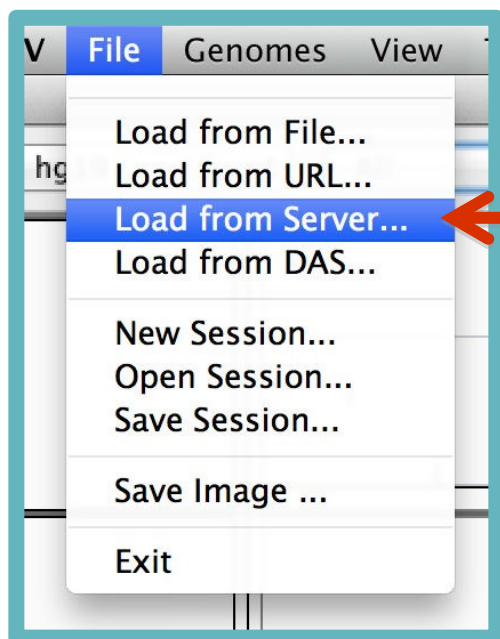
# Viewing SNPs



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

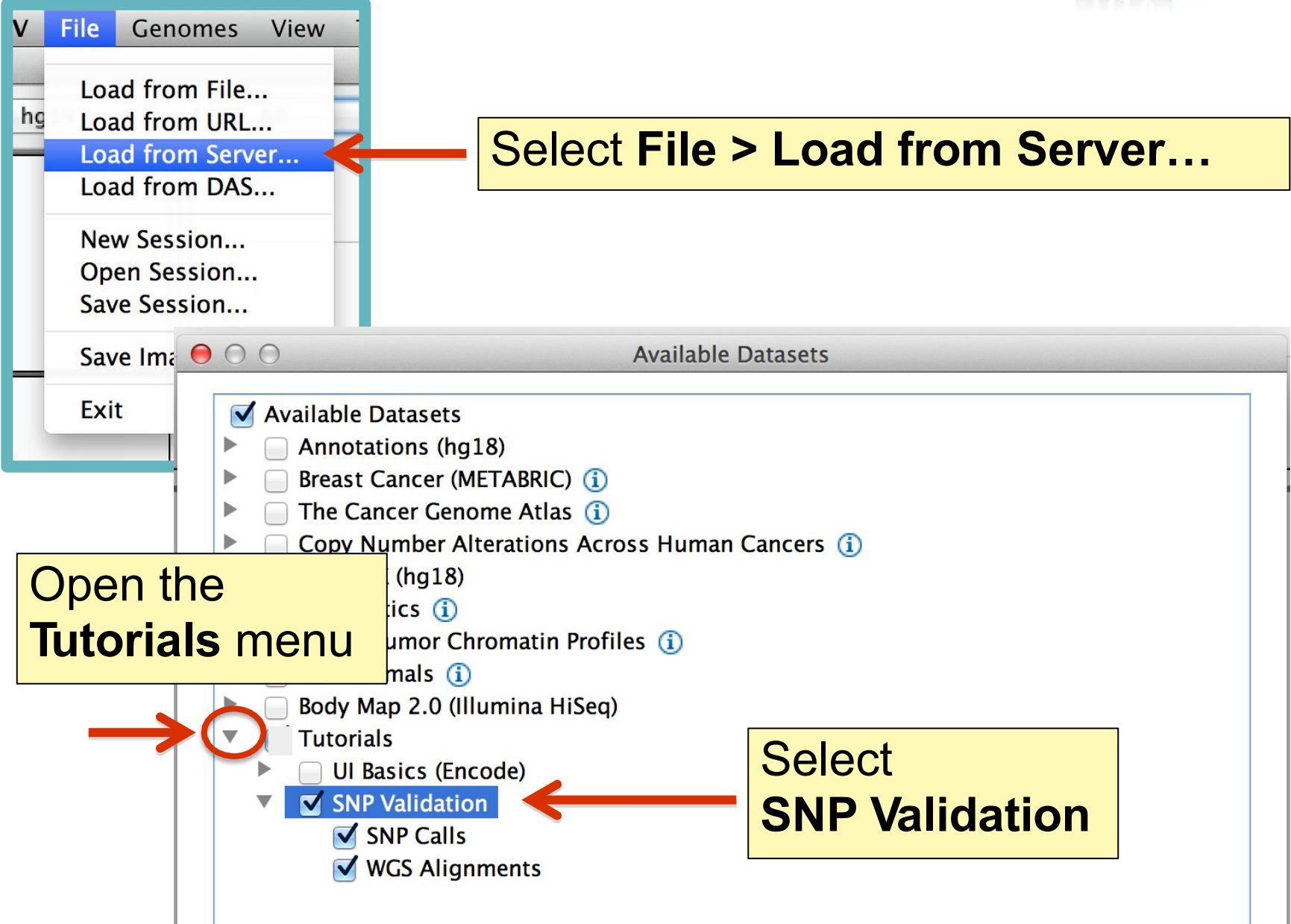


# Viewing SNPs



Select File > Load from Server...

# Viewing SNPs



The screenshot shows the IGV interface. The 'File' menu is open, with 'Load from Server...' selected. A yellow box with an arrow points to this option, containing the text 'Select File > Load from Server...'. Below the menu, the 'Available Datasets' dialog is open, showing a list of datasets. The 'Tutorials' section is expanded, and 'SNP Validation' is selected. A yellow box with an arrow points to this option, containing the text 'Select SNP Validation'. Another yellow box with an arrow points to the 'Tutorials' dropdown arrow, containing the text 'Open the Tutorials menu'.

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

**Open the Tutorials menu**

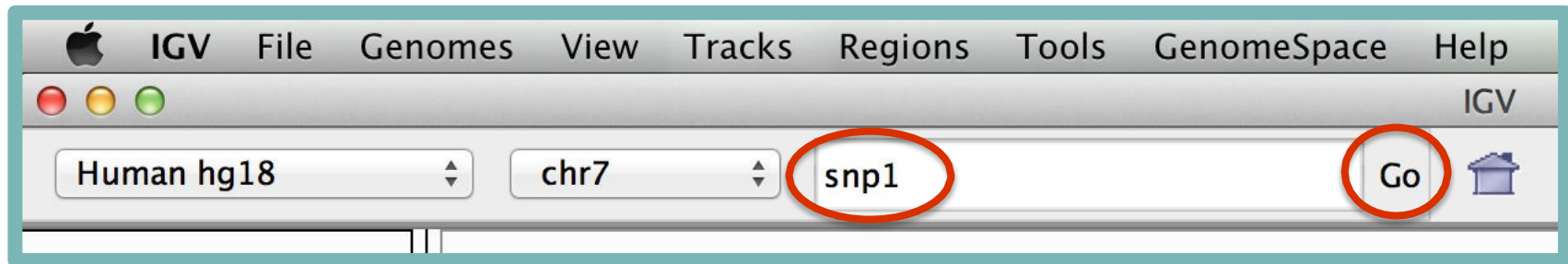
**Select SNP Validation**

- File
- Genomes
- View
- Load from File...
- Load from URL...
- Load from Server...**
- Load from DAS...
- New Session...
- Open Session...
- Save Session...
- Save Image...
- Exit

Available Datasets

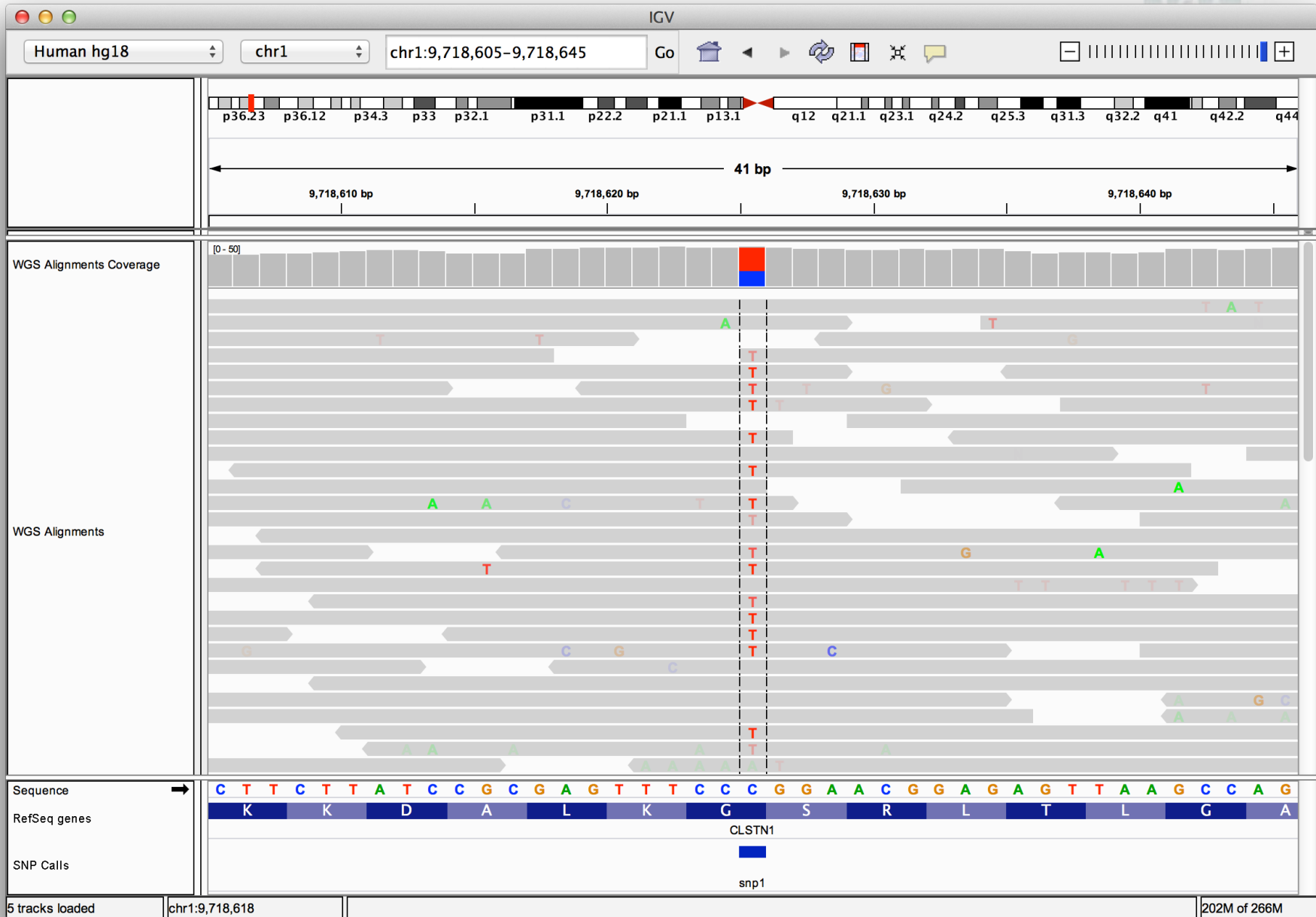
- Available Datasets
- Annotations (hg18)
- Breast Cancer (METABRIC) ⓘ
- The Cancer Genome Atlas ⓘ
- Copy Number Alterations Across Human Cancers ⓘ
- ... (hg18)
- ... ics ⓘ
- ... umor Chromatin Profiles ⓘ
- ... mals ⓘ
- Body Map 2.0 (Illumina HiSeq)
- Tutorials
  - UI Basics (Encode)
  - SNP Validation**
  - SNP Calls
  - WGS Alignments

# Viewing SNPs

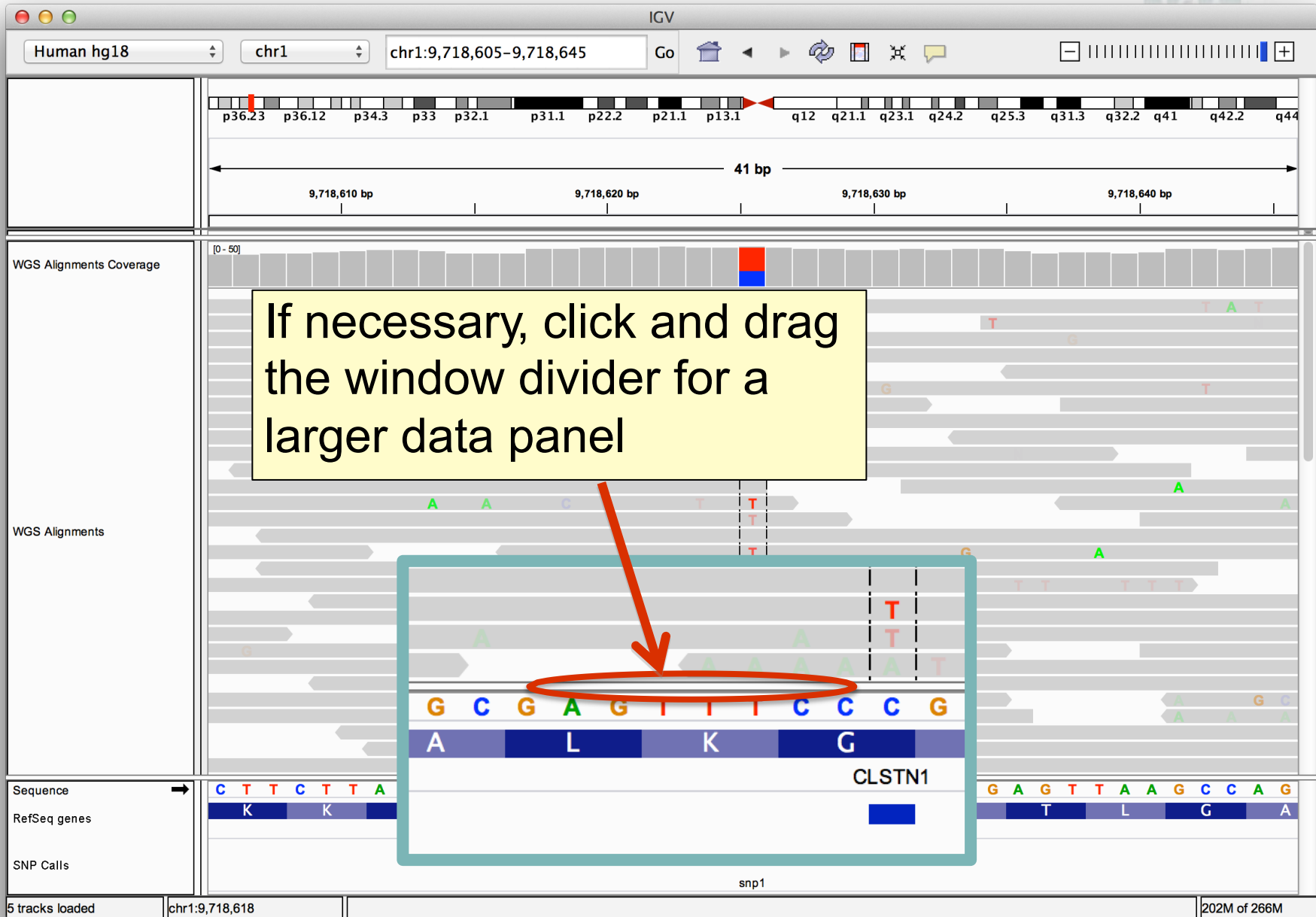


Type “snp1” in the **Search Box** and click **Go**

# Viewing SNPs



# Viewing SNPs



IGV

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

p36.23 p36.12 p34.3 p33 p32.1 p31.1 p22.2 p21.1 p13.1 q12 q21.1 q23.1 q24.2 q25.3 q31.3 q32.2 q41 q42.2 q44

9,718,610 bp 9,718,620 bp 9,718,630 bp 9,718,640 bp

41 bp

WGS Alignments Coverage

WGS Alignments

Sequence →

RefSeq genes

SNP Calls

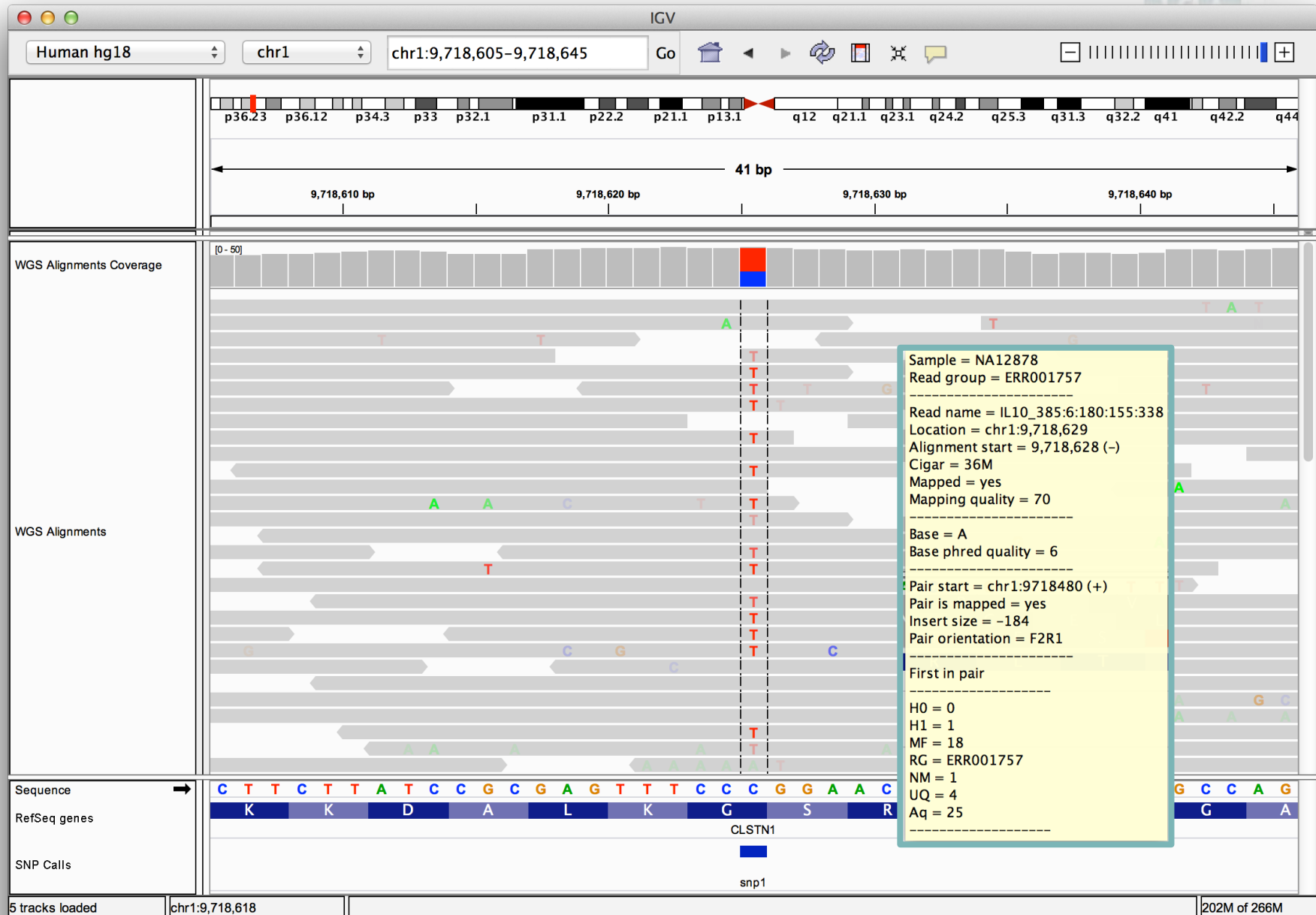
CLSTN1

snp1

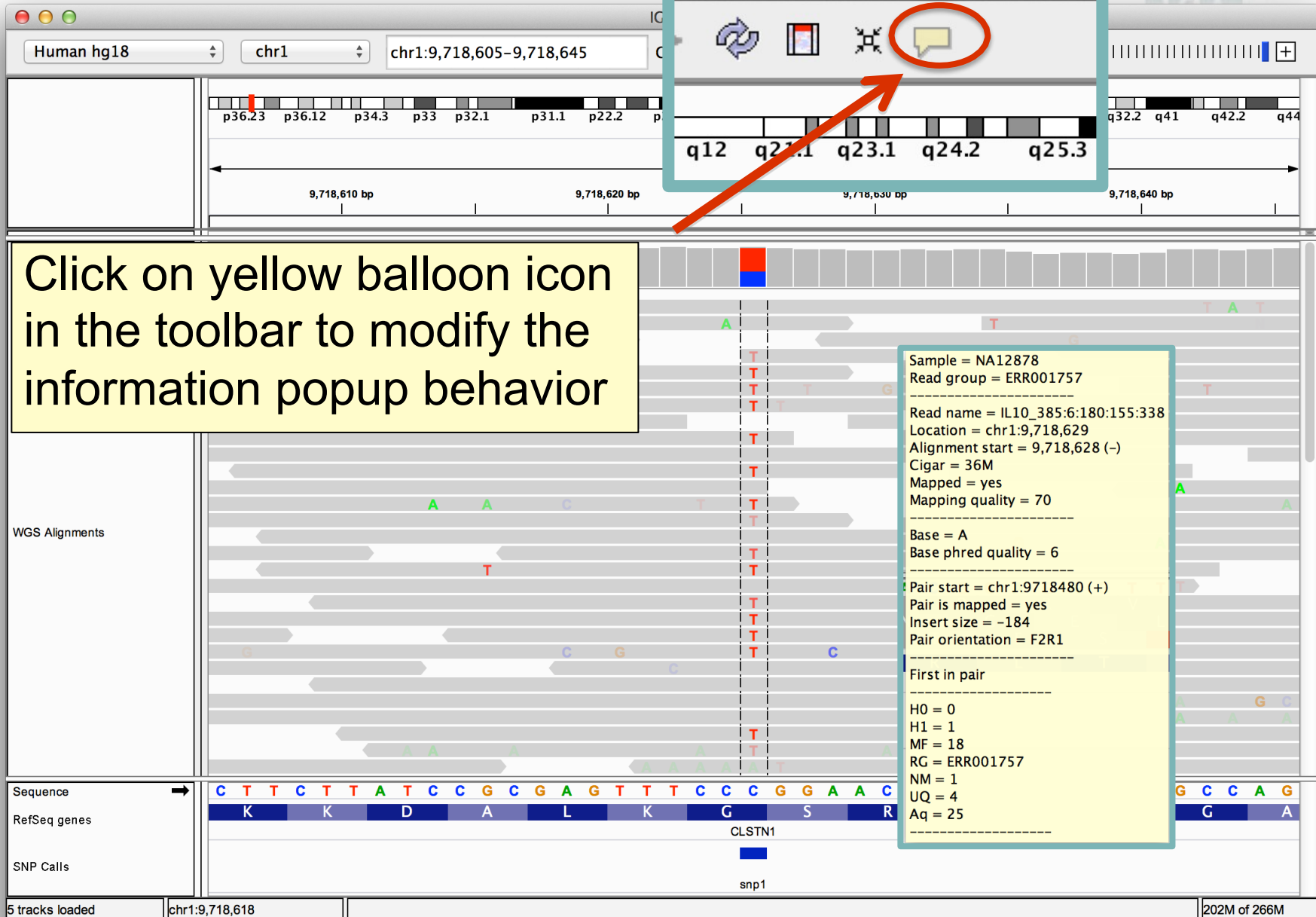
5 tracks loaded chr1:9,718,618 202M of 266M

If necessary, click and drag the window divider for a larger data panel

# Viewing SNPs



# Viewing SNPs



Human hg18 chr1 chr1:9,718,605-9,718,645

q12 q23.1 q23.1 q24.2 q25.3

9,718,610 bp 9,718,620 bp 9,718,630 bp 9,718,640 bp

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

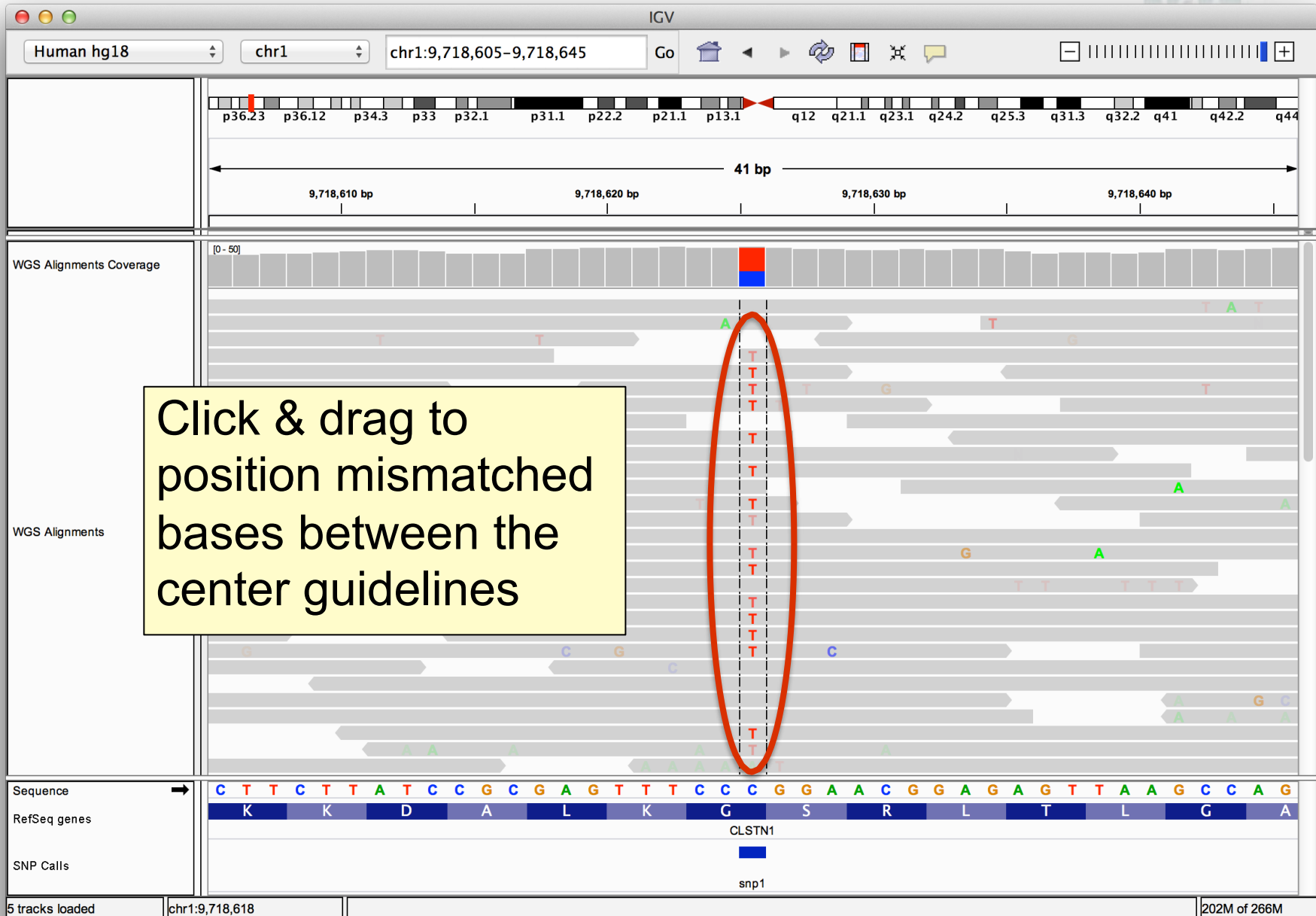
WGS Alignments

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

Sample = NA12878  
Read group = ERR001757  
Read name = IL10\_385:6:180:155:338  
Location = chr1:9,718,629  
Alignment start = 9,718,628 (-)  
Cigar = 36M  
Mapped = yes  
Mapping quality = 70  
Base = A  
Base phred quality = 6  
Pair start = chr1:9718480 (+)  
Pair is mapped = yes  
Insert size = -184  
Pair orientation = F2R1  
First in pair  
H0 = 0  
H1 = 1  
MF = 18  
RG = ERR001757  
NM = 1  
UQ = 4  
Aq = 25

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

# Viewing SNPs



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

WGS Alignments Coverage

WGS Alignments

Sequence → 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 A G T T A A G C C A G

RefSeq genes K K D A L K G S R L T L G A

CLSTN1

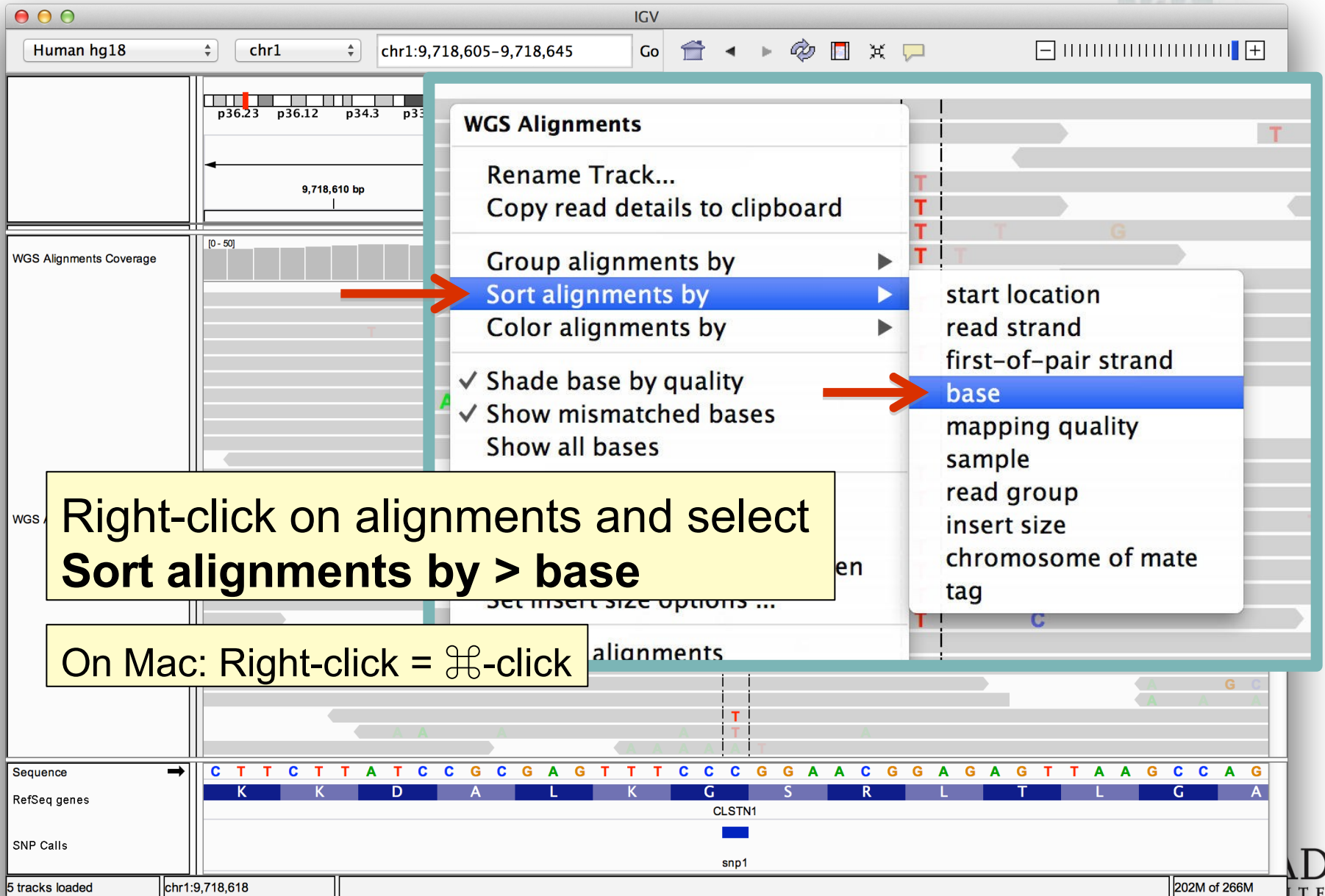
SNP Calls snp1

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

Click & drag to position mismatched bases between the center guidelines



# Viewing SNPs




Human hg18 chr1 chr1:9,718,605-9,718,645

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

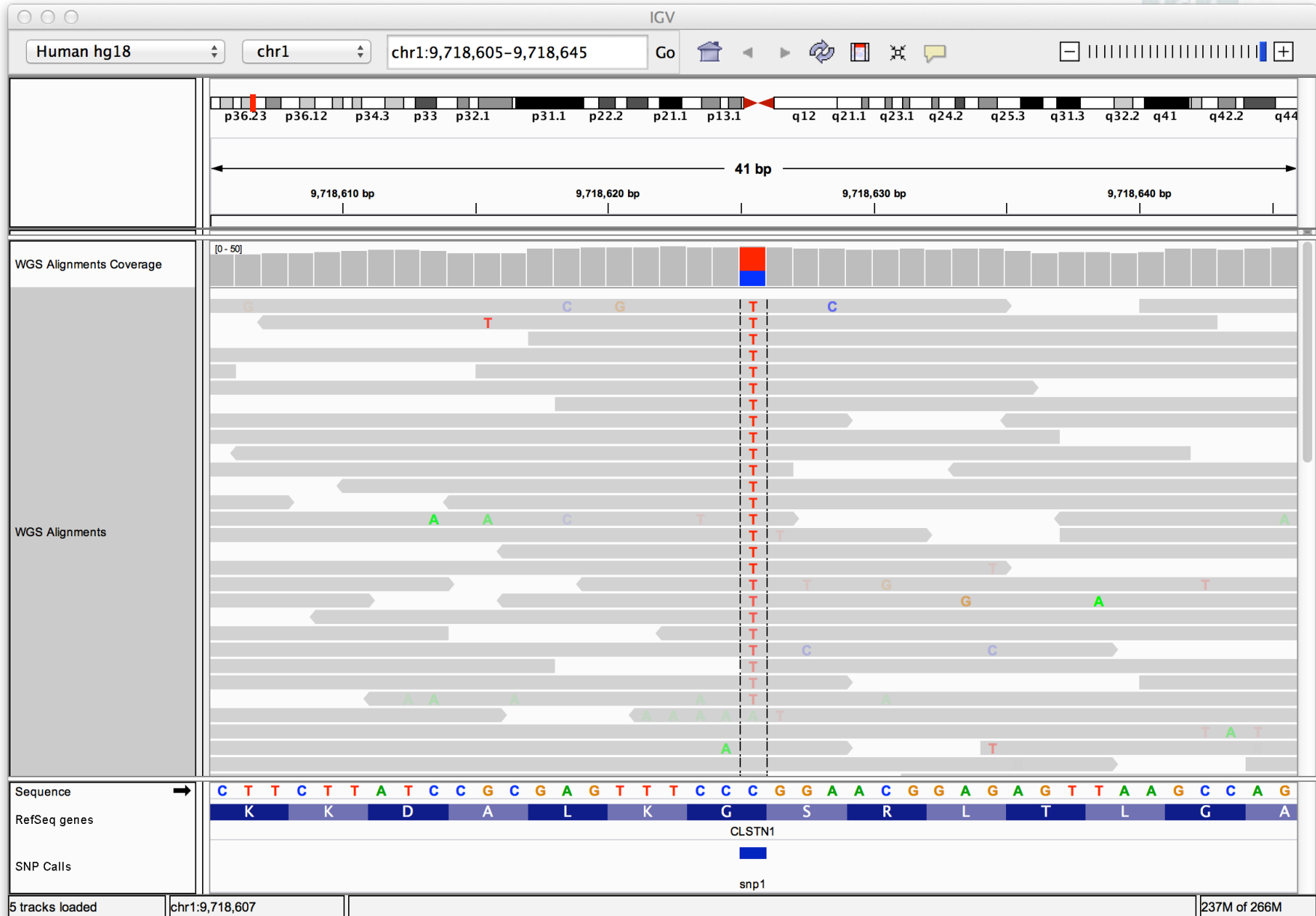
Right-click on alignments and select **Sort alignments by > base**

On Mac: Right-click = -click

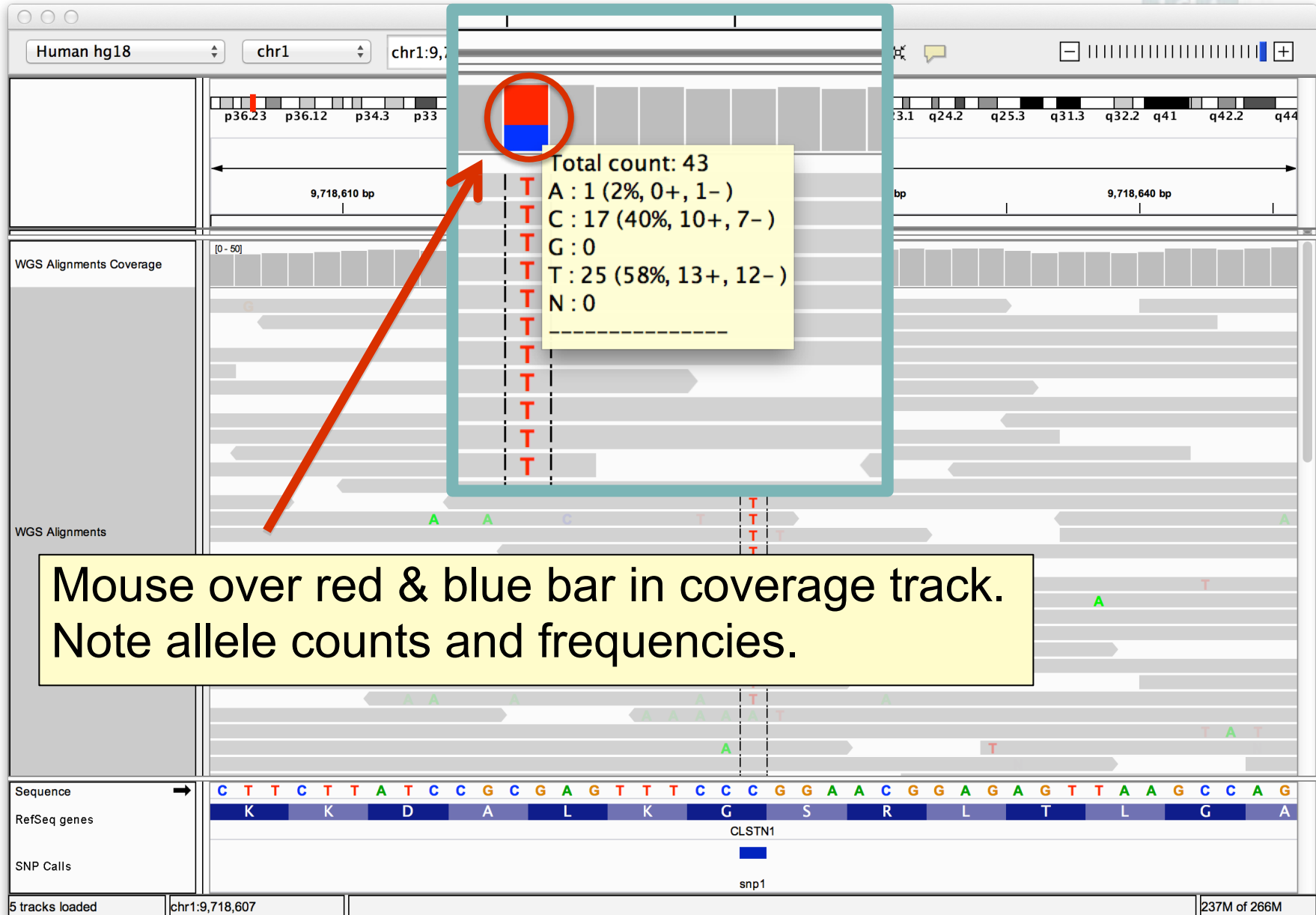
Sequence  
RefSeq genes  
SNP Calls

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

# Viewing SNPs

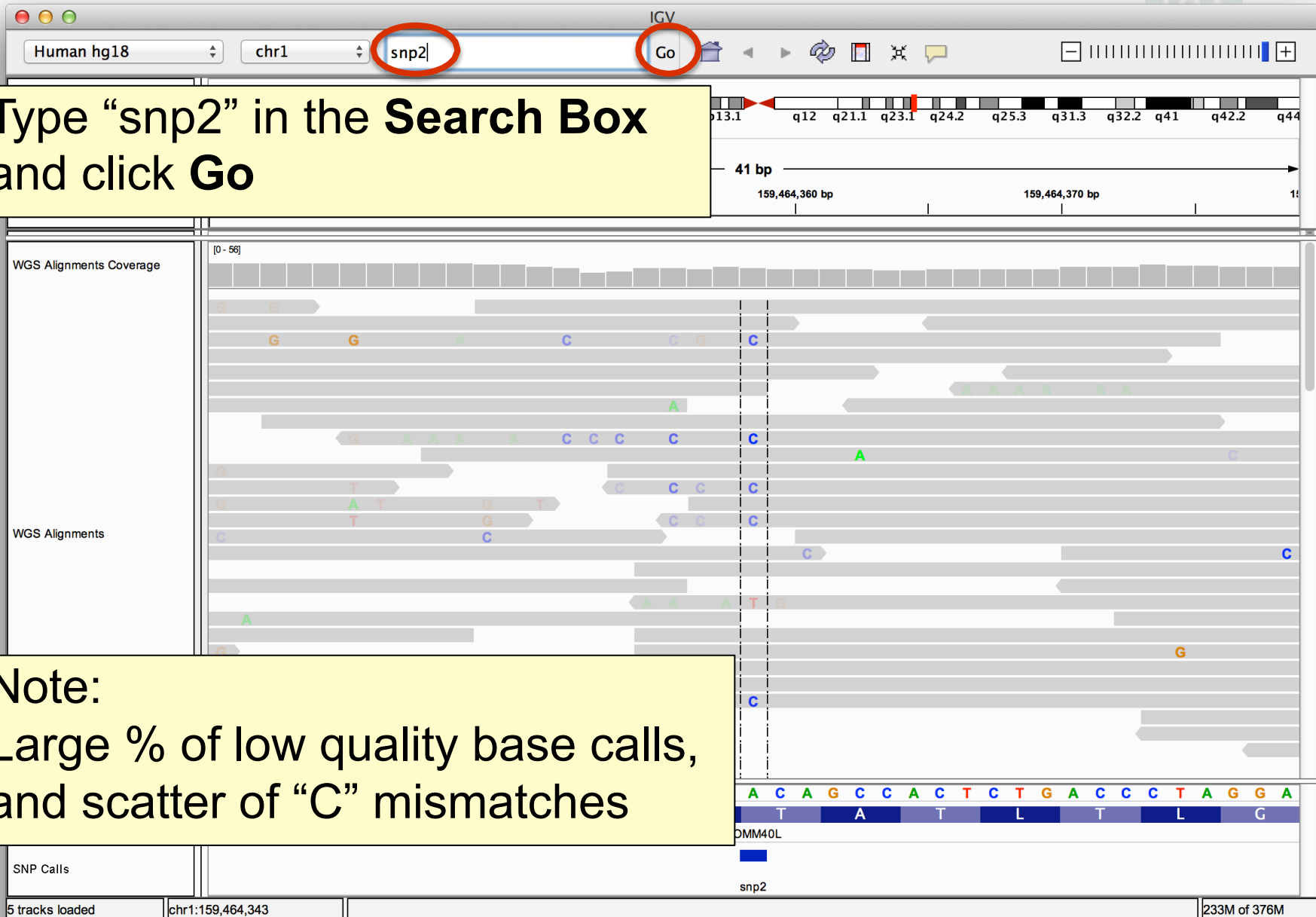


# Viewing SNPs



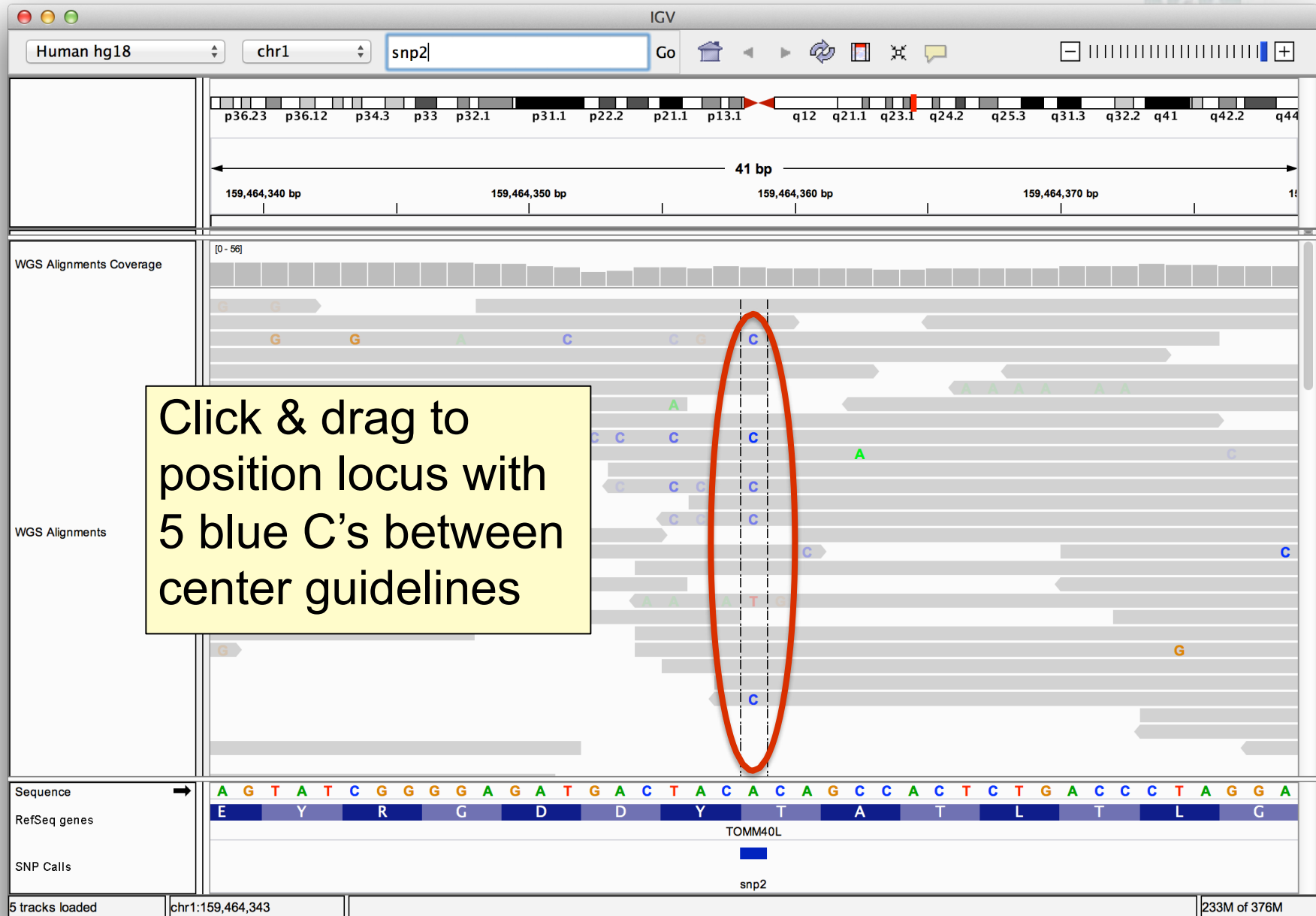
Mouse over red & blue bar in coverage track.  
Note allele counts and frequencies.

# Viewing SNPs



Note:  
Large % of low quality base calls,  
and scatter of "C" mismatches

# Viewing SNPs



Human hg18 chr1 snp2 Go

41 bp

159,464,340 bp 159,464,350 bp 159,464,360 bp 159,464,370 bp

WGS Alignments Coverage

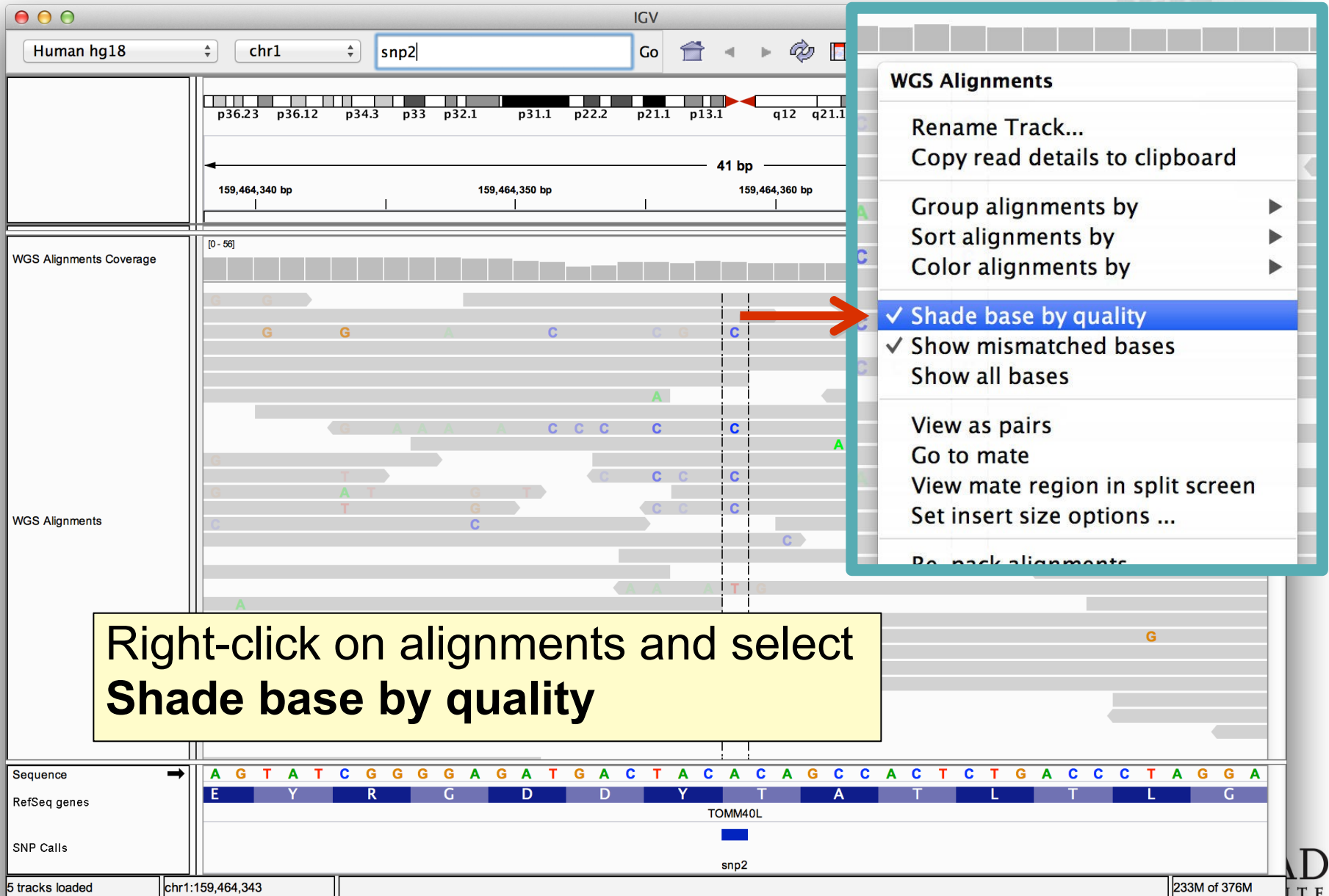
WGS Alignments

Click & drag to position locus with 5 blue C's between center guidelines

Sequence → A G T A T C G 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  
RefSeq genes E Y R G D D Y T A T L T L G  
SNP Calls TOMM40L  
snp2

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

# Viewing SNPs



The screenshot shows the IGV interface with a track titled 'WGS Alignments Coverage' and a track titled 'WGS Alignments'. A context menu is open over the alignments, listing various options. A red arrow points to the 'Shade base by quality' option, which is highlighted in blue. Below the alignments, the reference sequence is shown as 'A G T A T C G 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' and the RefSeq gene is 'E Y R G D D Y T A T L T L G'. The SNP is labeled 'snp2' and is located at position 159,464,343 on chromosome 1. The sequence 'TOMM40L' is also visible.

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
- Set insert size options ...
- Re-pack alignments

Right-click on alignments and select **Shade base by quality**

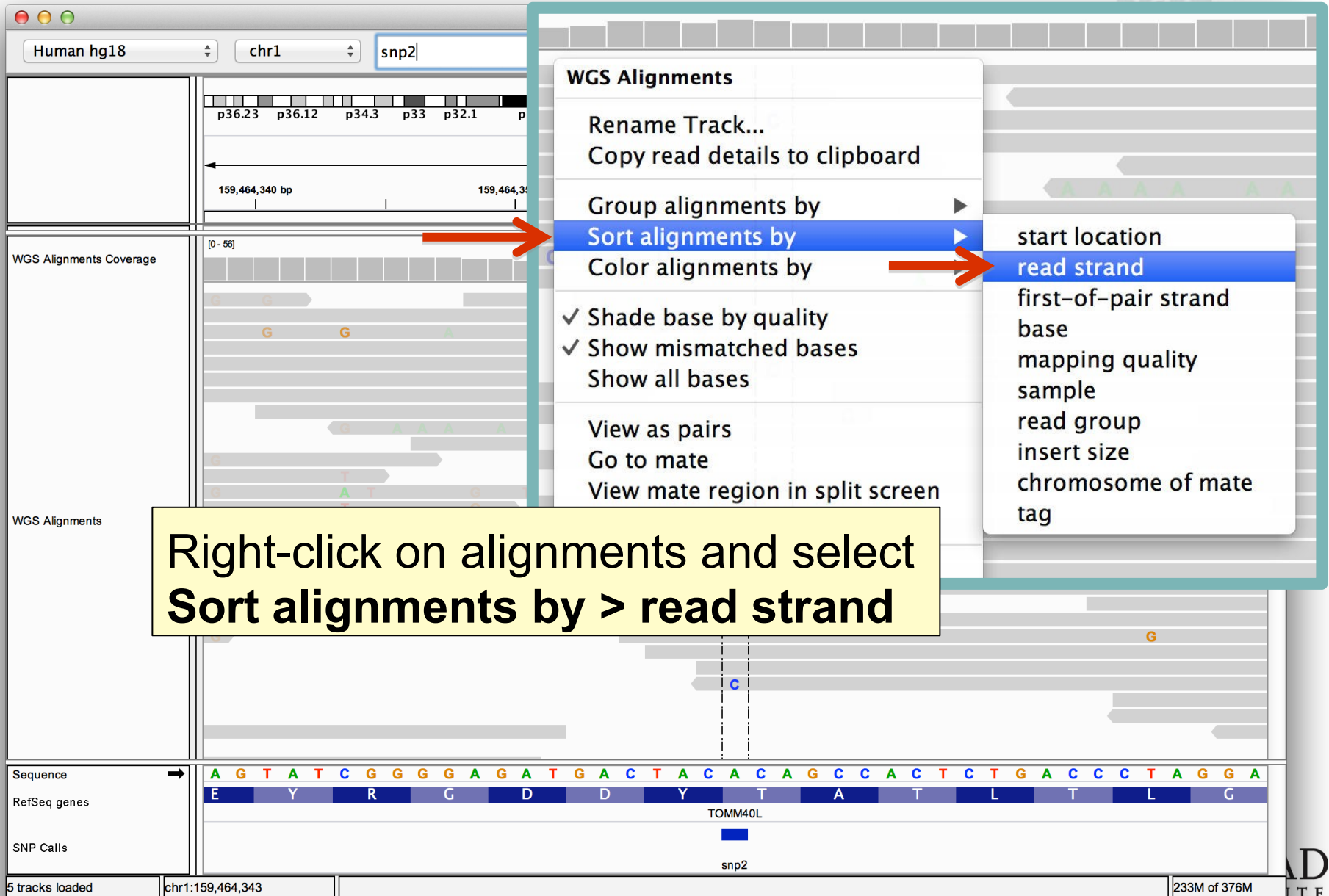
Sequence → A G T A T C G G G G A G A T G A C T A C A C A G C C A C T C T G A C C C T A G G A  
RefSeq genes E Y R G D D Y T A T L T L G  
SNP Calls  
TOMM40L  
snp2

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

# Viewing SNPs



# Viewing SNPs



Human hg18 chr1 snp2

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

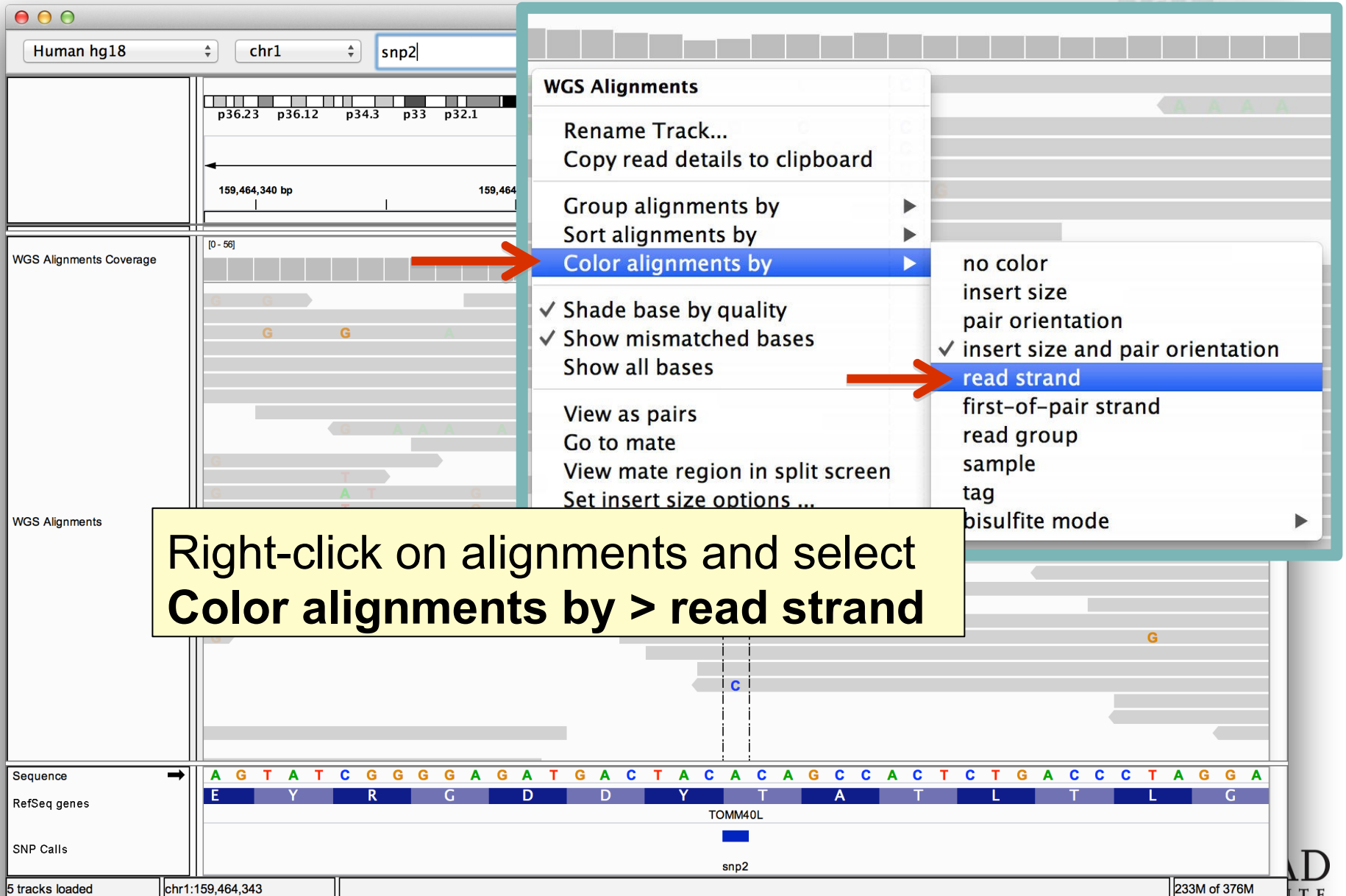
Right-click on alignments and select **Sort alignments by > read strand**

Sequence → A G T A T C G G G 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  
RefSeq genes E Y R G D D Y T A T L T L G  
SNP Calls TOMM40L  
snp2

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



# Viewing SNPs



Human hg18 chr1 snp2

WGS Alignments

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

WGS Alignments Coverage

WGS Alignments

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

RefSeq genes TOMM40L

SNP Calls

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

AD  
TE

Right-click on alignments and select  
**Color alignments by > read strand**

# Viewing SNPs



# Viewing Structural Events

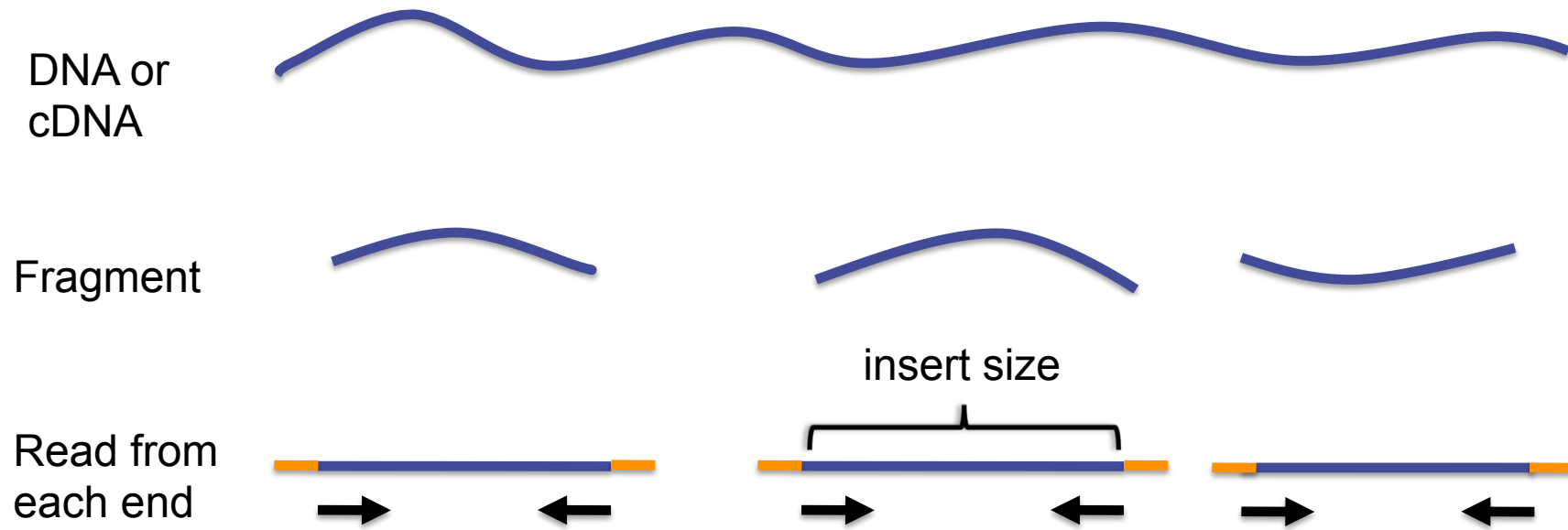
# Structural events

---

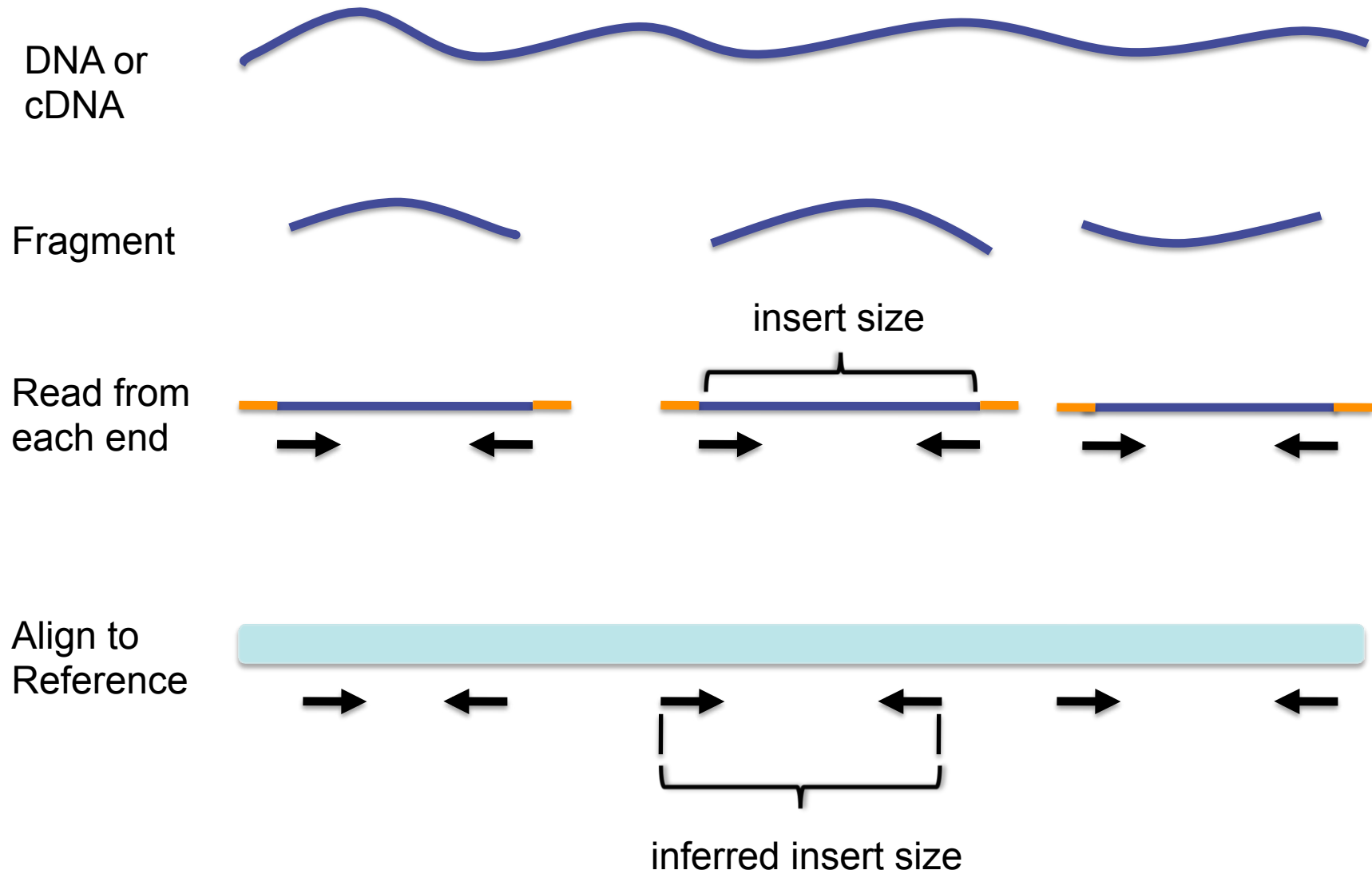


- Paired reads can yield evidence for genomic “structural events”, such as deletions, translocations, and inversions.
- Alignment coloring options help highlight these events based on:
  - Inferred insert size (template length)
  - Pair orientation (relative strand of pair)

# 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

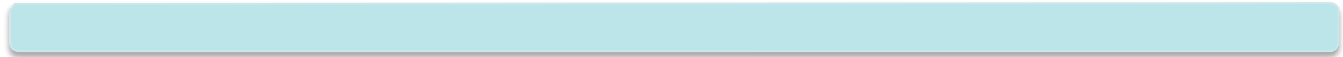


# Deletion

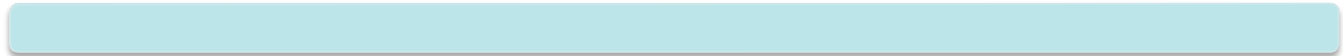
---



Reference  
Genome



Subject

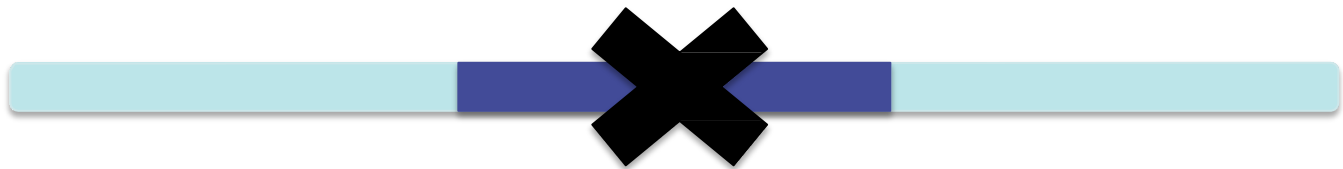


# Deletion

Reference  
Genome



Subject

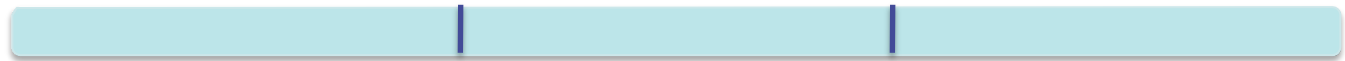


# Deletion

---



Reference  
Genome

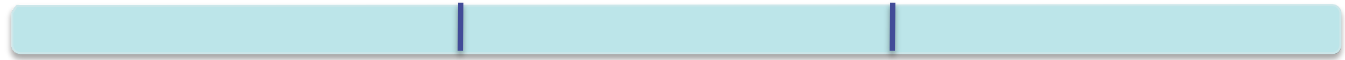


Subject



# Deletion

Reference  
Genome



Subject



# Deletion

Reference  
Genome

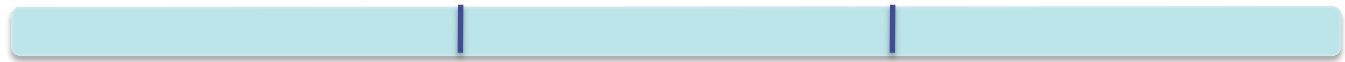


Subject

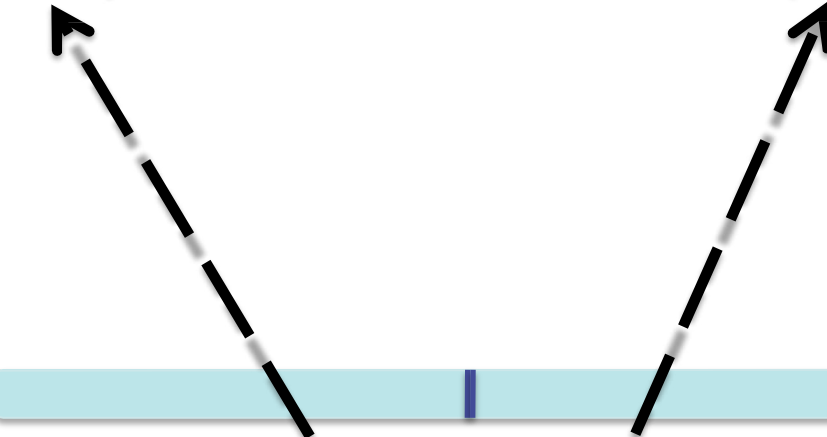
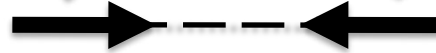


# Deletion

Reference  
Genome



Subject





# Deletion

Reference  
Genome



inferred insert size

Subject



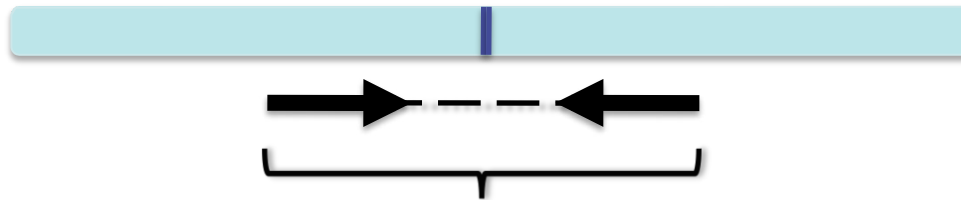
# Deletion

Reference  
Genome



inferred insert size

Subject



expected insert size

# Deletion

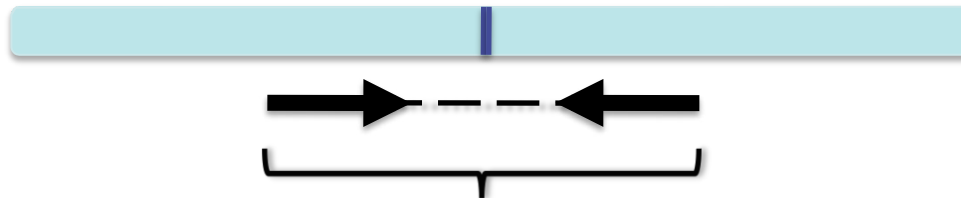
Inferred insert size is  $>$  expected value

Reference  
Genome



inferred insert size

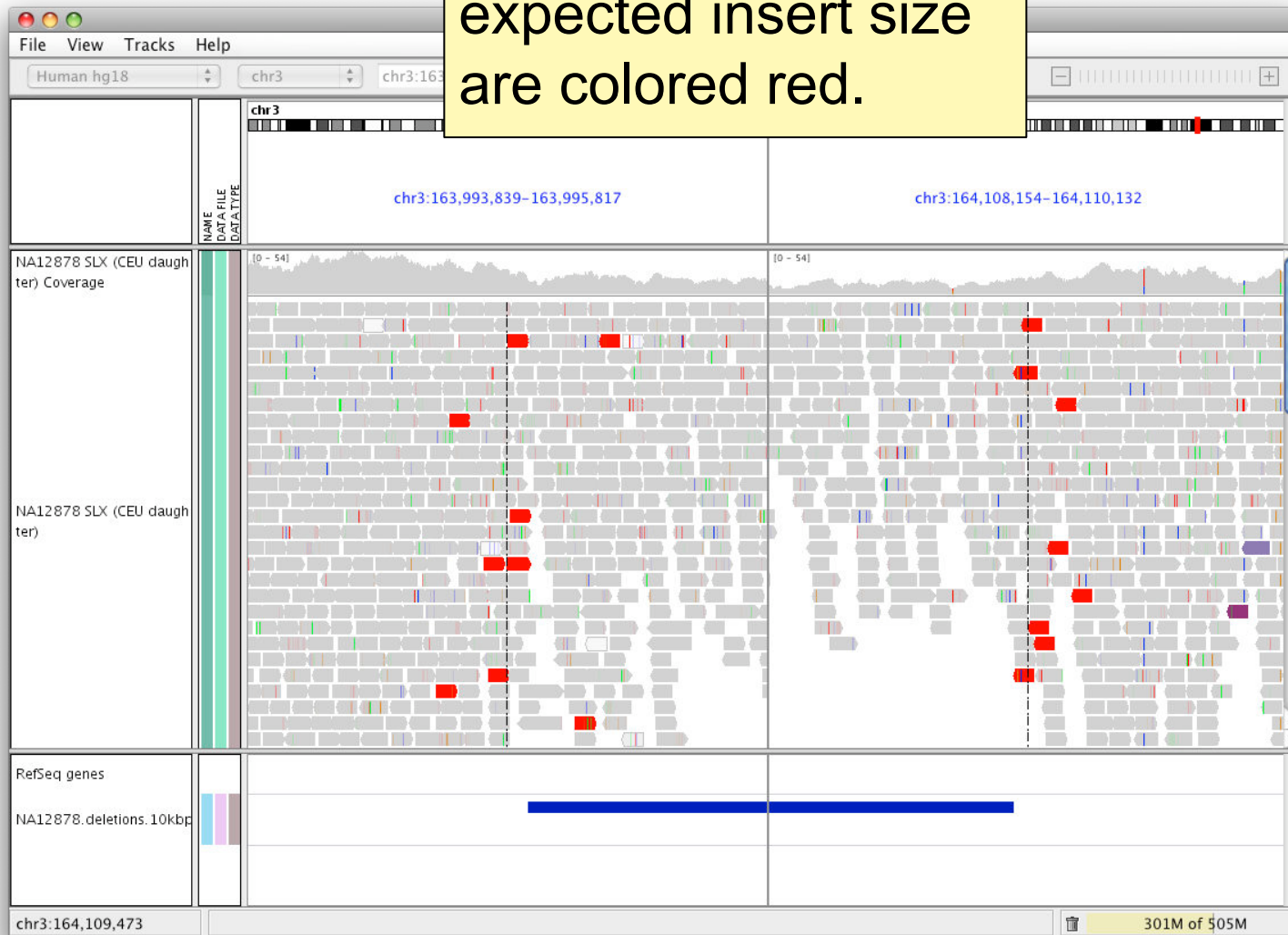
Subject



expected insert size

# Deletion

Pairs with larger than expected insert size are colored red.





# Deletion

Note drop in coverage



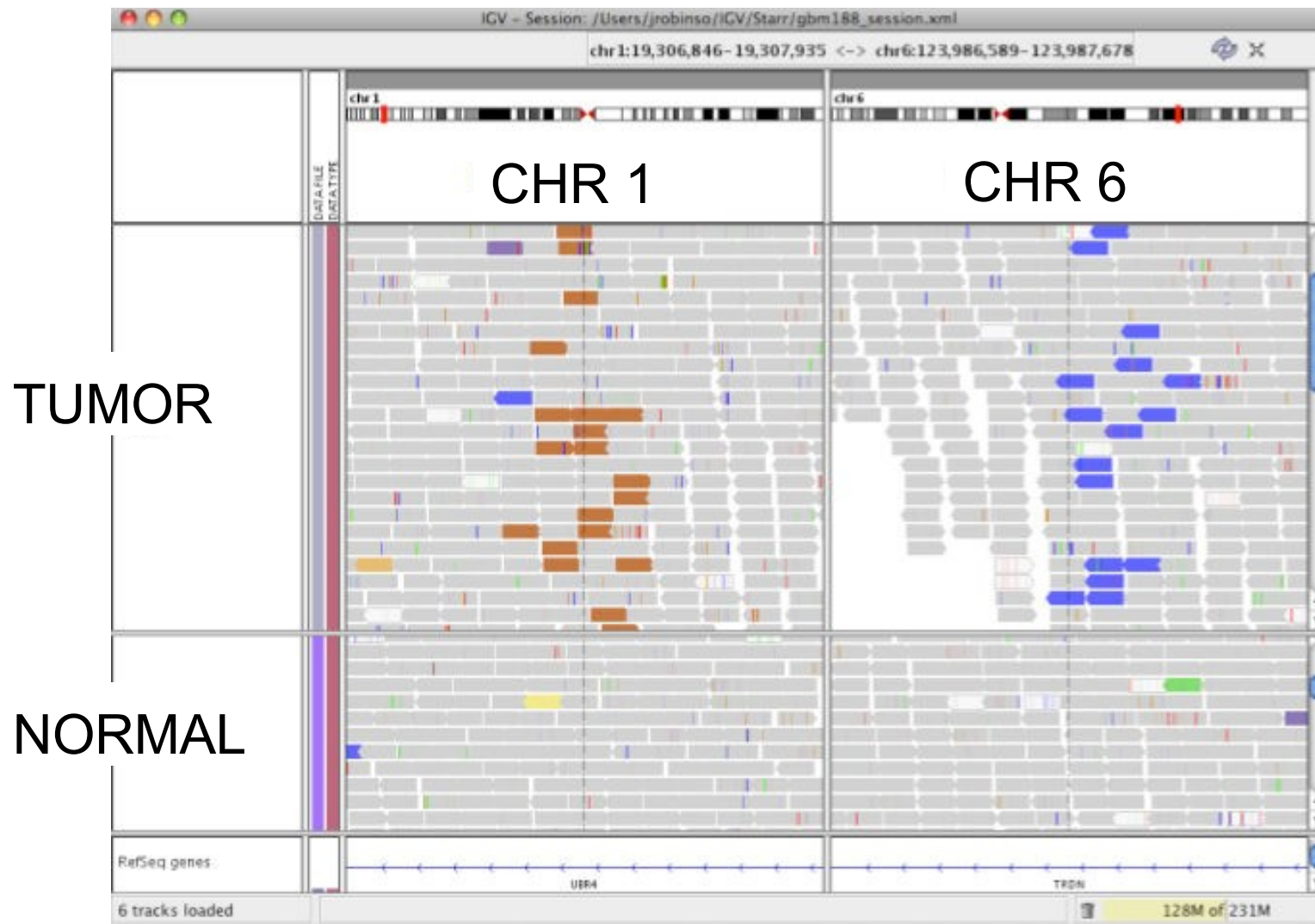
# Insert size color scheme

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

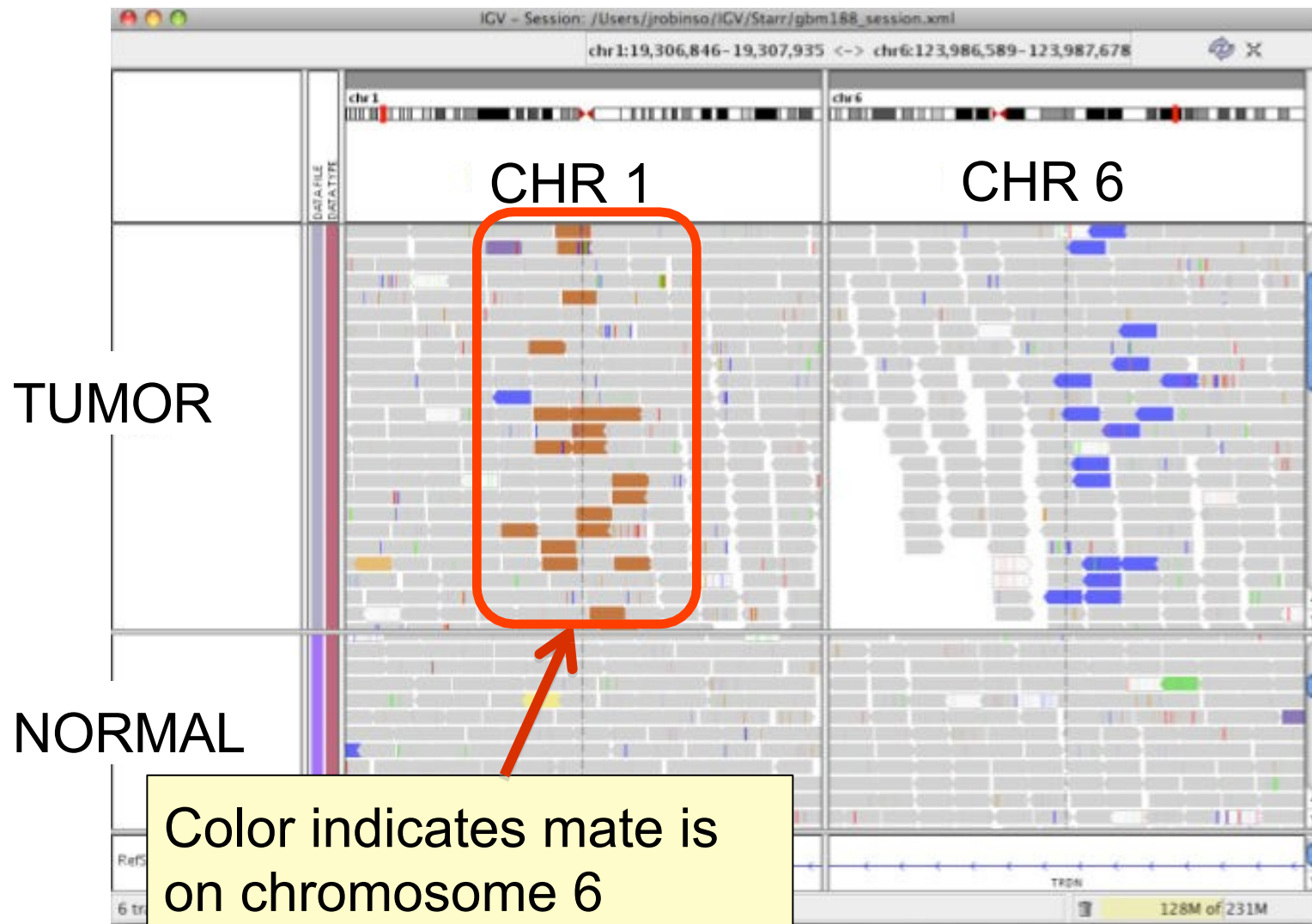
*Each end colored by chromosome of its mate*



# 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

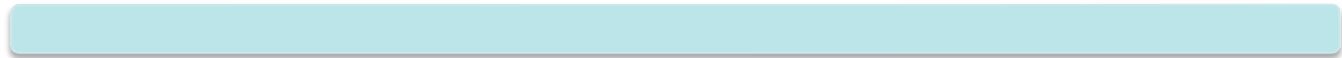
- read strand, left *vs* right, *and*
- read order, first *vs* second

# Inversion

---



Reference  
genome



# Inversion

---

Reference  
genome



# Inversion

Reference  
Genome



A

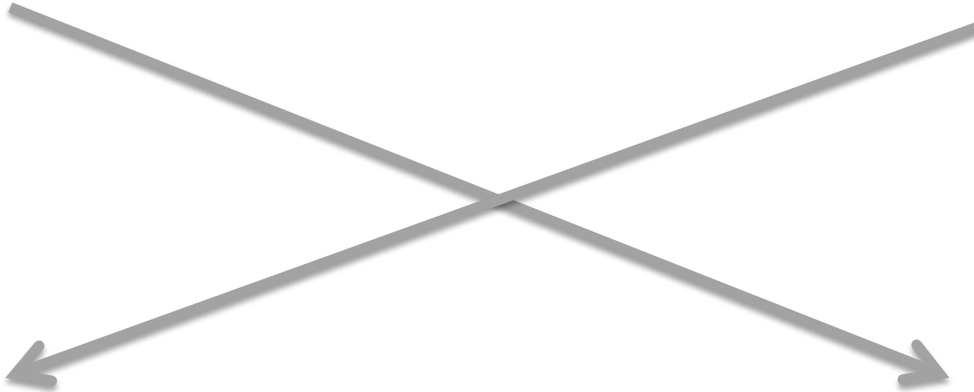
B

Subject



B

A



# Inversion

Reference  
Genome



Subject



# Inversion

Reference  
Genome



Subject

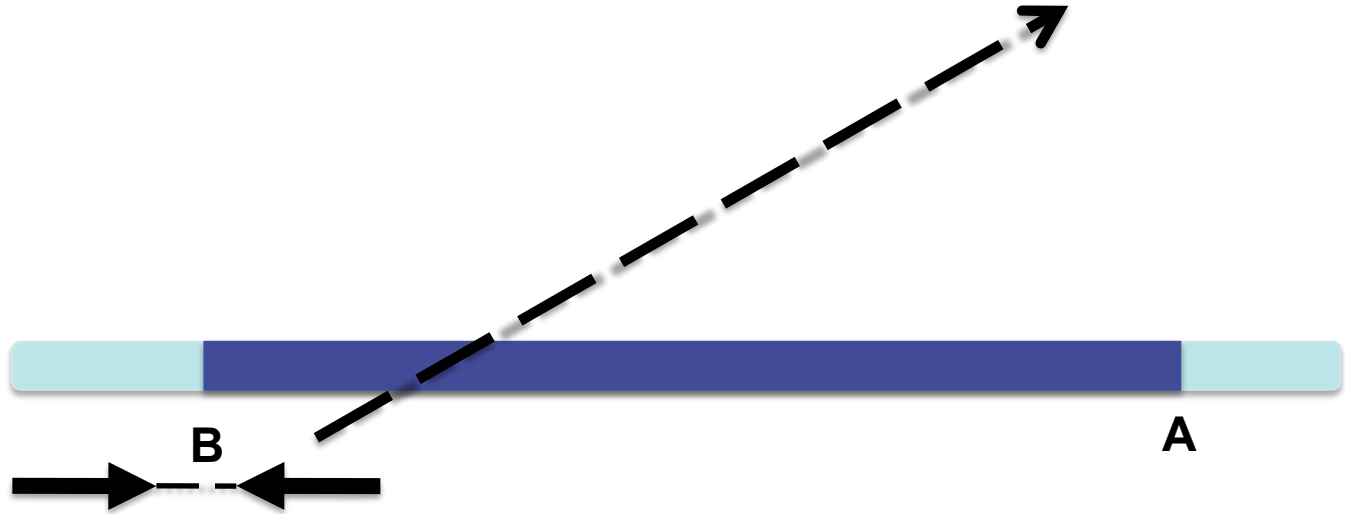


# Inversion

Reference  
Genome



Subject





# Inversion

Reference  
Genome



Subject



# Inversion

Reference  
Genome



Subject



# Inversion

Reference  
Genome



Subject



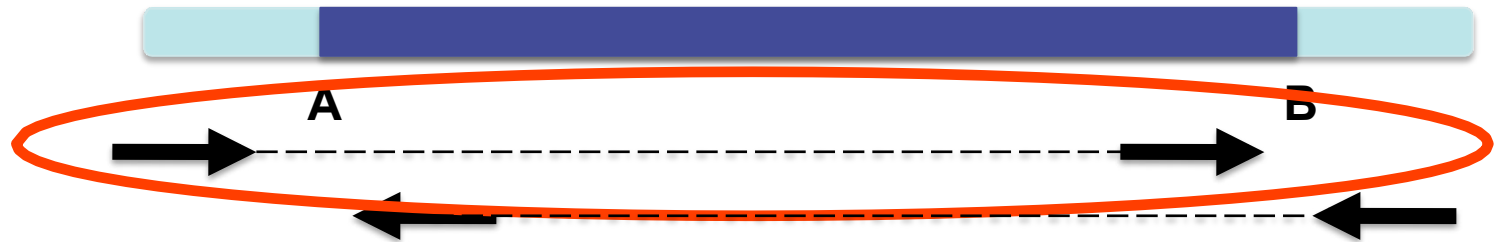
# Inversion


Reference  
Genome



# Inversion

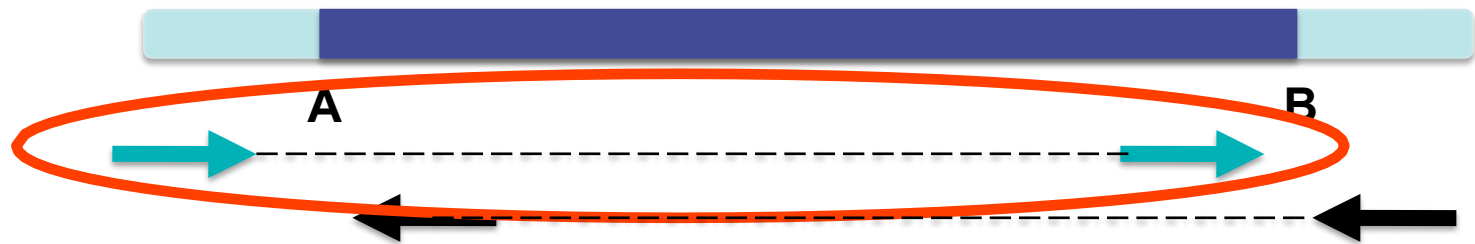
Reference  
Genome



Anomaly –  
Expected pair orientation is  
inward facing (  )

# Inversion

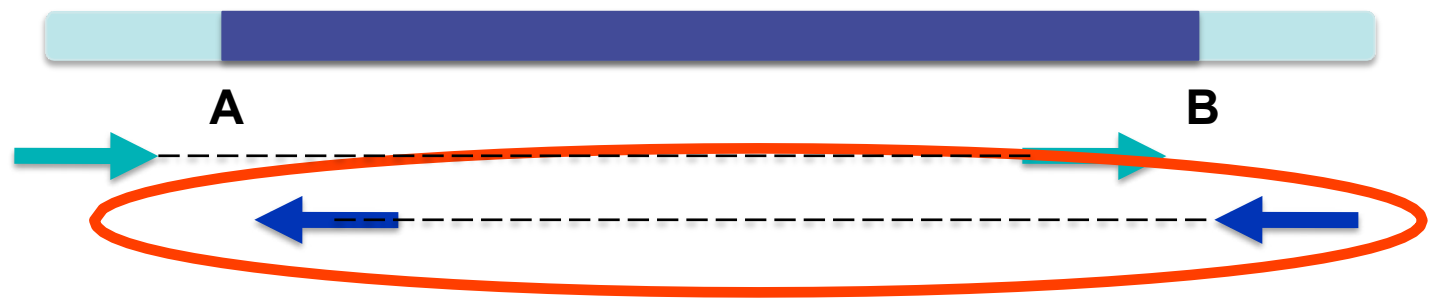
Reference  
Genome



“Left” side pair

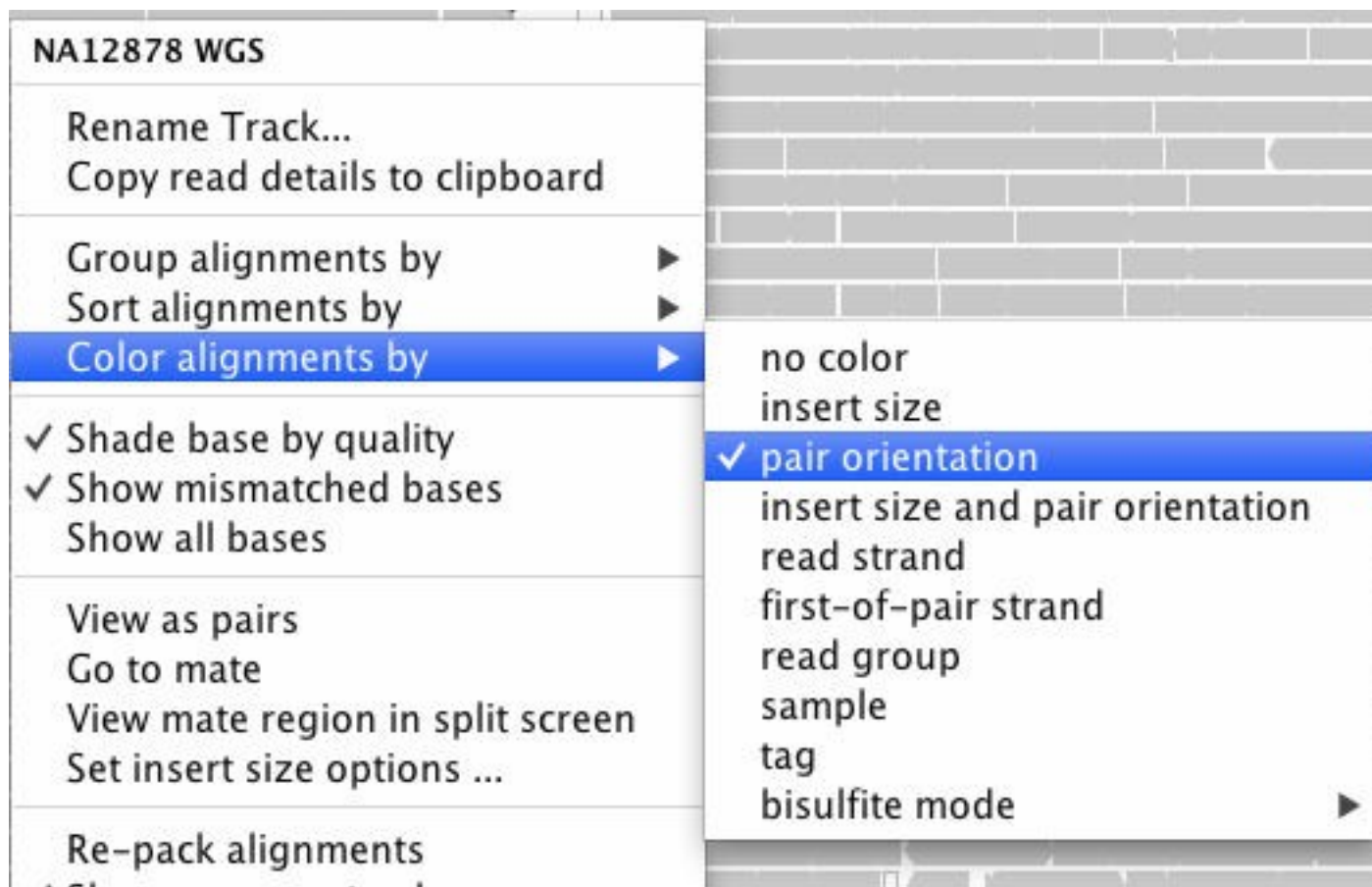
# Inversion

Reference  
Genome



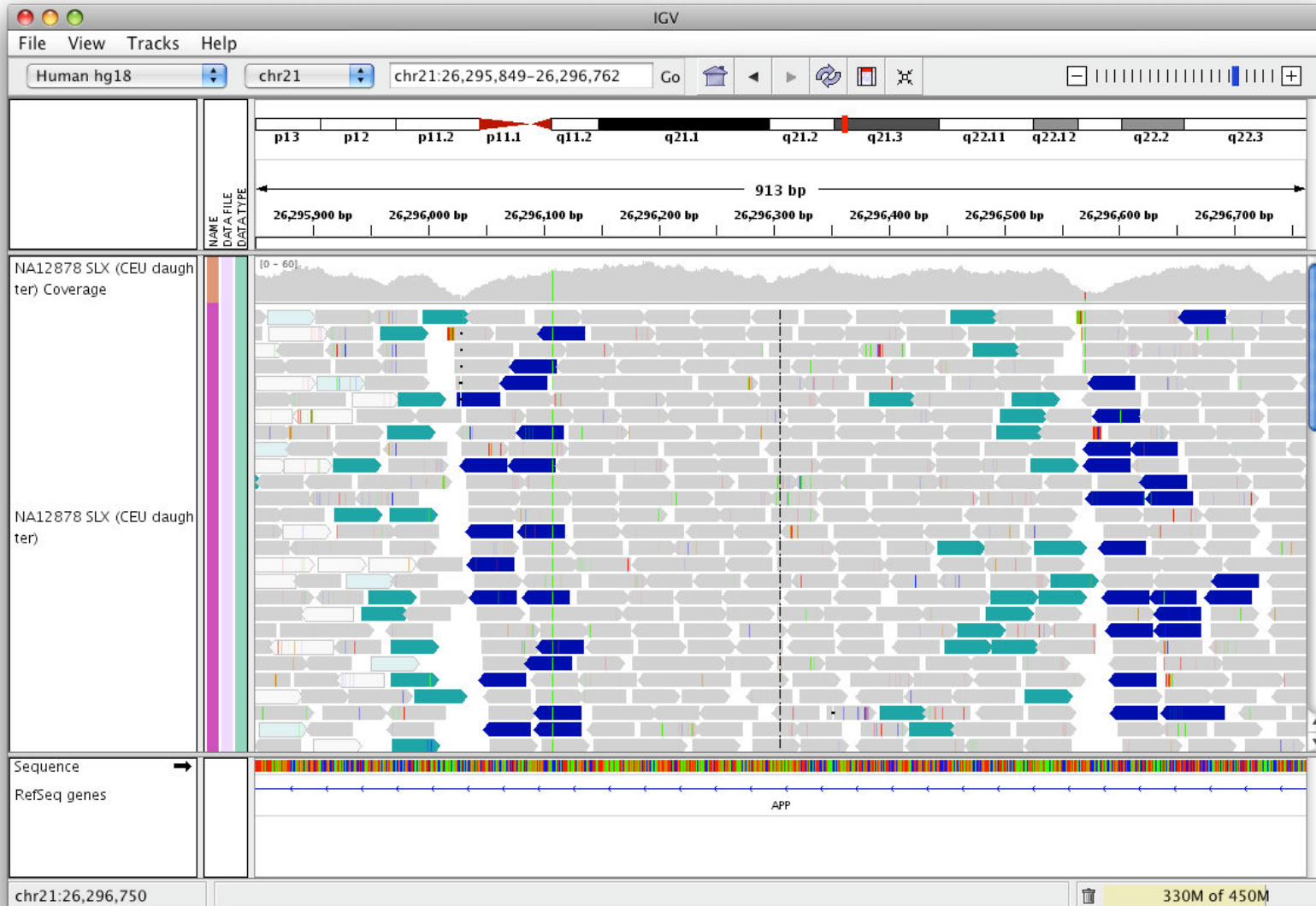
“Right” side pair

# Color by pair orientation



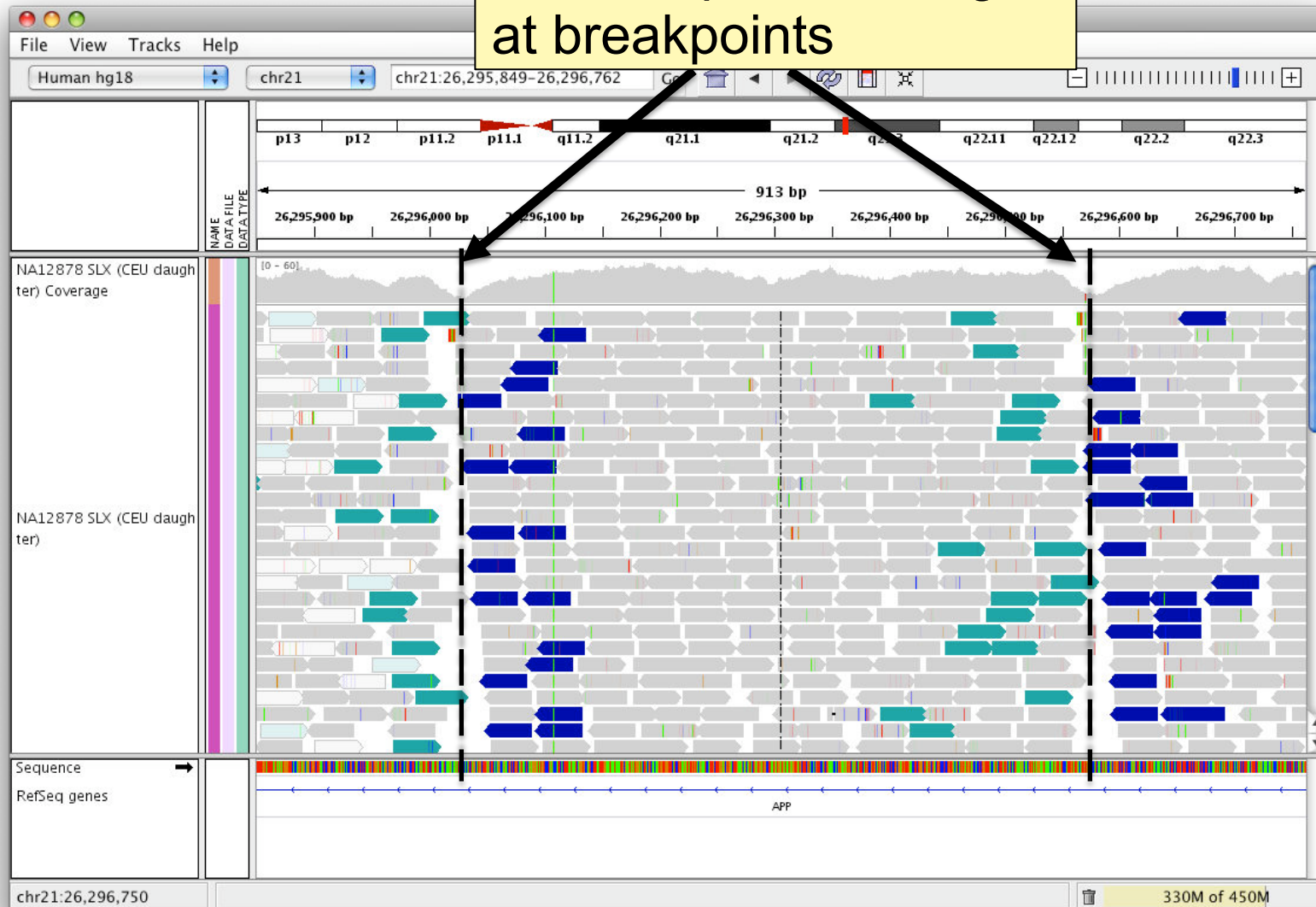


# Inversion

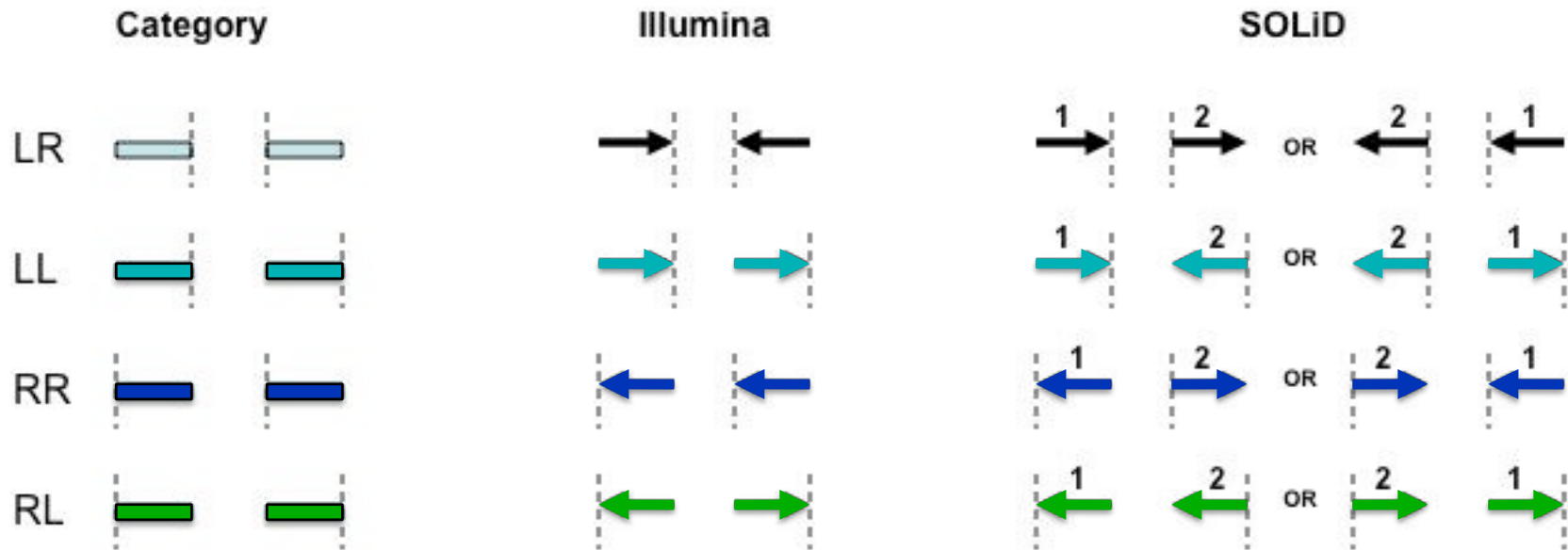


# Inversion

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*

# RNA-Seq

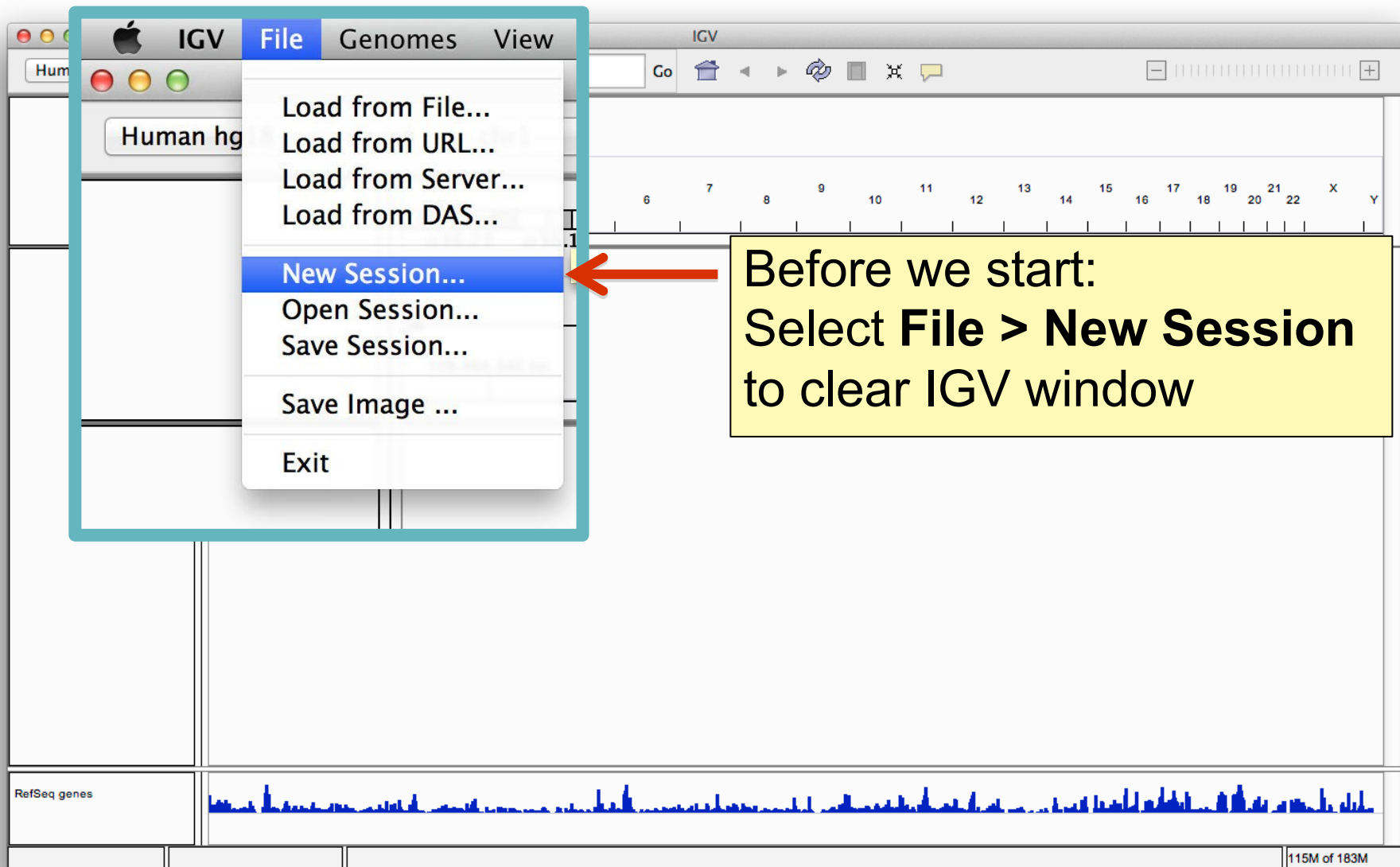
---



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

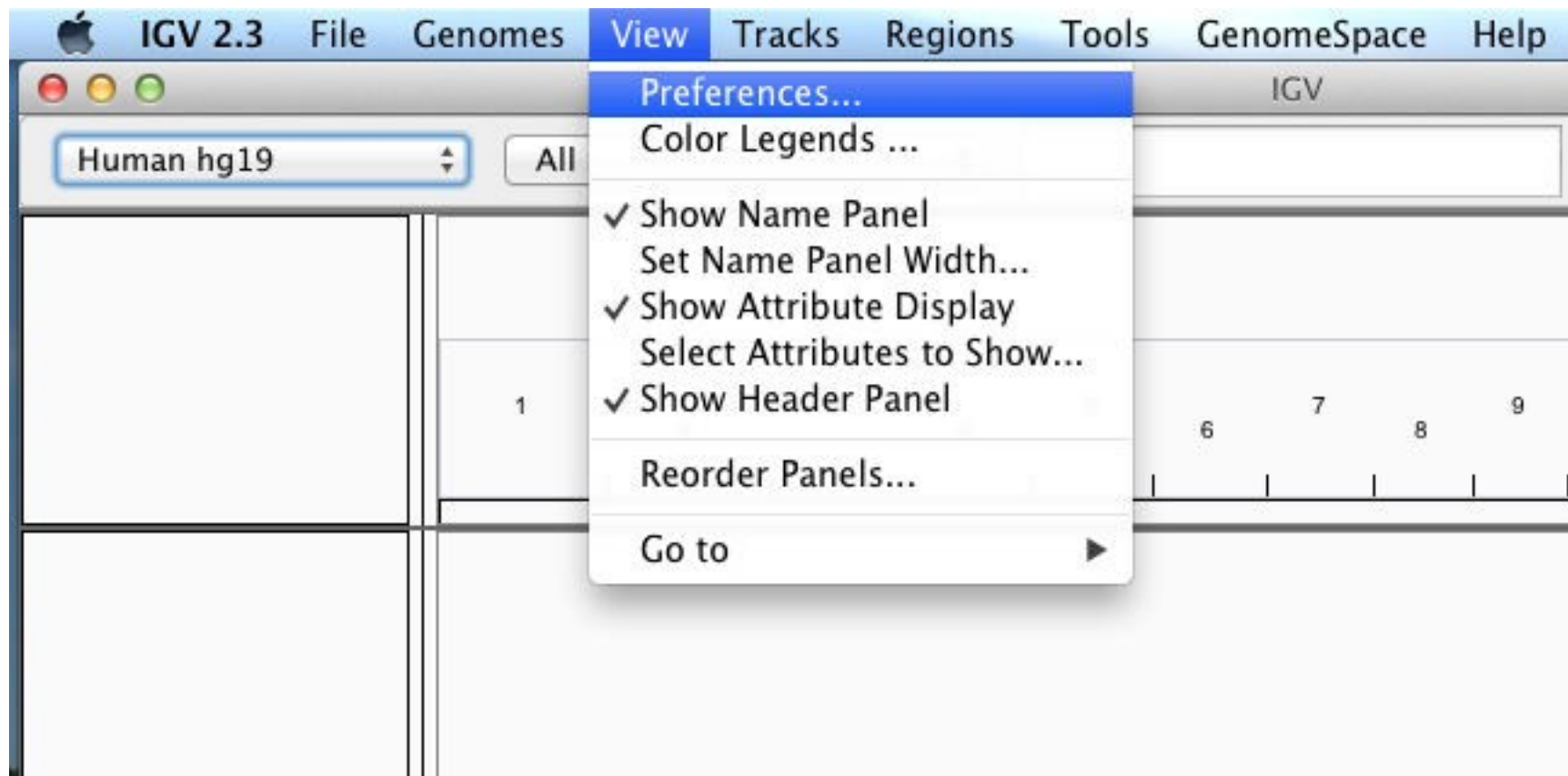
---



- Step 1: Tune settings for RNA.

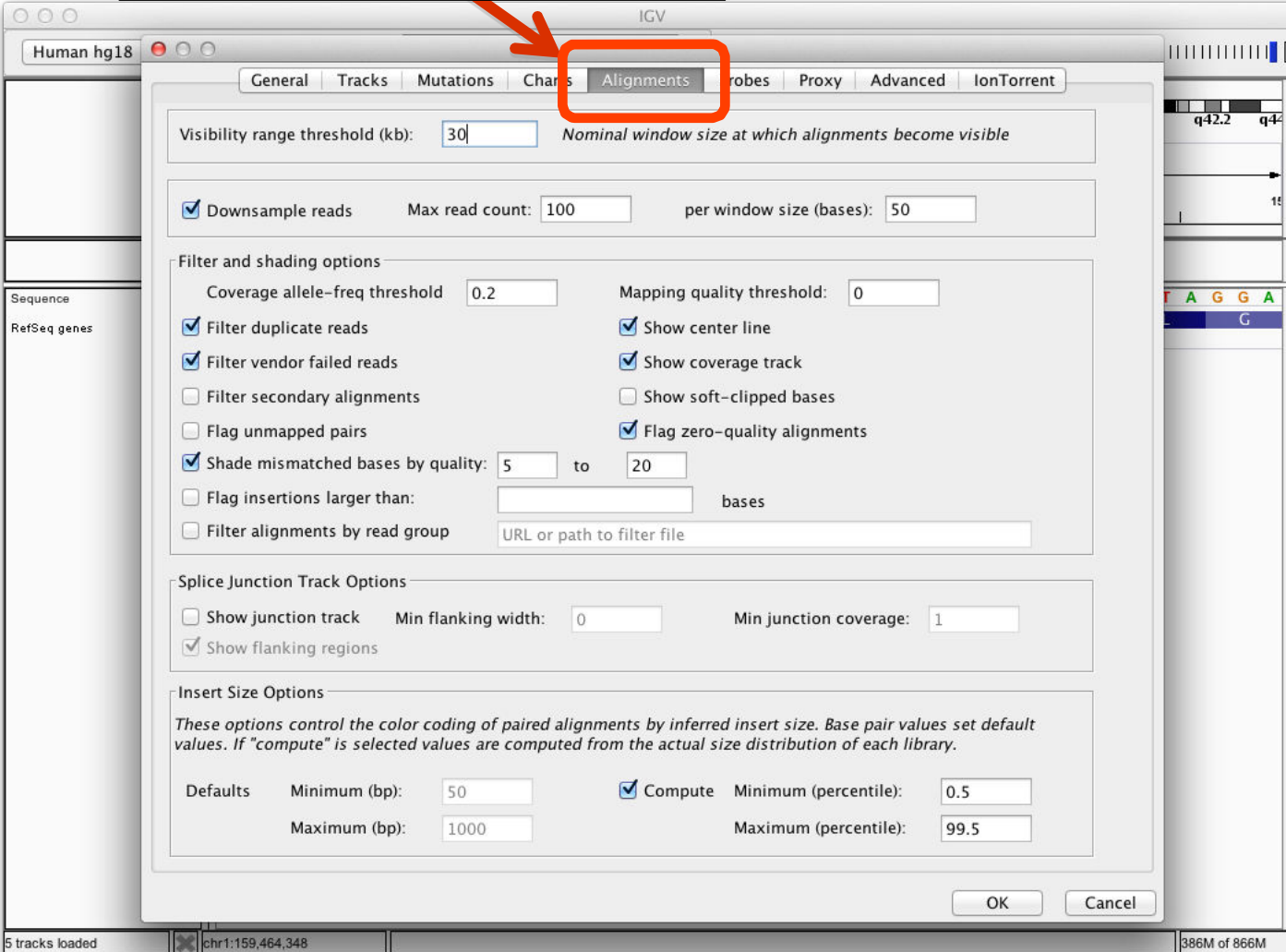
# RNA-seq alignments

Select **View > Preferences...**



# RNA-seq alignments

Click Alignments tab



The screenshot shows the IGV (Integrative Genomics Viewer) interface with the 'Alignments' tab selected. A yellow box highlights the 'Alignments' tab, and a red arrow points to it. The dialog box contains the following settings:

- Visibility range threshold (kb): 30 (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
  - 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
  - Show center line
  - Show coverage track
  - Show soft-clipped bases
  - Flag zero-quality alignments
- Splice Junction Track Options:
  - Show junction track: Min flanking width: 0, Min junction coverage: 1
  - Show flanking regions
- Insert Size Options:

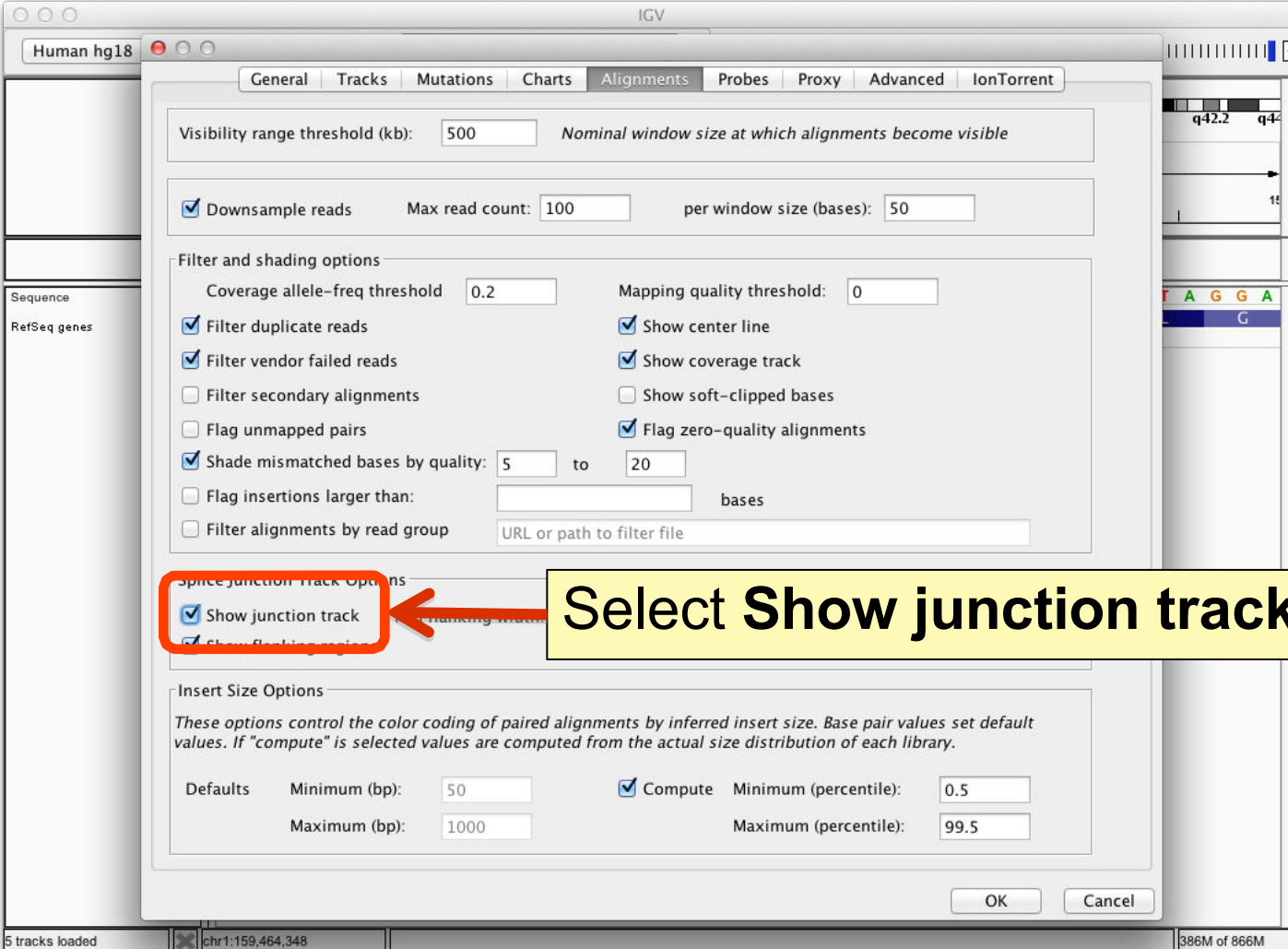
*These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.*

Defaults	Minimum (bp):	Maximum (bp):	<input checked="" type="checkbox"/> Compute	Minimum (percentile):	Maximum (percentile):
	50	1000		0.5	99.5

Buttons: OK, Cancel



# RNA-seq alignments



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

Show flanking region

Insert Size Options

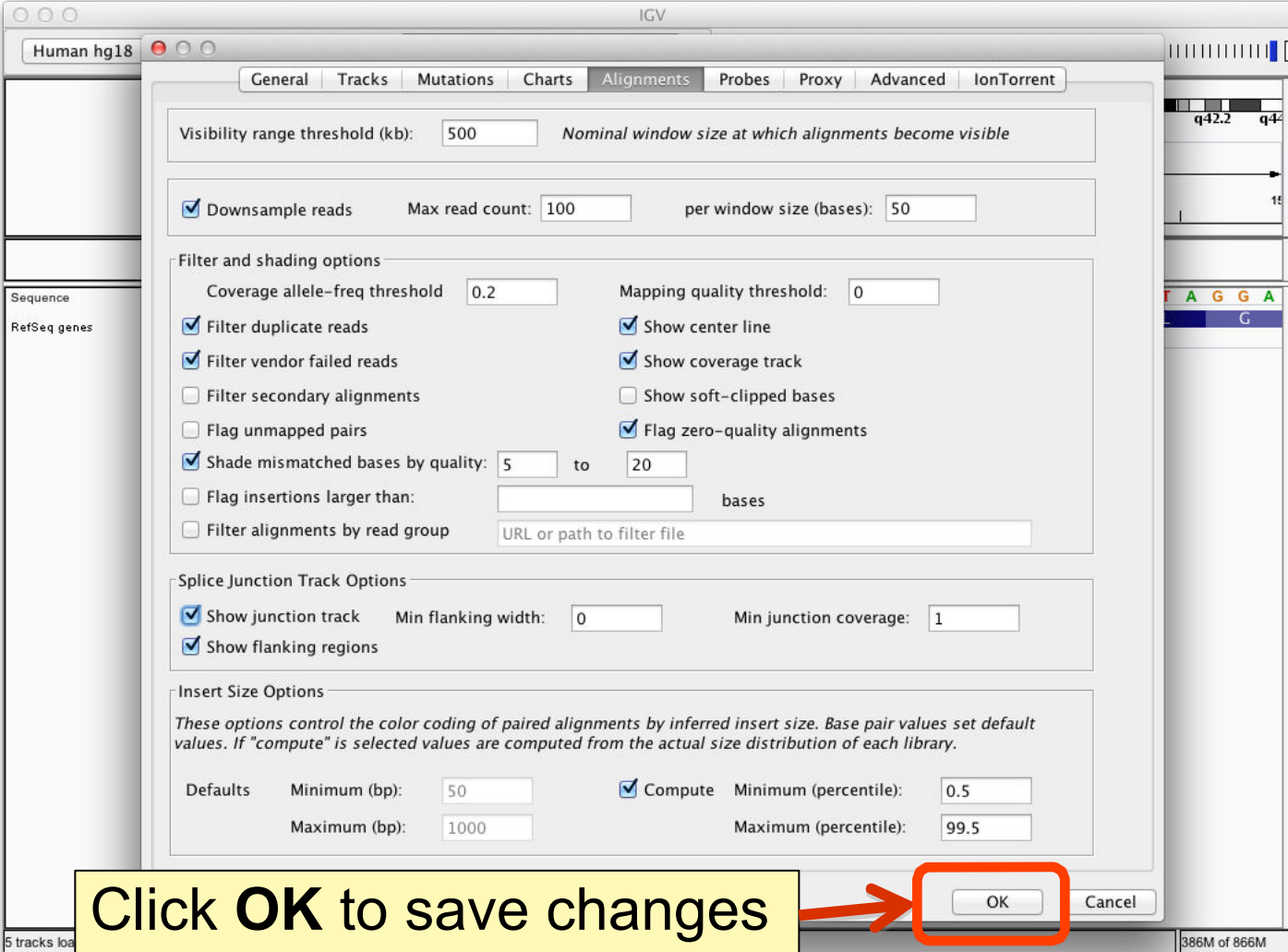
*These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.*

Defaults	Minimum (bp):	50	<input checked="" type="checkbox"/> Compute	Minimum (percentile):	0.5
	Maximum (bp):	1000		Maximum (percentile):	99.5

OK Cancel

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

# RNA-seq alignments



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

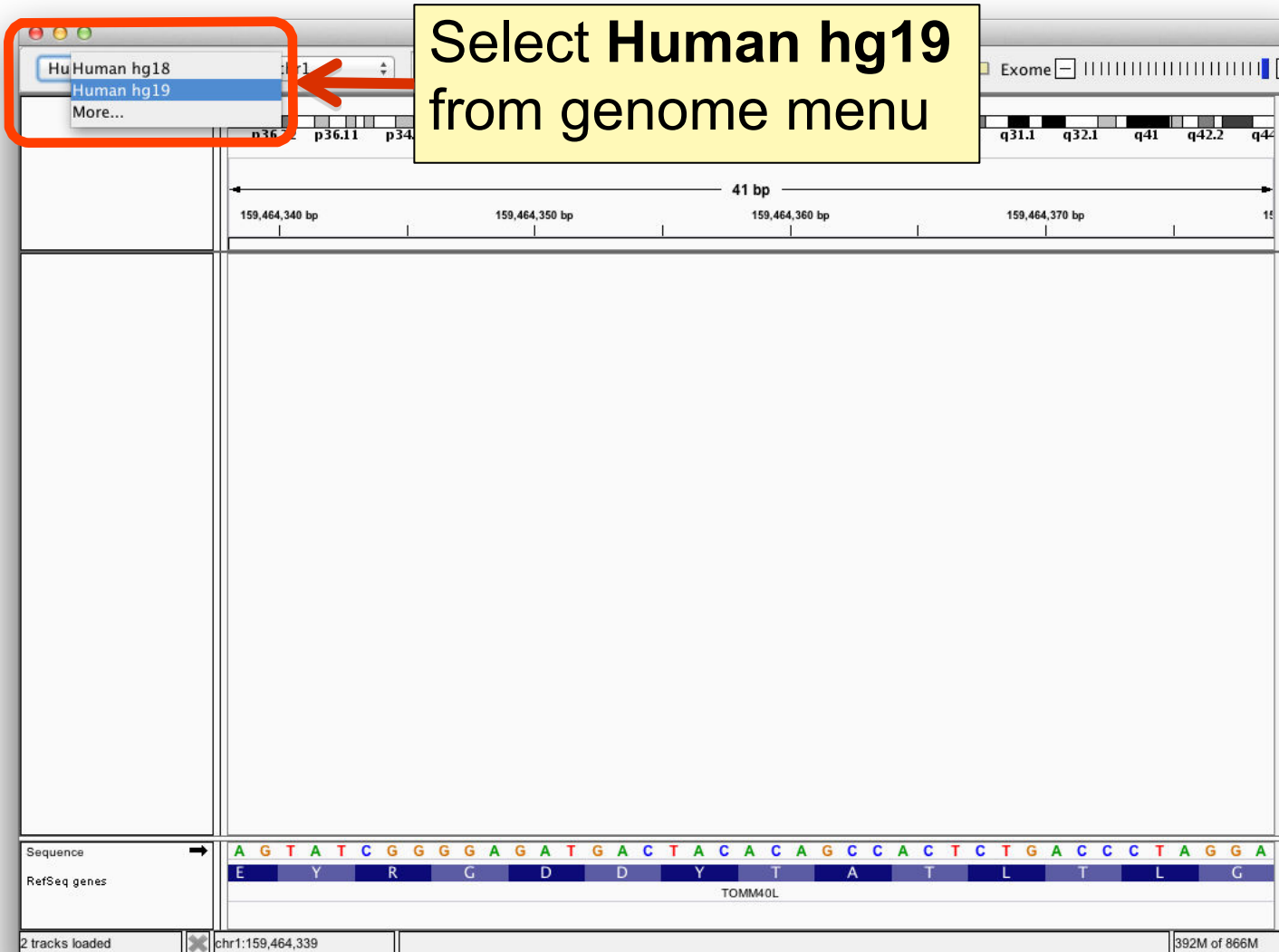
*These options control the color coding of paired alignments by inferred insert size. Base pair values set default values. If "compute" is selected values are computed from the actual size distribution of each library.*

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

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

Click **OK** to save changes →

# RNA-seq alignments

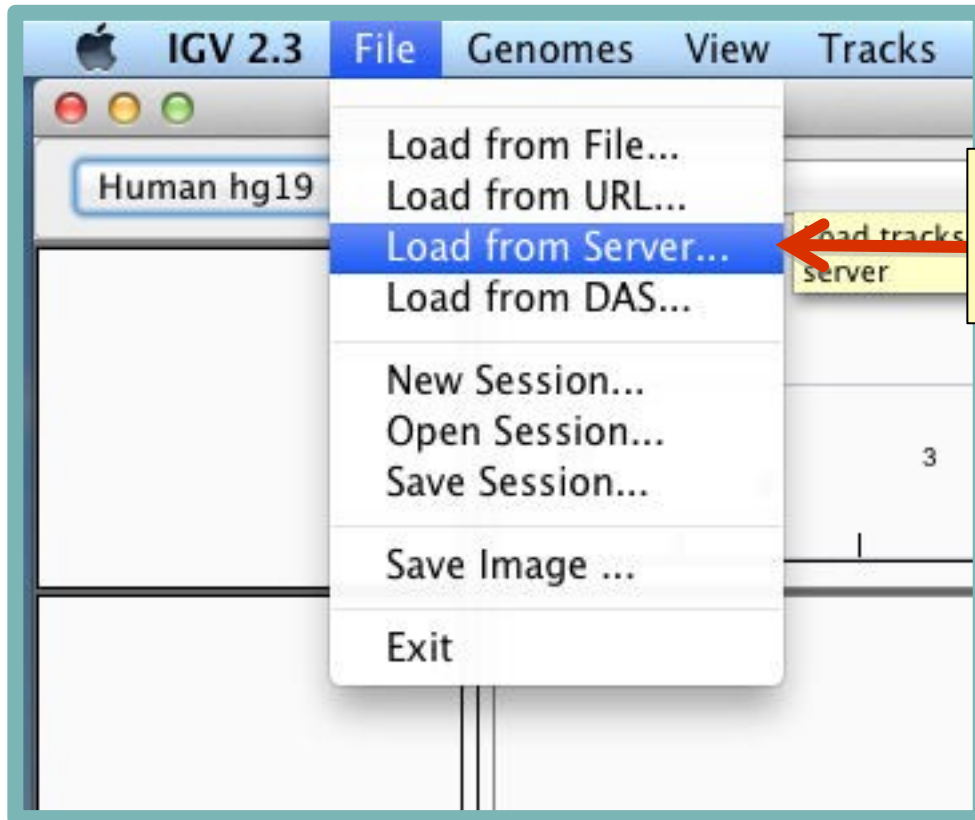


Select Human hg19 from genome menu

Sequence → A G T A T C G 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  
RefSeq genes E Y R G D D Y T A T L T L G  
TOMM40L

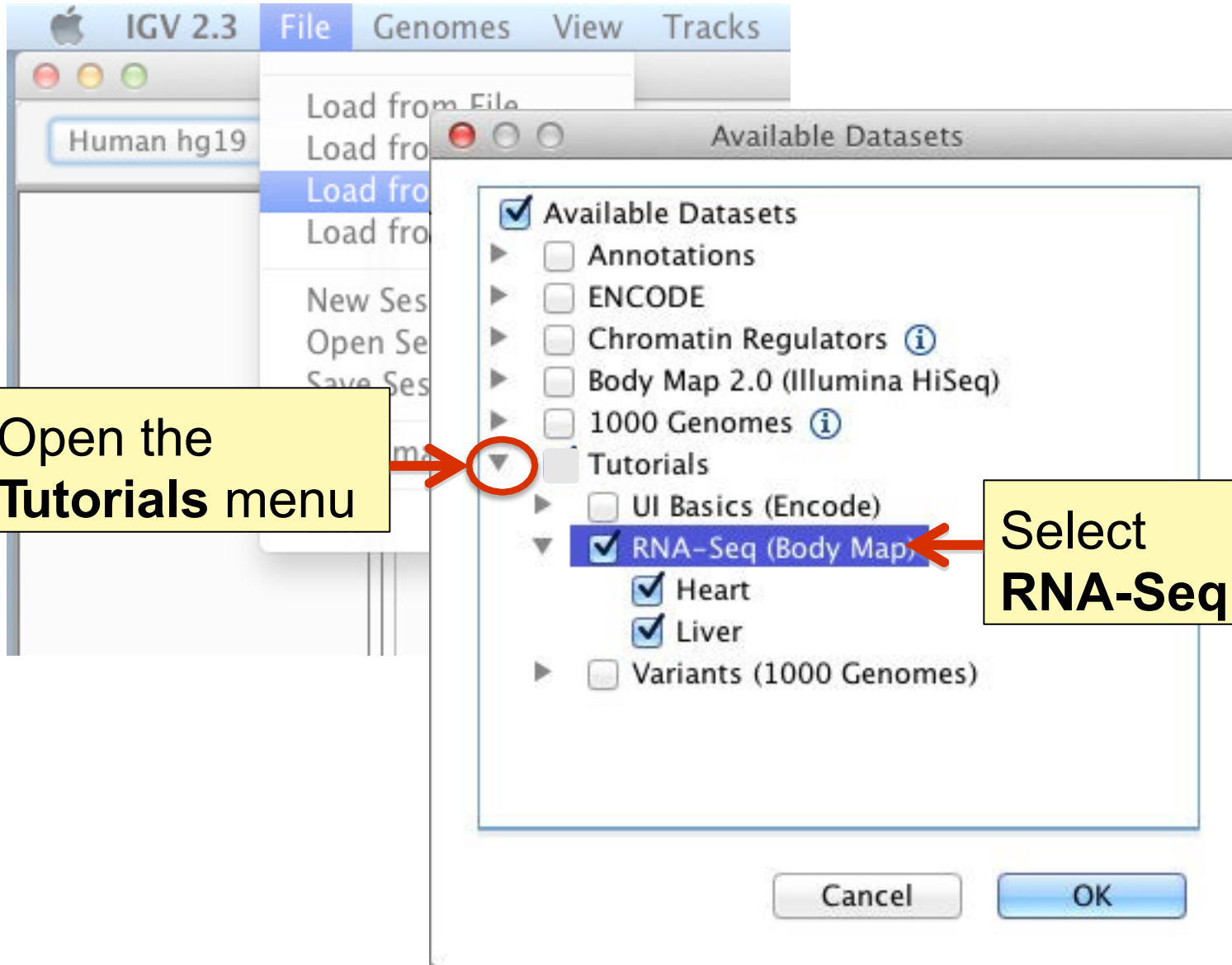
2 tracks loaded | chr1:159,464,339 | 392M of 866M

# RNA-seq alignments



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

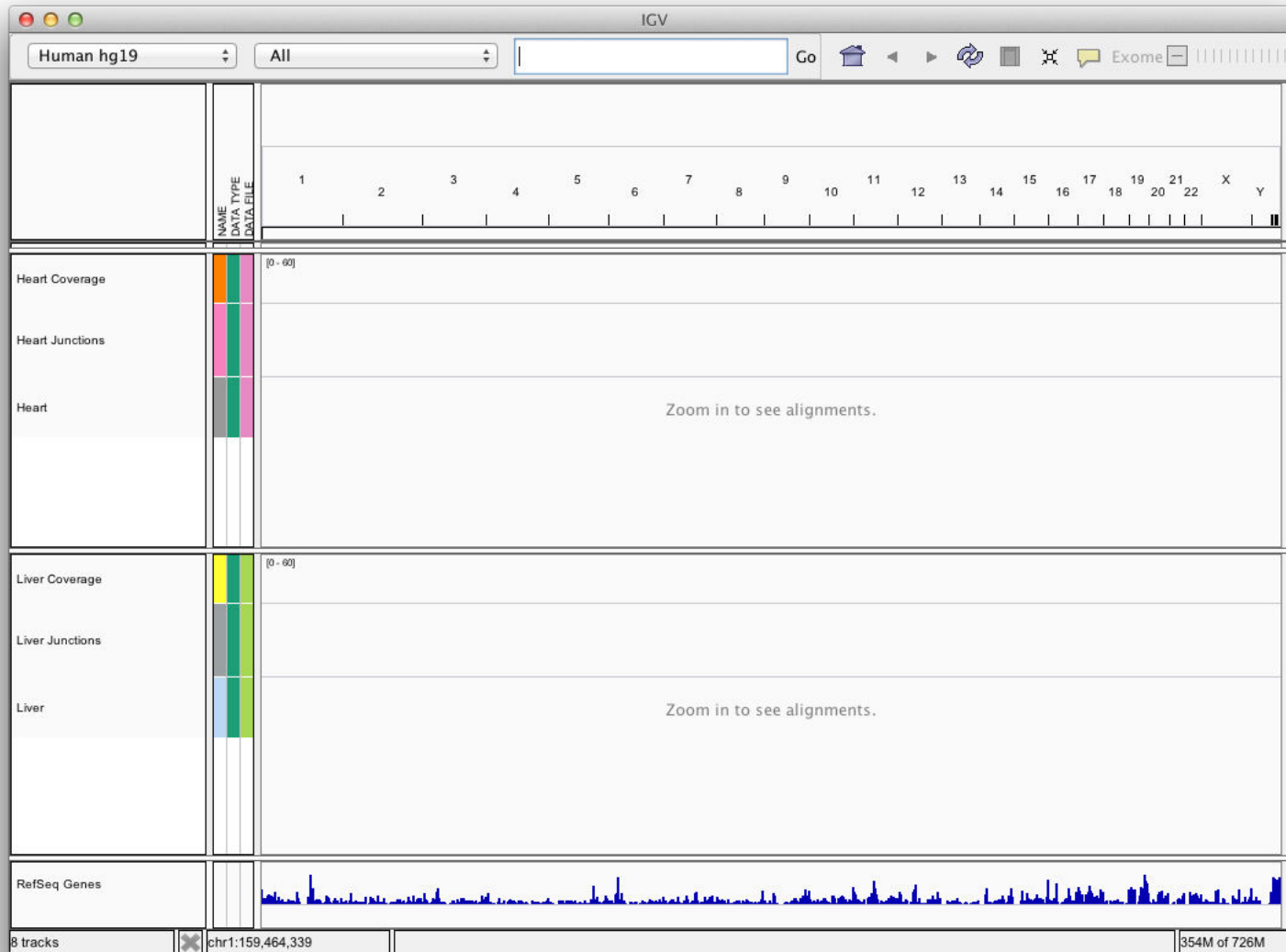
# RNA-seq alignments



Open the  
**Tutorials** menu

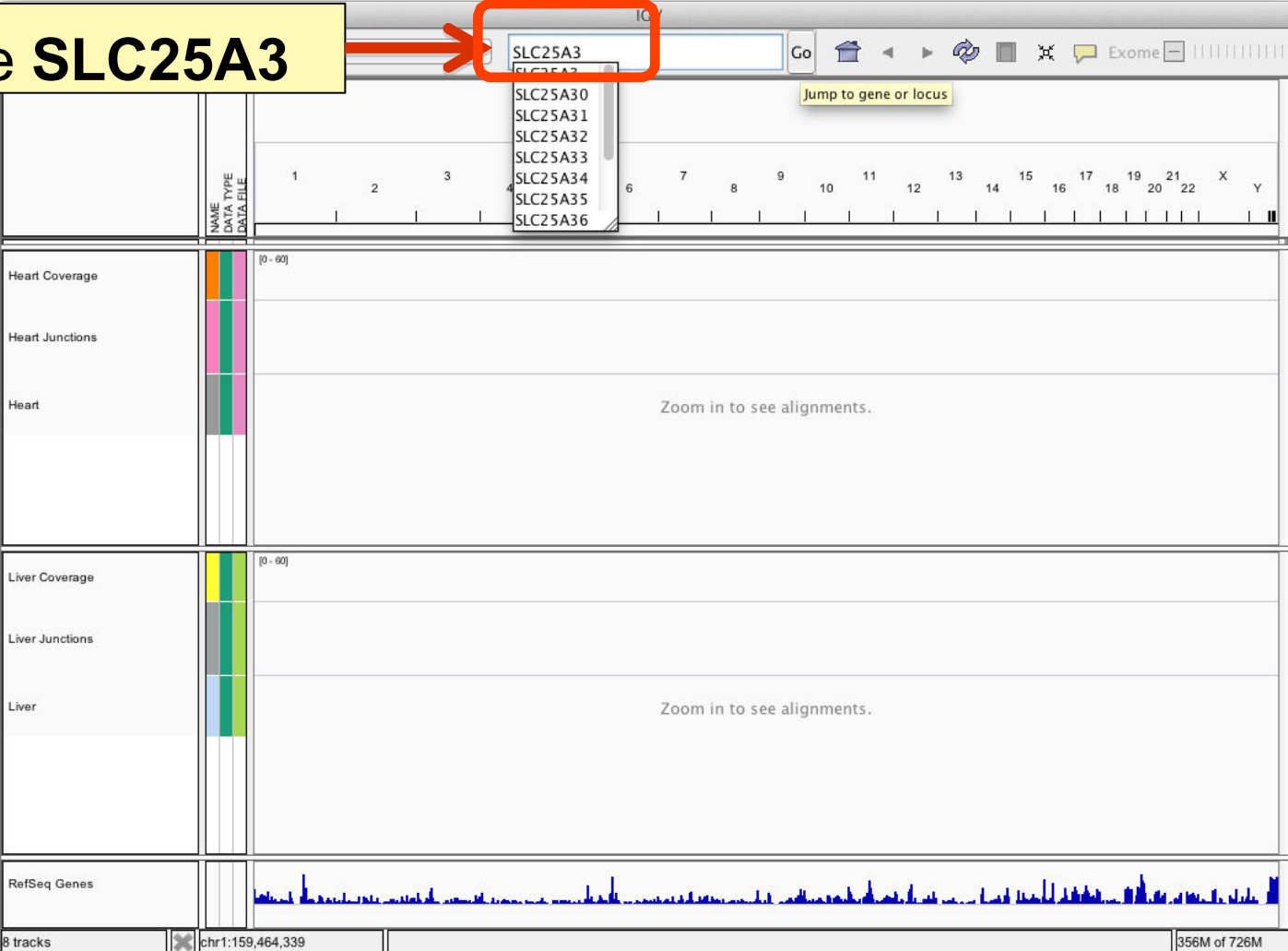
Select  
**RNA-Seq (Body Map)**

# RNA-seq alignments



# RNA-seq alignments

Type **SLC25A3** →



SLC25A3  
SLC25A30  
SLC25A31  
SLC25A32  
SLC25A33  
SLC25A34  
SLC25A35  
SLC25A36

Jump to gene or locus

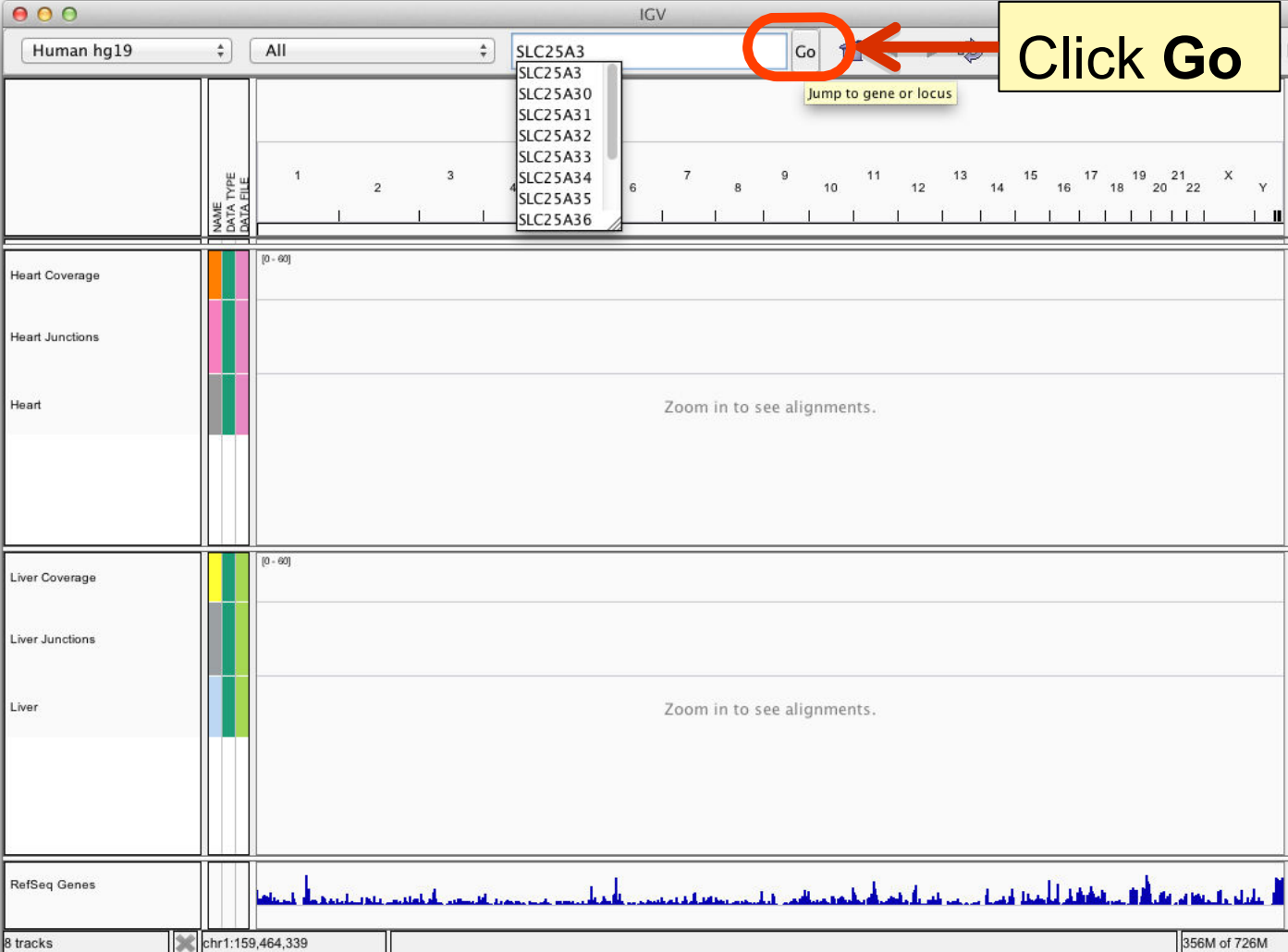
Heart Coverage  
Heart Junctions  
Heart

Liver Coverage  
Liver Junctions  
Liver

RefSeq Genes

8 tracks | chr1:159,464,339 | 356M of 726M

# RNA-seq alignments



Human hg19 All SLC25A3 SLC25A3 SLC25A30 SLC25A31 SLC25A32 SLC25A33 SLC25A34 SLC25A35 SLC25A36

Go Jump to gene or locus

Click Go

Heart Coverage [0 - 60]

Heart Junctions

Heart

Liver Coverage [0 - 60]

Liver Junctions

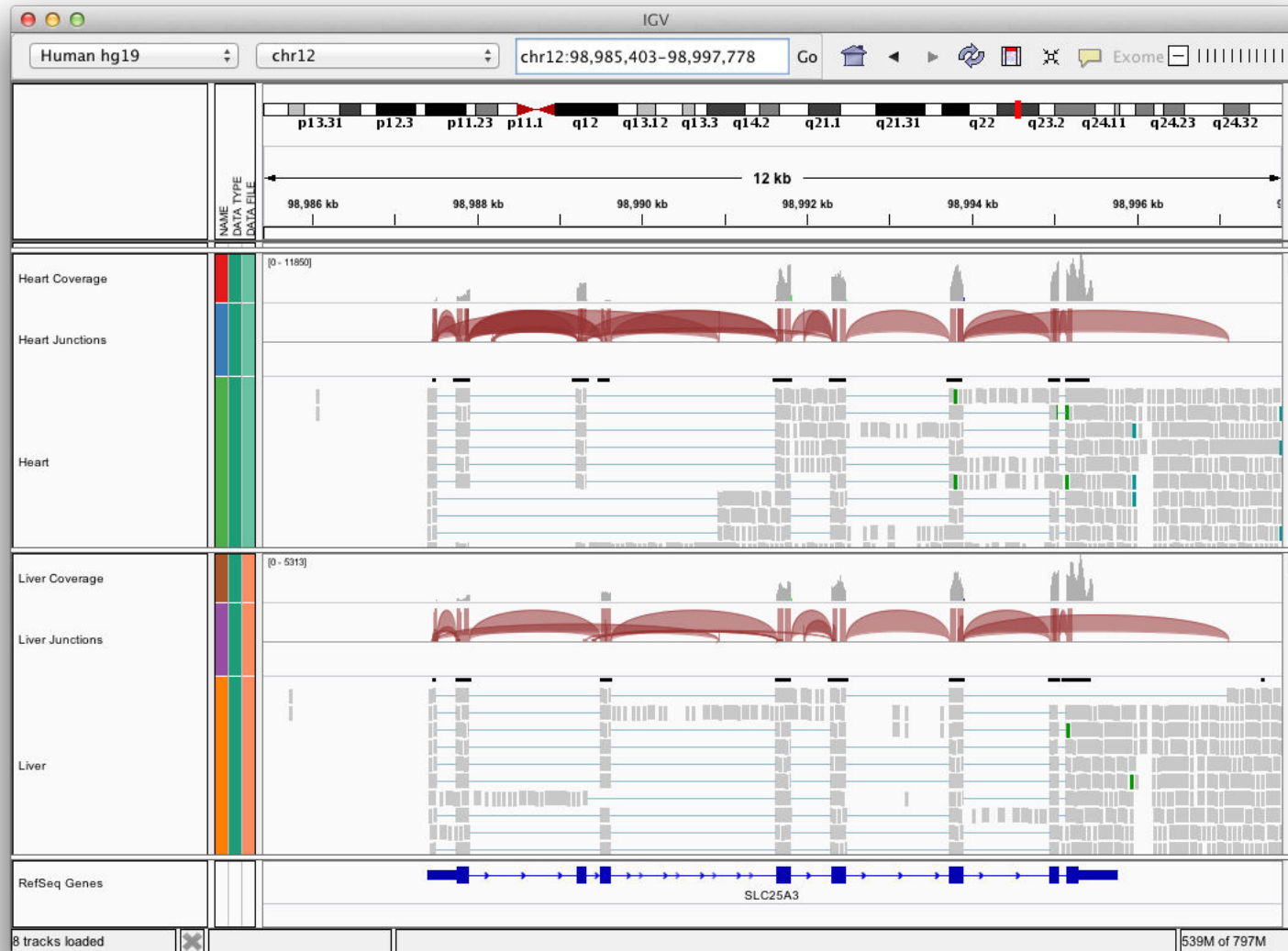
Liver

RefSeq Genes

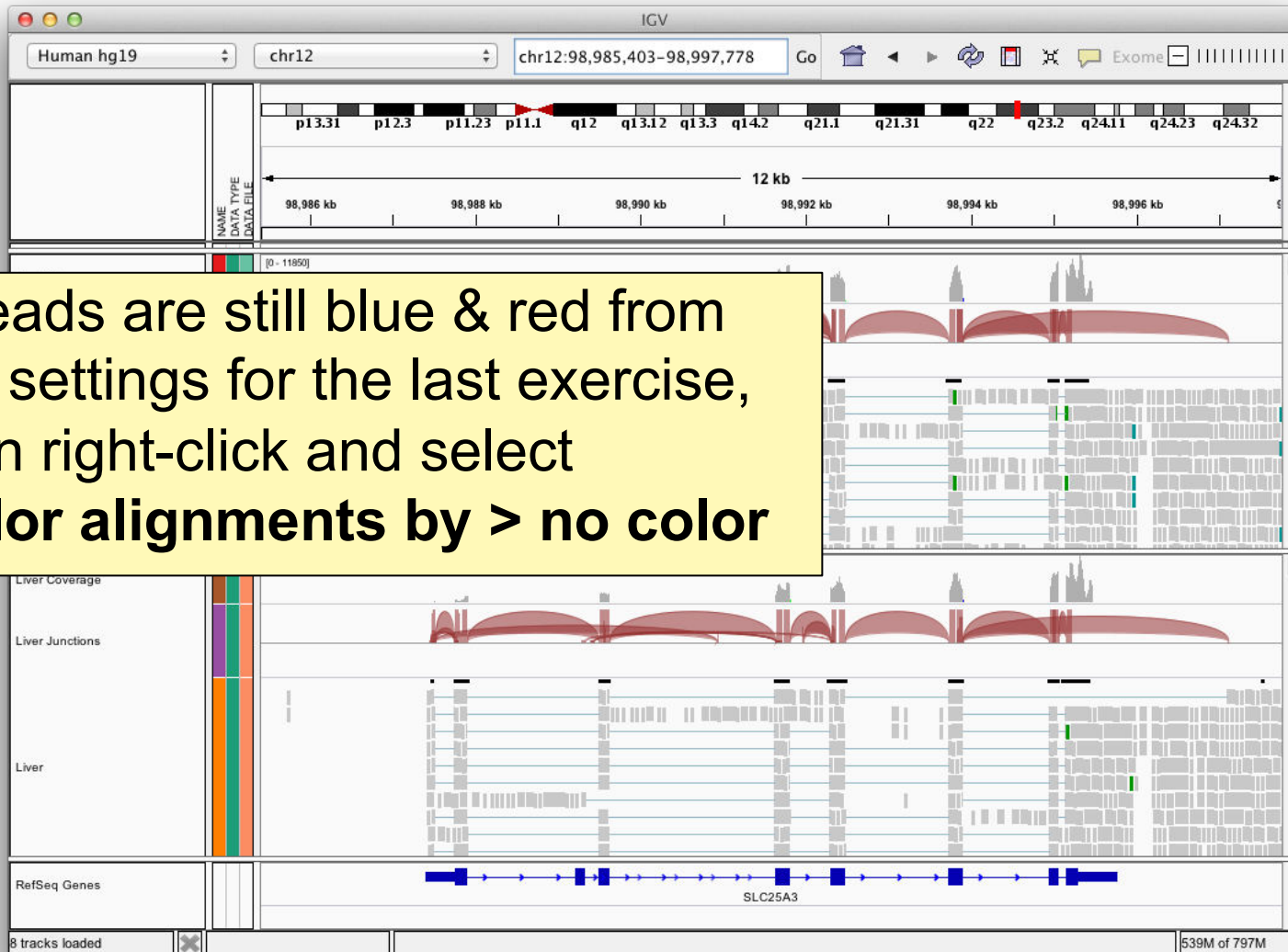
8 tracks chr1:159,464,339 356M of 726M



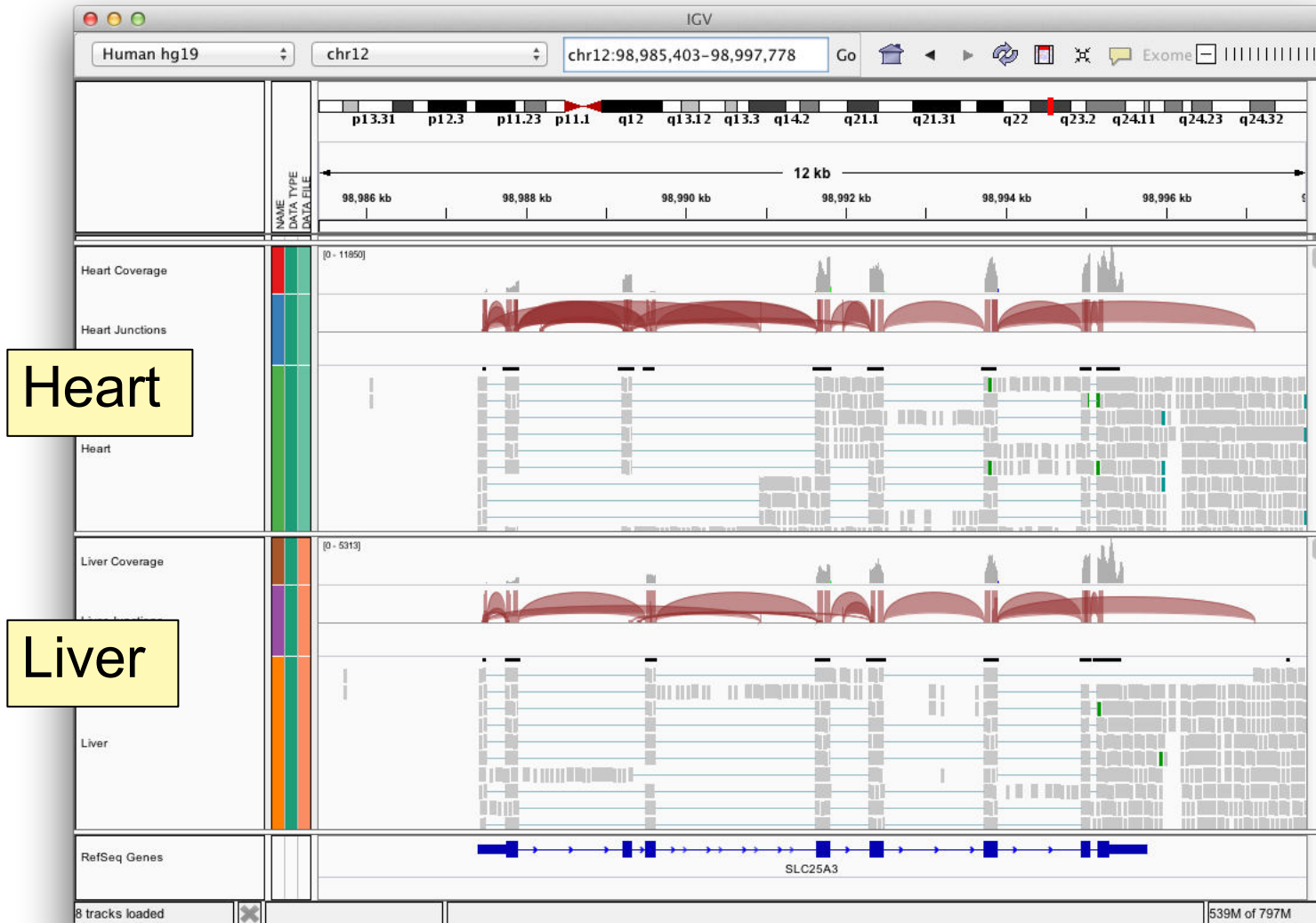
# RNA-seq alignments



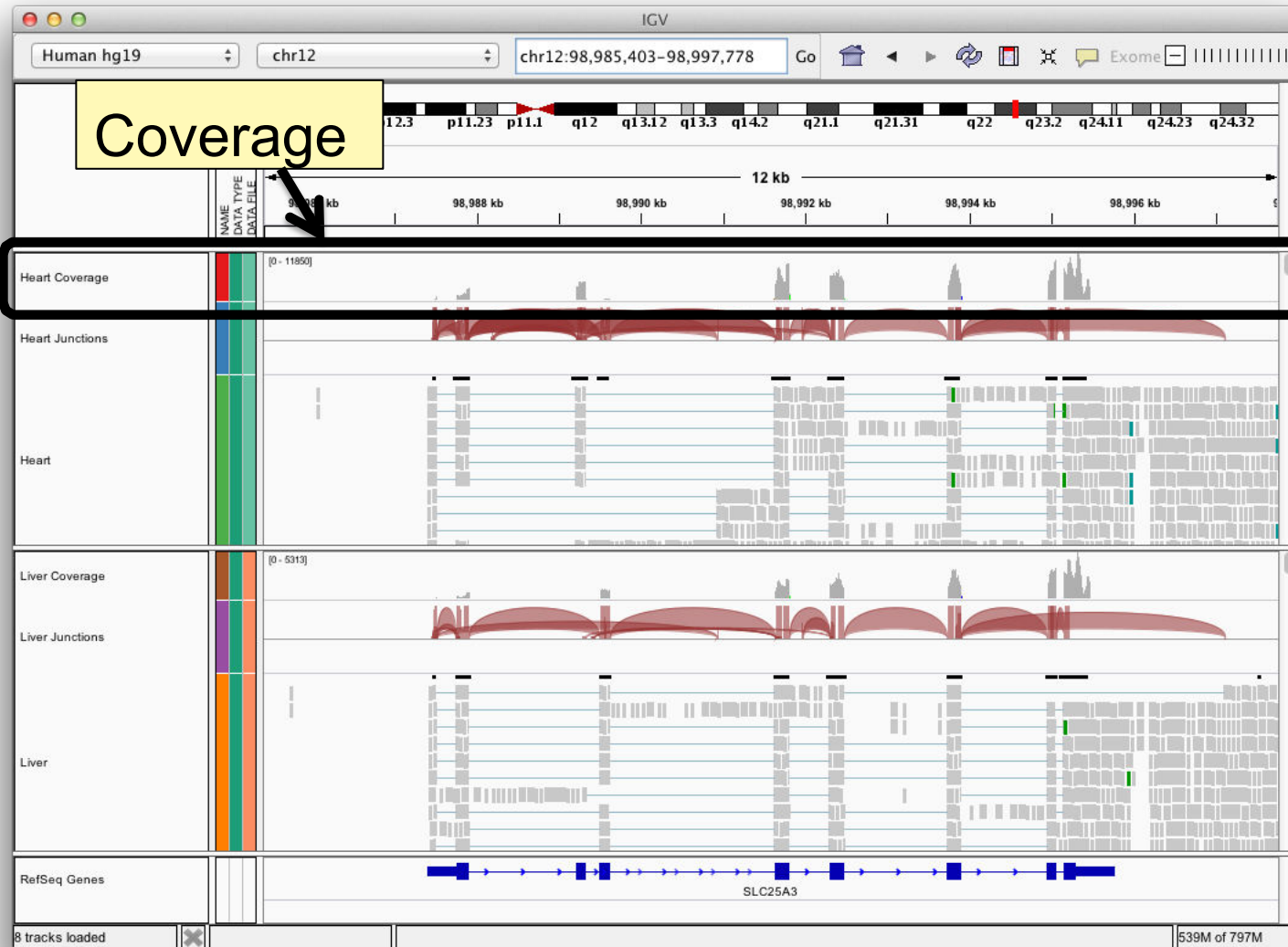
# RNA-seq alignments



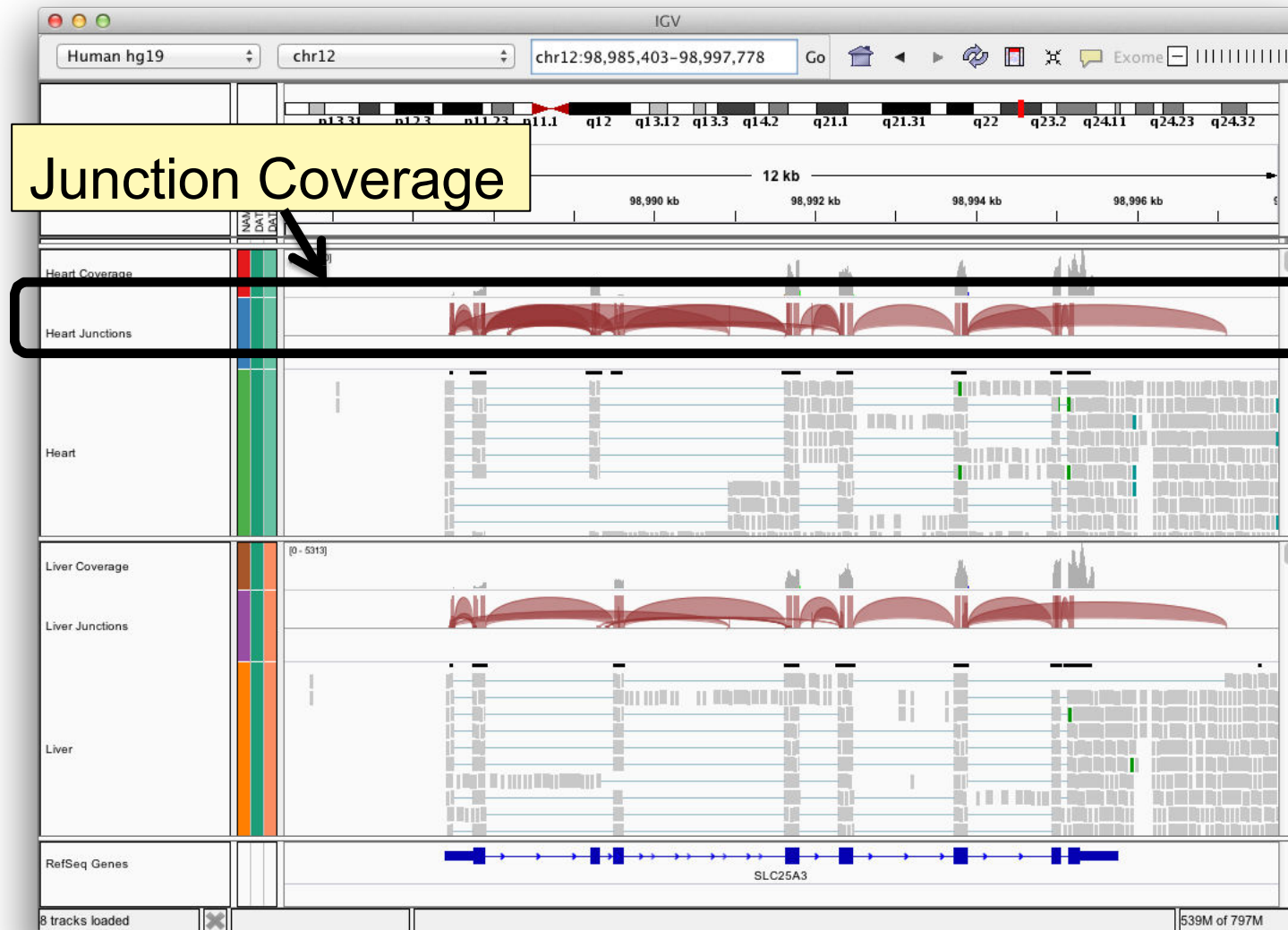
# RNA-seq alignments



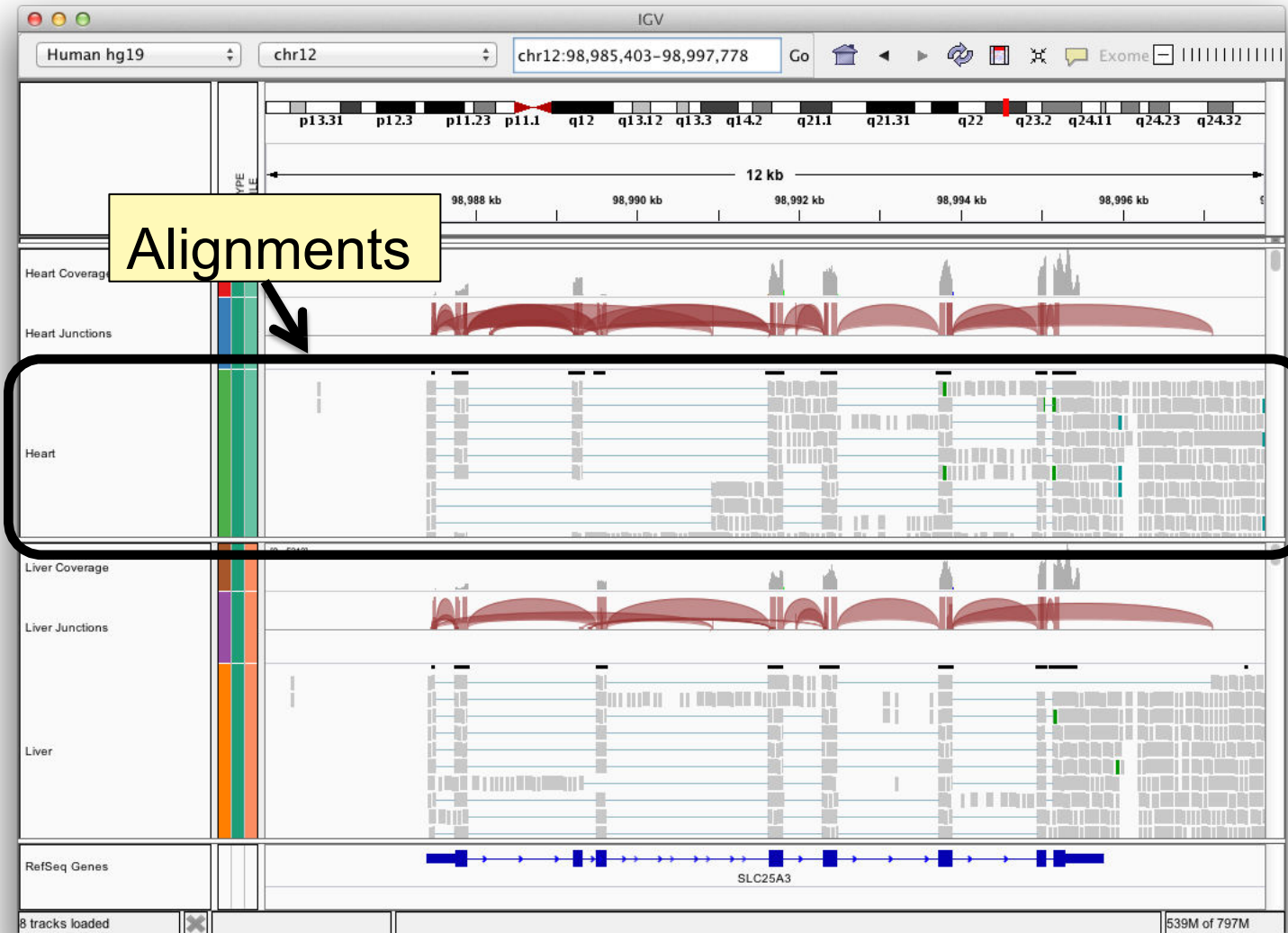
# RNA-seq alignments



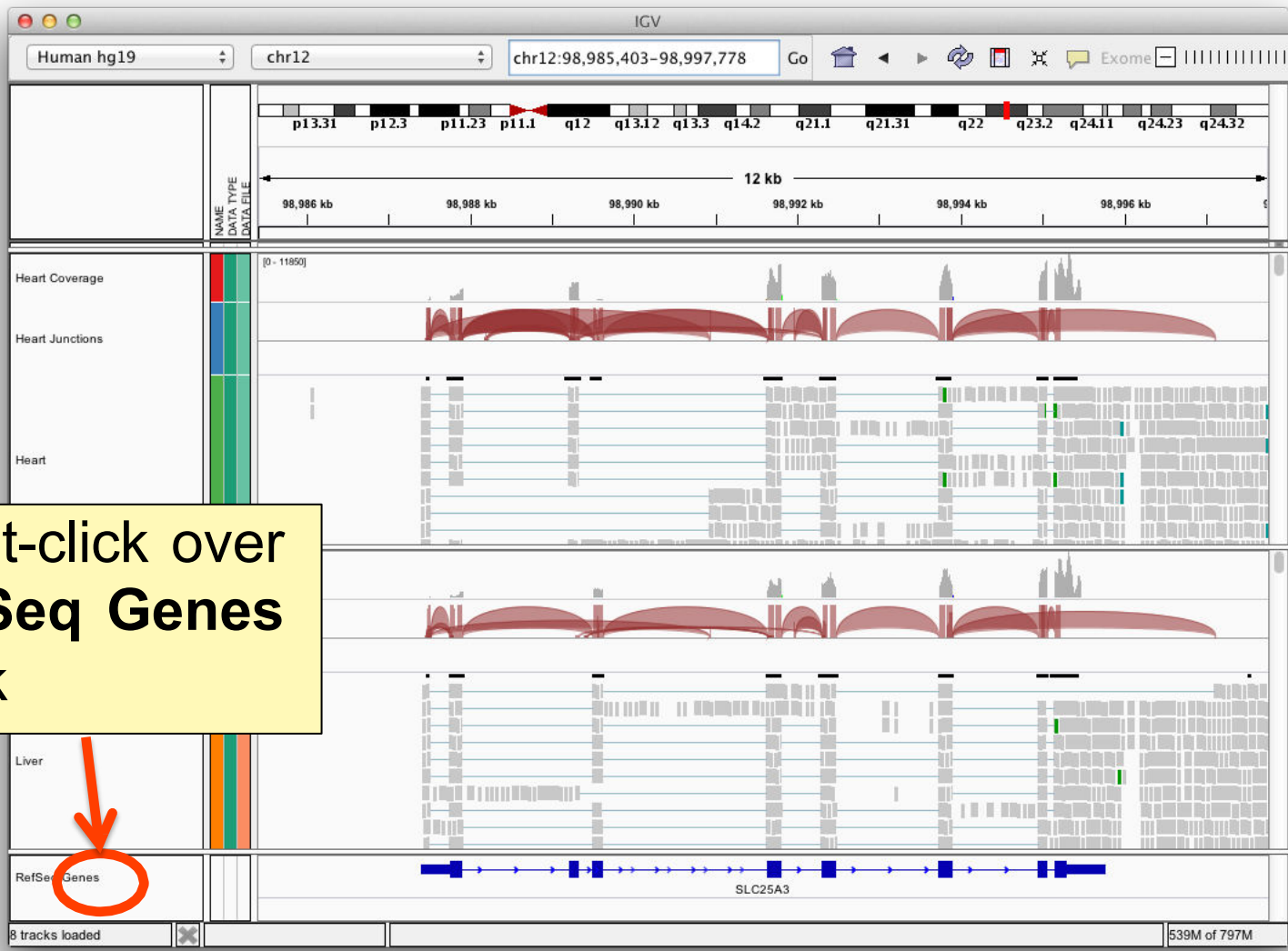
# RNA-seq alignments



# RNA-seq alignments

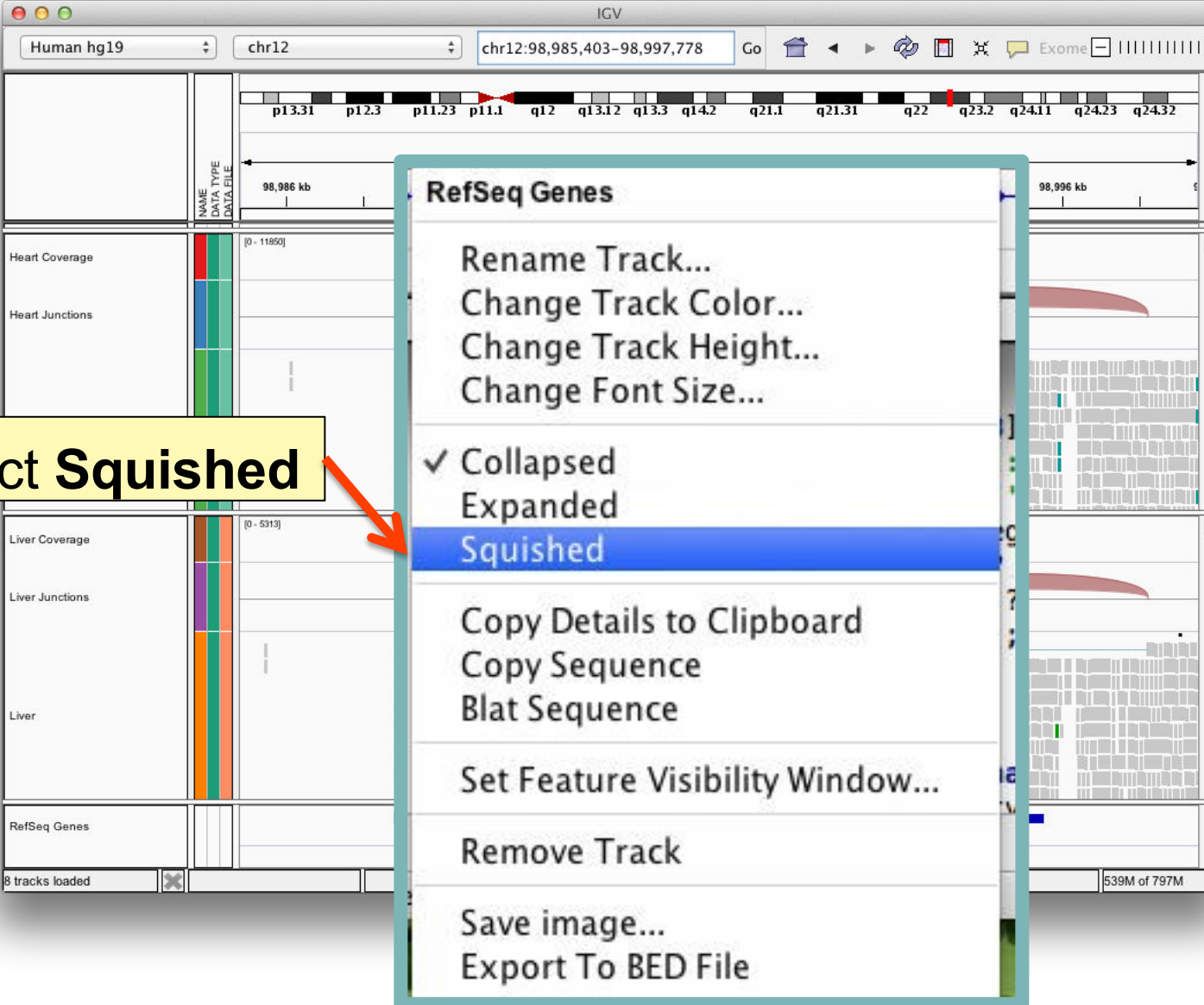


# RNA-seq alignments



Right-click over  
**RefSeq Genes**  
track

# RNA-seq alignments



Human hg19 chr12 chr12:98,985,403-98,997,778

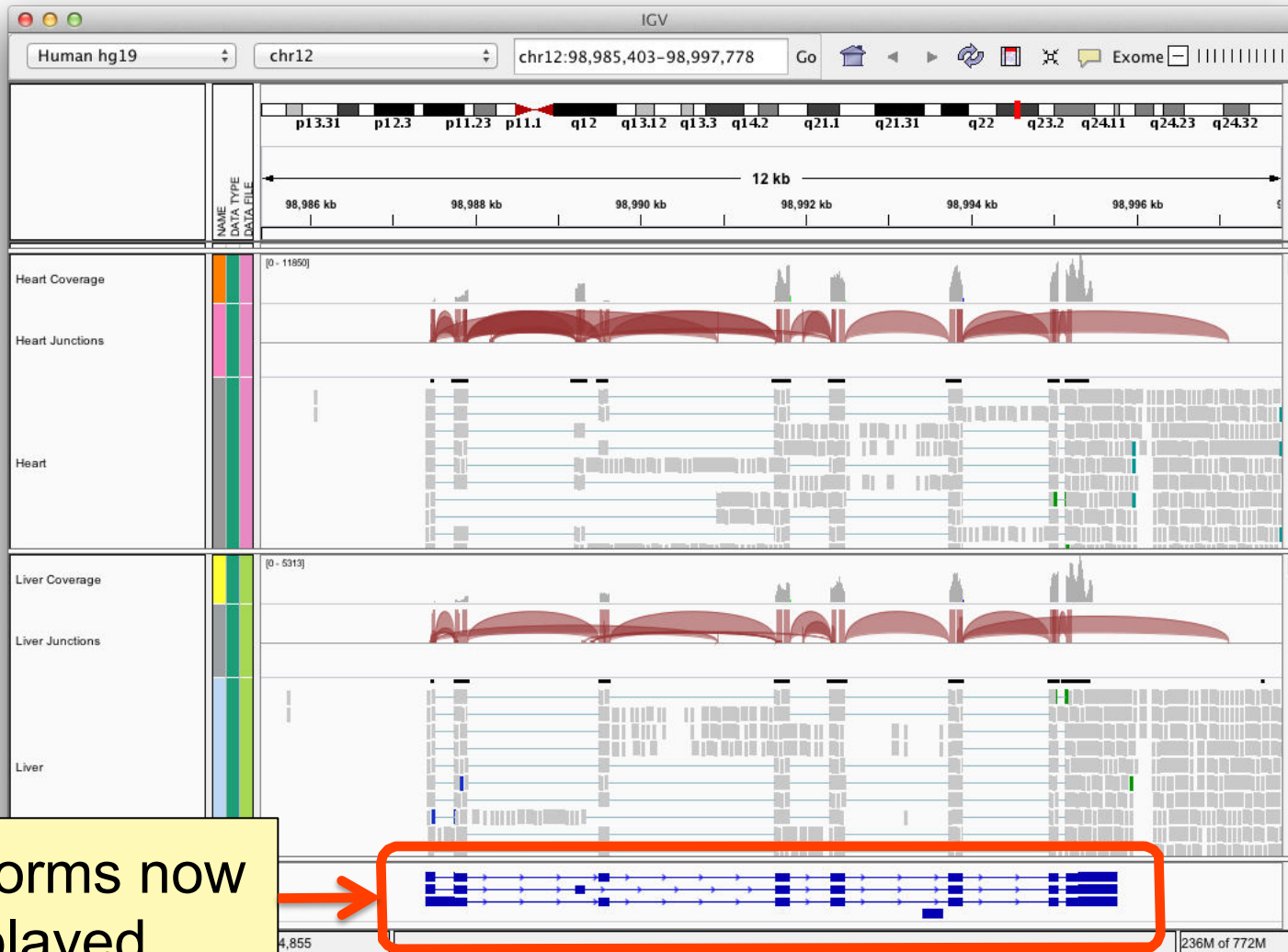
RefSeq Genes

- 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

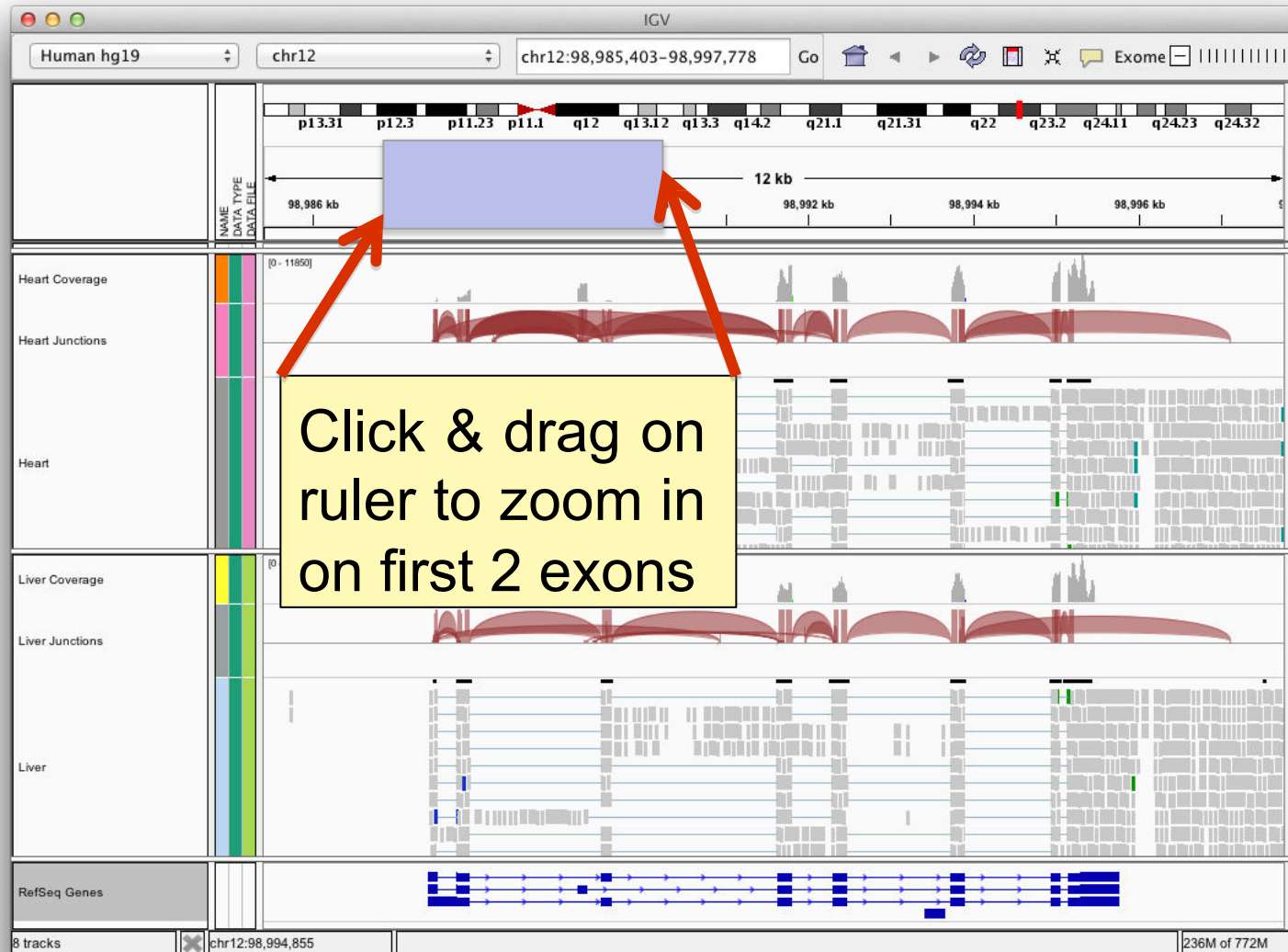
Select Squished



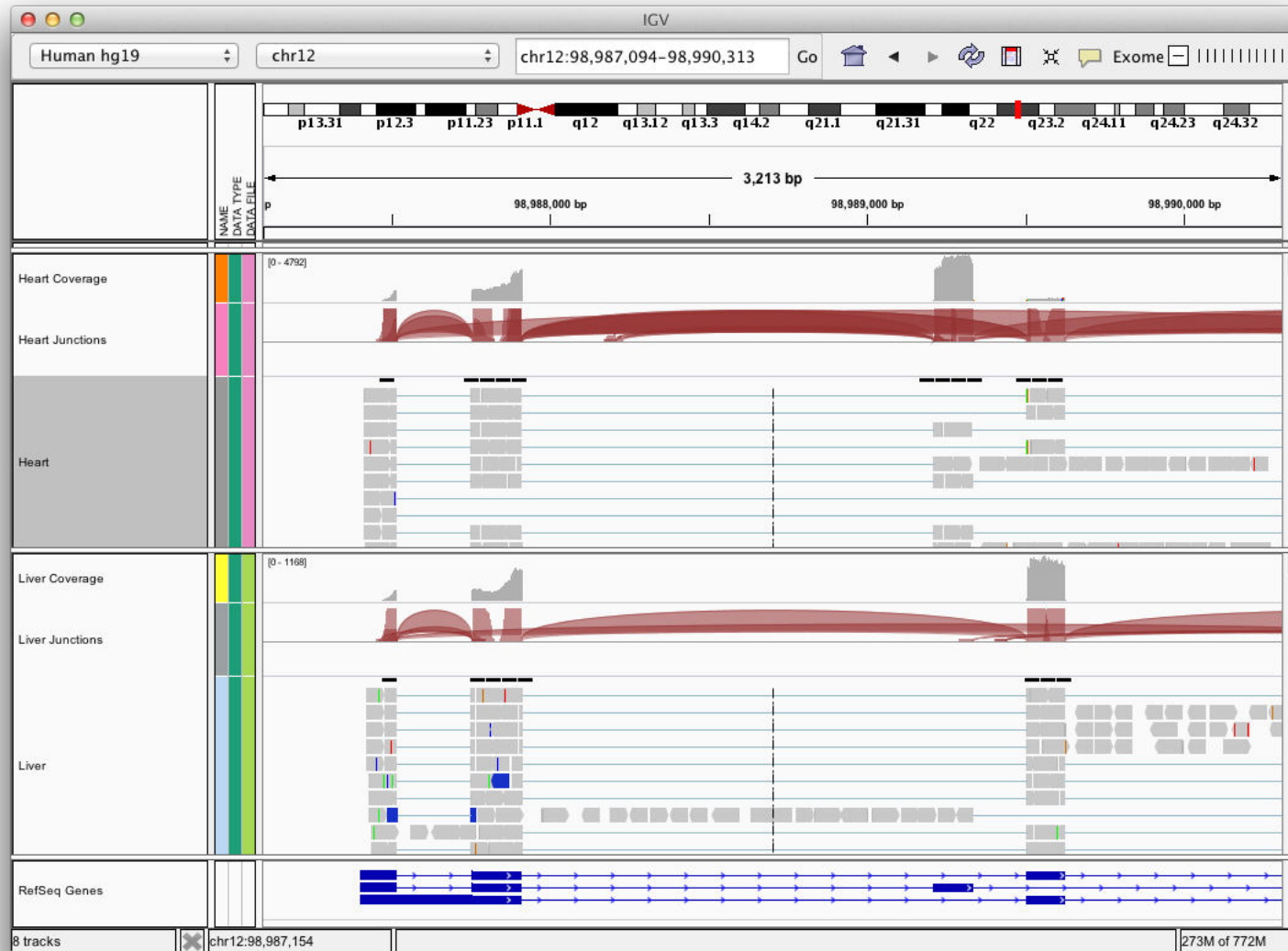
# RNA-seq alignments



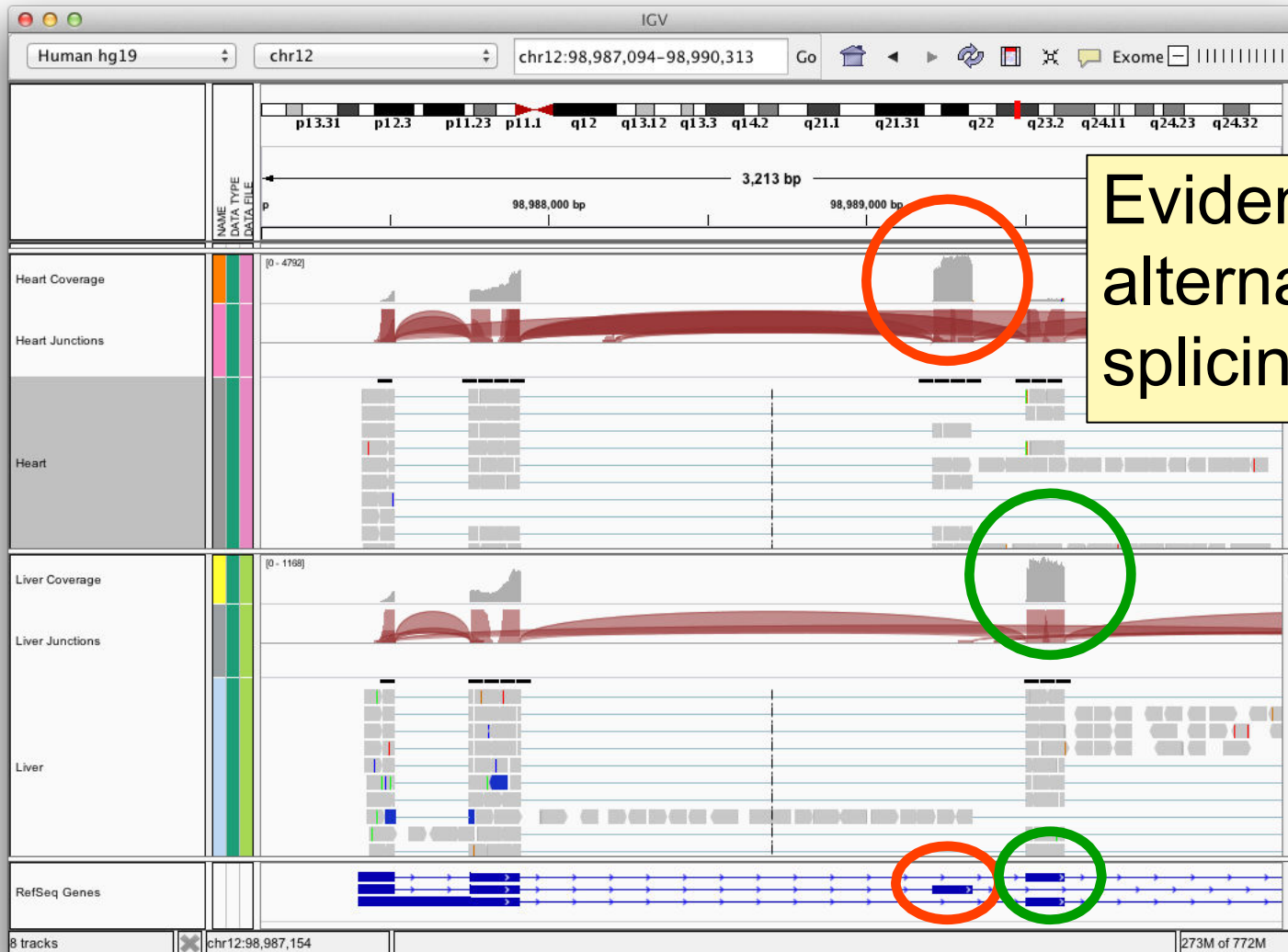
# RNA-seq alignments



# RNA-seq alignments



# RNA-seq alignments



# Sashimi plot

---

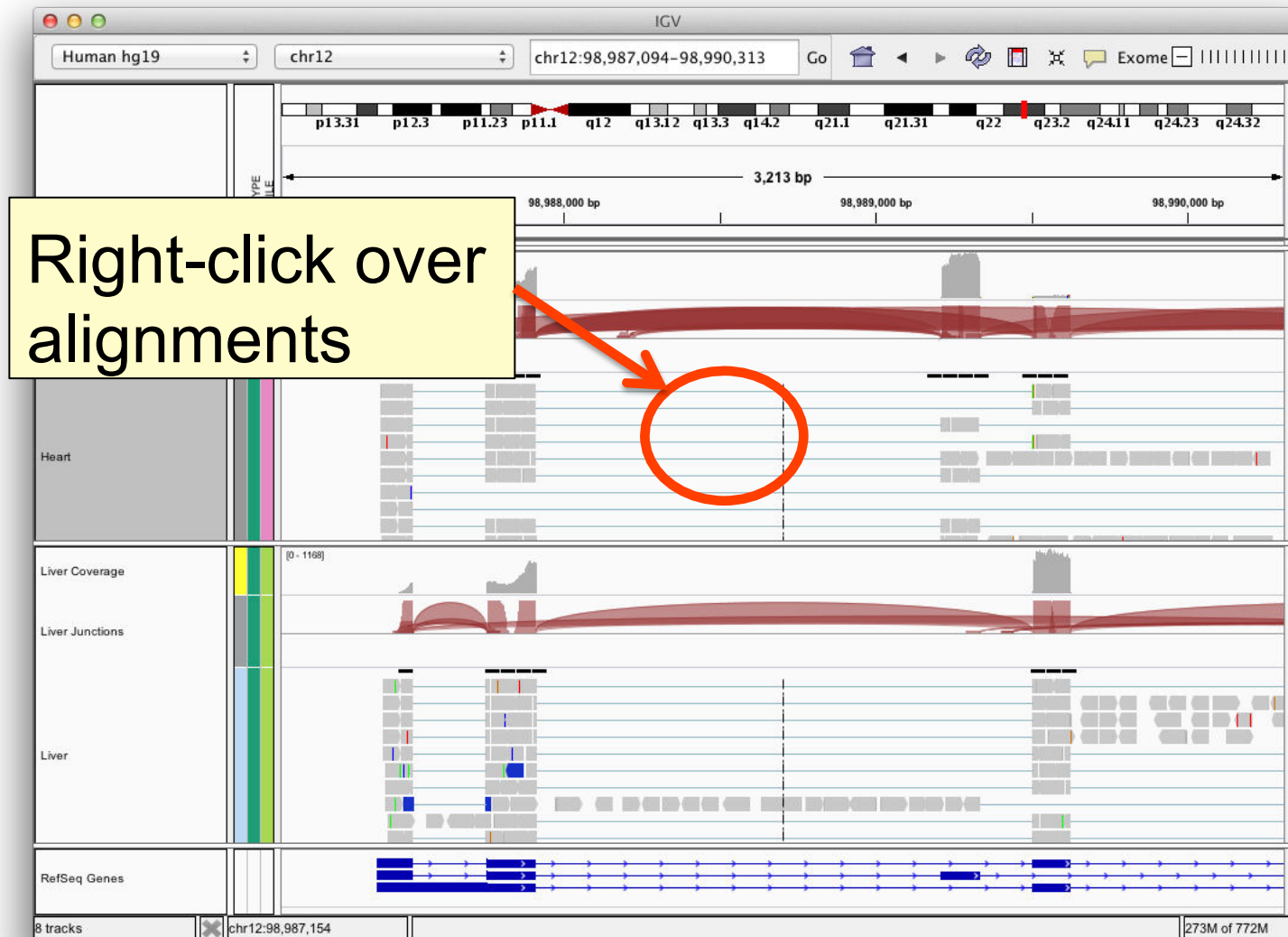


## Viewing RNA splicing with Sashimi Plots

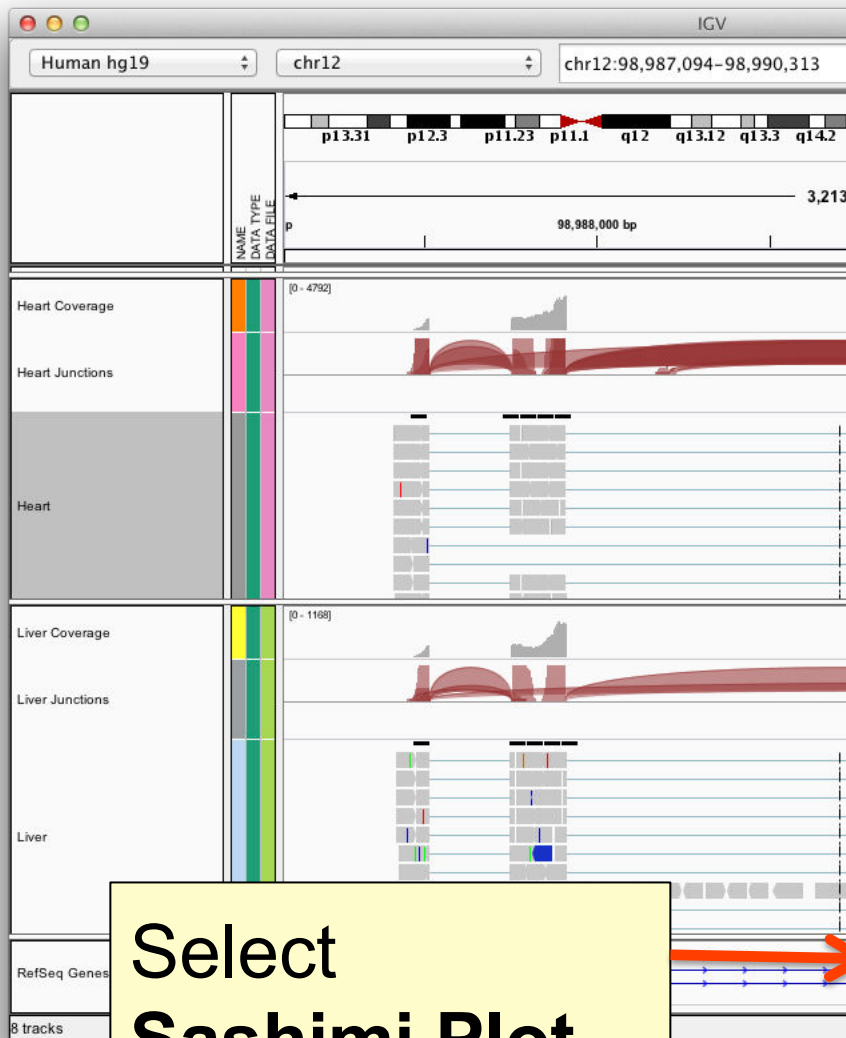
Reference: Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airoidi EM, Burge, CB.

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

# RNA-seq alignments



# RNA-seq alignments

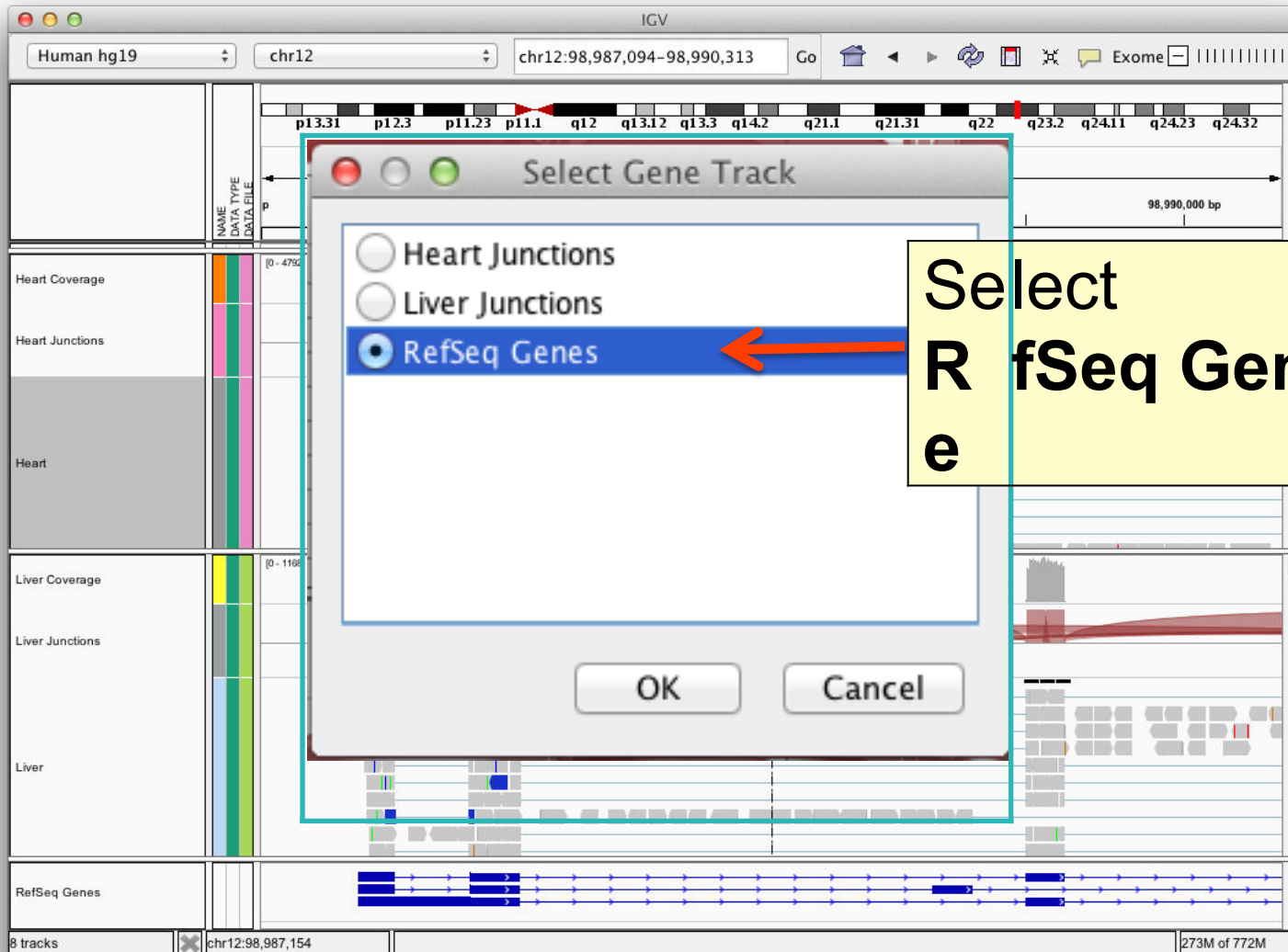


Select  
Sashimi Plot

- Heart
- 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
- ✓ Show coverage track
- Load coverage data...
- Collapsed
- ✓ Expanded
- Squished
- Select by name...
- Clear selections
- Copy read sequence
- Copy consensus sequence
- Sashimi Plot**
- Remove Track
- Save image...



# RNA-seq alignments



Human hg19 chr12 chr12:98,987,094-98,990,313 Go Exome

Heart Coverage  
Heart Junctions  
Heart

Liver Coverage  
Liver Junctions  
Liver

RefSeq Genes

8 tracks chr12:98,987,154 273M of 772M

Select Gene Track

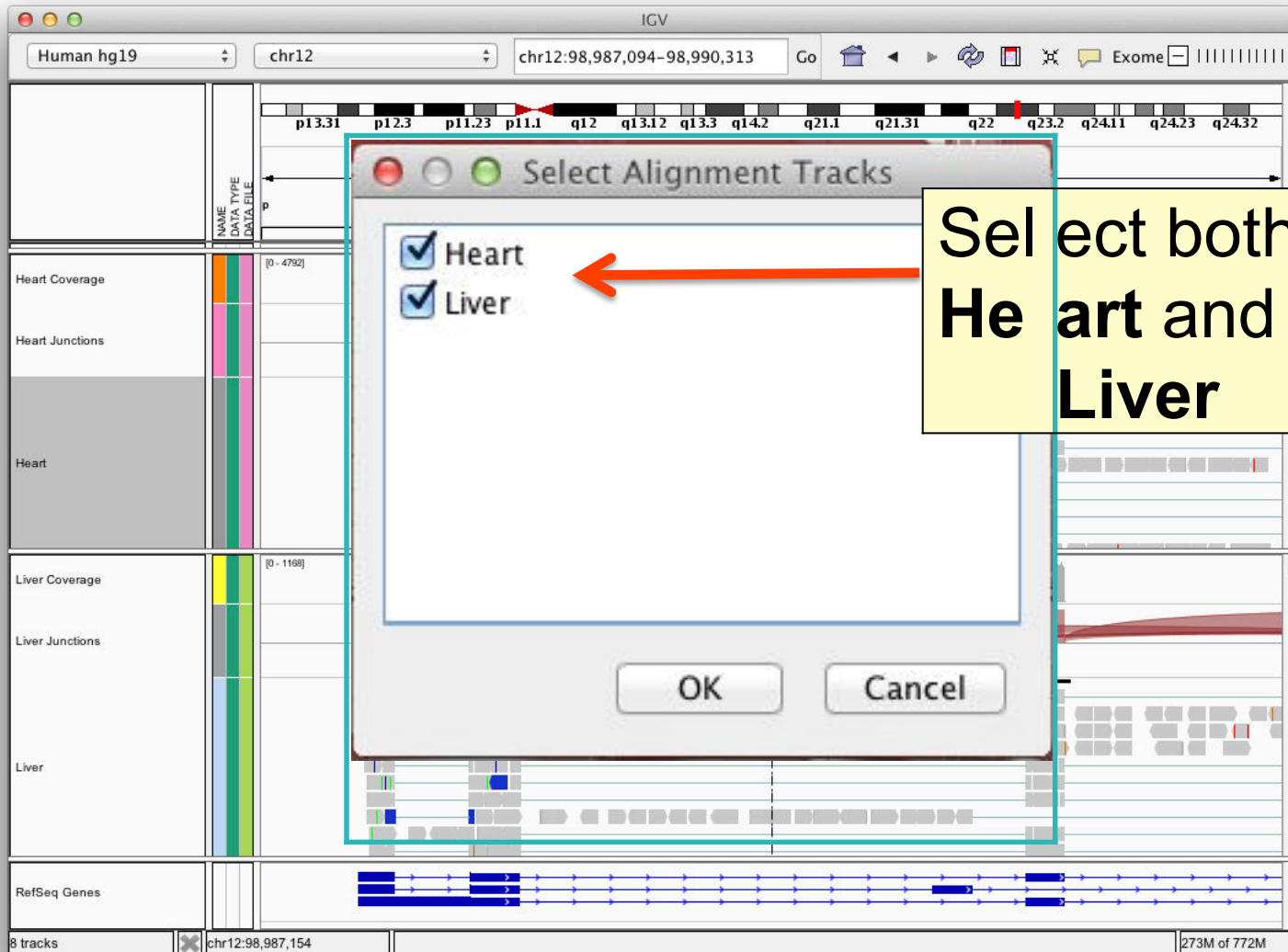
- Heart Junctions
- Liver Junctions
- RefSeq Genes

OK Cancel

Select RefSeq Genes

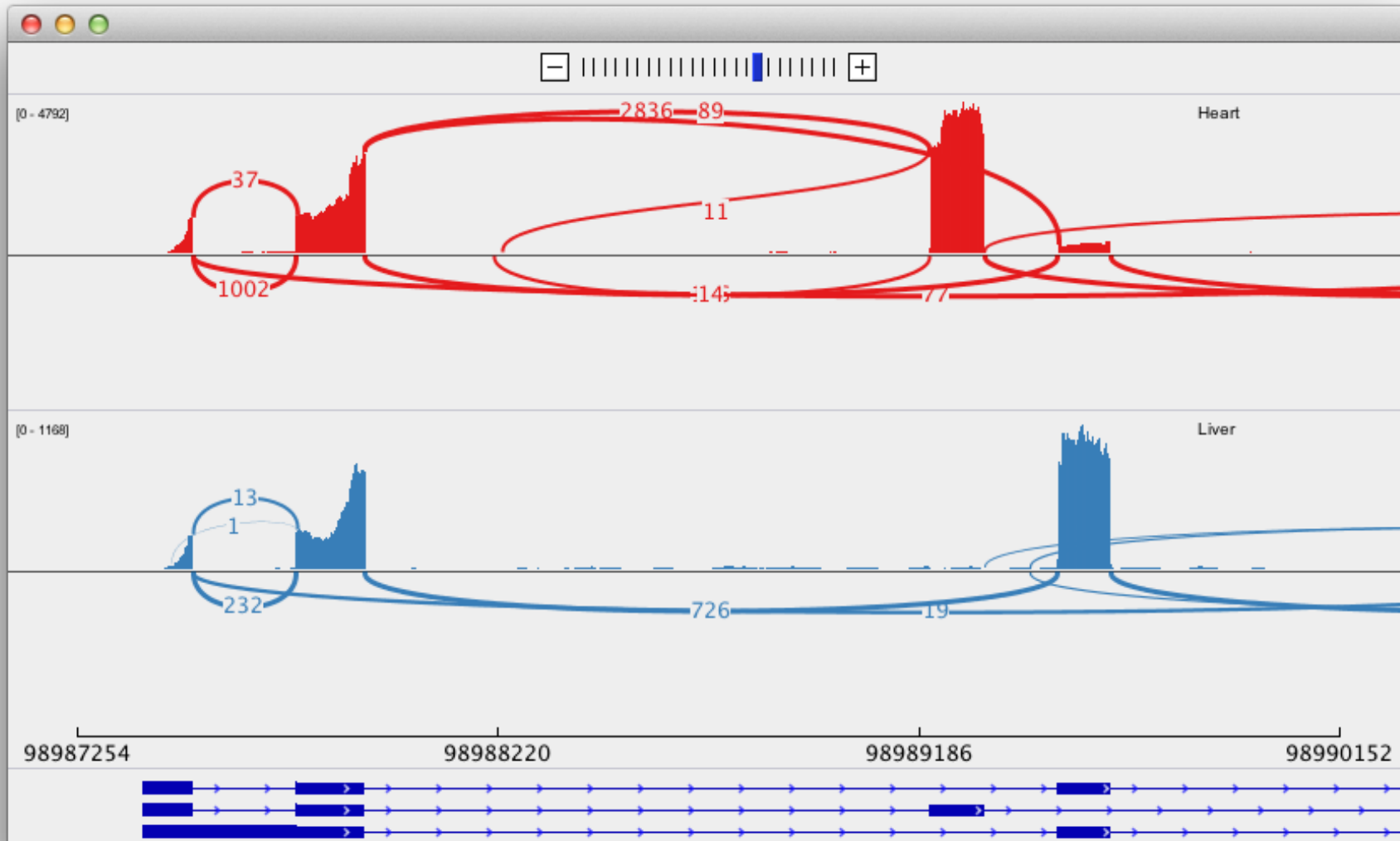


# RNA-seq alignments



The screenshot shows the IGV interface with a 'Select Alignment Tracks' dialog box. The dialog box has a title bar with 'Select Alignment Tracks' and three window control buttons. It contains two checked items: 'Heart' and 'Liver'. Below the list are 'OK' and 'Cancel' buttons. A yellow callout box with a red arrow points to the 'Heart' and 'Liver' items, containing the text 'Select both Heart and Liver'. The background shows the IGV tracks for 'Heart Coverage', 'Heart Junctions', 'Heart', 'Liver Coverage', 'Liver Junctions', 'Liver', and 'RefSeq Genes'.

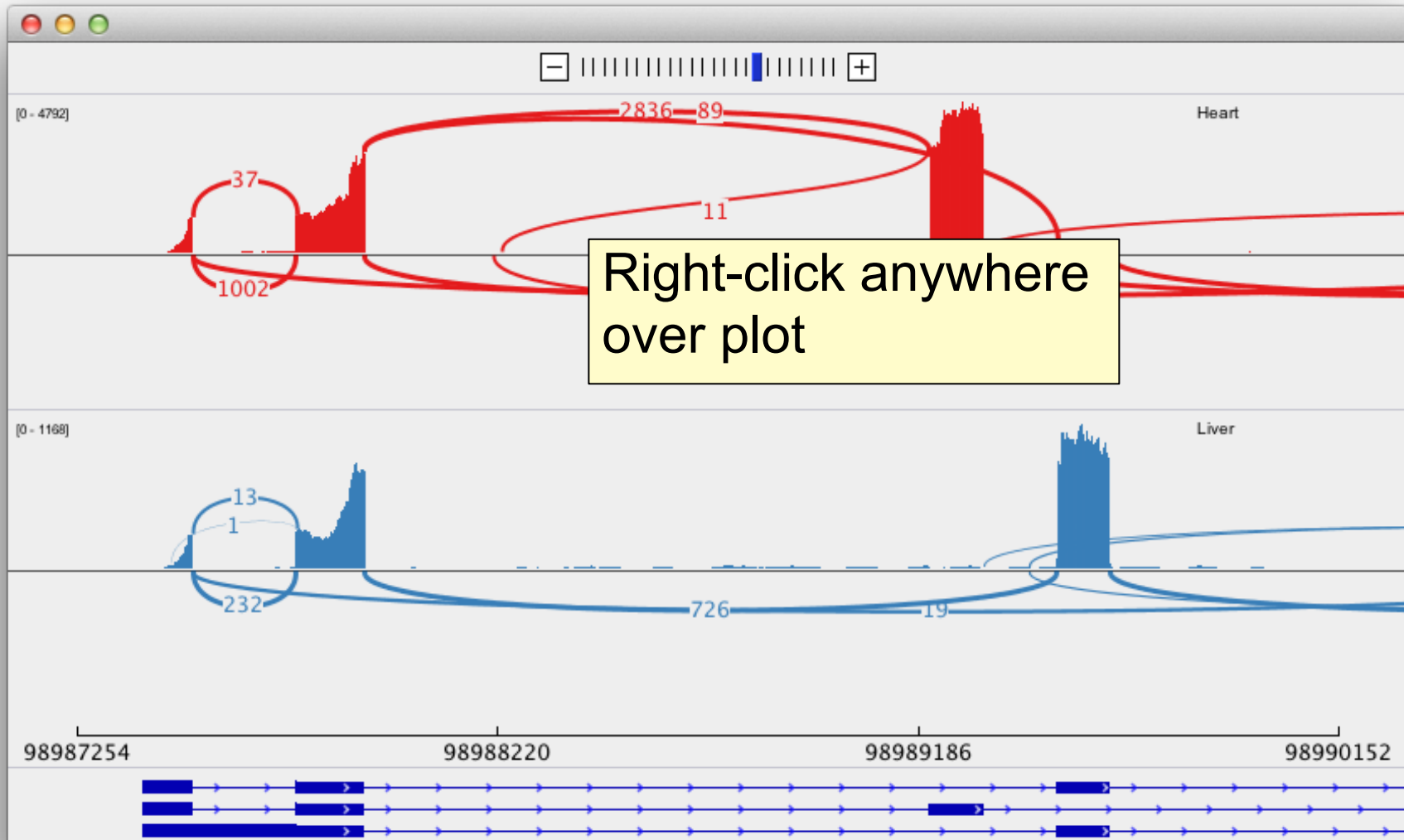
# RNA-seq alignments



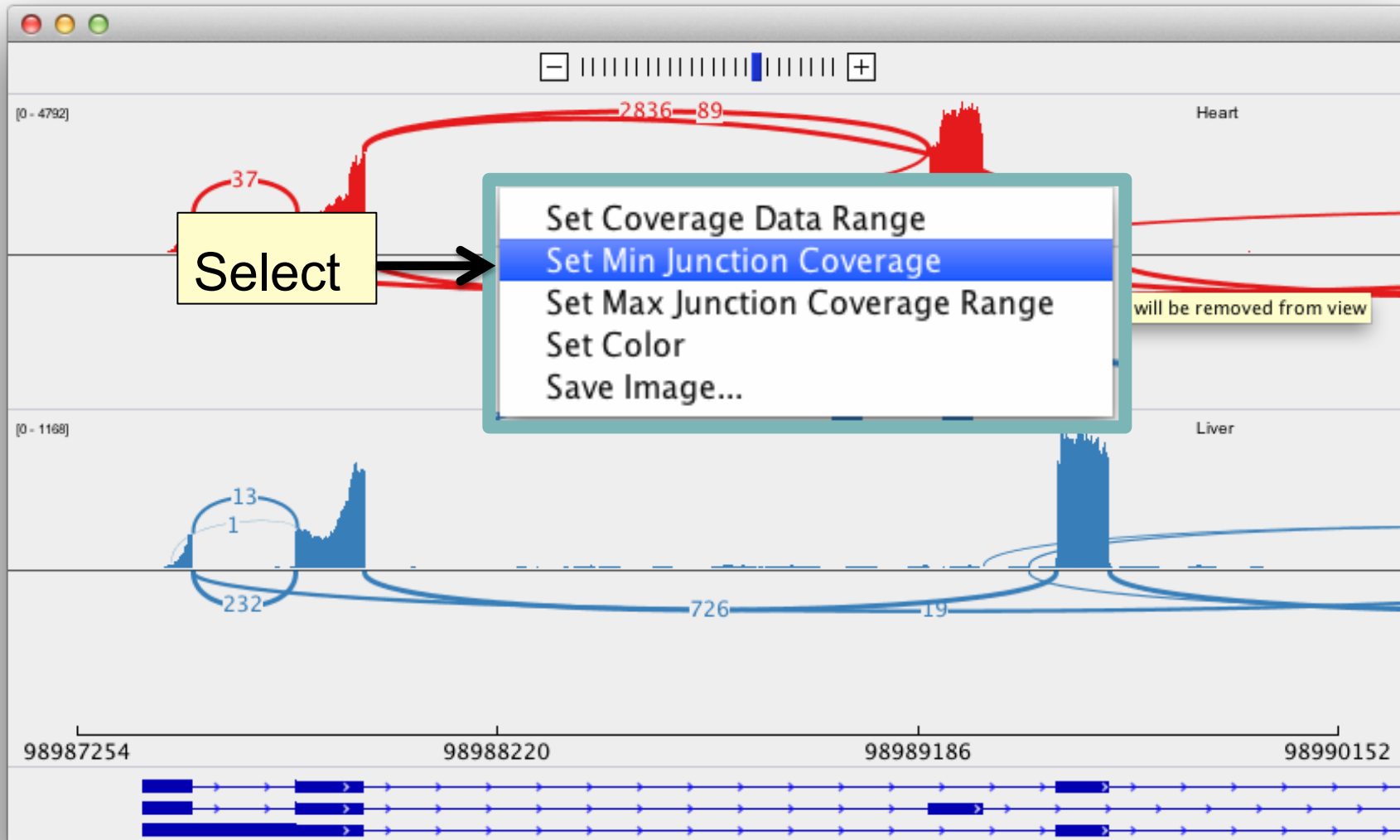
# RNA-seq alignments



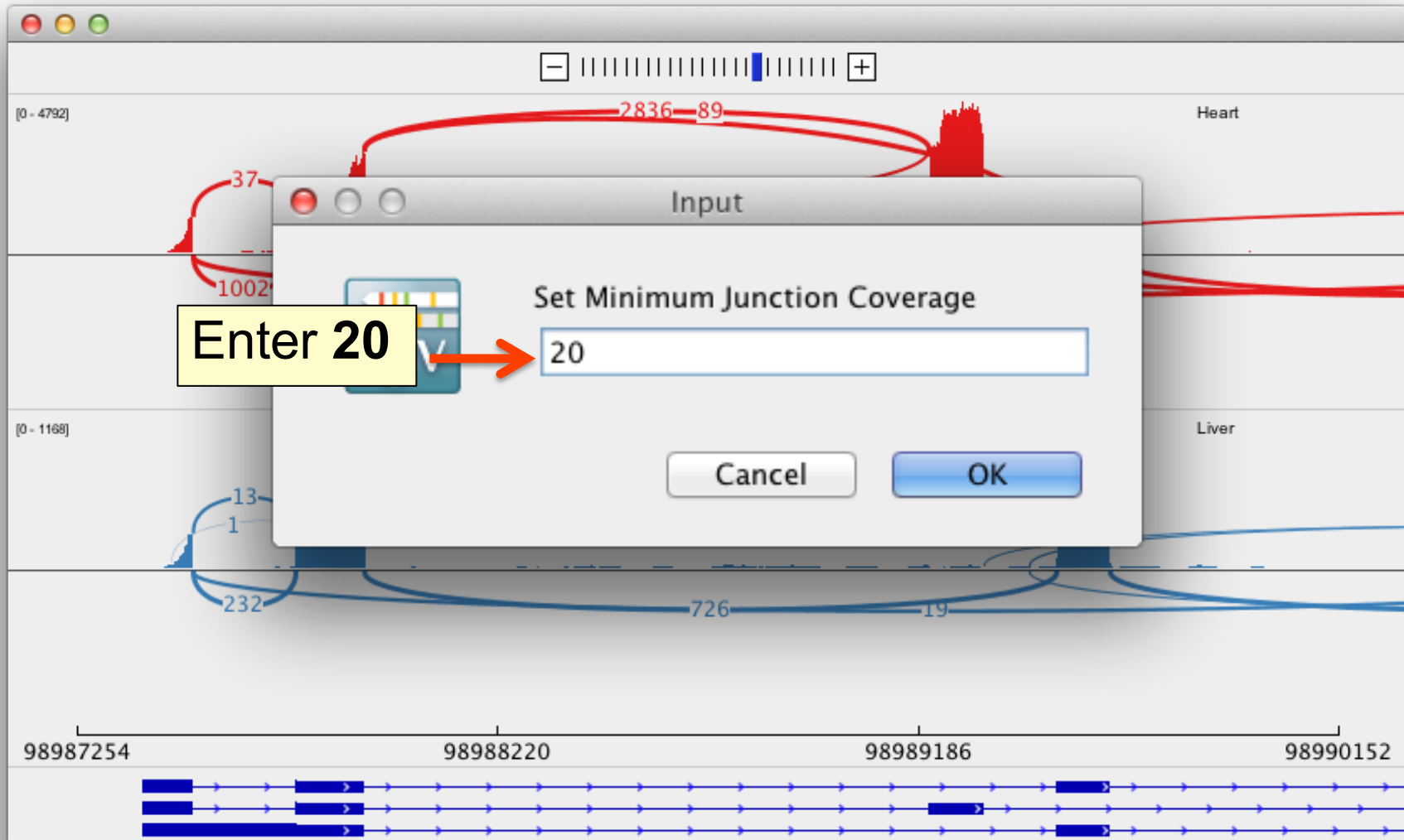
# RNA-seq alignments



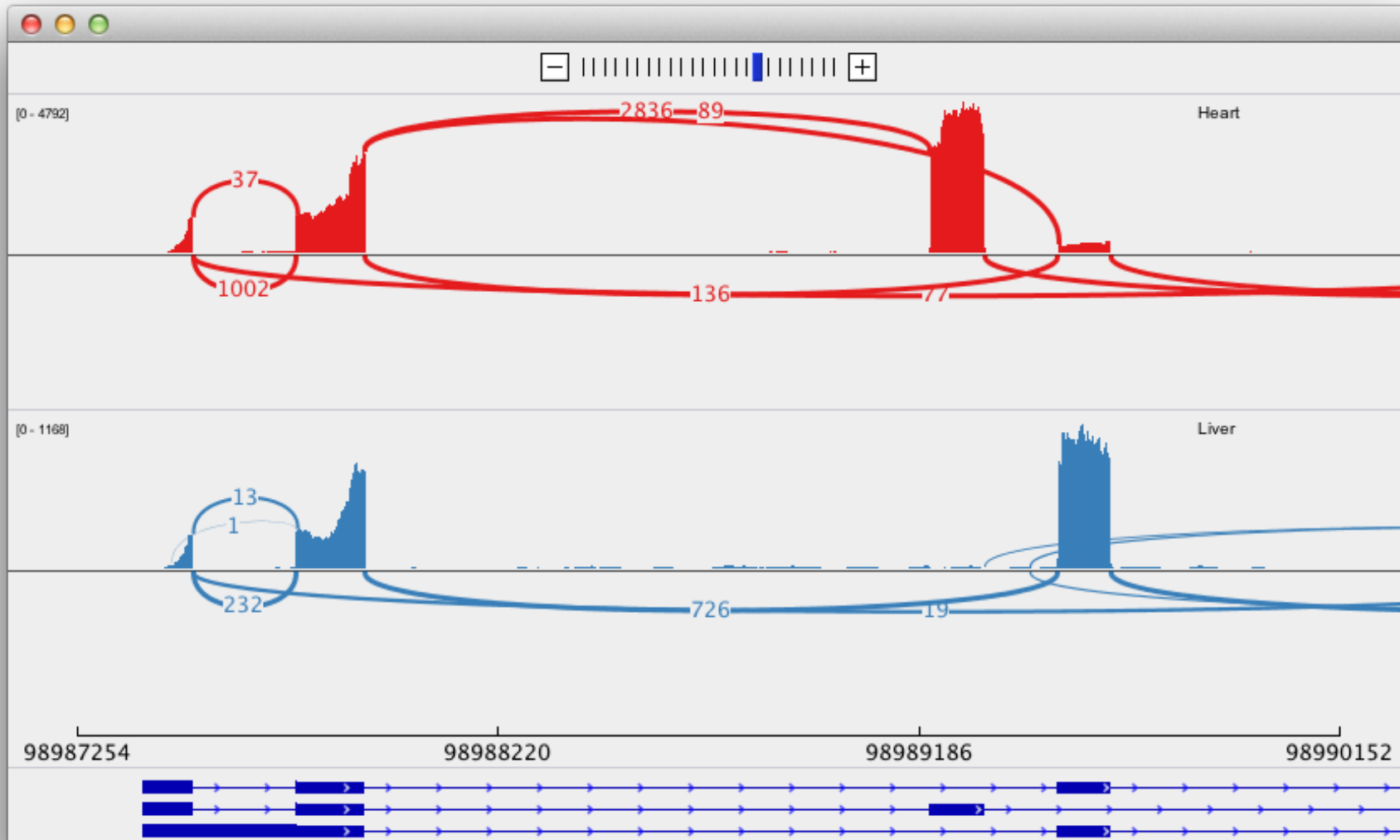
# RNA-seq alignments



# RNA-seq alignments



# RNA-seq alignments



# igvtools



# igvtools

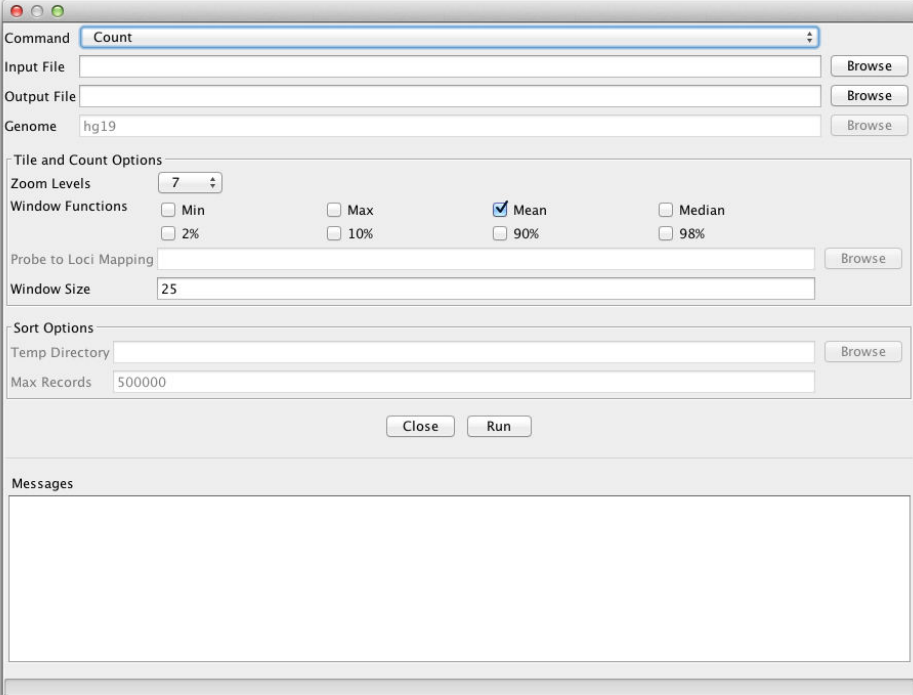


A set of utilities for preparing files for efficient display.

<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



The screenshot shows the 'igvtools' command-line interface window. The 'Command' field is set to 'Count'. The 'Input File', 'Output File', and 'Genome' fields are empty, each with a 'Browse' button. The 'Genome' field is set to 'hg19'. The 'Tile and Count Options' section includes a 'Zoom Levels' dropdown set to '7', and checkboxes for 'Window Functions': 'Min', 'Max', 'Mean' (checked), 'Median', '2%', '10%', '90%', and '98%'. There is a 'Probe to Loci Mapping' field with a 'Browse' button and a 'Window Size' field set to '25'. The 'Sort Options' section includes a 'Temp Directory' field with a 'Browse' button and a 'Max Records' field set to '500000'. At the bottom, there are 'Close' and 'Run' buttons. A 'Messages' section is visible at the bottom of the window.

# 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.

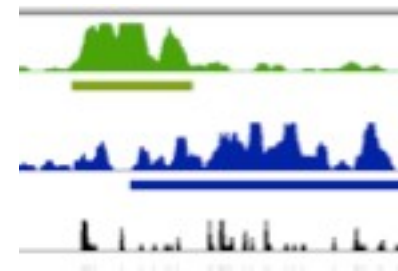
# igvtools count

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](ftp://ftp.broadinstitute.org/pub/igv/CSH_2013/files.zip)

Files included in the zip:

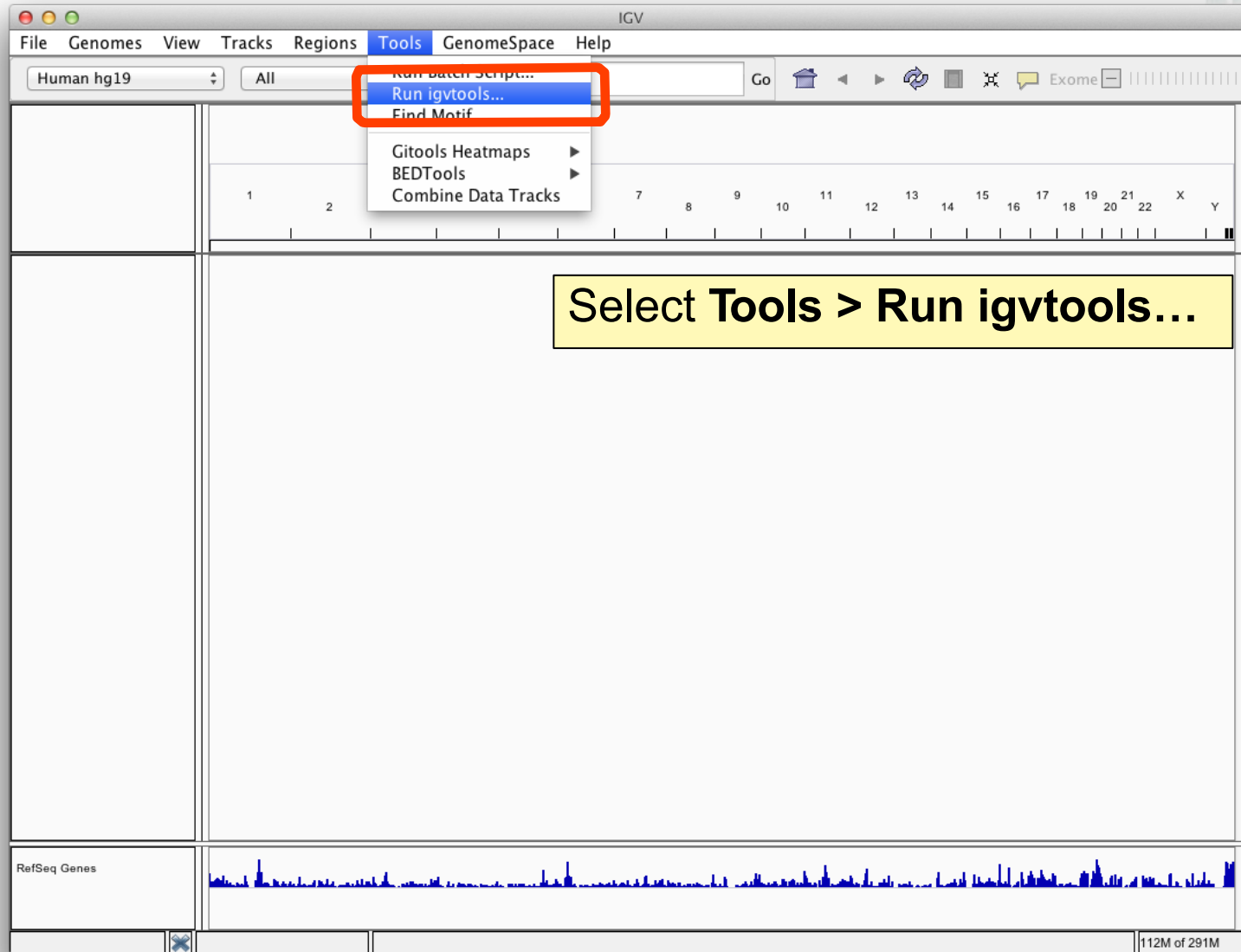
heart.bodyMap.bam

heart.bodyMap.bam.bai

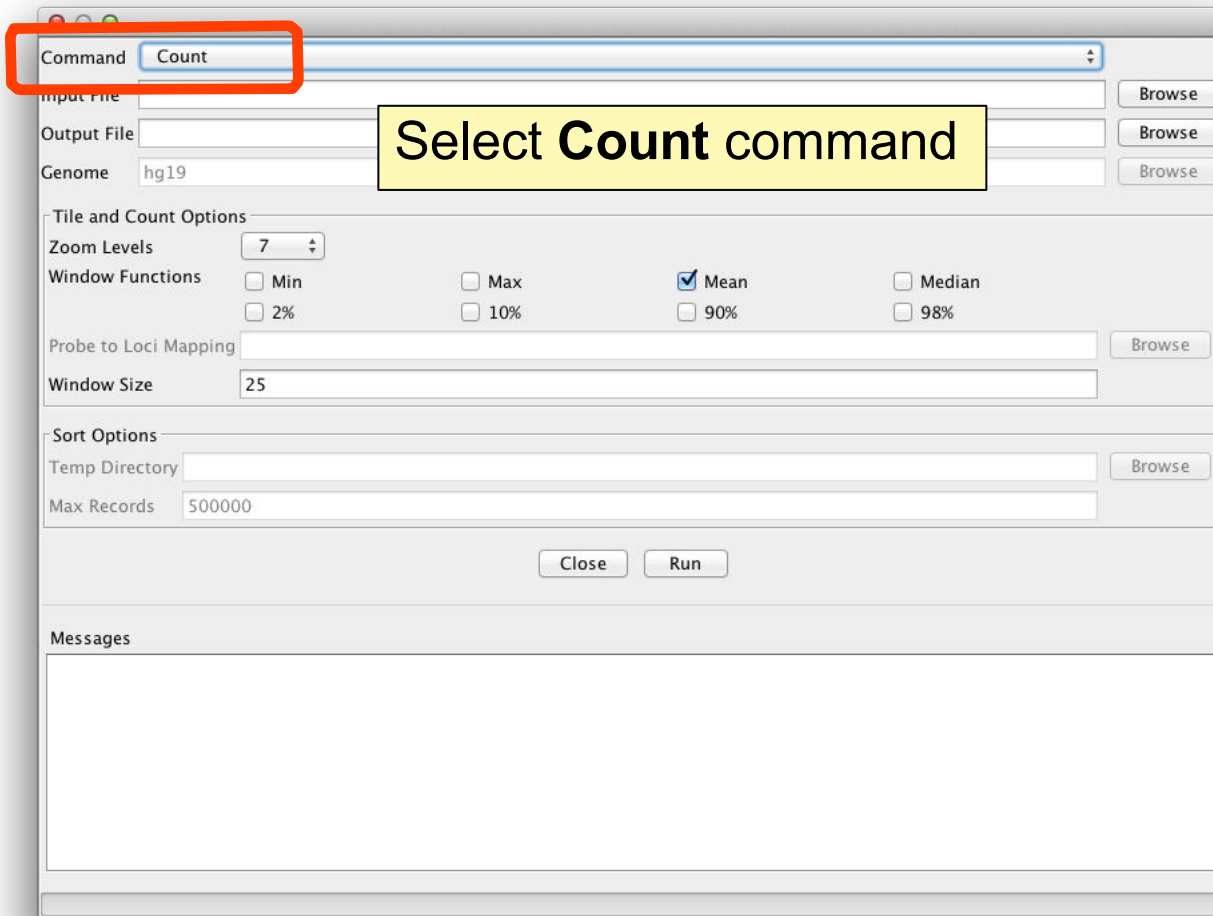
sacCer3.fa (used in next exercise)



# Computing coverage: igvtools



# Computing coverage: igvtools



Command **Count**

Input File  Browse

Output File  Browse

Genome hg19  Browse

Tile and Count Options

Zoom Levels

Window Functions  Min  Max  Mean  Median  
 2%  10%  90%  98%

Probe to Loci Mapping  Browse

Window Size

Sort Options

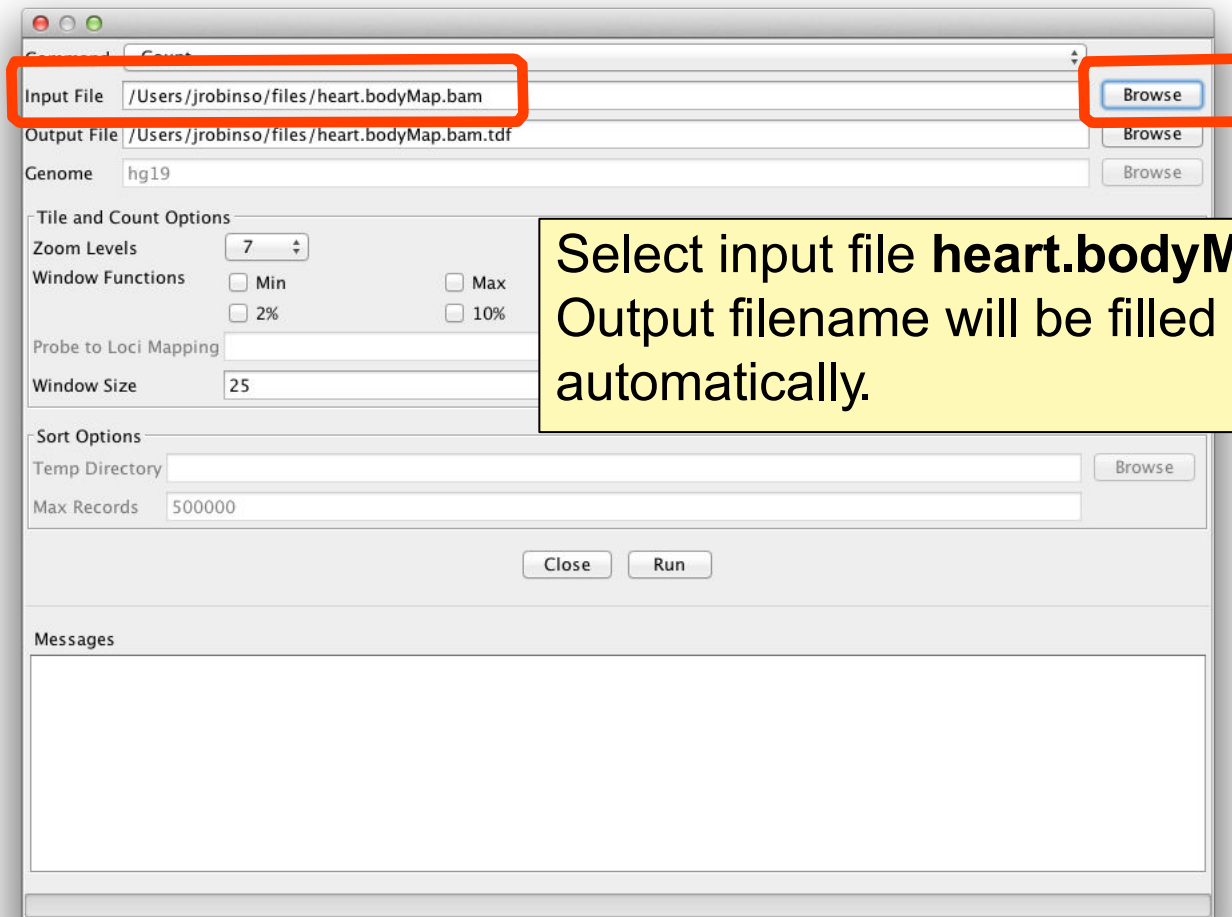
Temp Directory  Browse

Max Records

Close Run

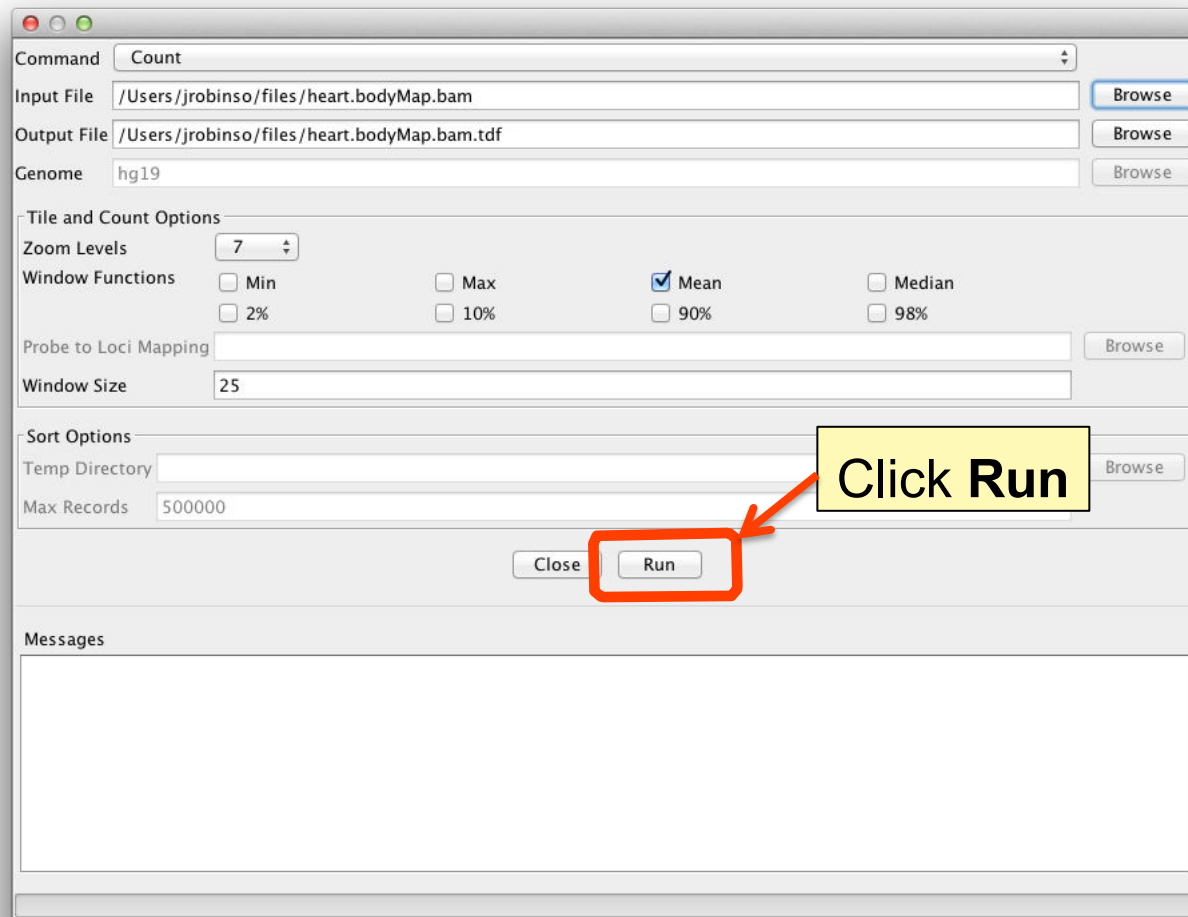
Messages

# Computing coverage: igvtools

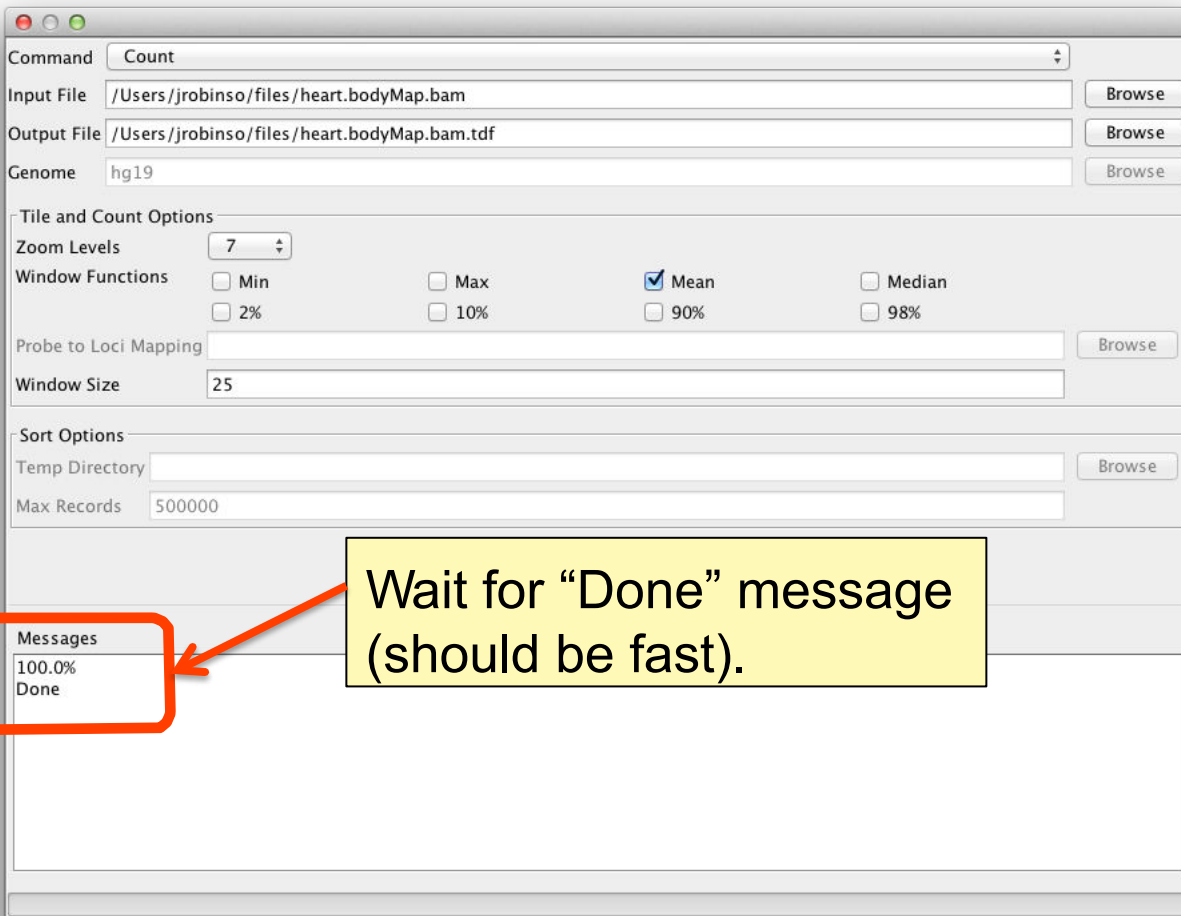


Select input file **heart.bodyMap.bam**  
Output filename will be filled in automatically.

# Computing coverage: igvtools



# Computing coverage: igvtools



Command: Count

Input File: /Users/jrobinso/files/heart.bodyMap.bam Browse

Output File: /Users/jrobinso/files/heart.bodyMap.bam.tdf Browse

Genome: hg19 Browse

**Tile and Count Options**

Zoom Levels: 7

Window Functions:  Min  Max  Mean  Median  
 2%  10%  90%  98%

Probe to Loci Mapping: Browse

Window Size: 25

**Sort Options**

Temp Directory: Browse

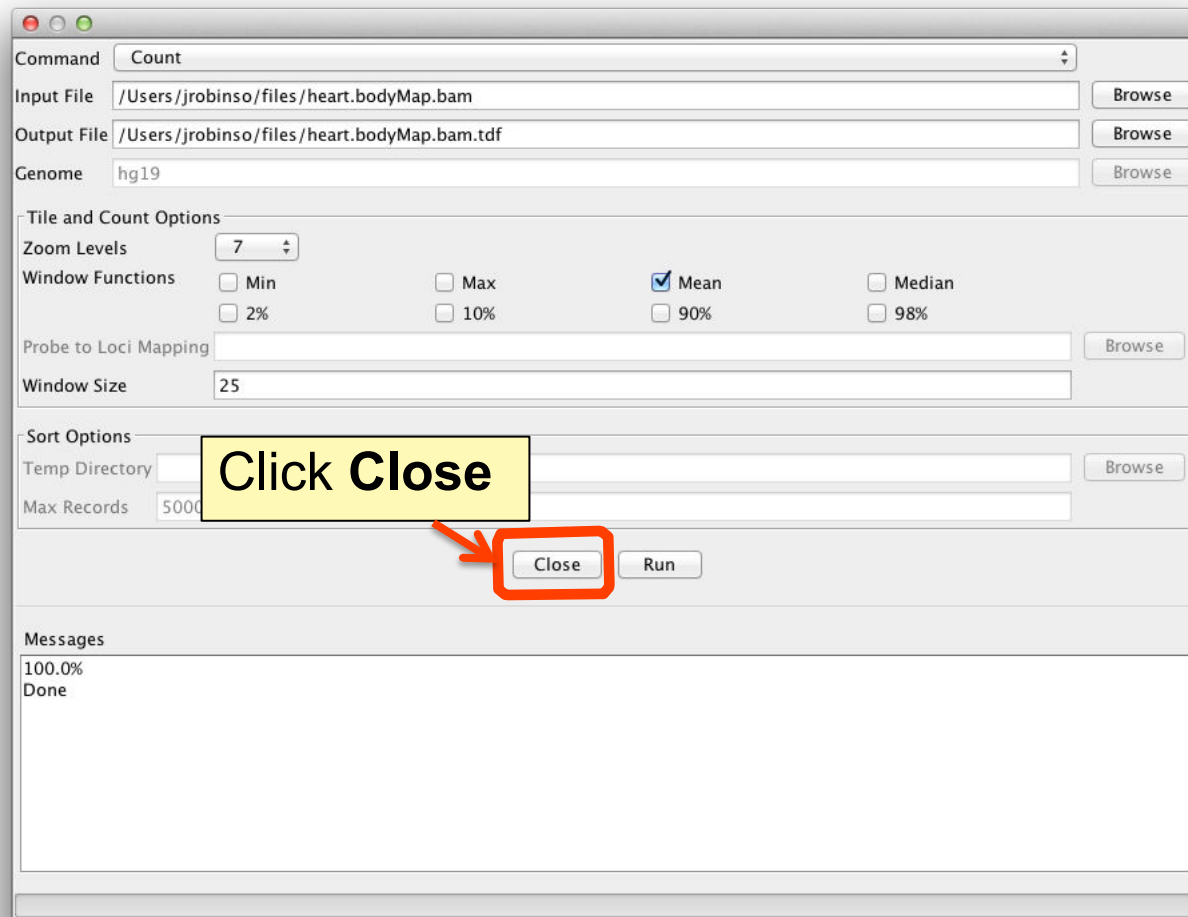
Max Records: 500000

**Messages**

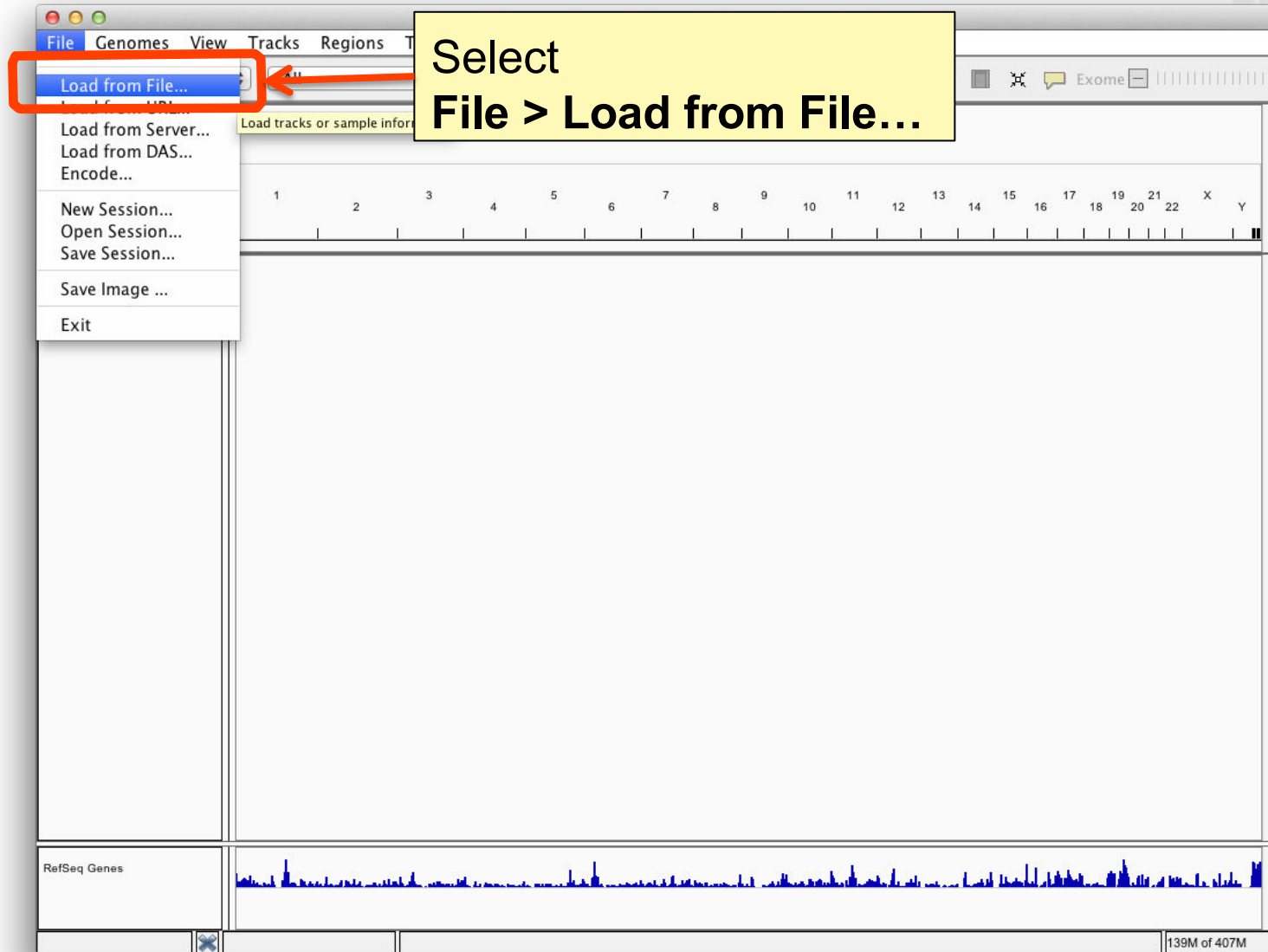
100.0%  
Done

Wait for "Done" message (should be fast).

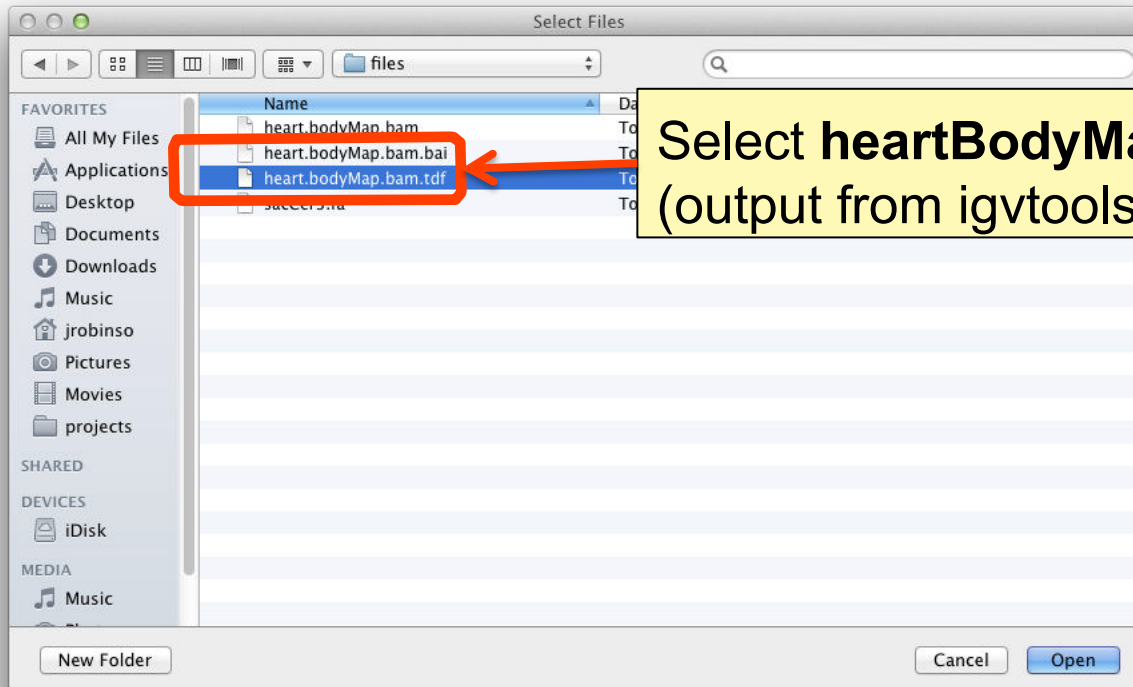
# Computing coverage: igvtools



# Computing coverage: igvtools



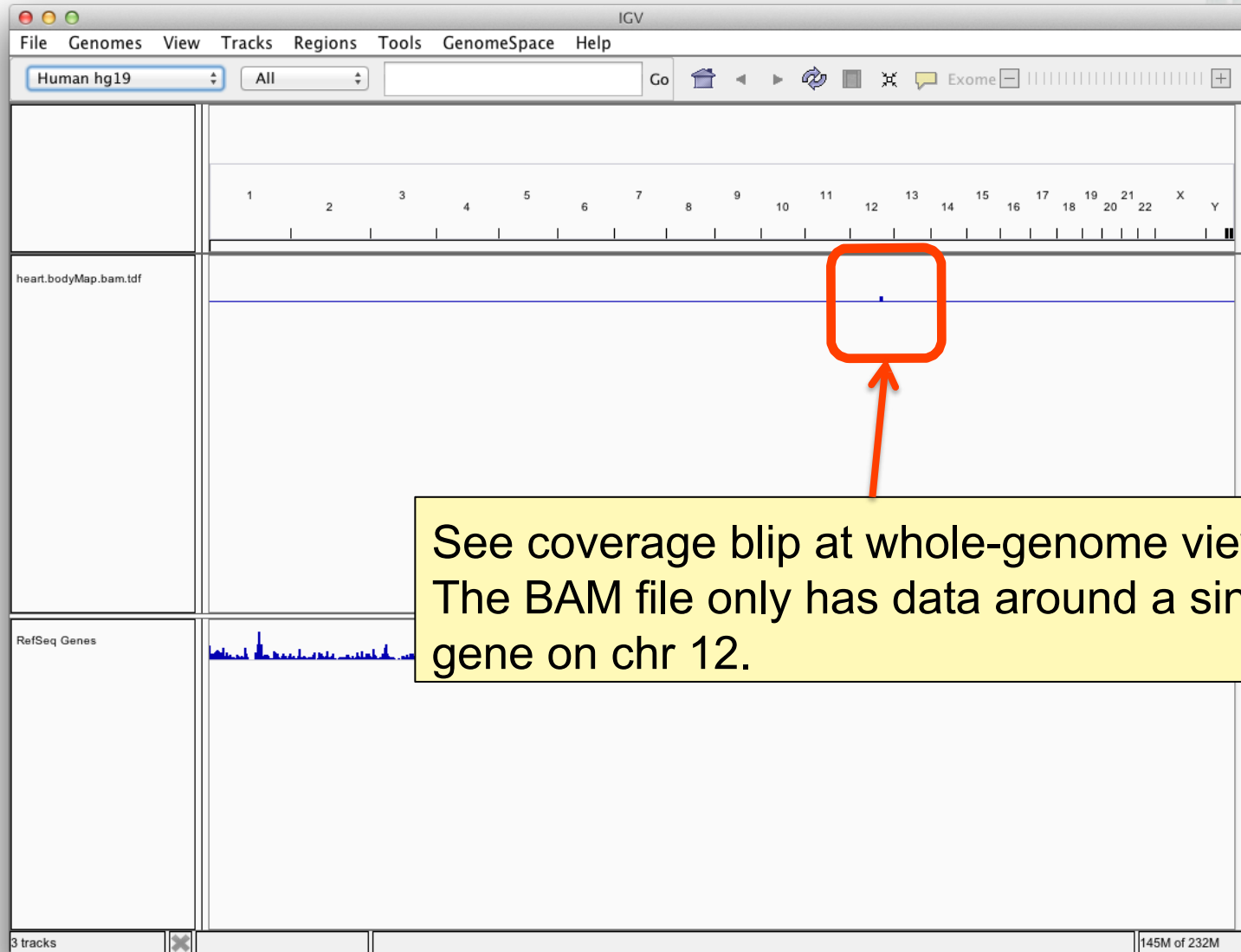
# Computing coverage: igvtools



Select **heartBodyMap.bam.tdf**  
(output from igvtools)



# Computing coverage: igvtools



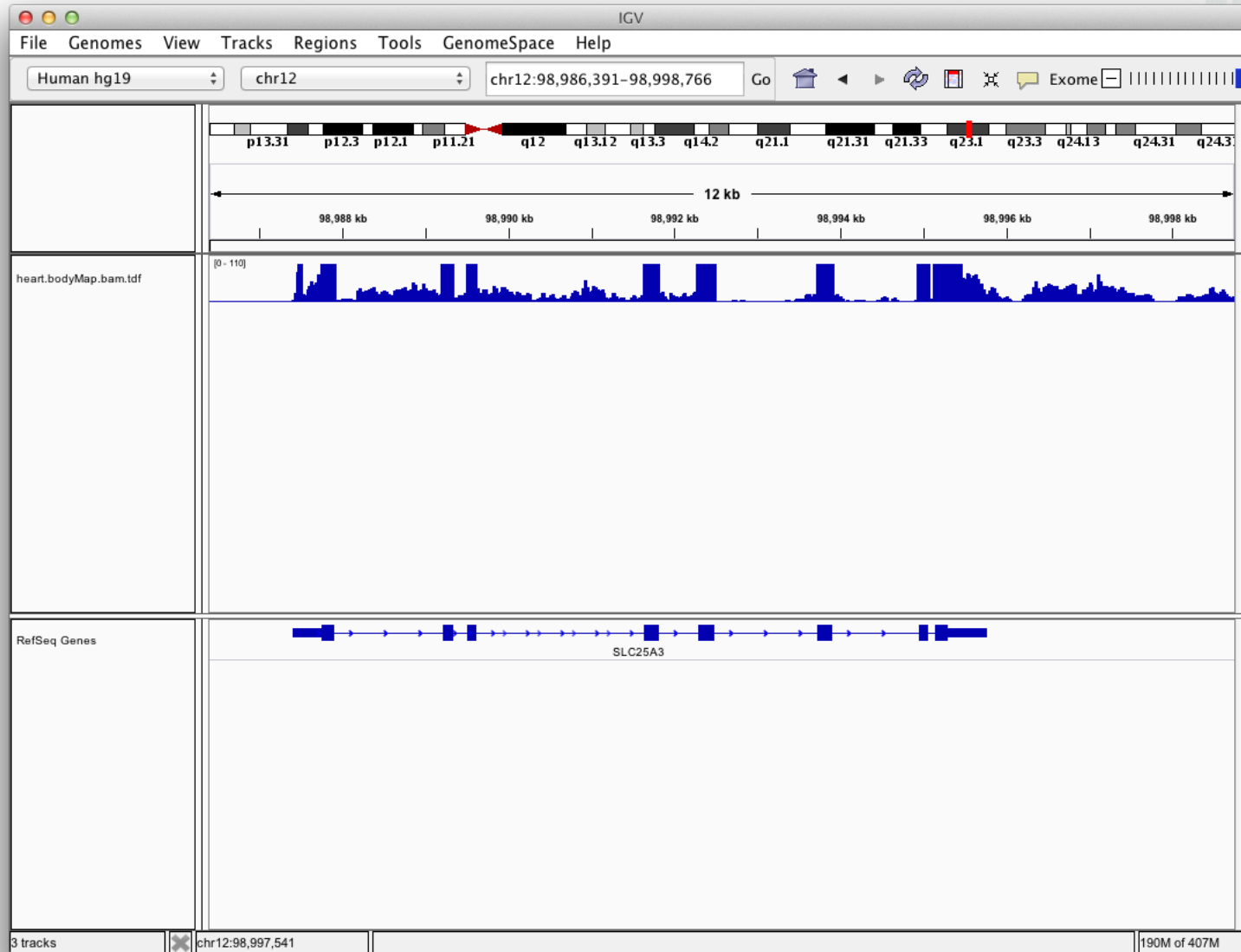
See coverage blip at whole-genome view. The BAM file only has data around a single gene on chr 12.

# Computing coverage: igvtools



A screenshot of the Integrative Genomics Viewer (IGV) interface. The window title is 'IGV'. The menu bar includes 'File', 'Genomes', 'View', 'Tracks', 'Regions', 'Tools', 'GenomeSpace', and 'Help'. The 'Genomes' dropdown is set to 'Human hg19'. A search box contains 'SLC25A3' and a 'Go' button. A dropdown menu is open below the search box, listing 'SLC25A3', 'SLC25A31', 'SLC25A32', 'SLC25A33', 'SLC25A34', 'SLC25A35', and 'SLC25A36'. A yellow tooltip above the search box says 'Enter a gene or locus, e.f. EGFR, chr1, or chr1:100,000-200,000'. A yellow callout box with a black border and text says 'Enter SLC25A3 in the search box and click Go'. The main area shows a genomic track with chromosomes 1-22, X, and Y. Below the track, there are two tracks: 'heart.bodyMap.bam.tdf' and 'RefSeq Genes'. The bottom status bar shows '3 tracks' on the left and '149M of 407M' on the right.

# Computing coverage: igvtools



# More about reference genomes



IGV doesn't host the genome you need?

Use any genome you want, if you have the sequence in FASTA format.

Optionally, package genome annotations with the sequence.

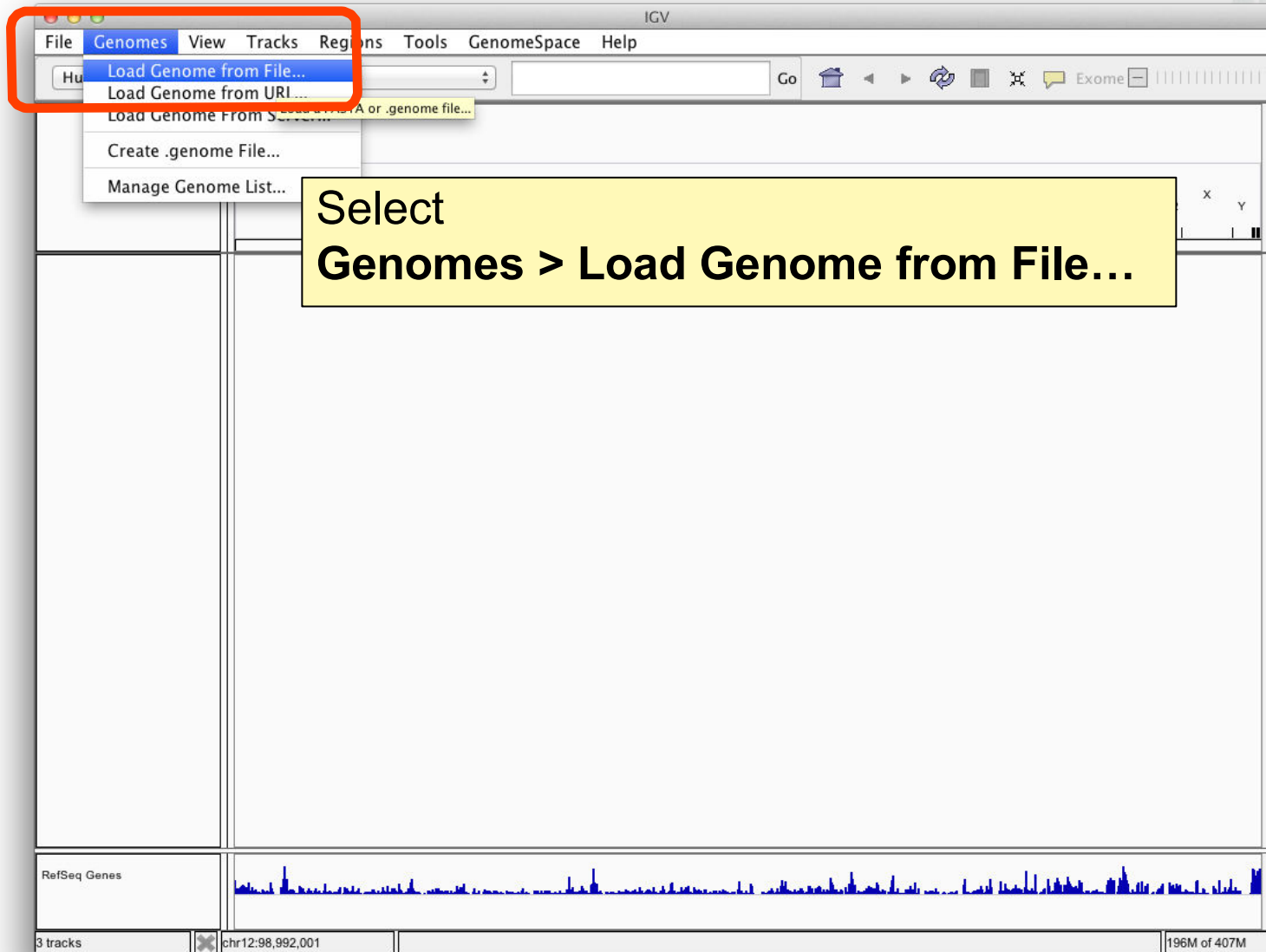
# Loading a genome

---

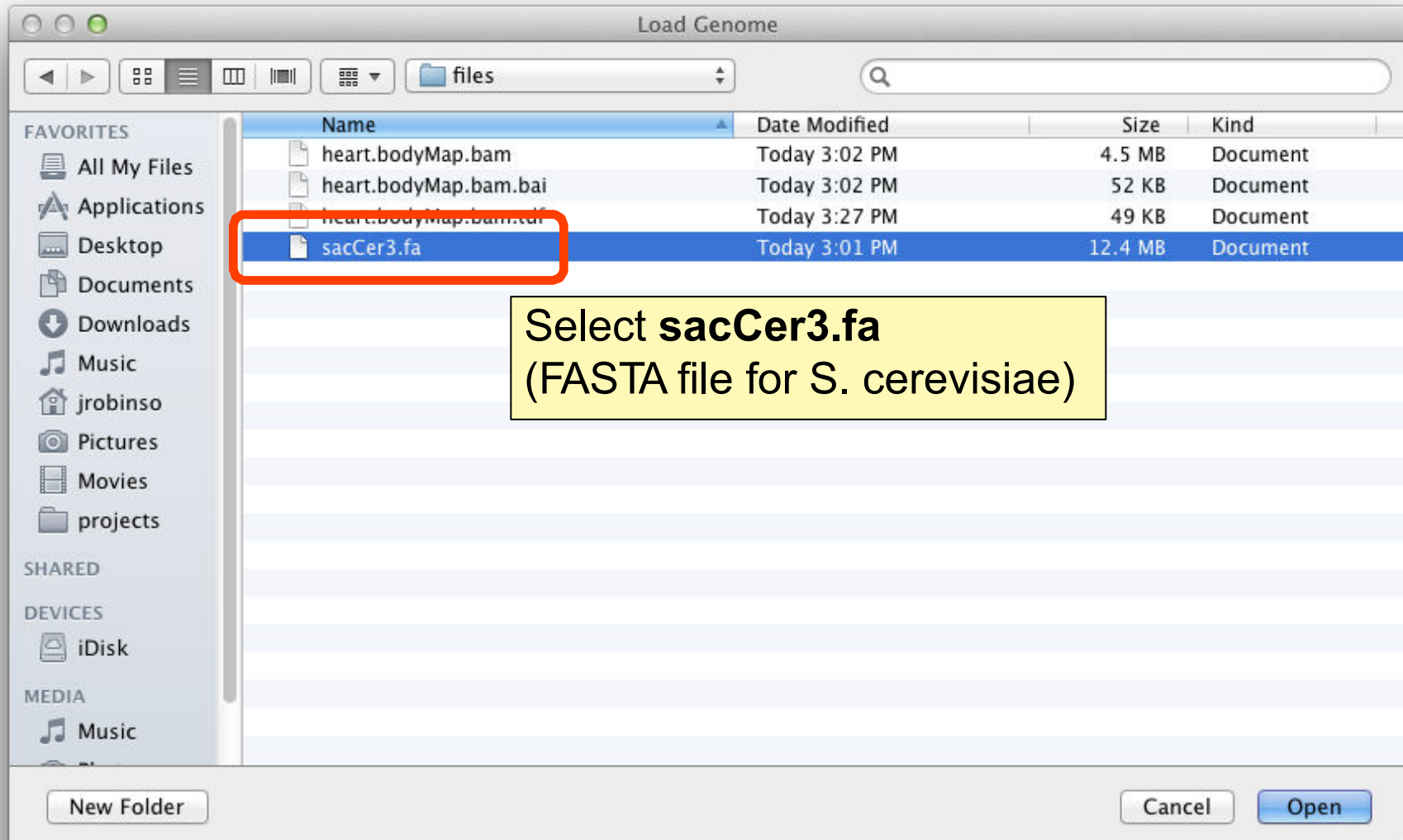


Hands-on exercise

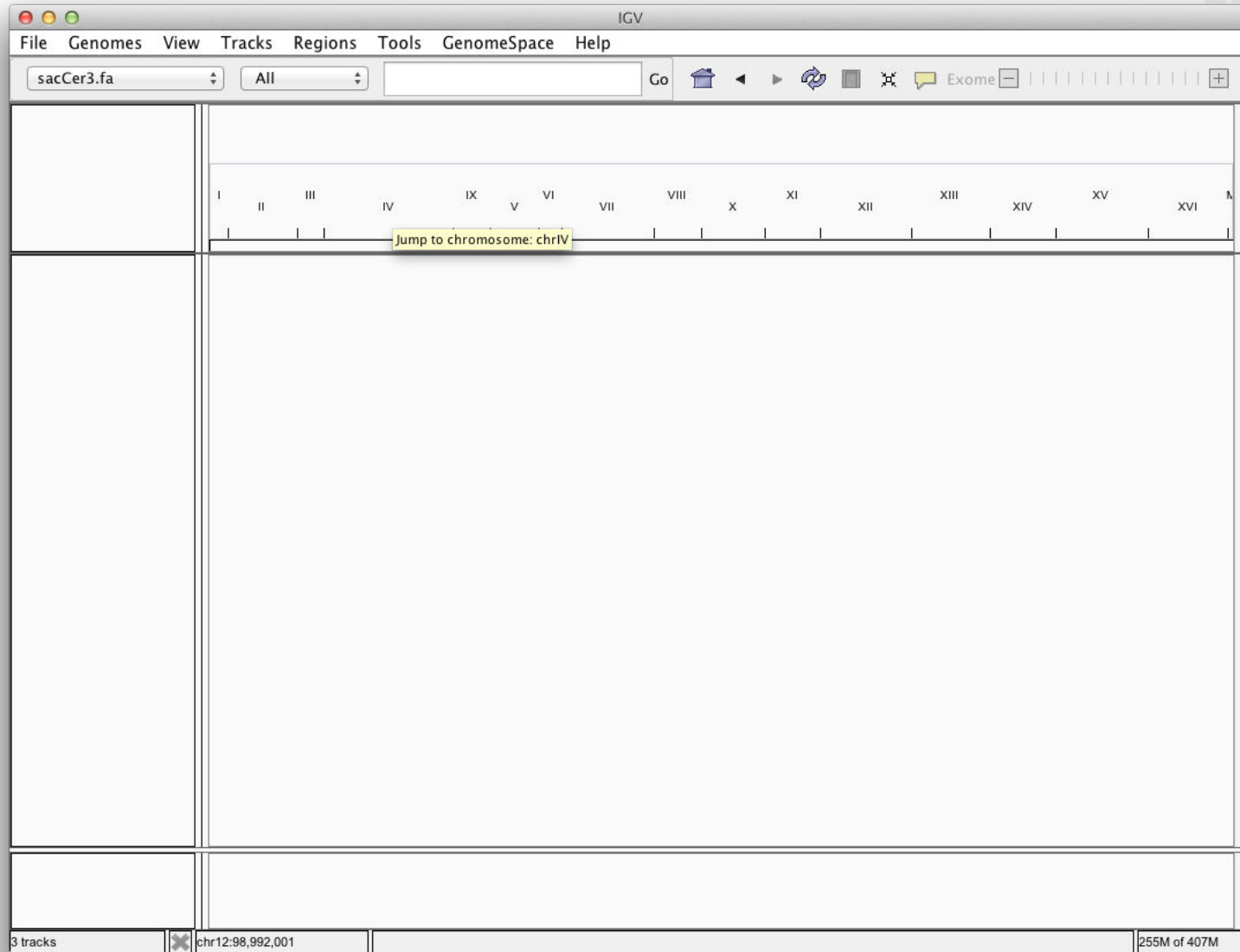
# Loading a genome



# Loading a genome

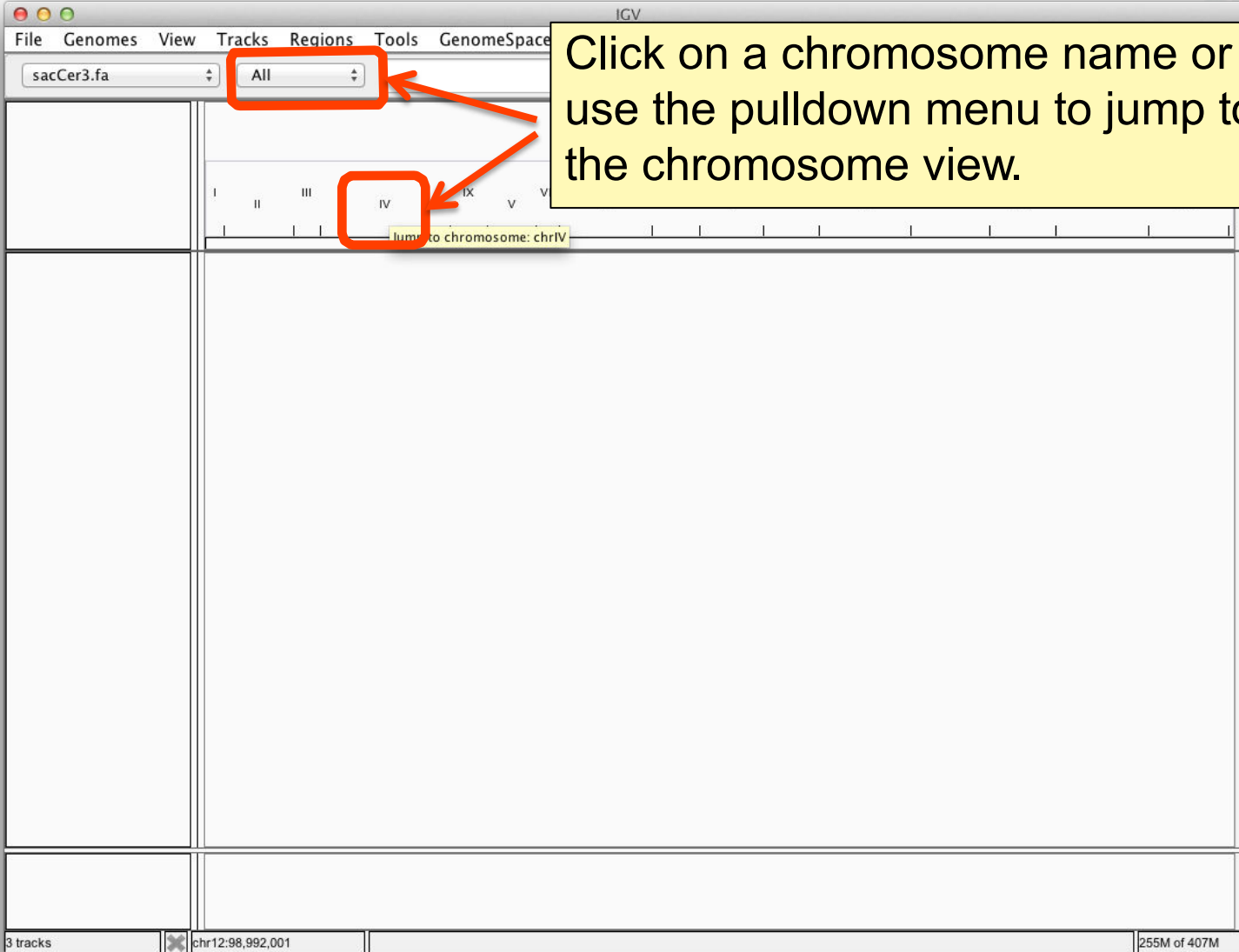


# Loading a genome





# Loading a genome



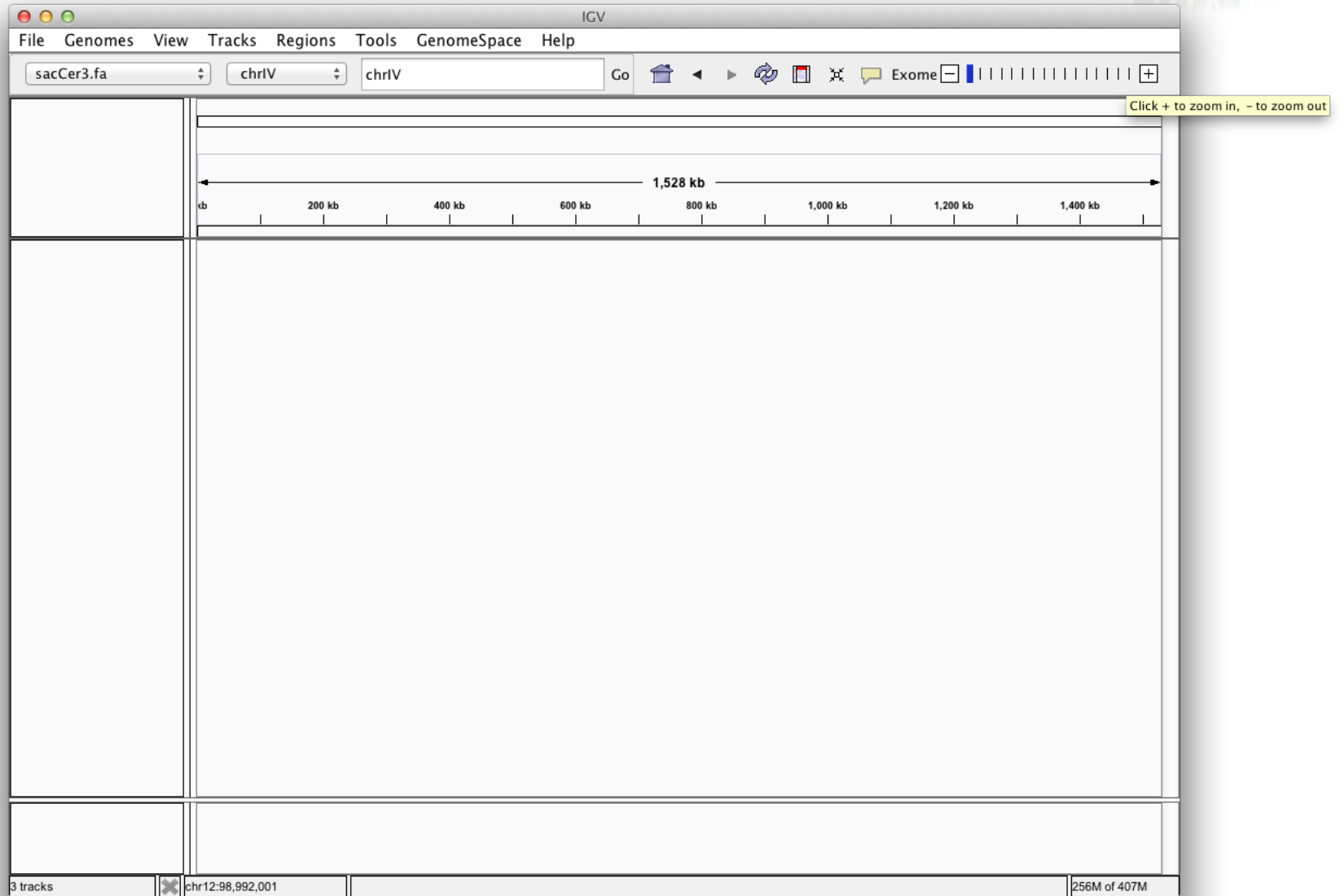
The screenshot shows the IGV interface with the following elements:

- Menu bar: File, Genomes, View, Tracks, Regions, Tools, GenomeSpace
- Genome dropdown: sacCer3.fa
- Chromosome dropdown: All
- Chromosome list: I, II, III, IV, IX, V, VI
- Chromosome IV is highlighted with a yellow box and a tooltip that says "Jump to chromosome: chrIV".
- Status bar: 3 tracks, chr12:98,992,001, 255M of 407M

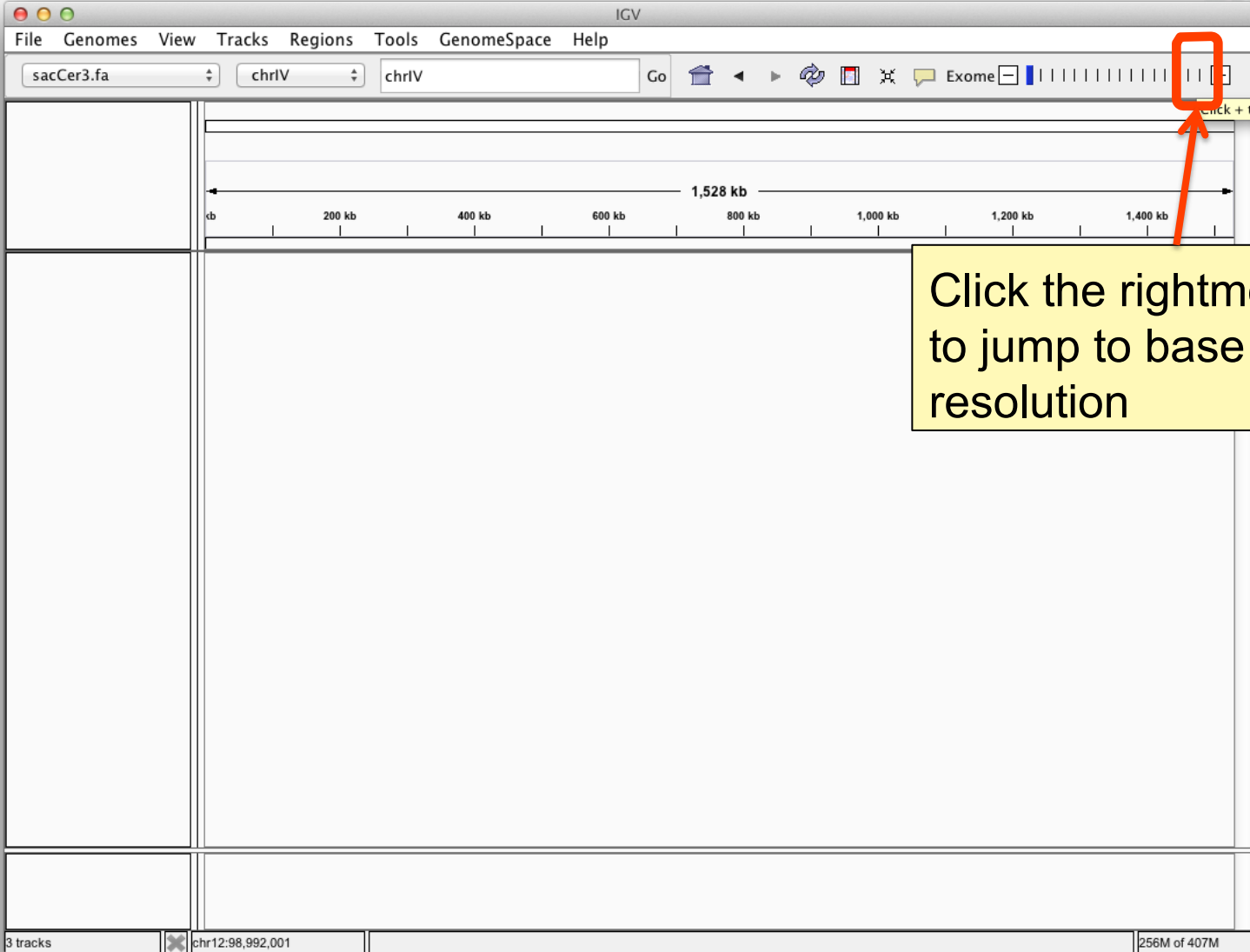
Annotations in the image include red boxes around the 'All' dropdown and chromosome 'IV', and red arrows pointing from a yellow text box to these elements.

Click on a chromosome name or use the pulldown menu to jump to the chromosome view.

# Loading a genome



# Loading a genome



File Genomes View Tracks Regions Tools GenomeSpace Help

sacCer3.fa chrV chrV Go Exome

1,528 kb

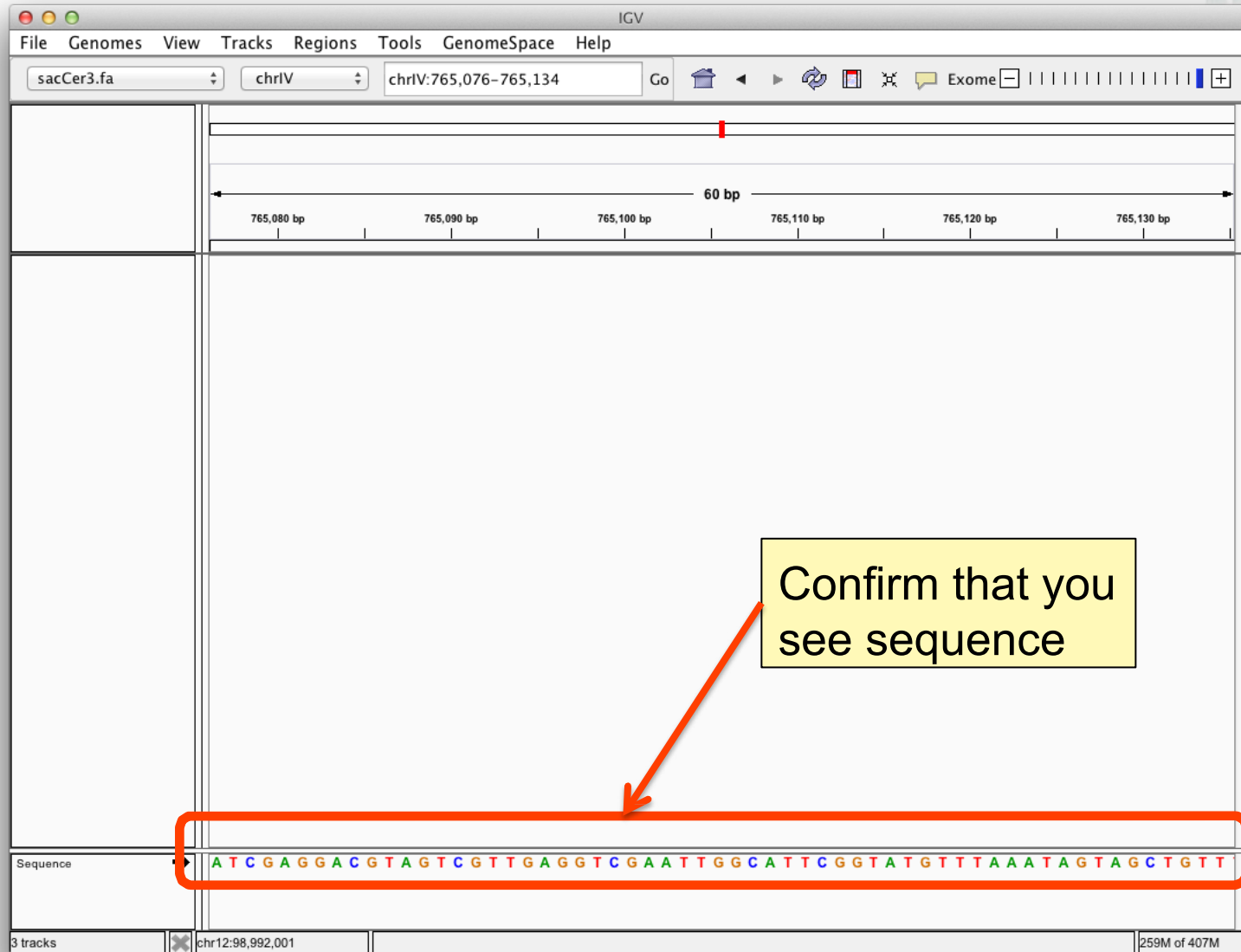
200 kb 400 kb 600 kb 800 kb 1,000 kb 1,200 kb 1,400 kb

Click + to zoom in, - to zoom out

Click the rightmost "tick" to jump to base pair resolution

3 tracks chr12:98,992,001 256M of 407M

# Loading a genome



The screenshot shows the IGV interface with the following details:

- File: sacCer3.fa
- Genomes: chrIV
- View: chrIV:765,076-765,134
- Tools: Exome
- Scale: 60 bp
- Sequence: ATCGAGGACGTAGTCGTTGAGGTCGAATTGGCATTCCGGTATGTTTAAATAGTAGCTGTT

A yellow box with the text "Confirm that you see sequence" and an orange arrow points to the sequence track. The sequence track is also highlighted with an orange box.

# Acknowledgments

---



## IGV Team

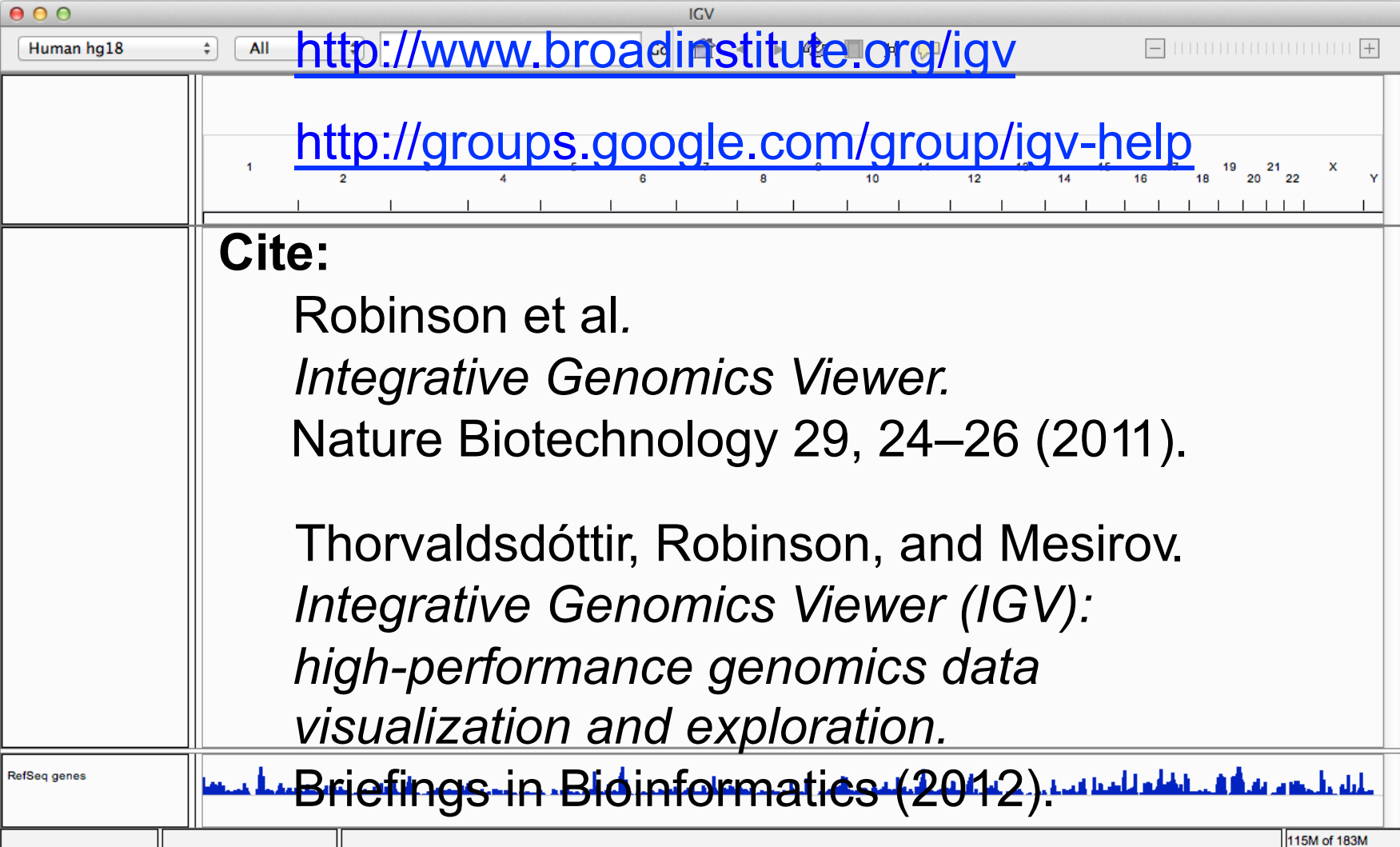
Jim Robinson, Jacob Silterra, Helga Thorvaldsdóttir, Jill Mesirov (PI)

## 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).

RefSeq genes

115M of 183M