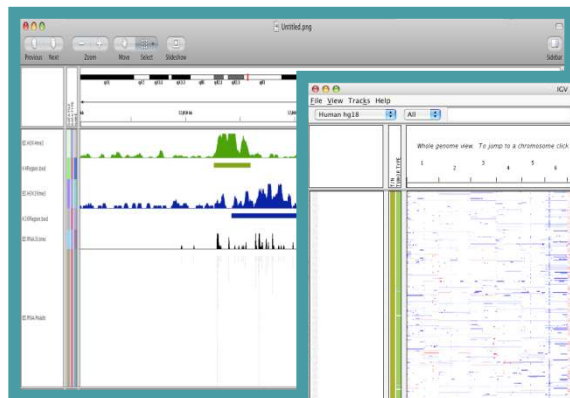


Introduction to NGS Visualization with the Integrative Genomics Viewer (IGV)

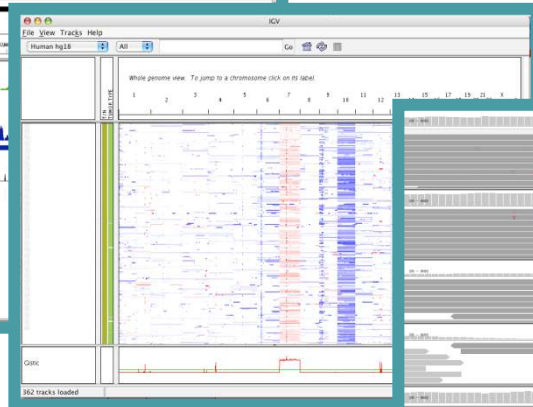
Integrative Genomics Viewer (IGV)



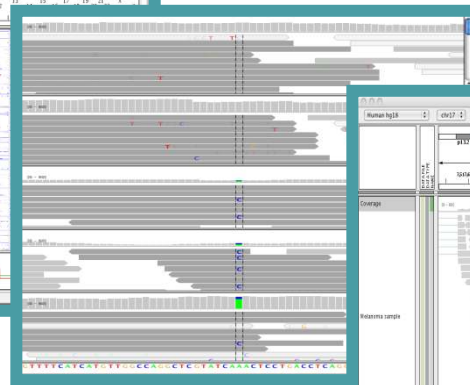
Desktop application for the interactive visual exploration of integrated genomic datasets



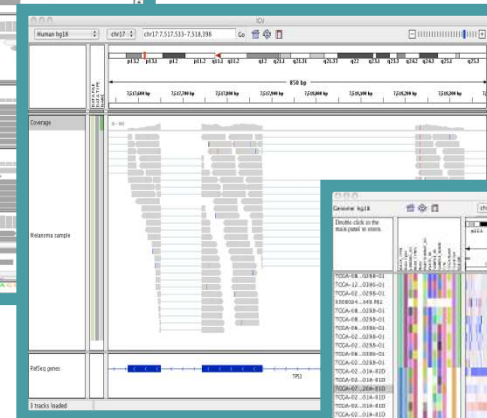
Epigenomics



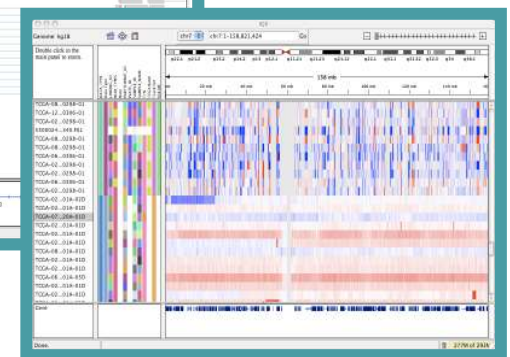
Microarrays



NGS alignments



RNA-Seq



mRNA, CNV, Seq

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

65,000 registrations

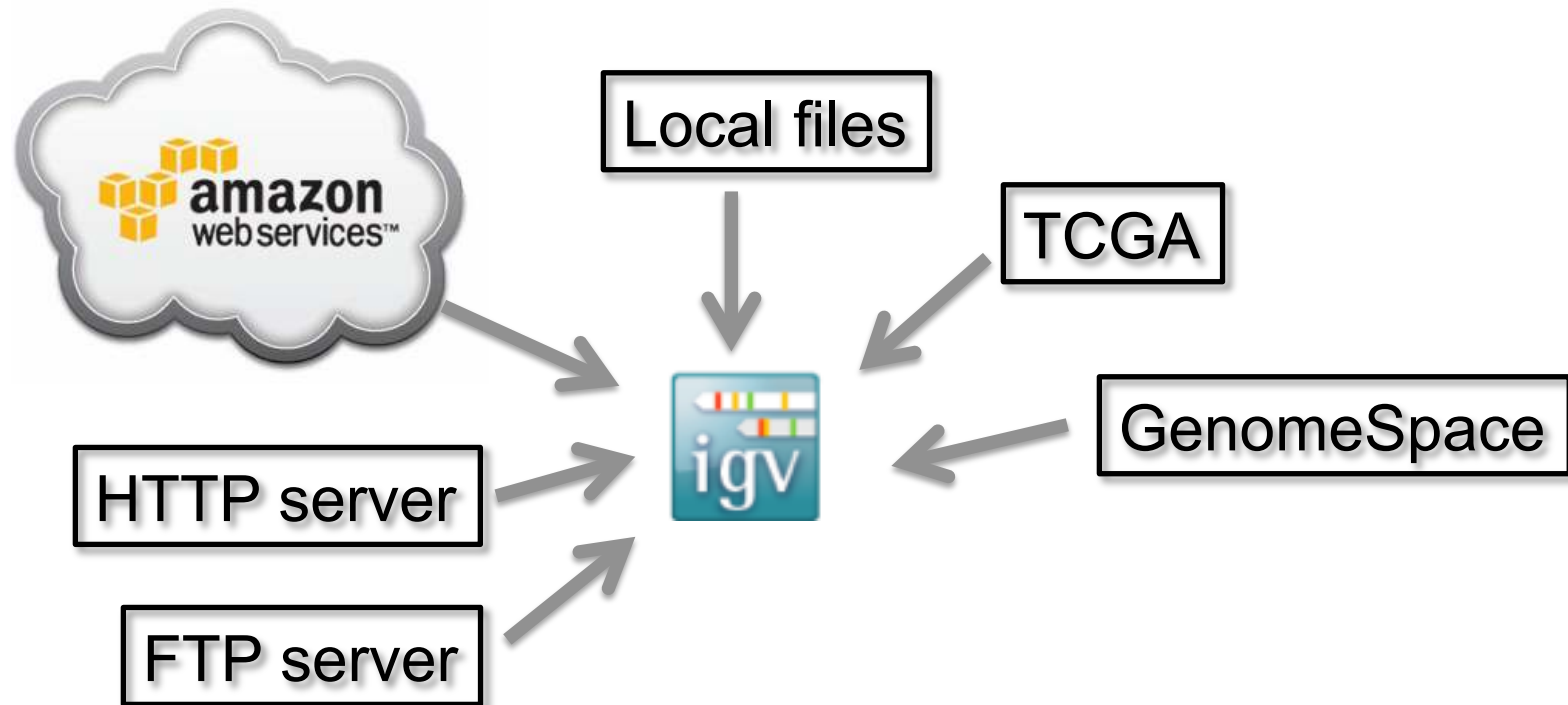
Features



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>



Launch IGV



Registration | Integrative Genomics Viewer

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

Reader

Integrative Genomics Viewer

- 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

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

Launch IGV



The screenshot shows the IGV download page on the Broad Institute website. The browser address bar shows www.broadinstitute.org/igv/download. The page has a sidebar on the left with navigation links: Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. The main content area is titled 'Downloads | Integrative Genomics Viewer' and 'Downloads'. It features a section for 'Integrative Genomics Viewer (Version 2.3)' with instructions for Mac, Java, Chrome, and Windows users. Below the instructions are four 'Launch' buttons with different memory requirements: 750 MB, 1.2 GB, 2 GB, and 10 GB. Red arrows point to the 'Launch' buttons and the sidebar. Below the launch buttons are links for 'Nightly Build' and 'Archived Versions'. Further down are sections for 'igvtools', 'Download', and 'Source Code'.

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 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/>

Launch IGV

A screenshot of a web browser window showing the 'Downloads | Integrative Genomics Viewer' page. The browser's address bar shows 'www.broadinstitute.org/igv/download'. The page has a sidebar on the left with navigation links: Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below these is a search bar and the Broad Institute logo with the copyright notice '© 2013 Broad Institute'. The main content area is titled 'Downloads' and features a section for 'Integrative Genomics Viewer (Version 2.3)'. This section contains instructions for Mac, Java, Chrome, and Windows users. Below the instructions is a table with four 'Launch' buttons, each with a specific memory requirement. The second button, 'Launch with 1.2 GB', is circled in red. Below the table are links for 'Nightly Build' and 'Archived Versions'. Further down, there is a section for 'igvtools' with a link to 'igvtools 2.3.20.zip'. A 'Download' section follows, with a link to 'IGV 2.3.20.zip'. Finally, a 'Source Code' section includes a link to the GitHub repository: 'https://github.com/broadinstitute/IGV/'.

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	Launch	Launch
Launch with 750 MB	Launch with 1.2 GB	Launch with 2 GB	Launch with 10 GB
	Maximum usable memory for Windows OS with 32-bit Java.	Maximum usable memory for 32-bit MacOS.	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](#)

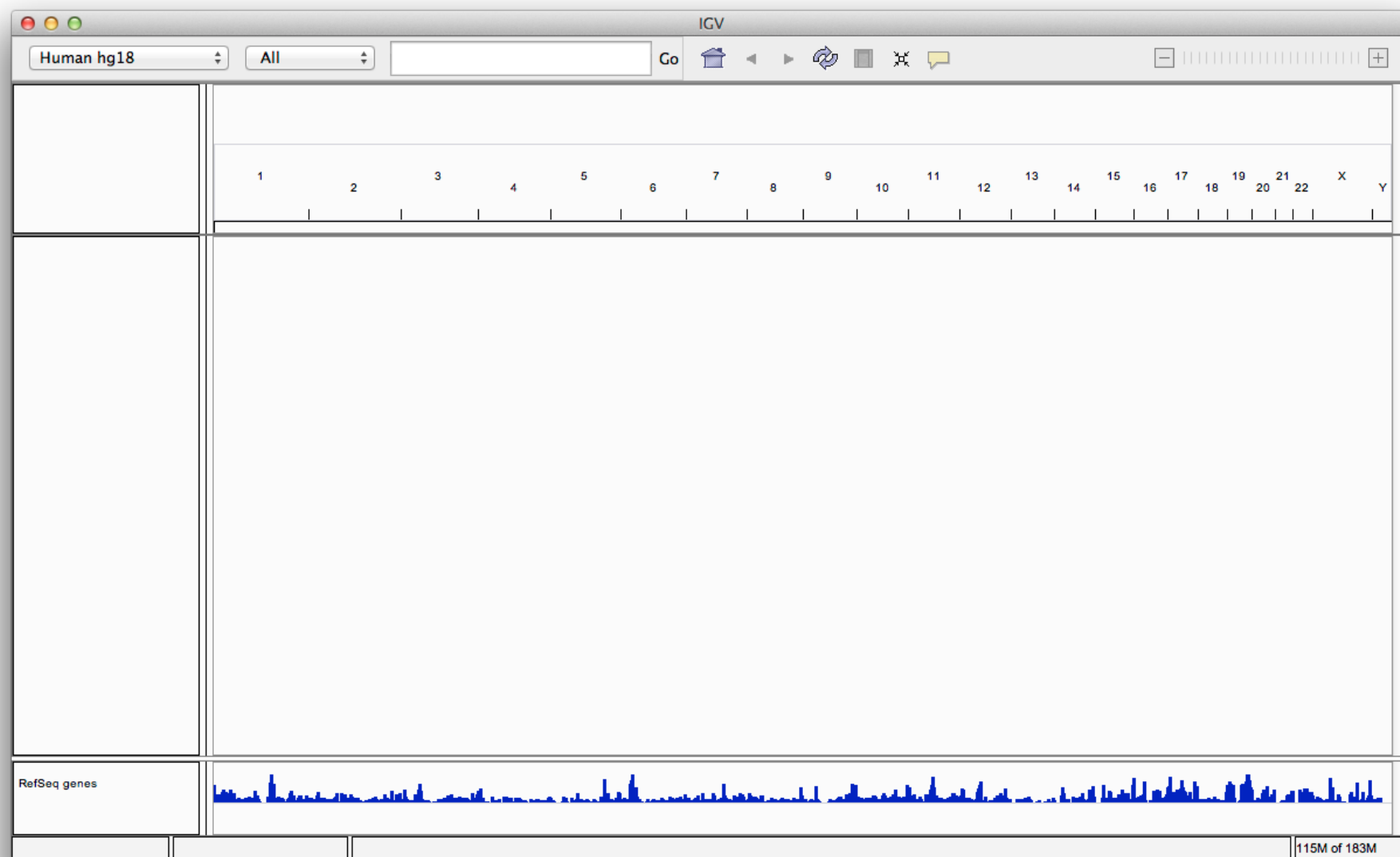
Source Code

Source distribution archive:

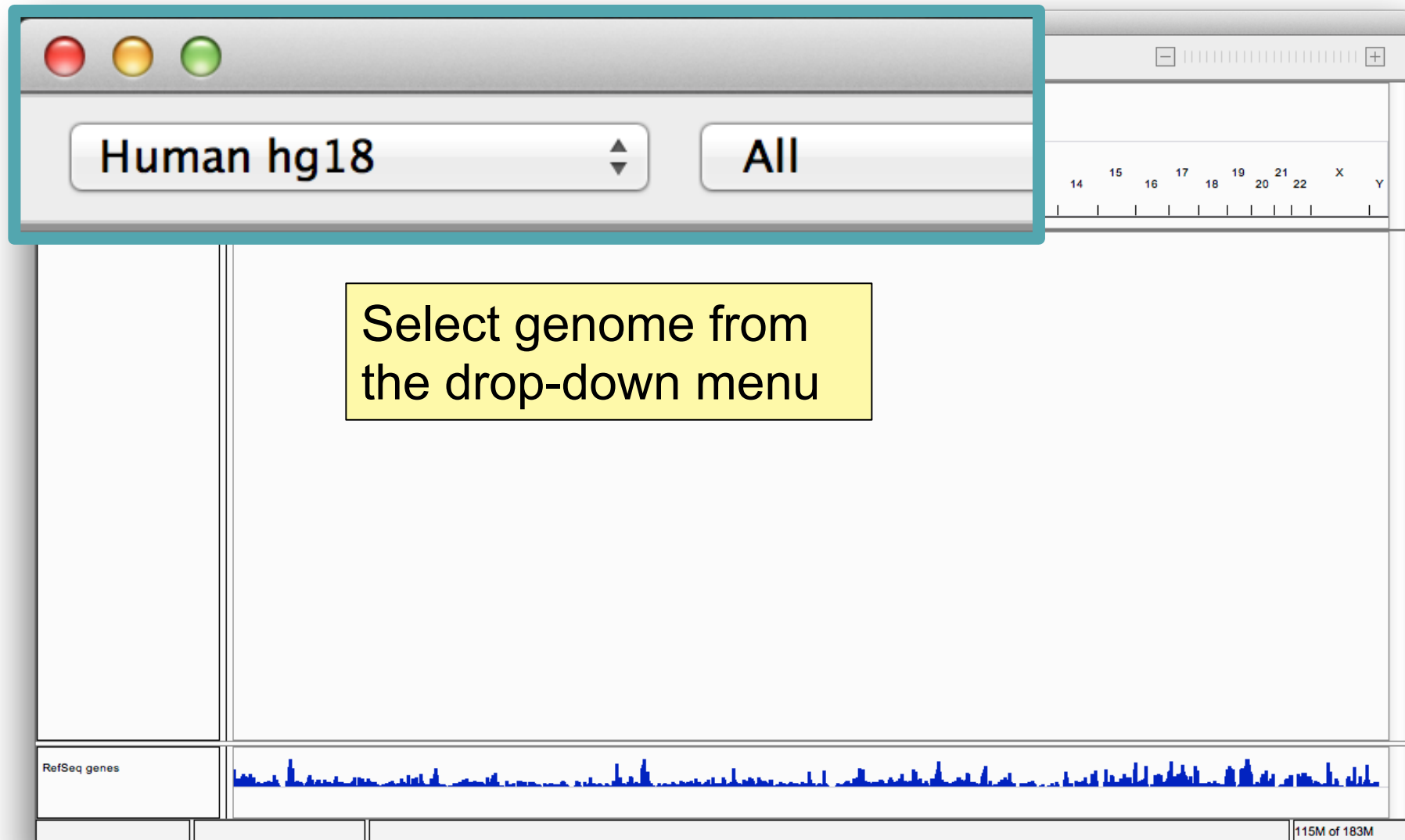
- [v2.3.20.zip](#)

Source code repository is hosted at github:

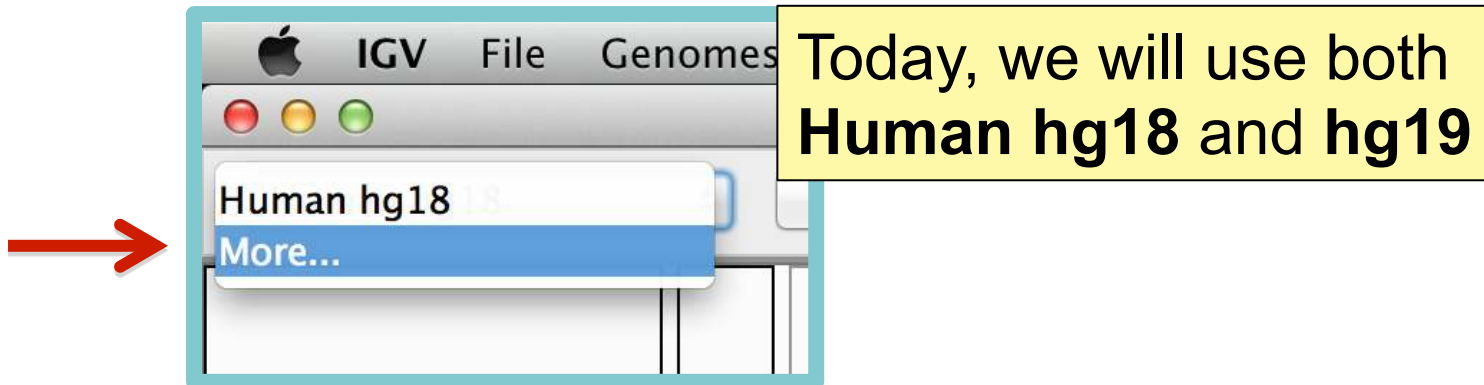
- <https://github.com/broadinstitute/IGV/>



Select the reference genome

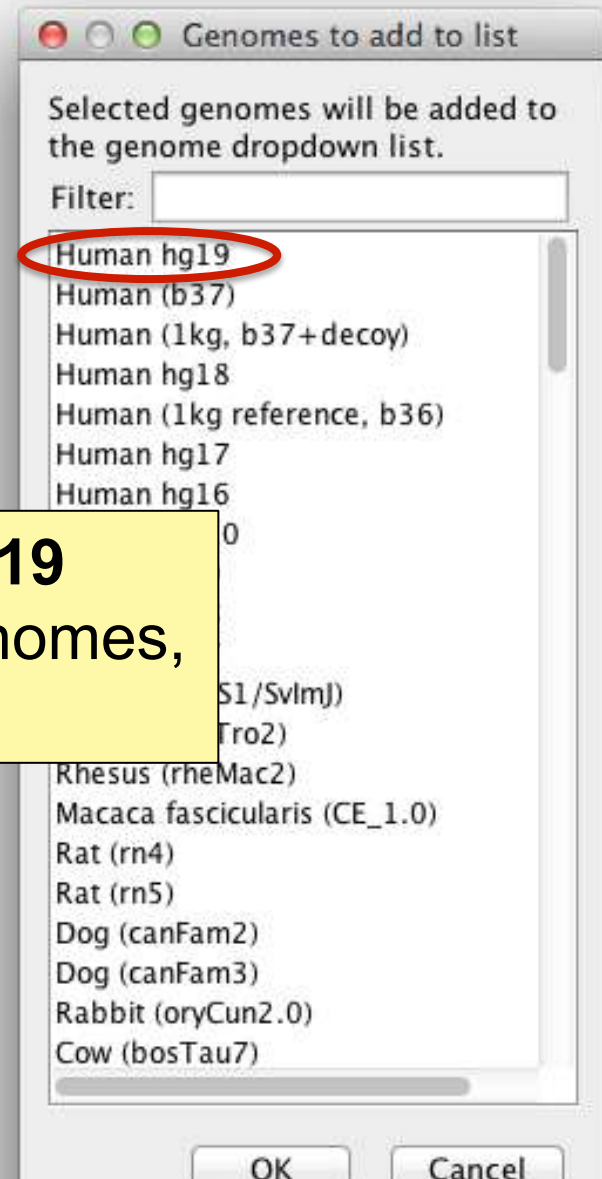
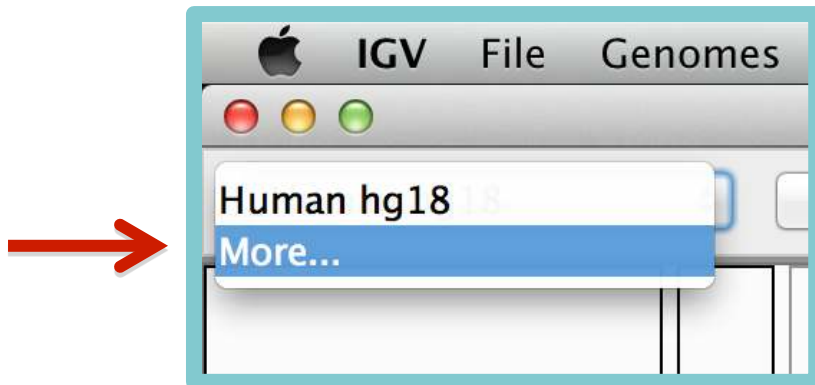


Select the reference genome



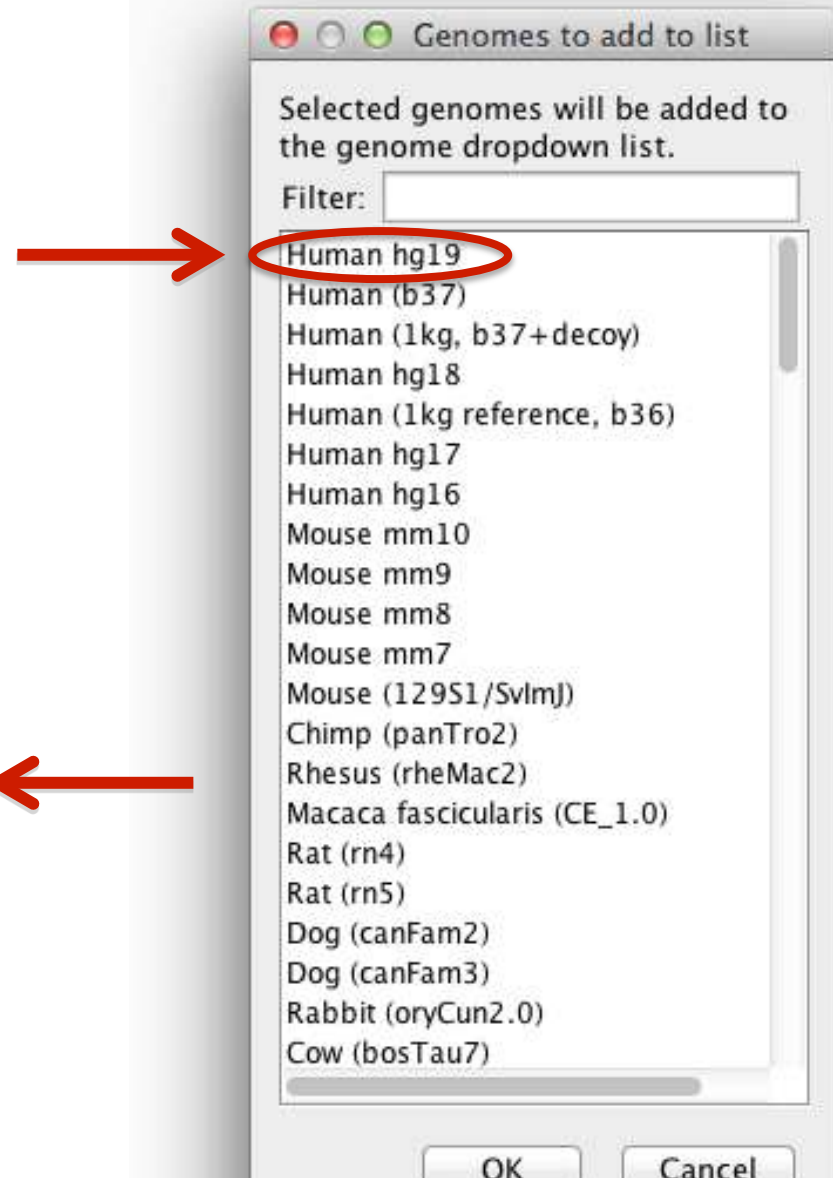
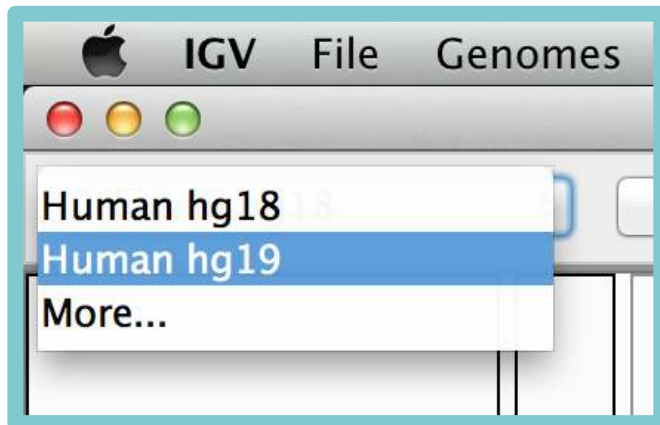
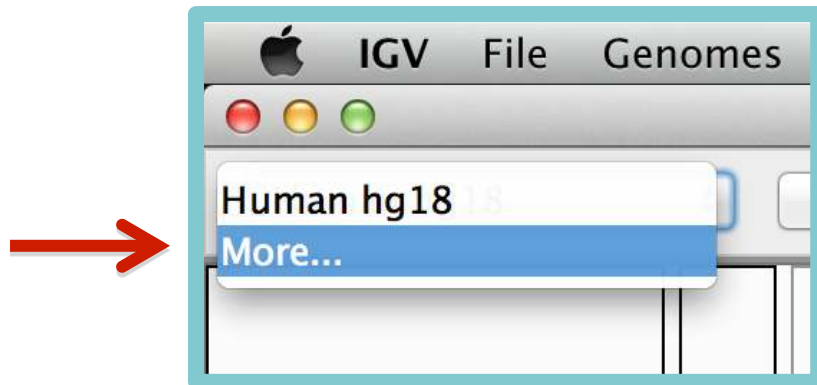
If **Human hg19** is not in the menu,
then click on ***More...***

Select the reference genome



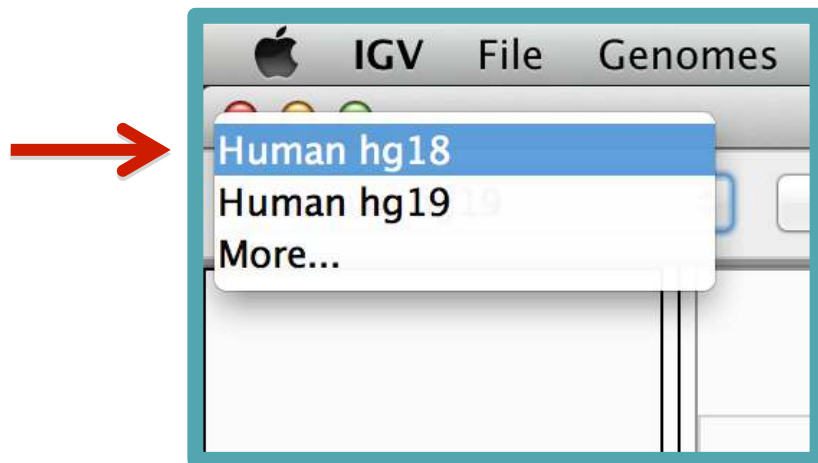
Select **Human hg19**
from the list of genomes,
and click **OK**

Select the reference genome



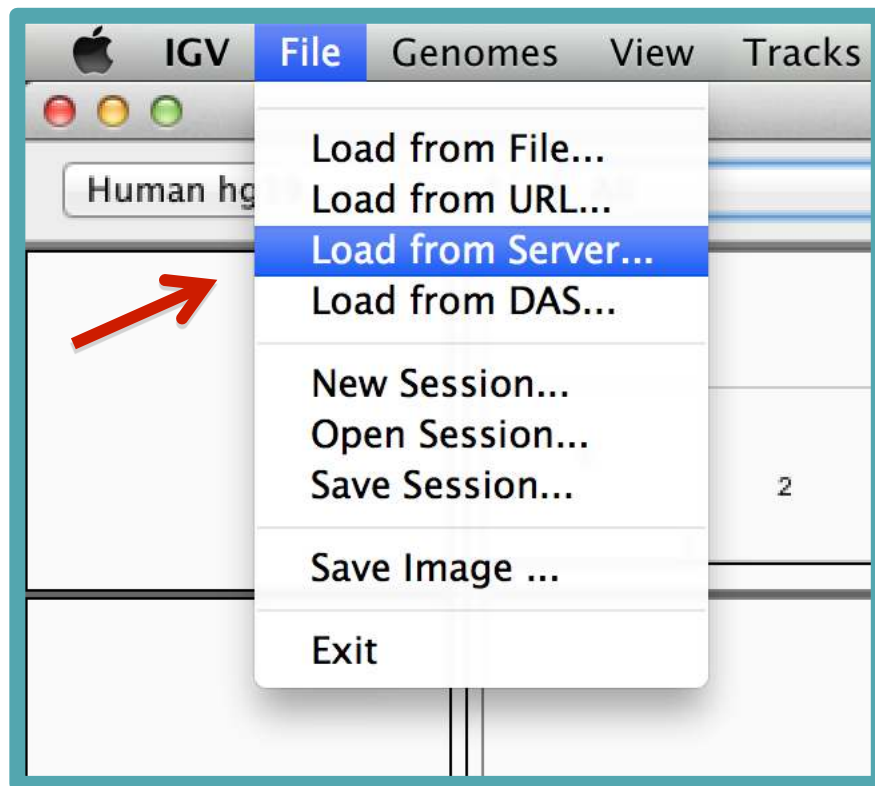
Select the reference genome

Select **Human hg18**



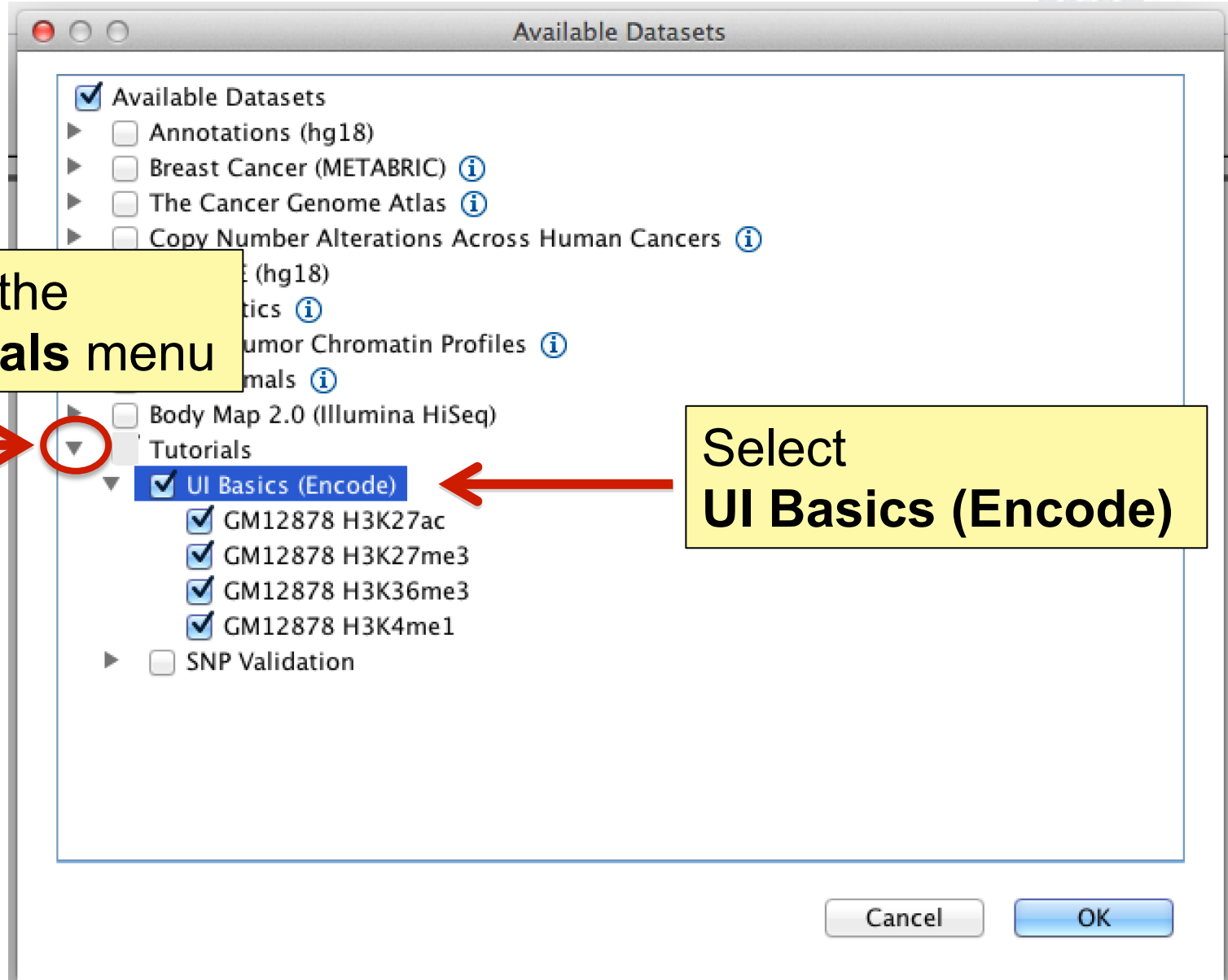
Load data

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



Load data

Open the
Tutorials menu



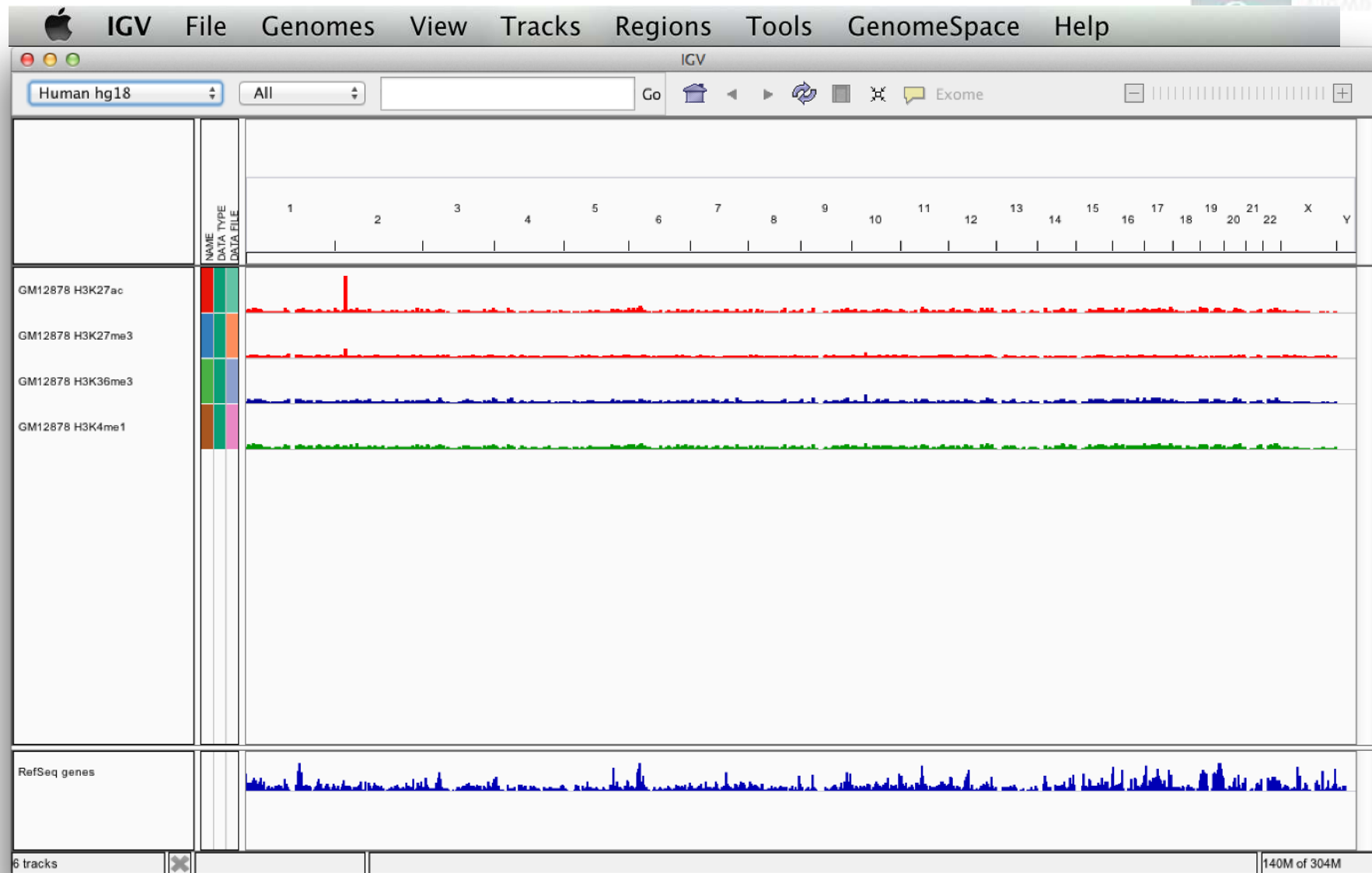
Select
UI Basics (Encode)



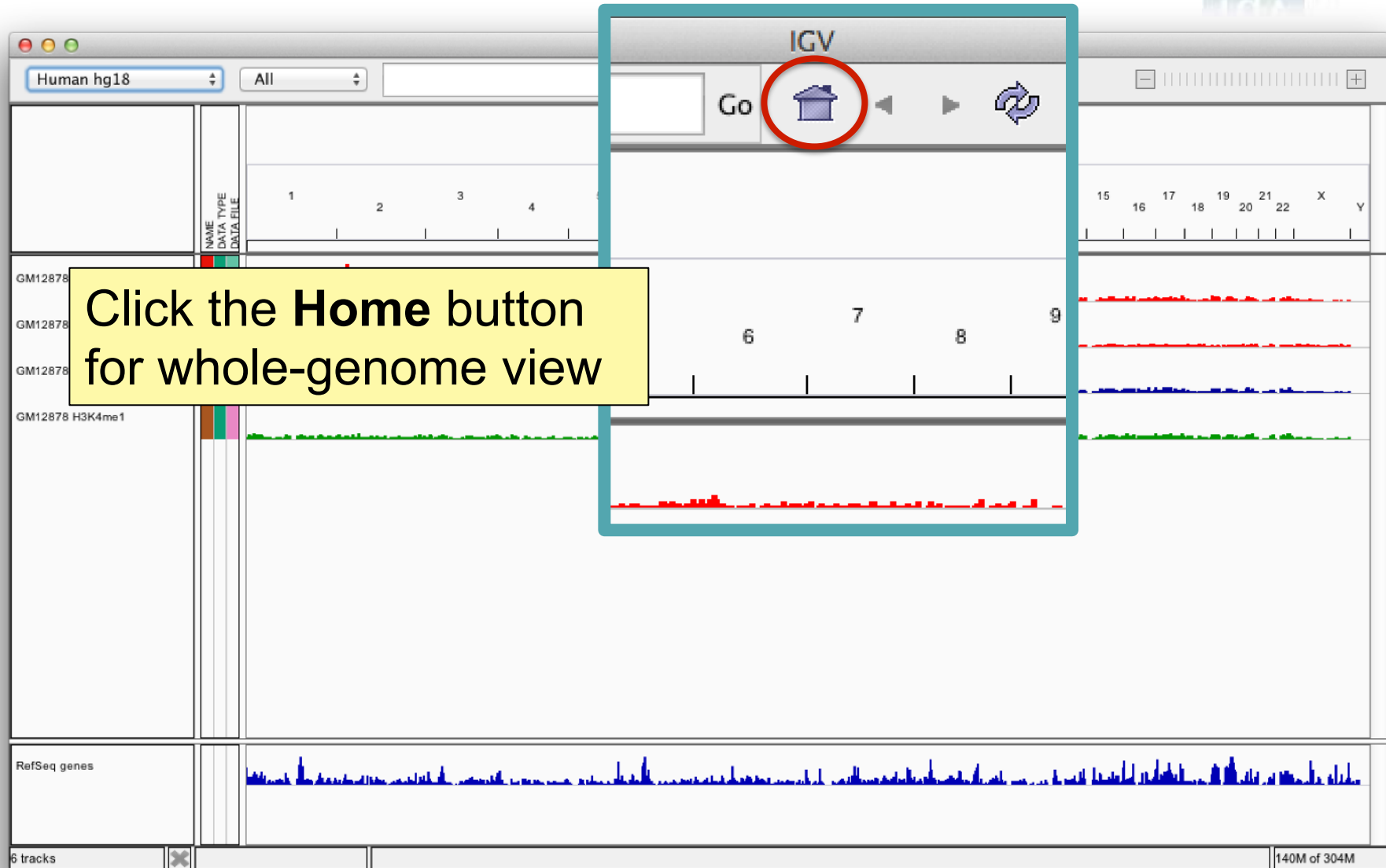
Cancel

OK

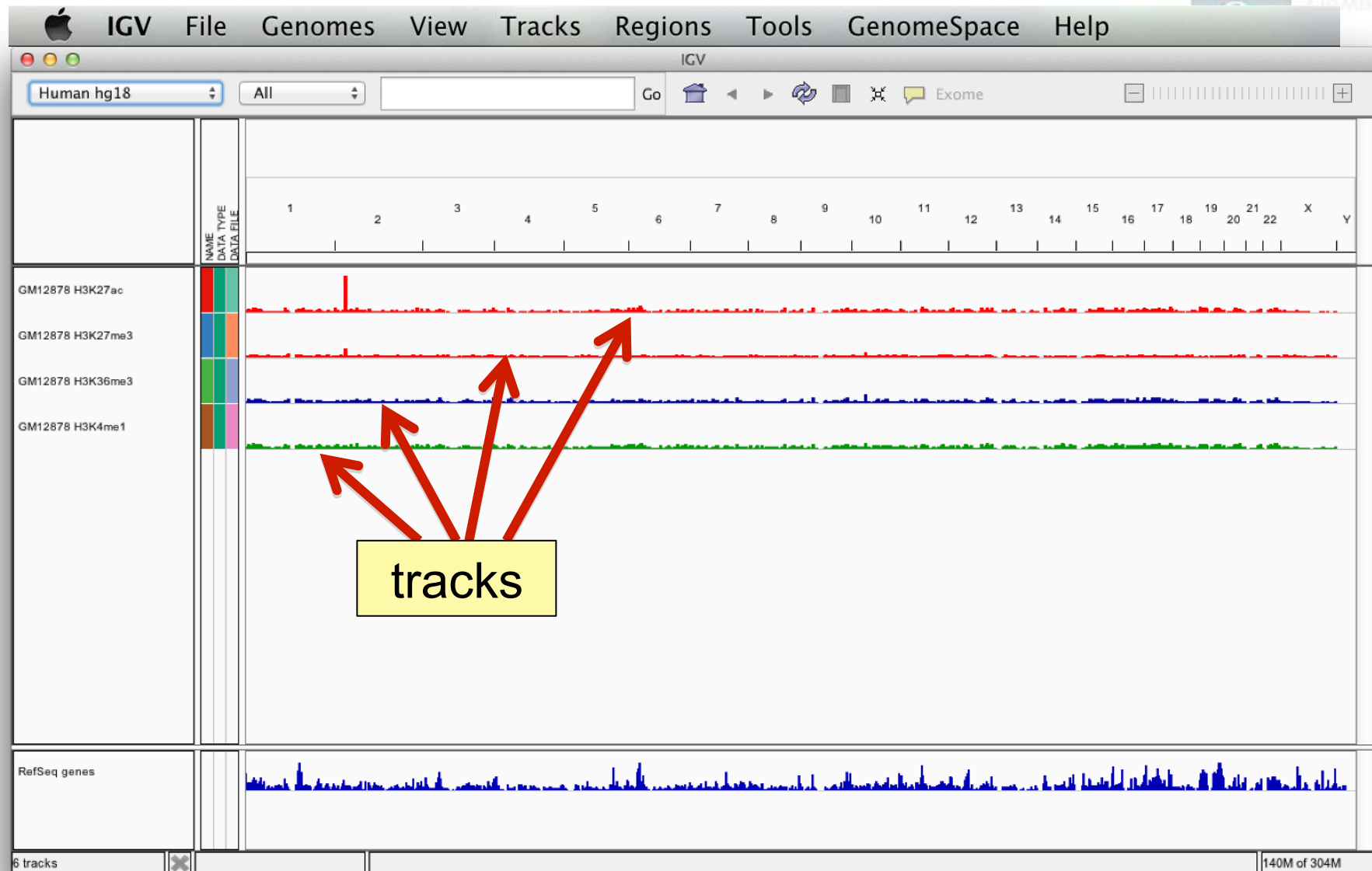
Screen layout



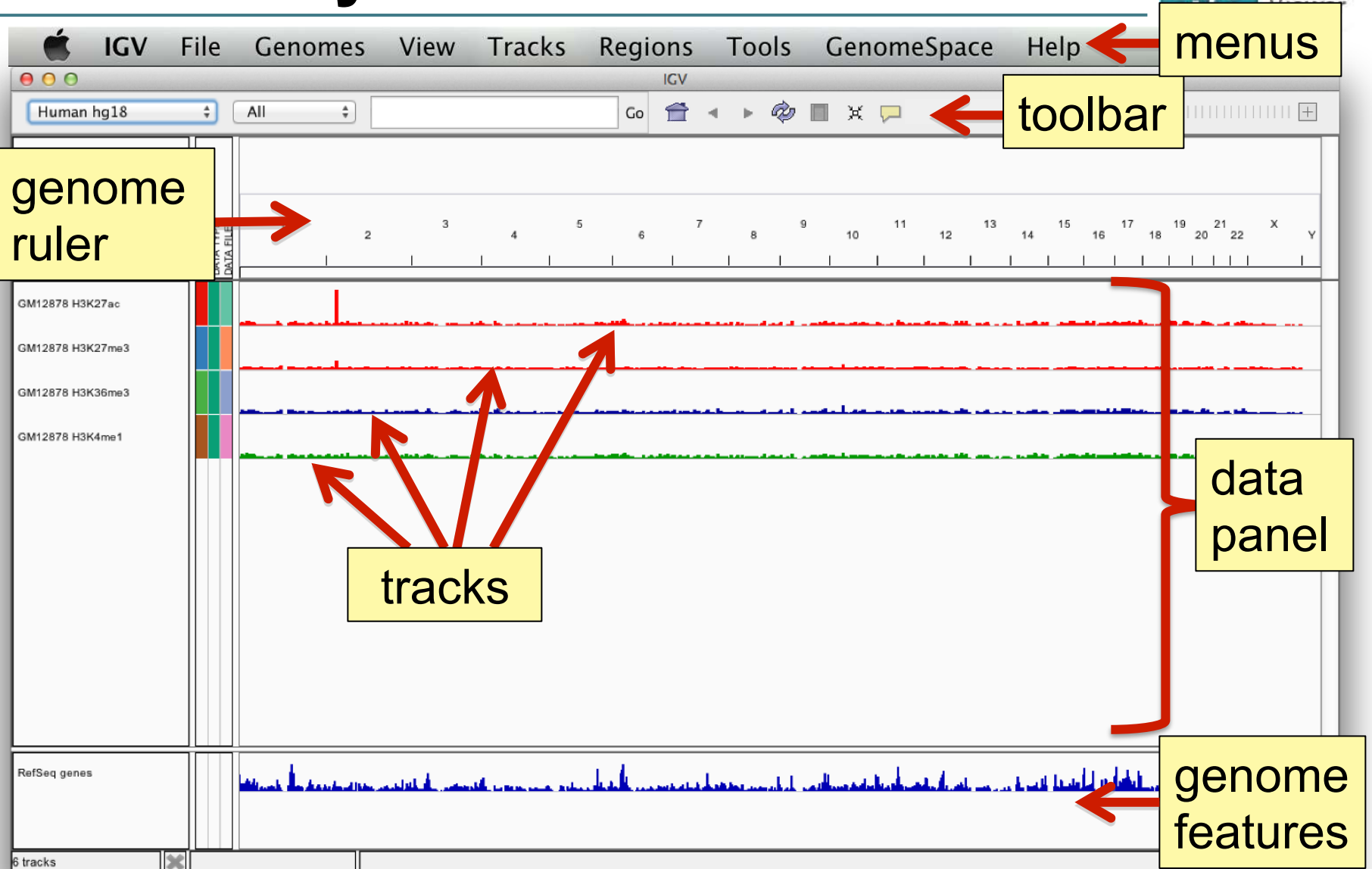
Screen layout



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

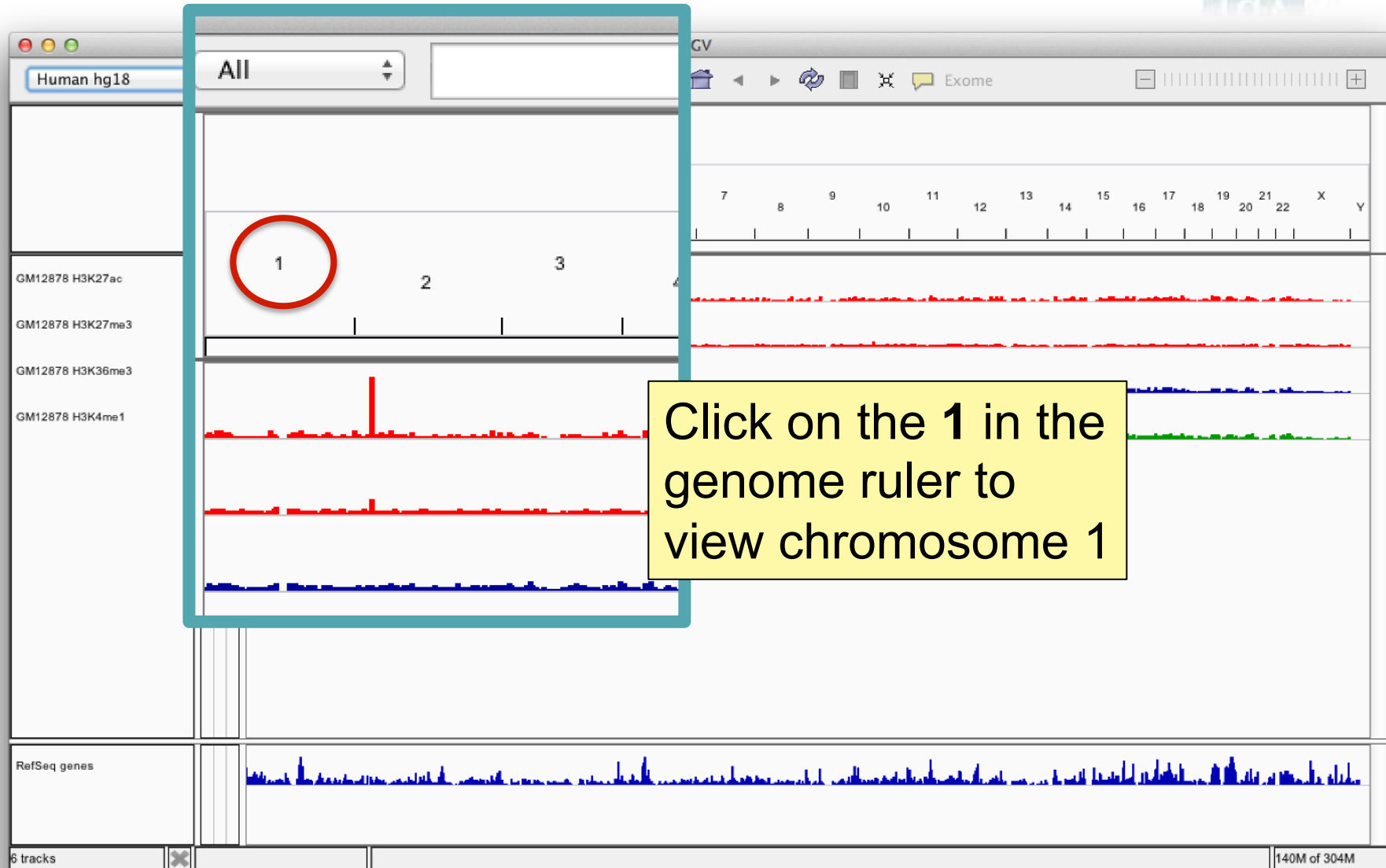


- The **file format** defines the track type.
- The **track type** determines the display options
- IGV supports many different file formats.

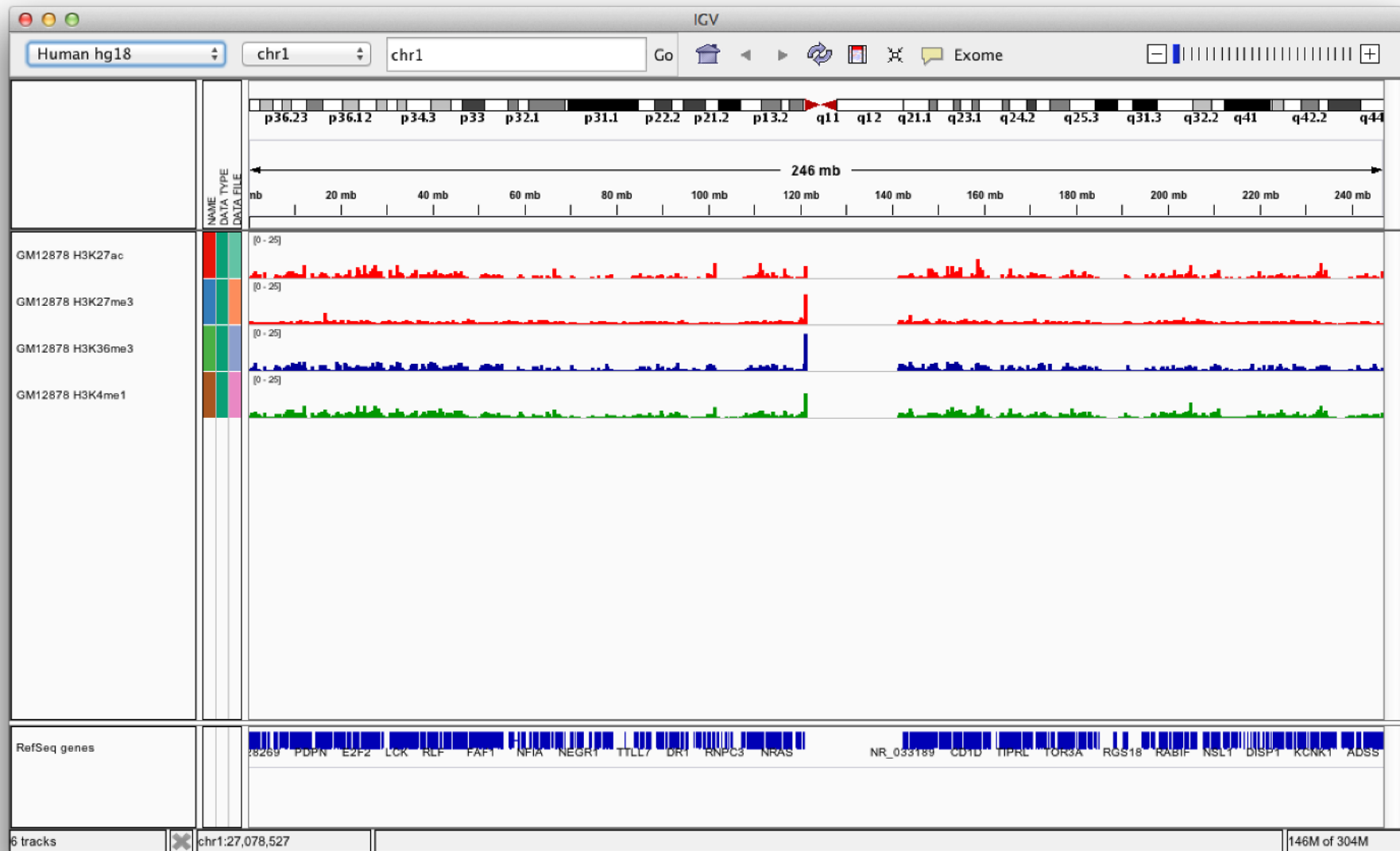
- | | | |
|---------------------------------------|--|--------------------------------------|
| ▪ BAM | ▪ GCT | ▪ PSL |
| ▪ BED | ▪ genePred | ▪ RES |
| ▪ BedGraph | ▪ GFF | ▪ SAM |
| ▪ bigBed | ▪ GISTIC | ▪ Sample Information |
| ▪ bigWig | ▪ Goby | ▪ SEG |
| ▪ Birdsuite Files | ▪ GWAS | ▪ SNP |
| ▪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

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

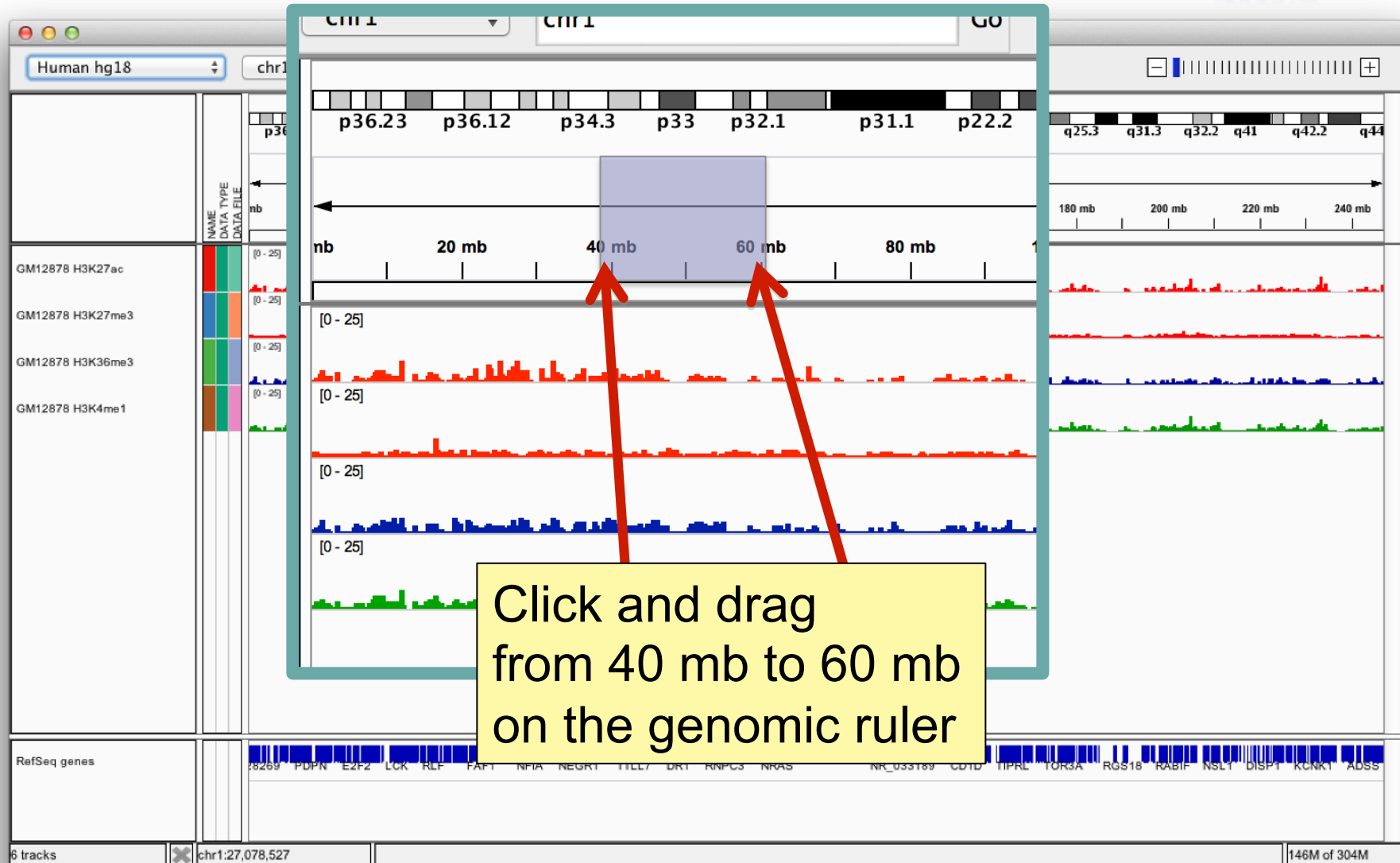
Navigate



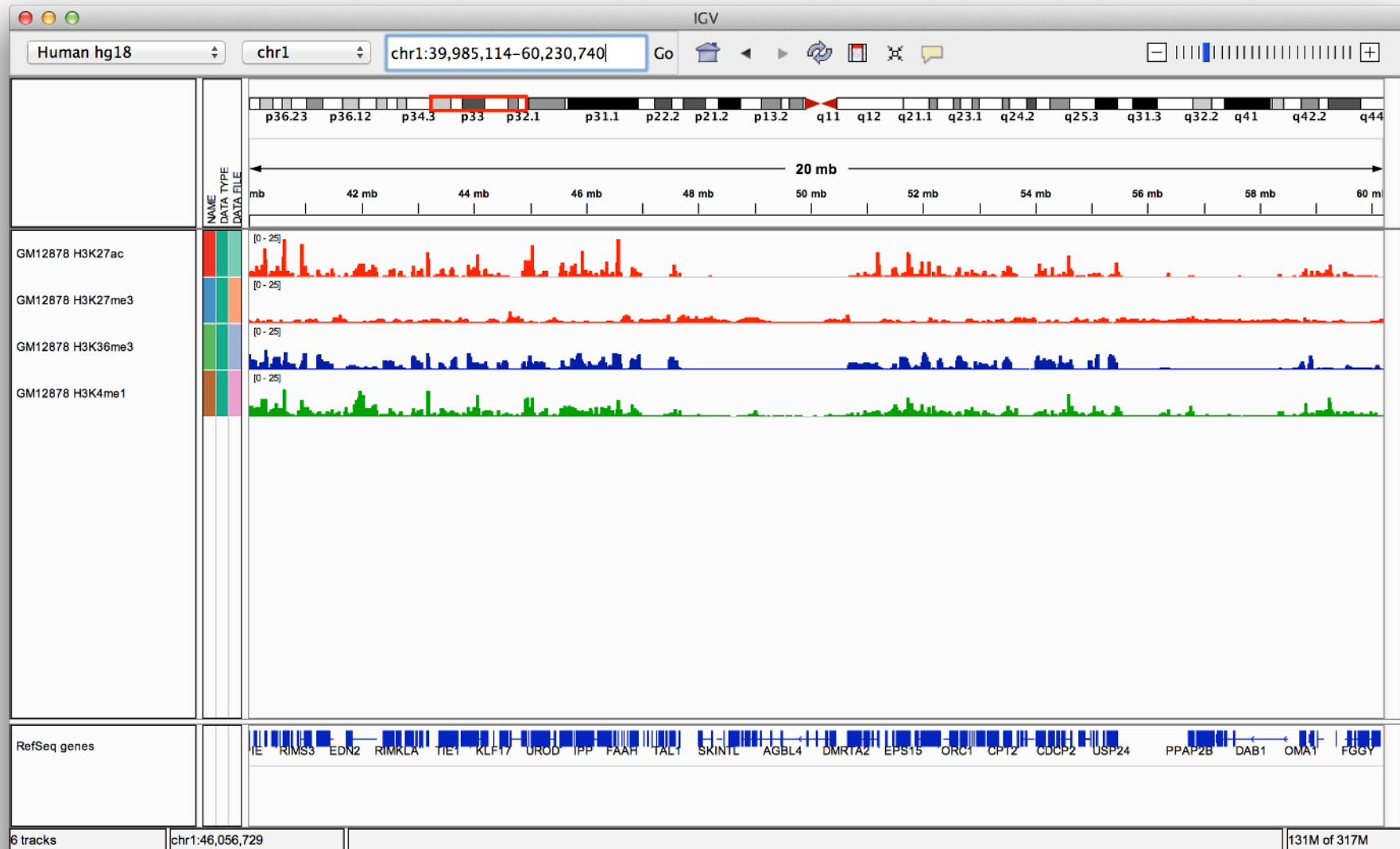
Navigate



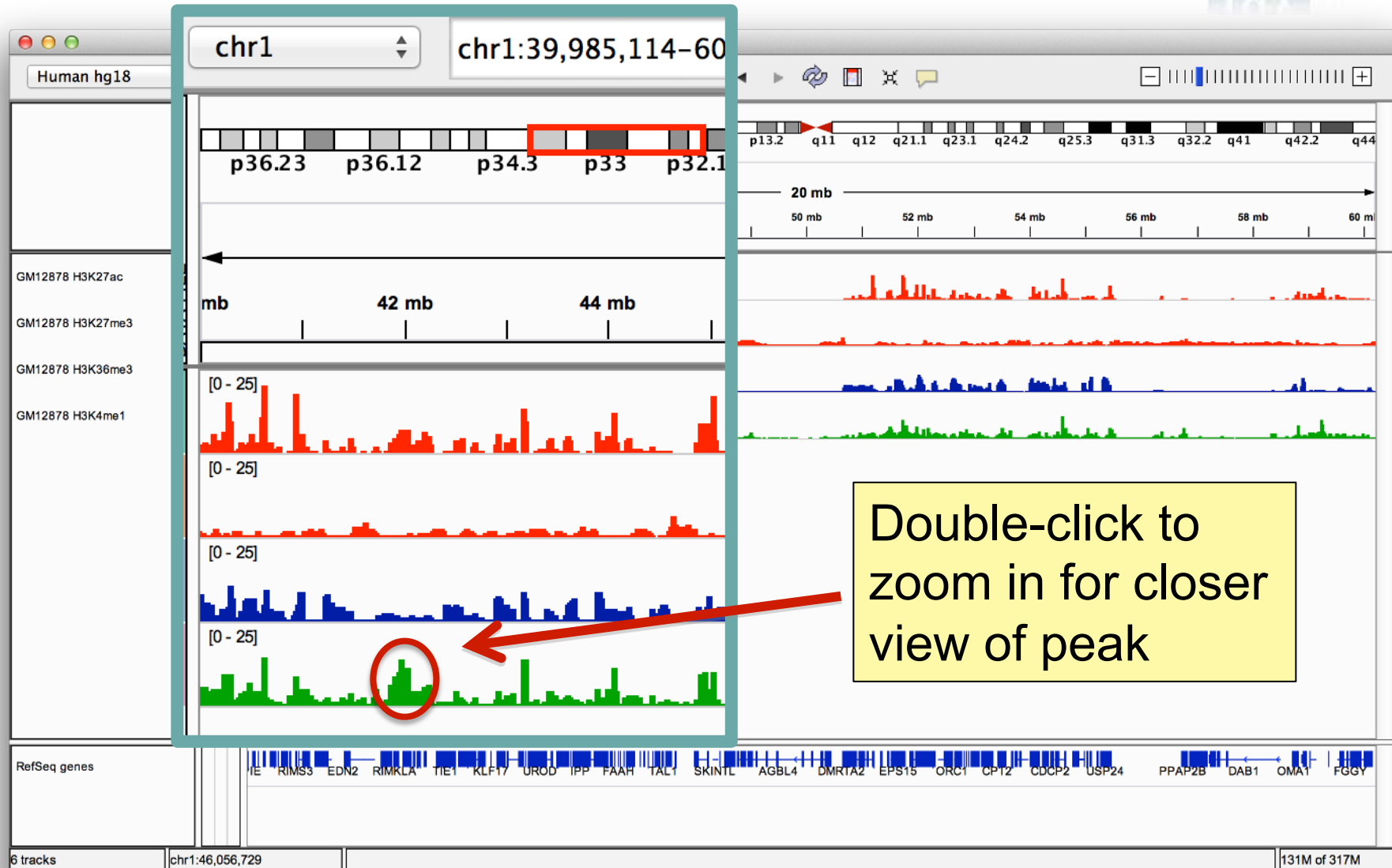
Navigate



Navigate



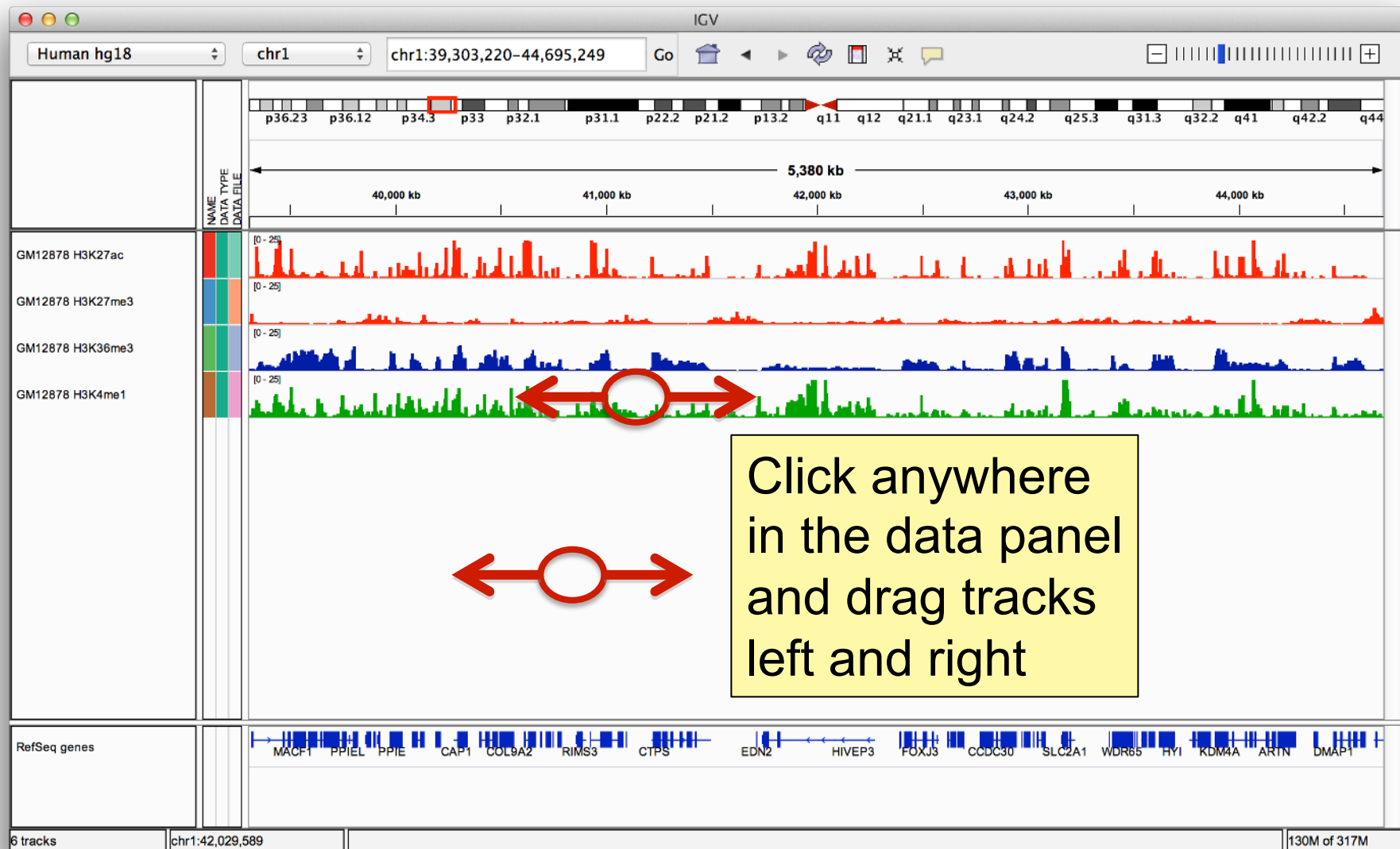
Navigate



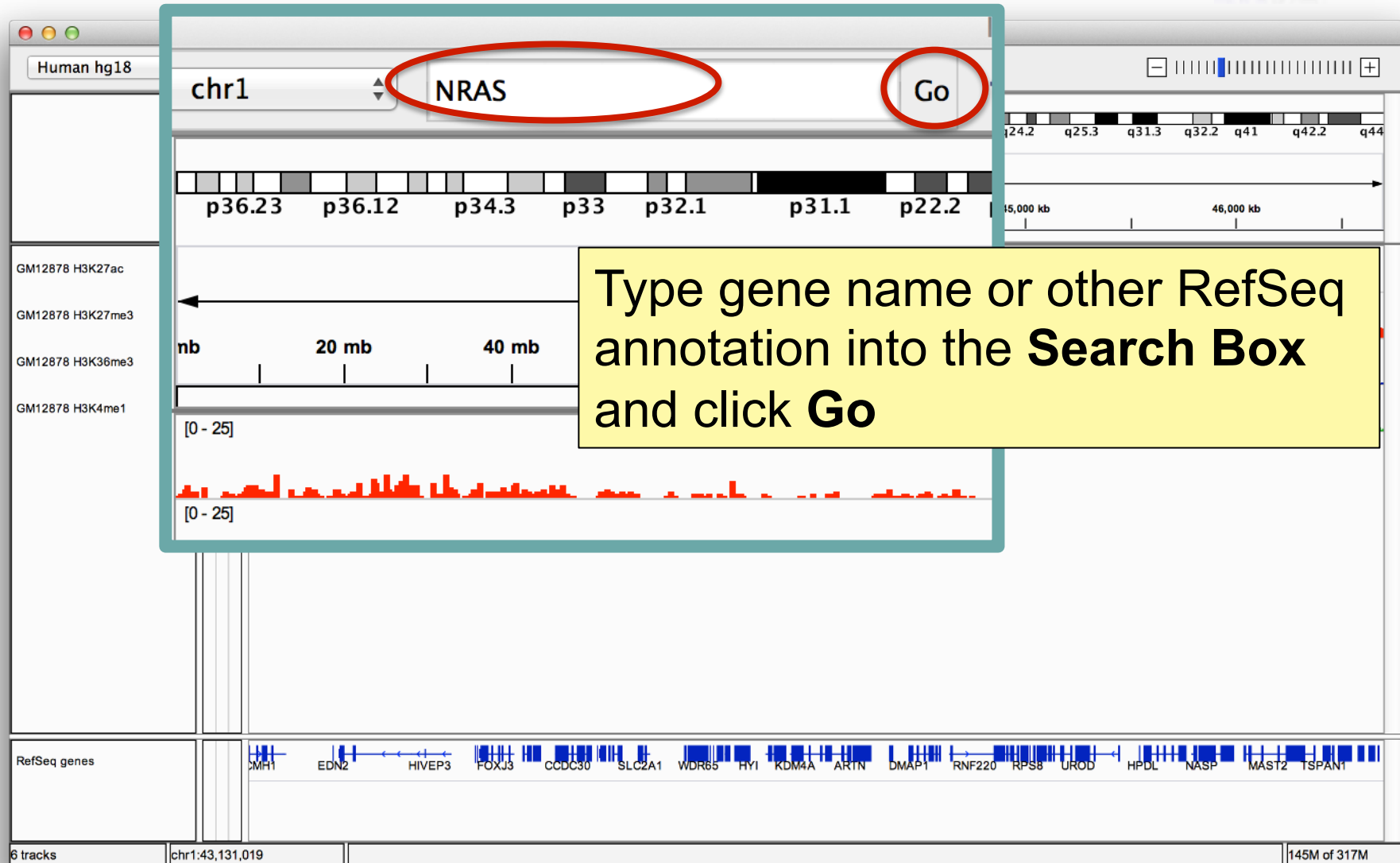
Navigate



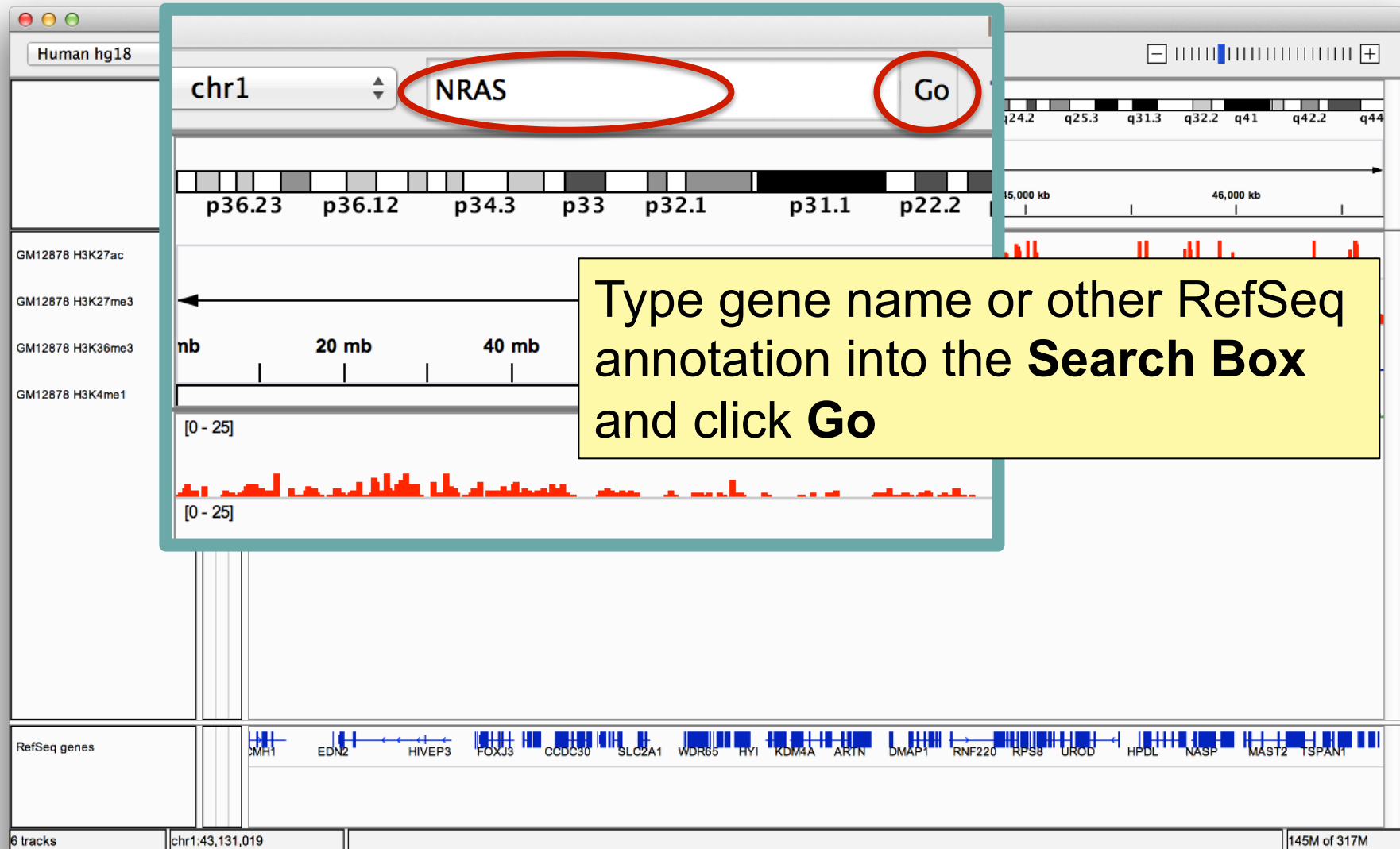
Navigate



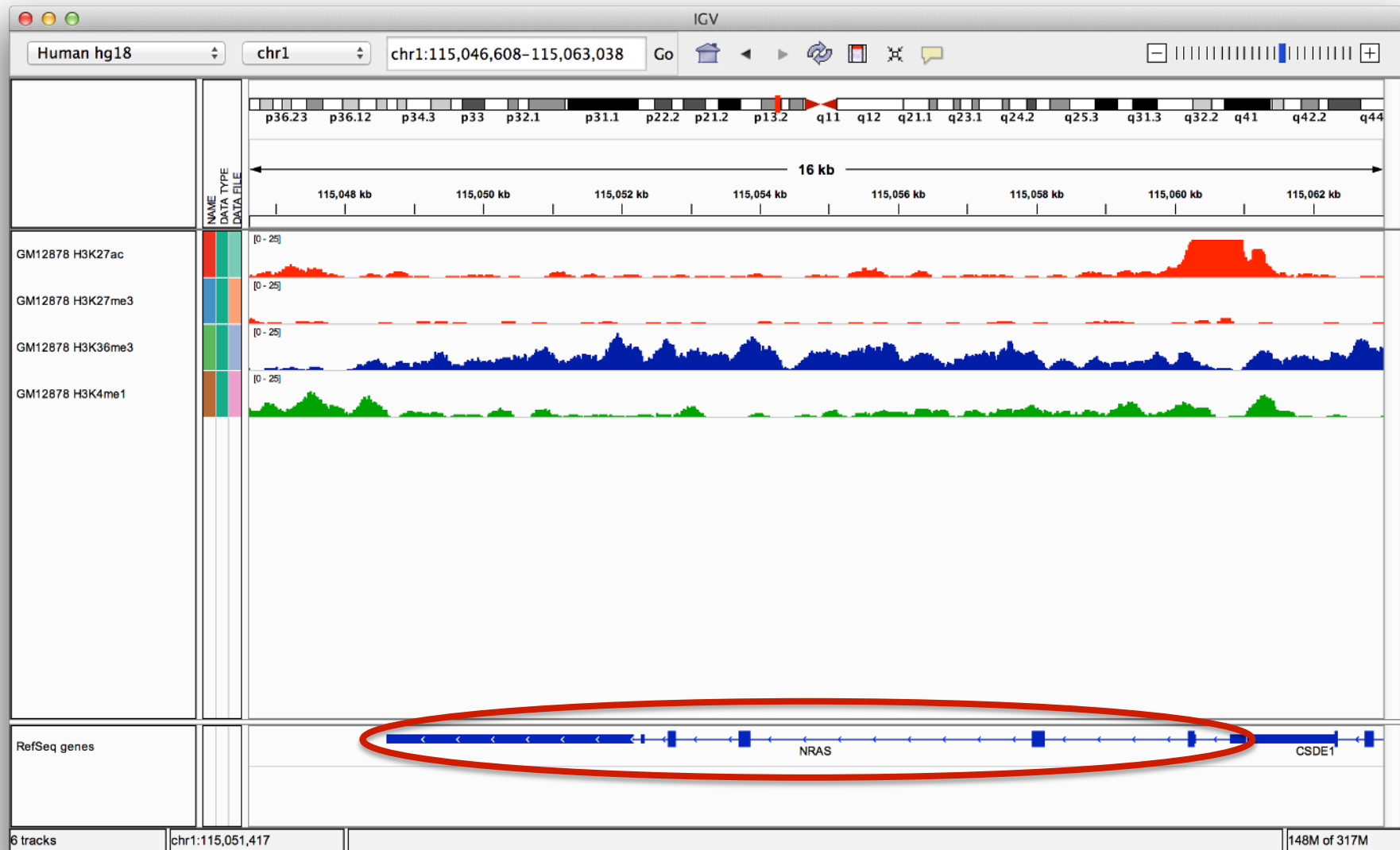
Navigate



Navigate



Navigate

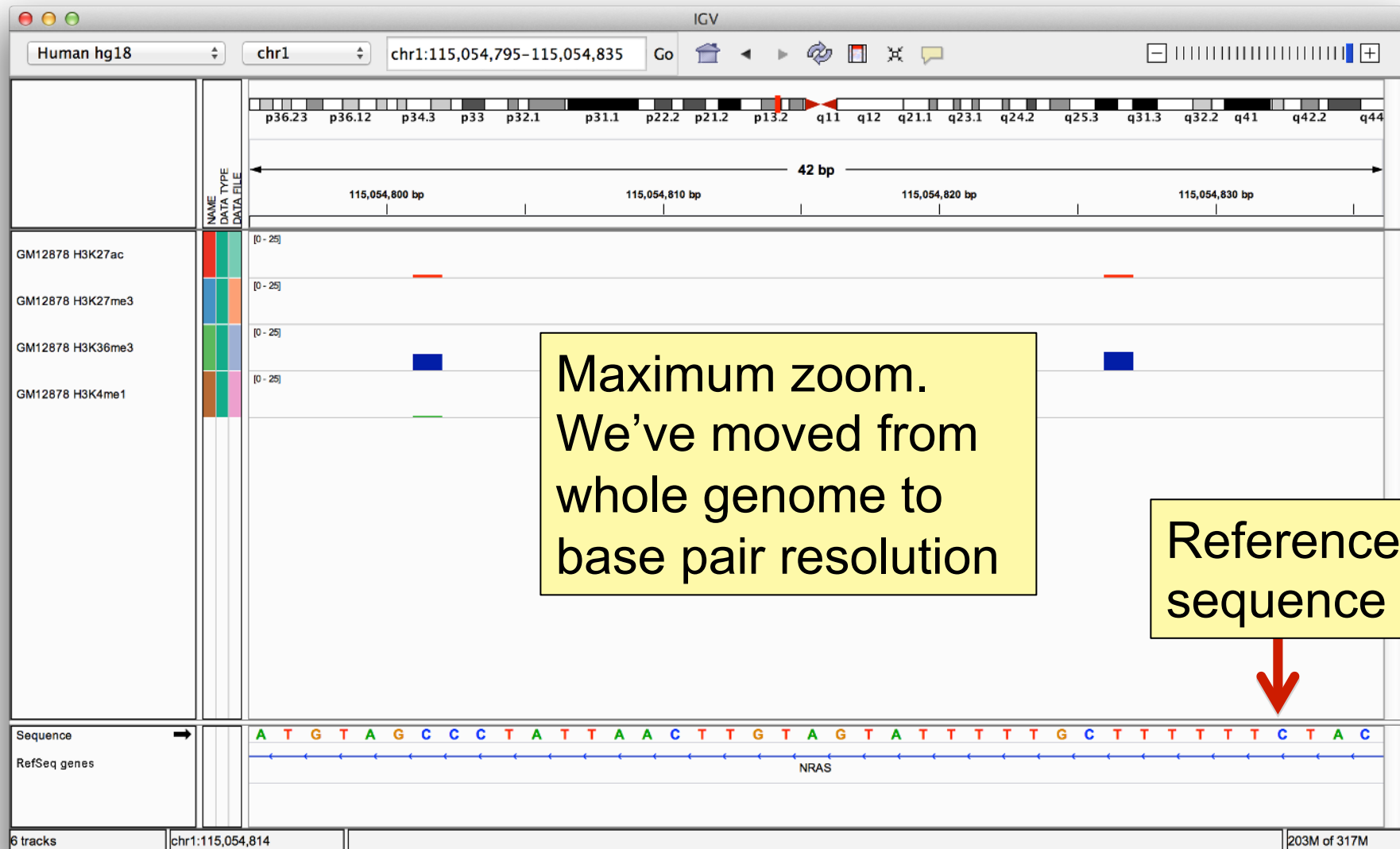


Navigate



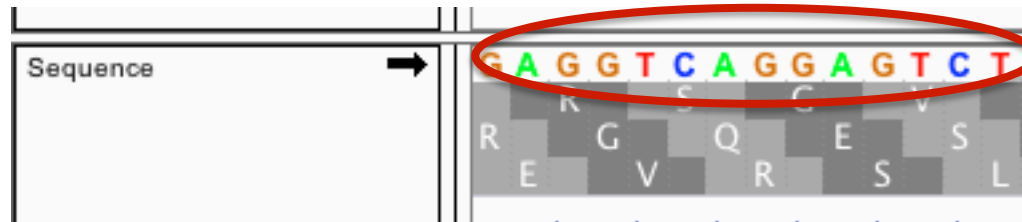
Click on the last tick on the
“railroad track” to zoom in
to maximum resolution

Navigate



Reference sequence

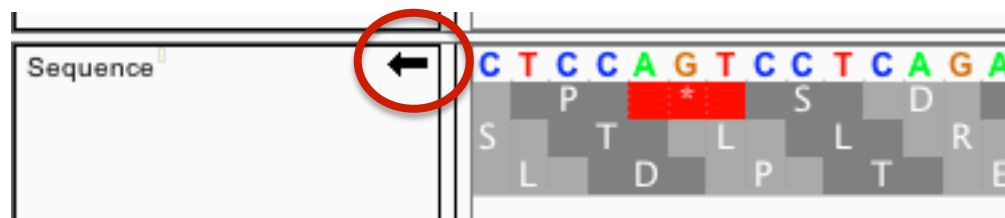
Click anywhere on the sequence to see a 3 frame translation.



By default the sequence for the forward strand is shown.



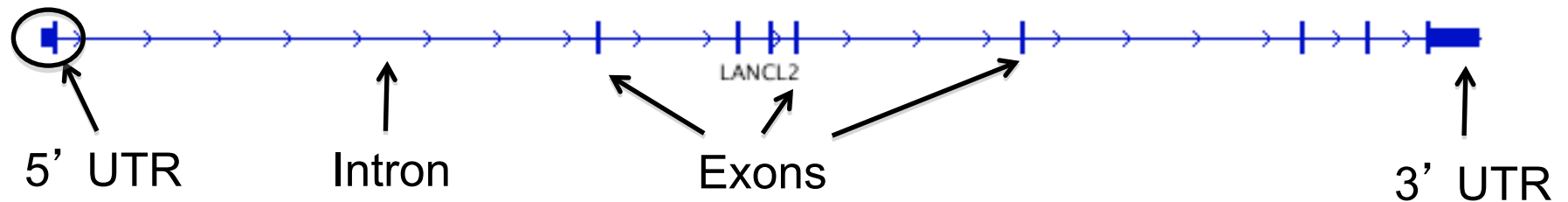
Click the arrow on the left to reverse the strand.



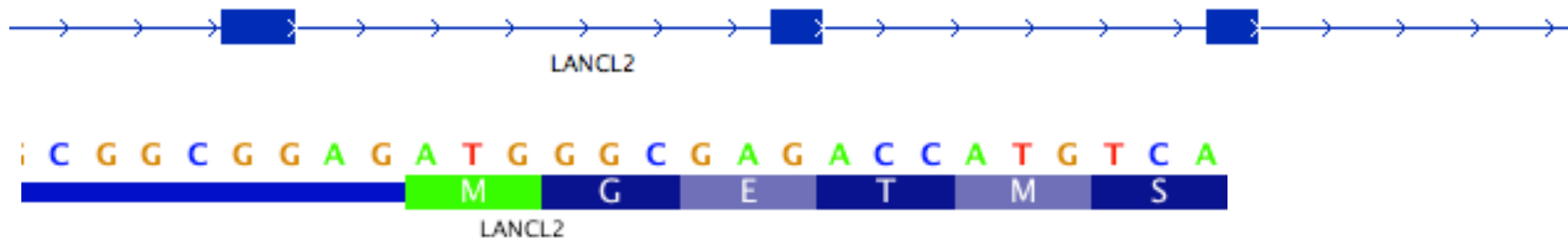
Genome annotation track



UCSC style gene representation



Zoomed in views



Zoomed out views

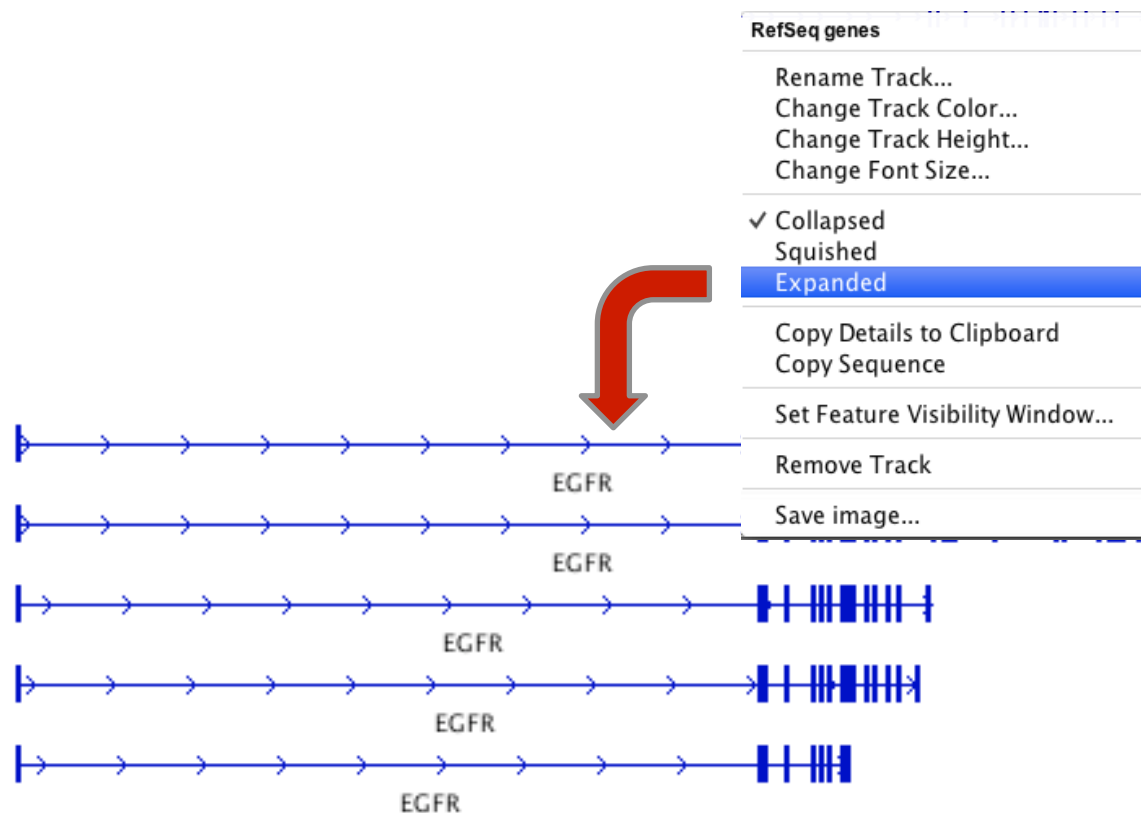


Annotation display mode

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

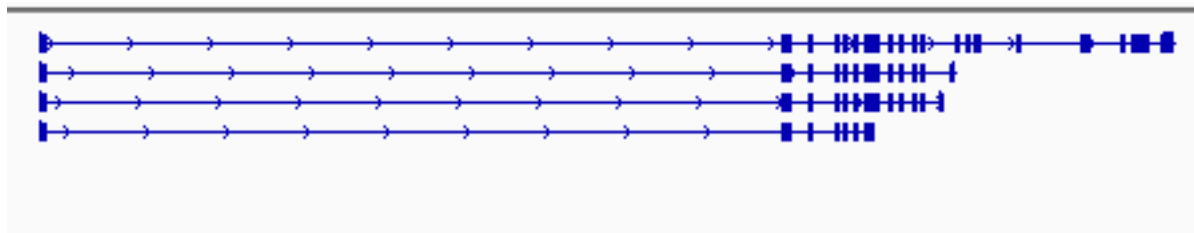
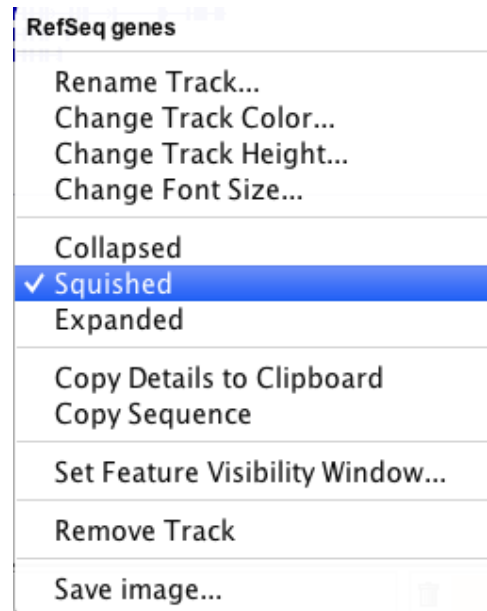


2. Expand the track using the popup menu



Annotation display mode

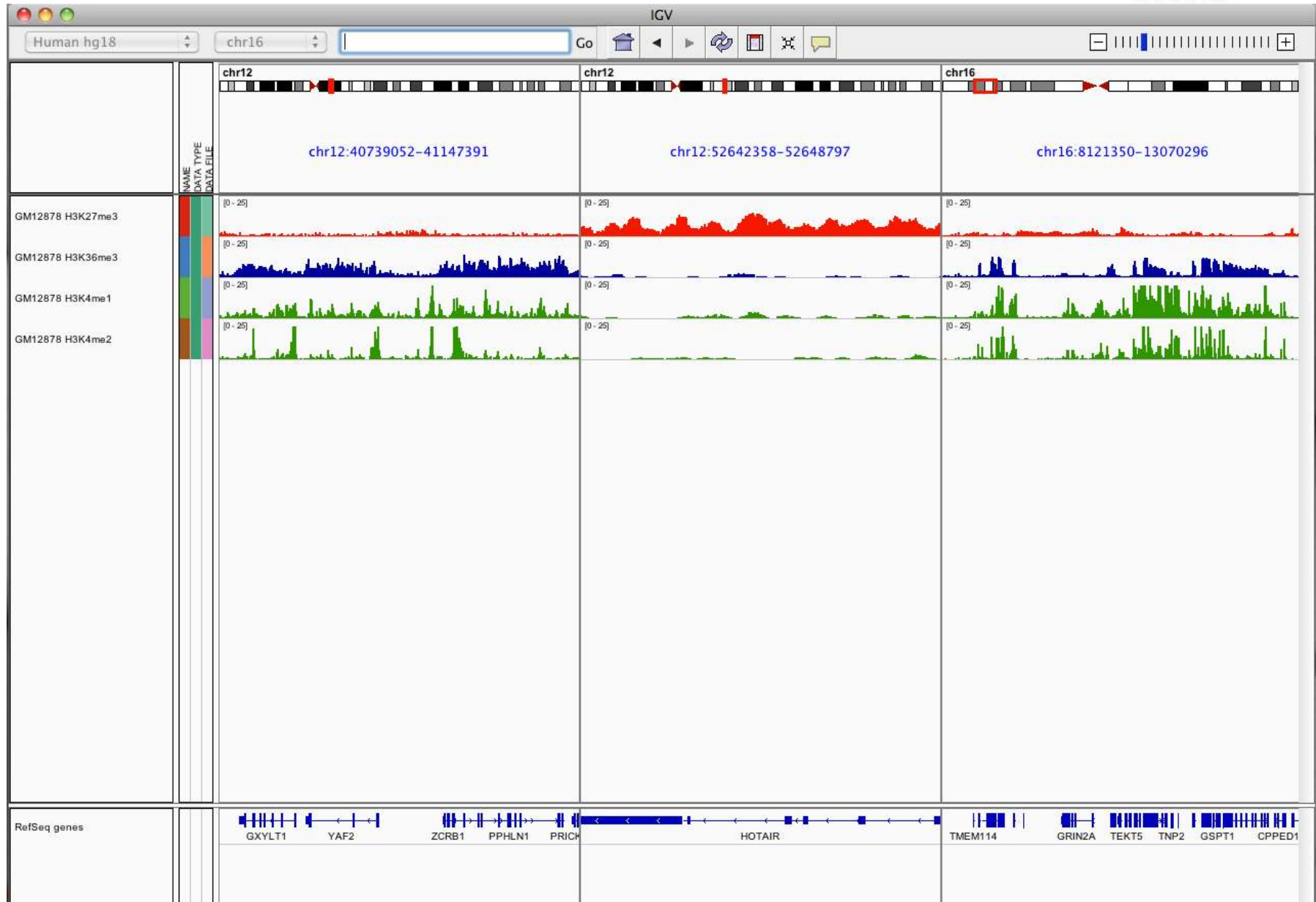
3. For a compact view of all variants use “Squished”



Viewing multiple regions



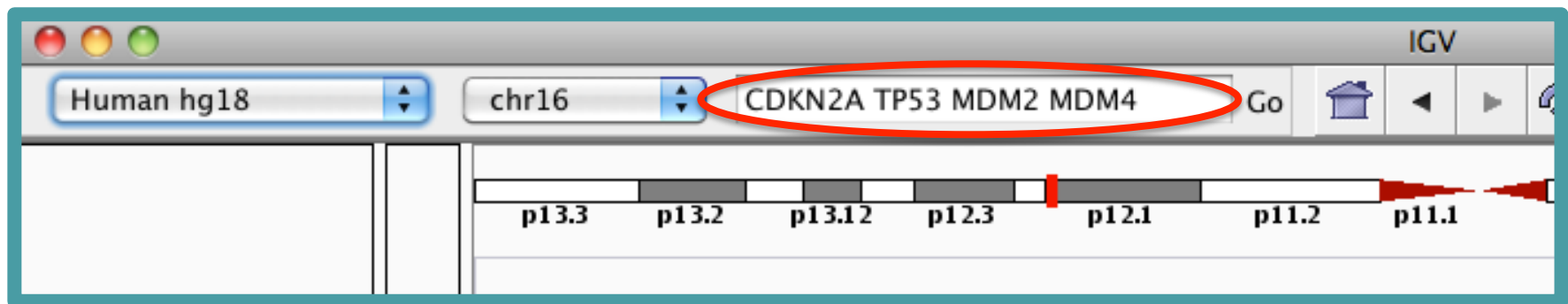
Viewing multiple regions



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:

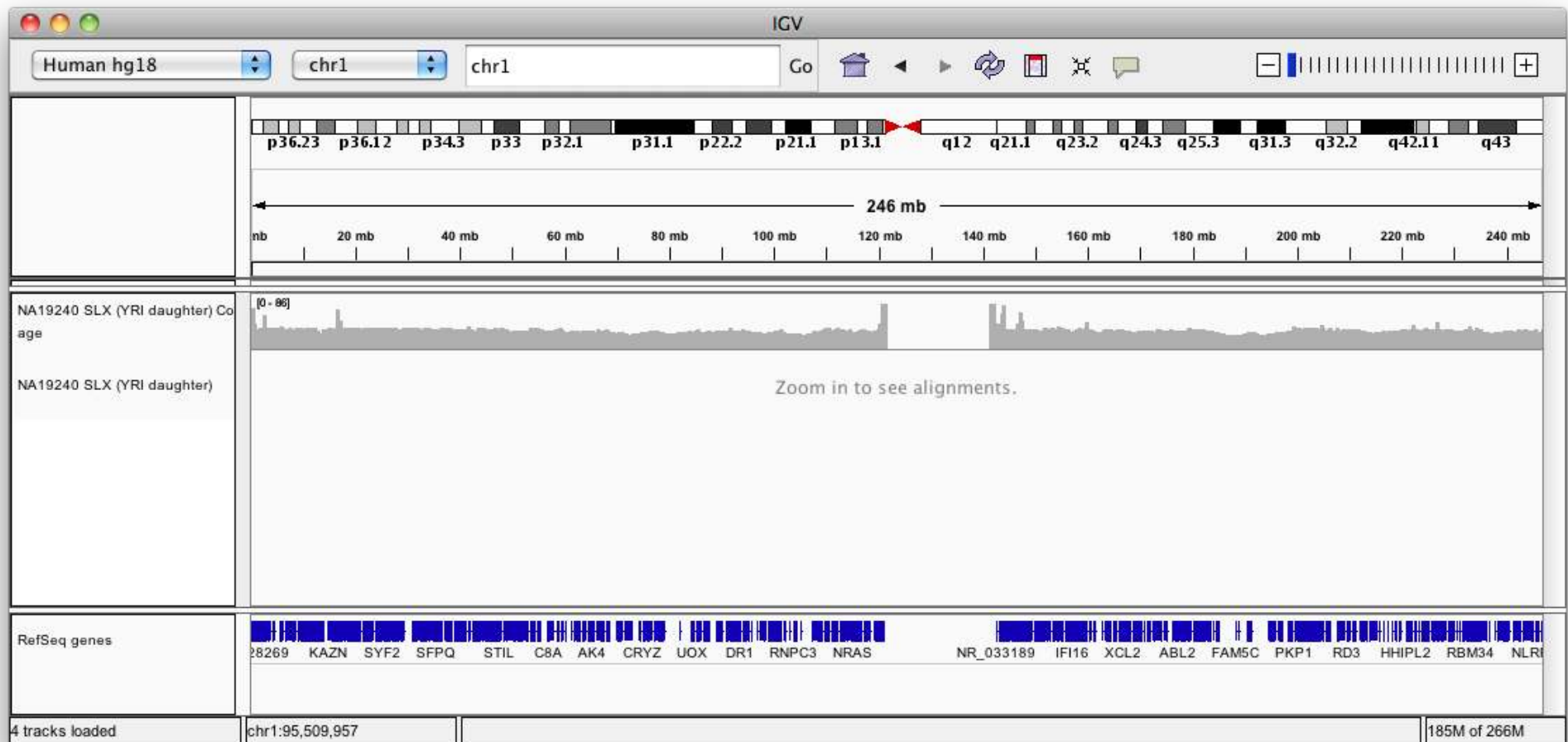
- *double-click* on a region label – or –
- *right-click* and select “Switch to standard view”



Viewing NGS Data

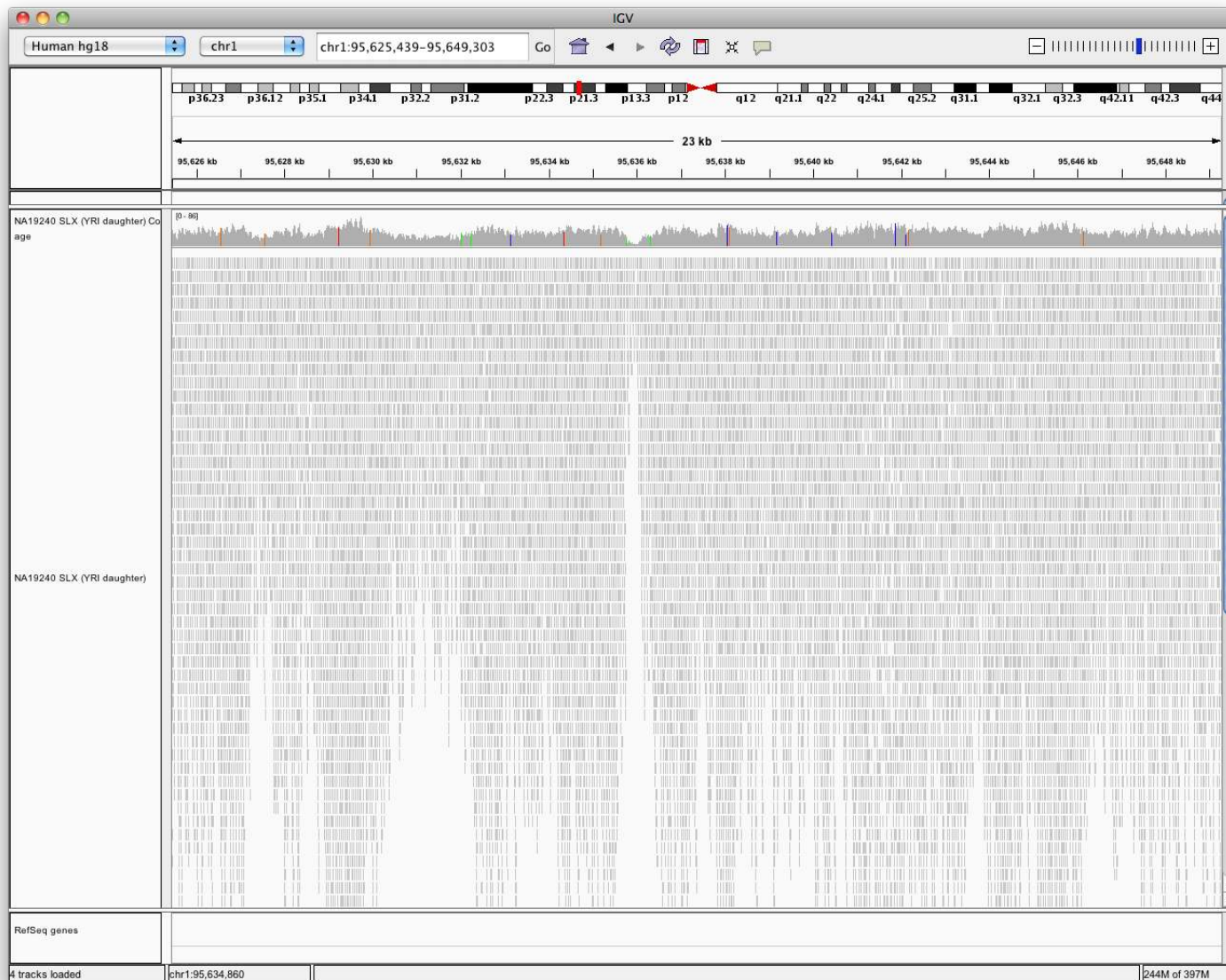
Viewing alignments

Whole chromosome view



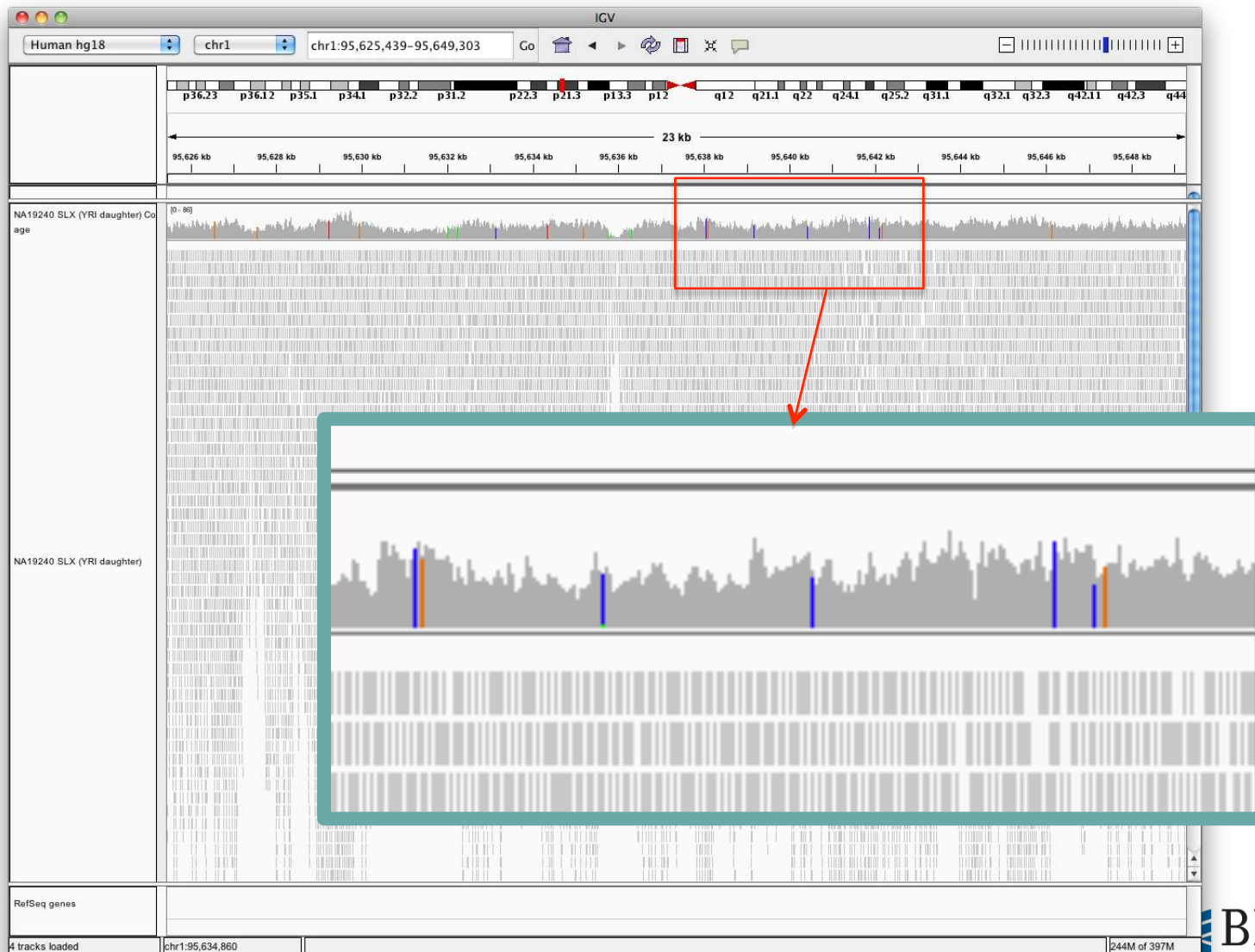
Viewing alignments

Zoom in to view alignments



Viewing alignments

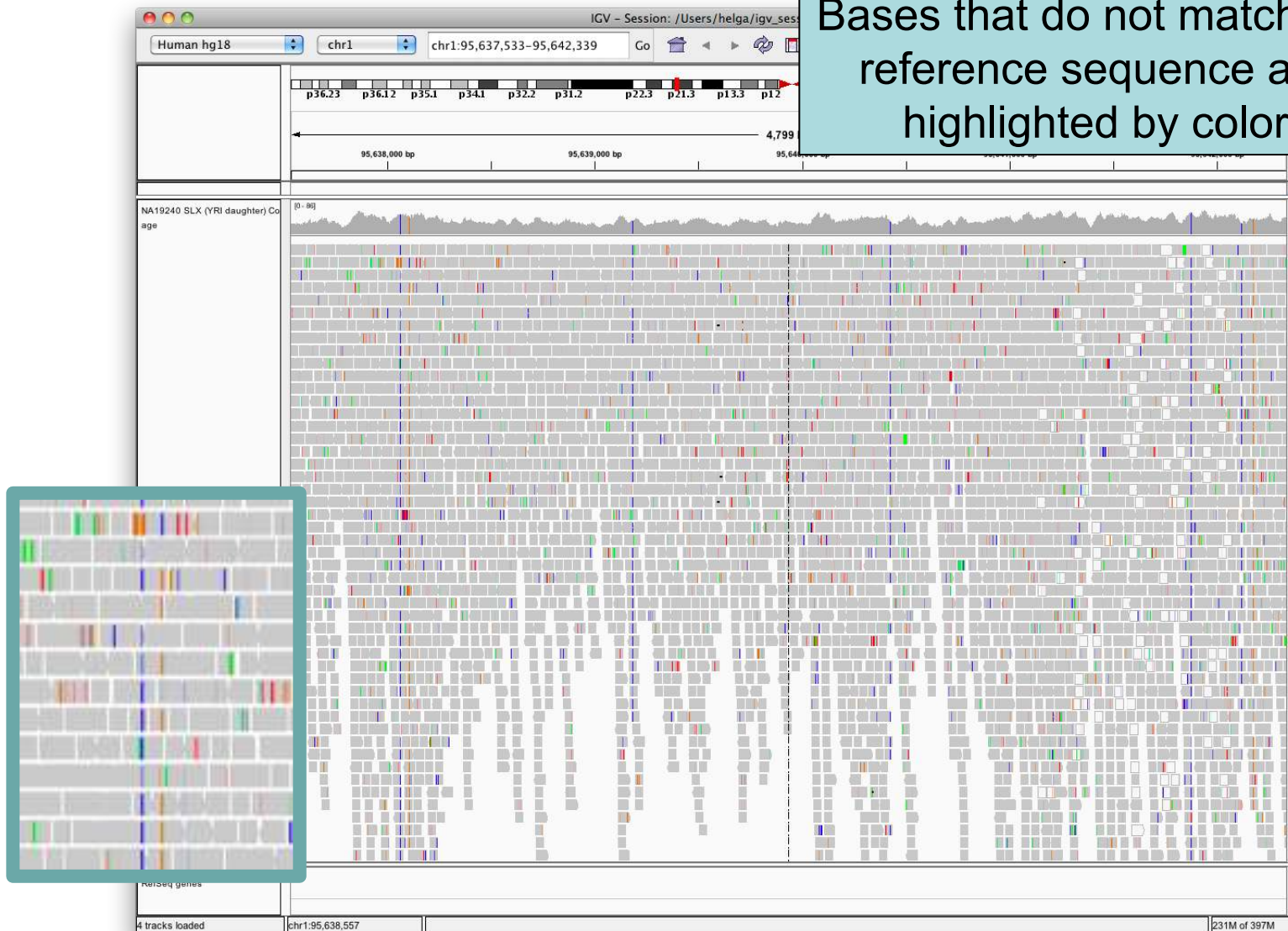
Coverage track now has more detail



Viewing alignments

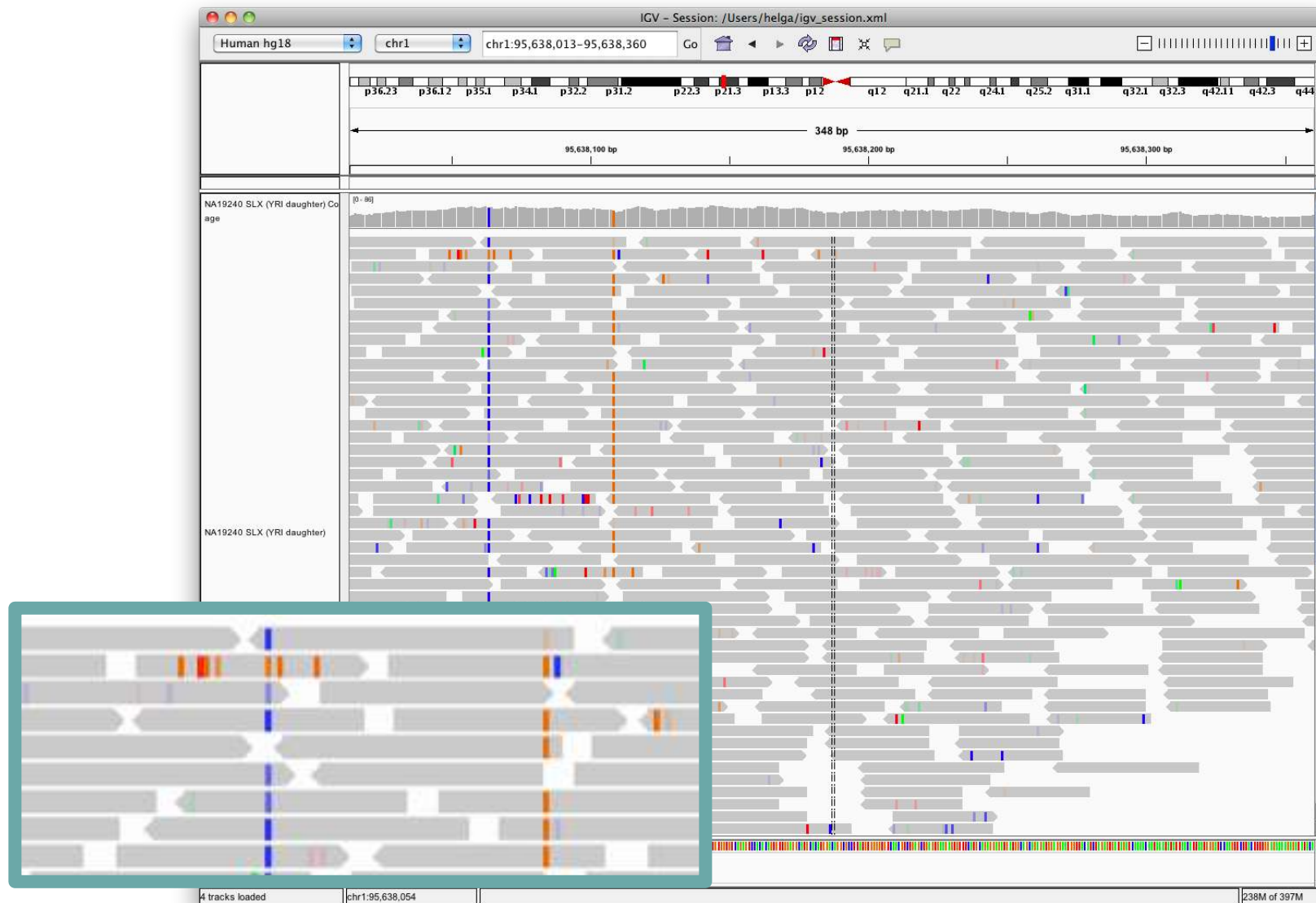
Zoom in to see more detail

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



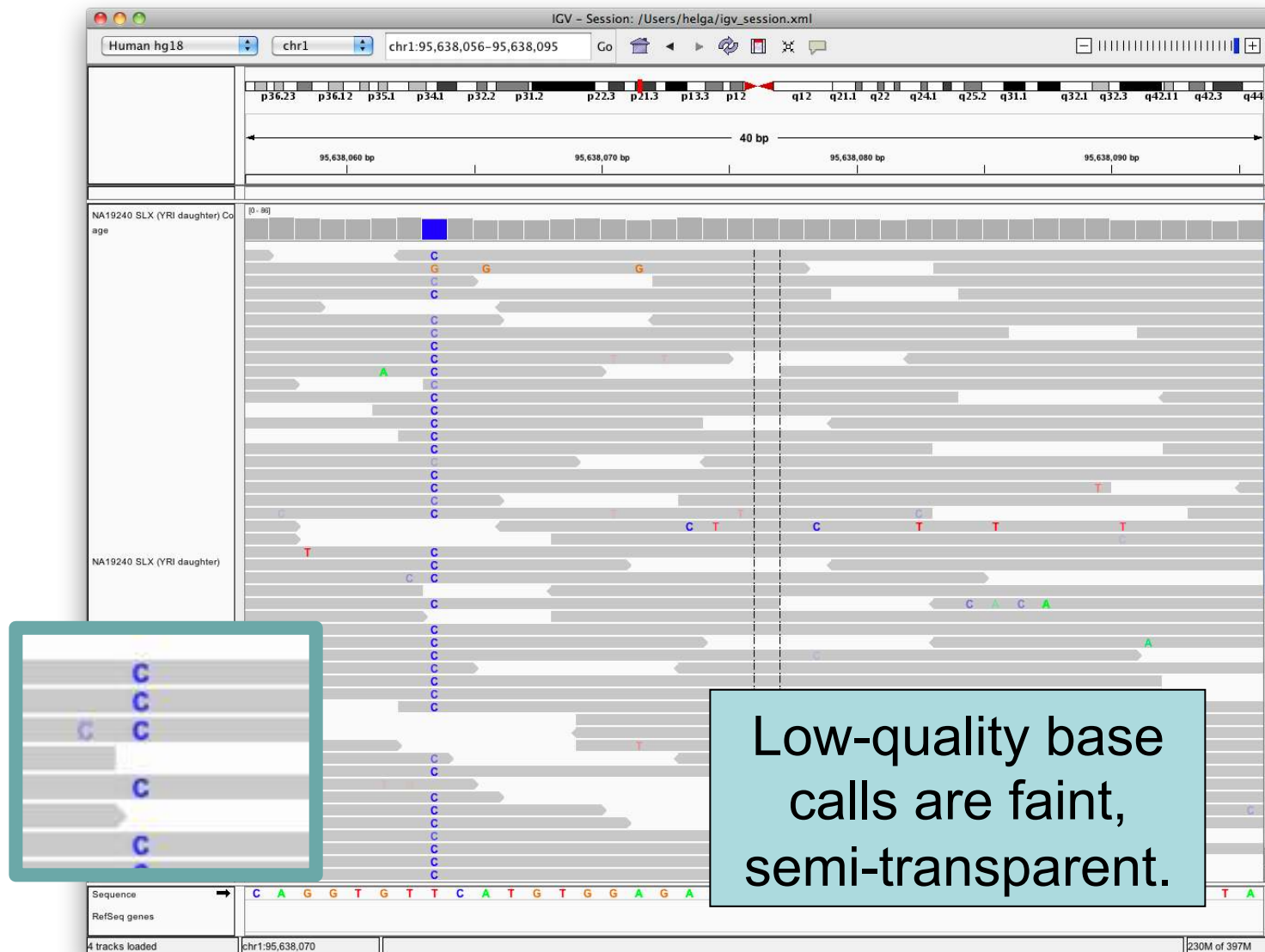
Viewing alignments

Zoom in to see more detail

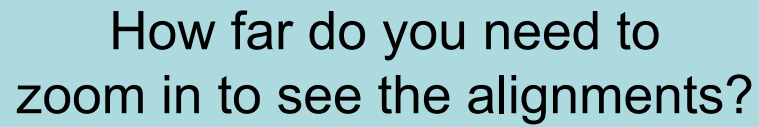


Viewing alignments

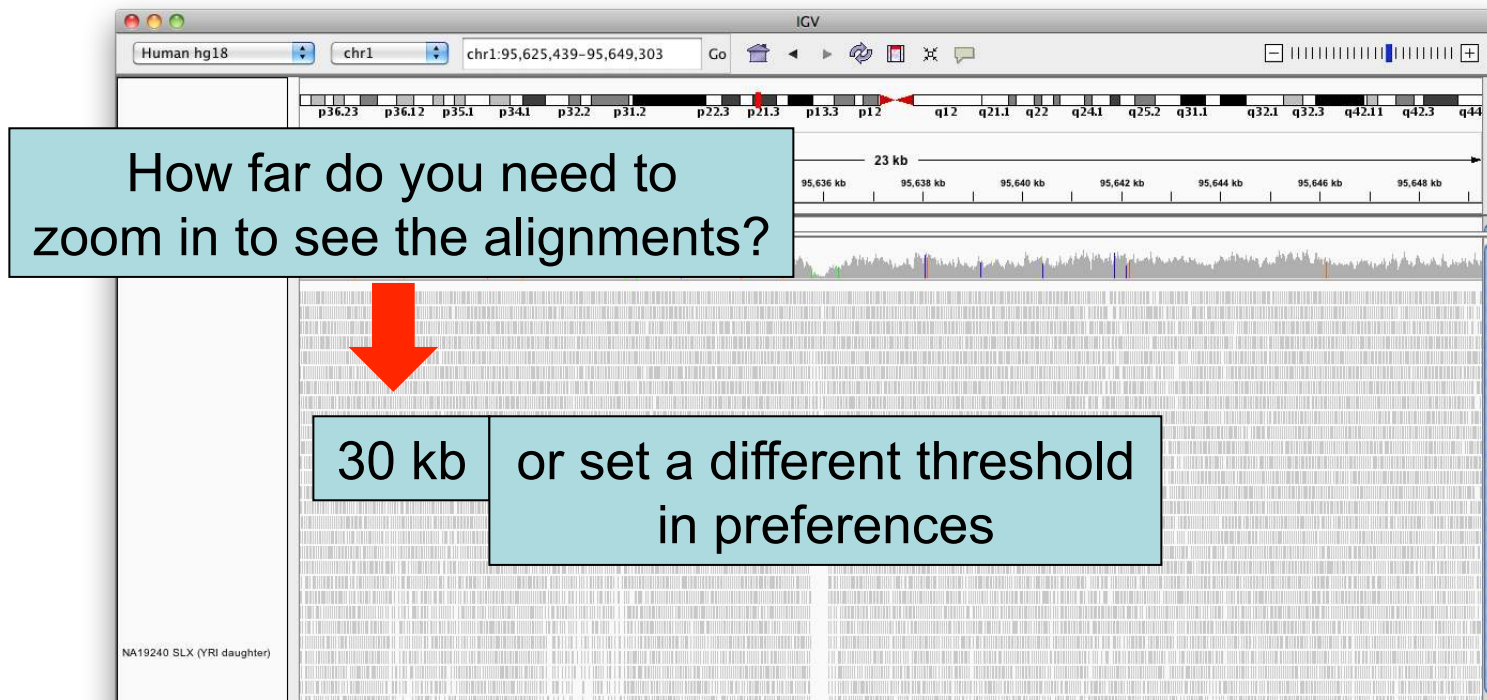
Zoom in to see more detail



Integrative
Genomics
Viewer

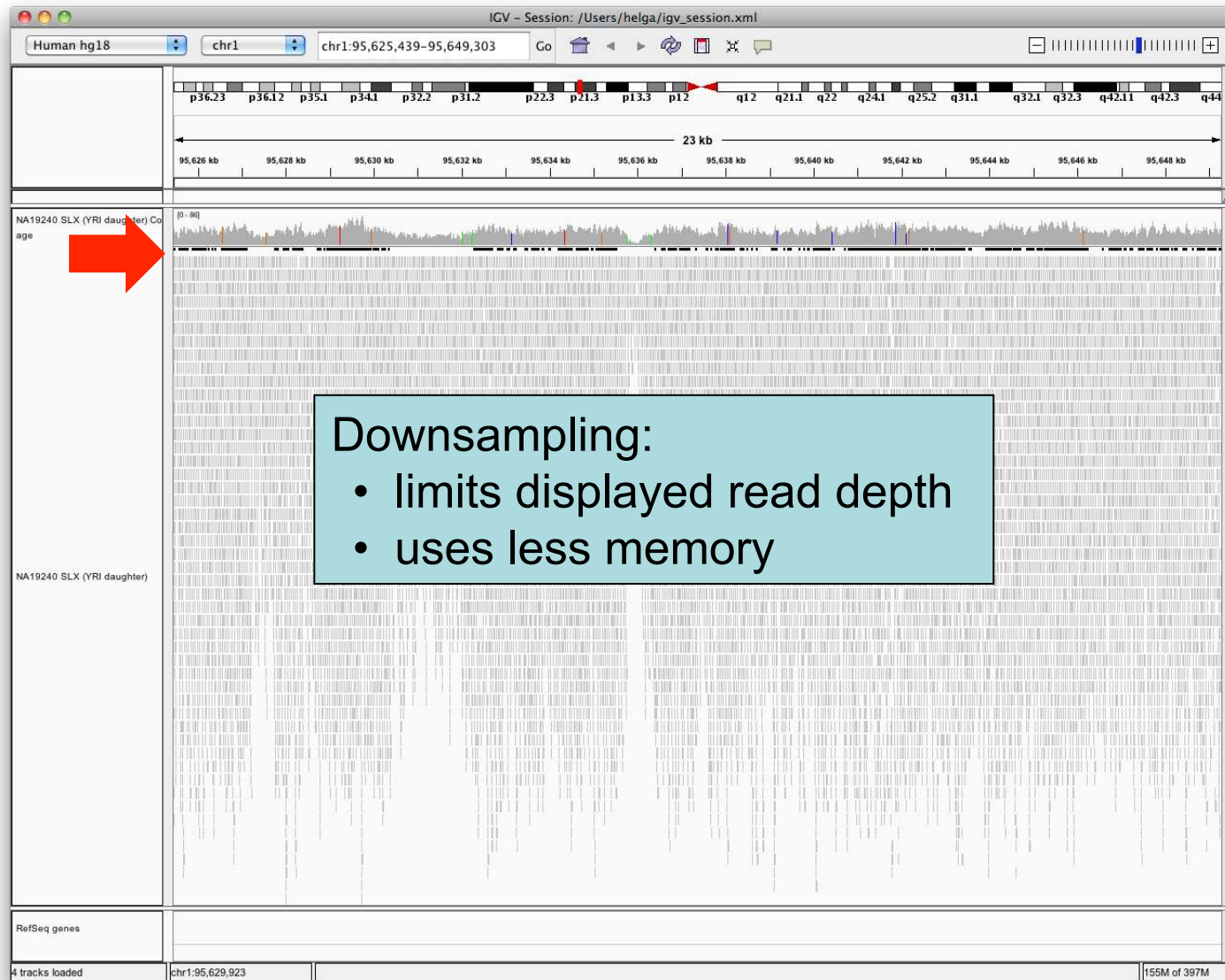


Viewing alignments



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

Viewing alignments



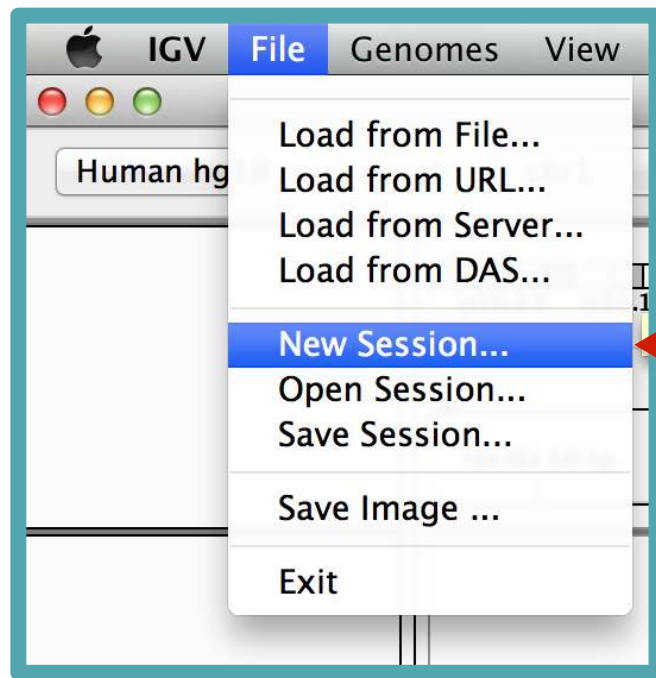
Viewing SNPs



Hands-on exercise

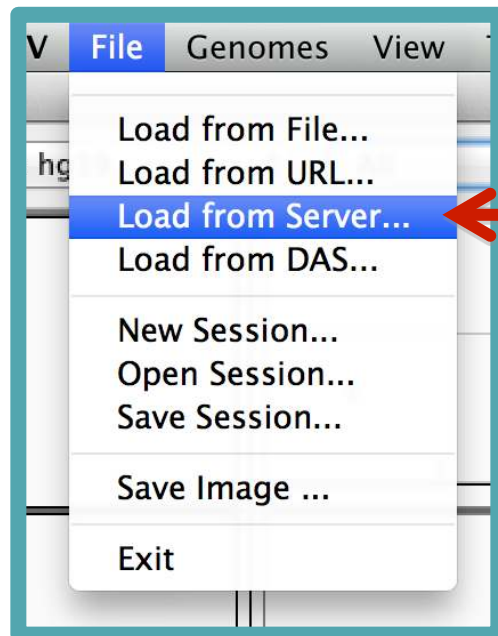
- Load alignments from whole genome sequencing
- View sites where SNPs were called
- Sort and color to highlight patterns

Viewing SNPs



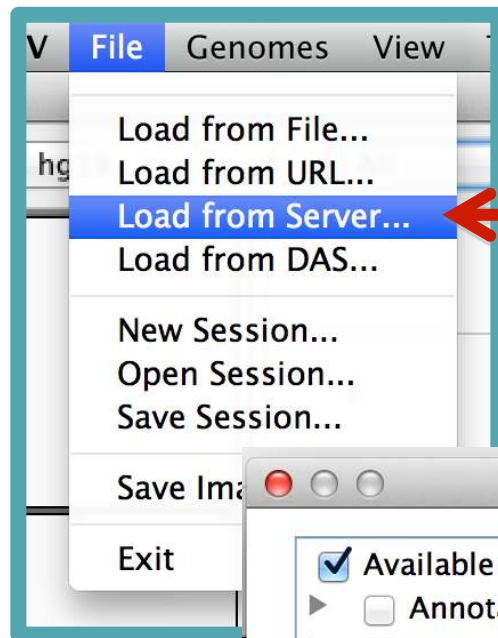
Before we start:
Select **File > New Session**
to clear IGV window

Viewing SNPs



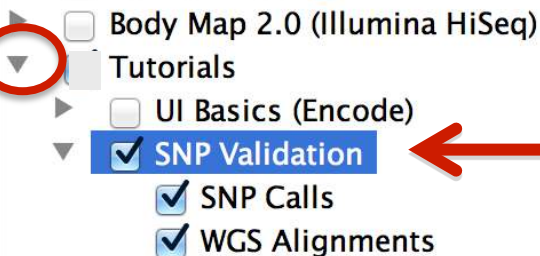
Select File > Load from Server...

Viewing SNPs



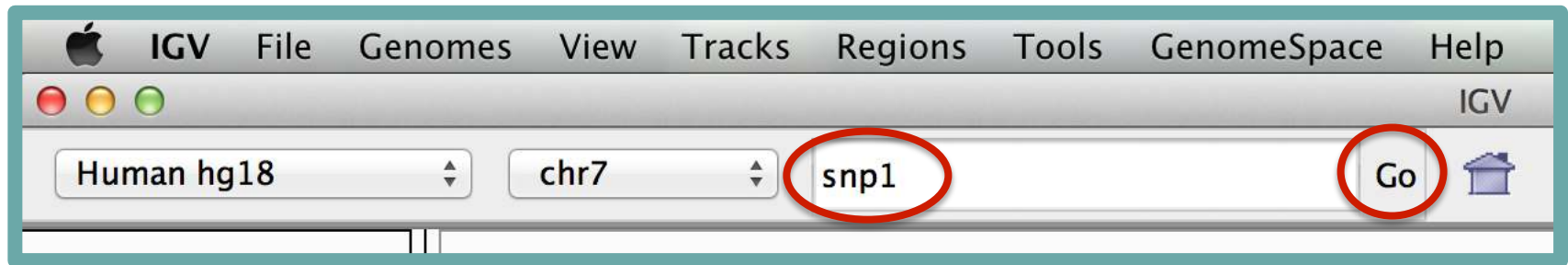
Select File > Load from Server...

Open the
Tutorials menu



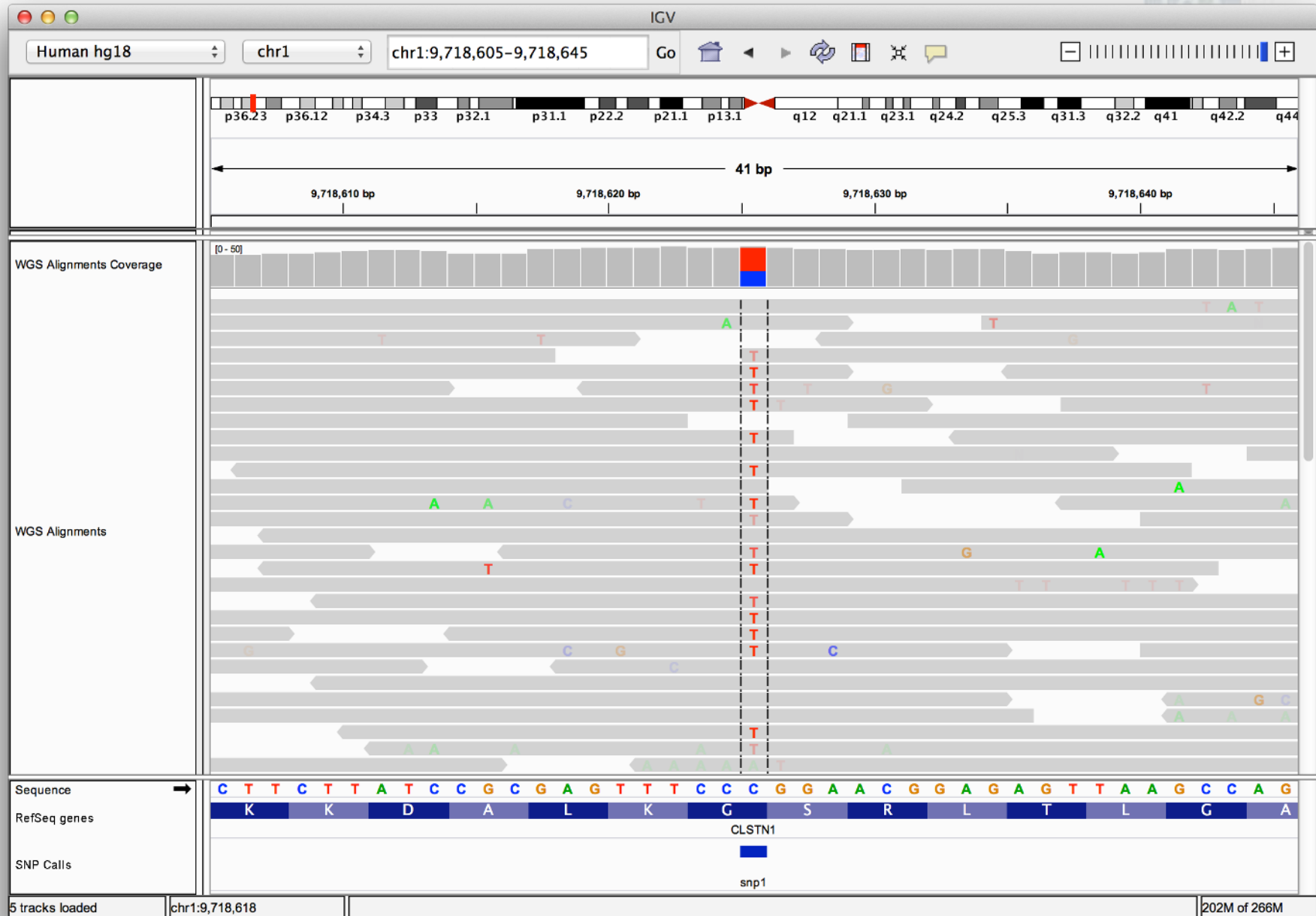
Select
SNP Validation

Viewing SNPs

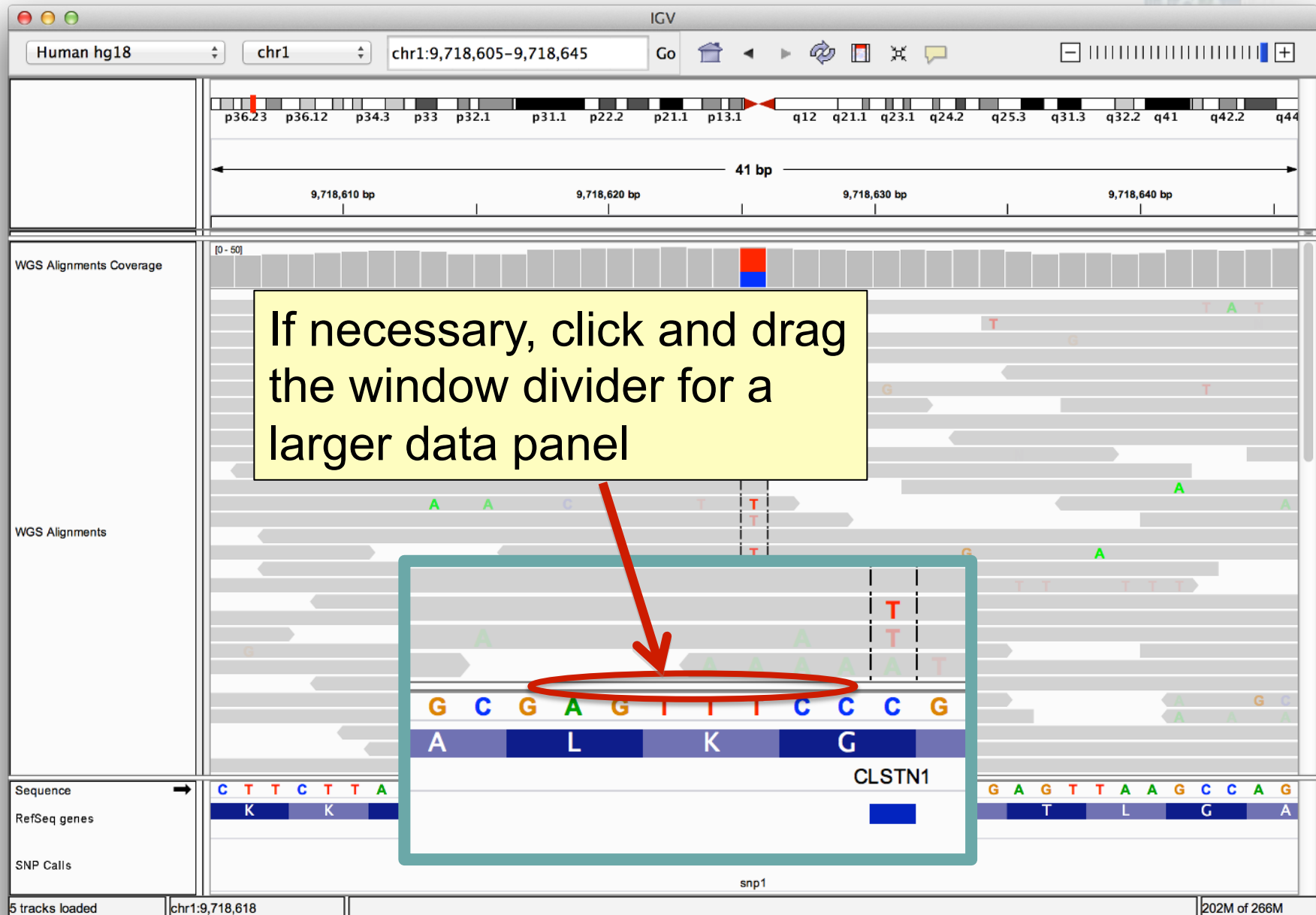


Type “snp1” in the **Search Box**
and click **Go**

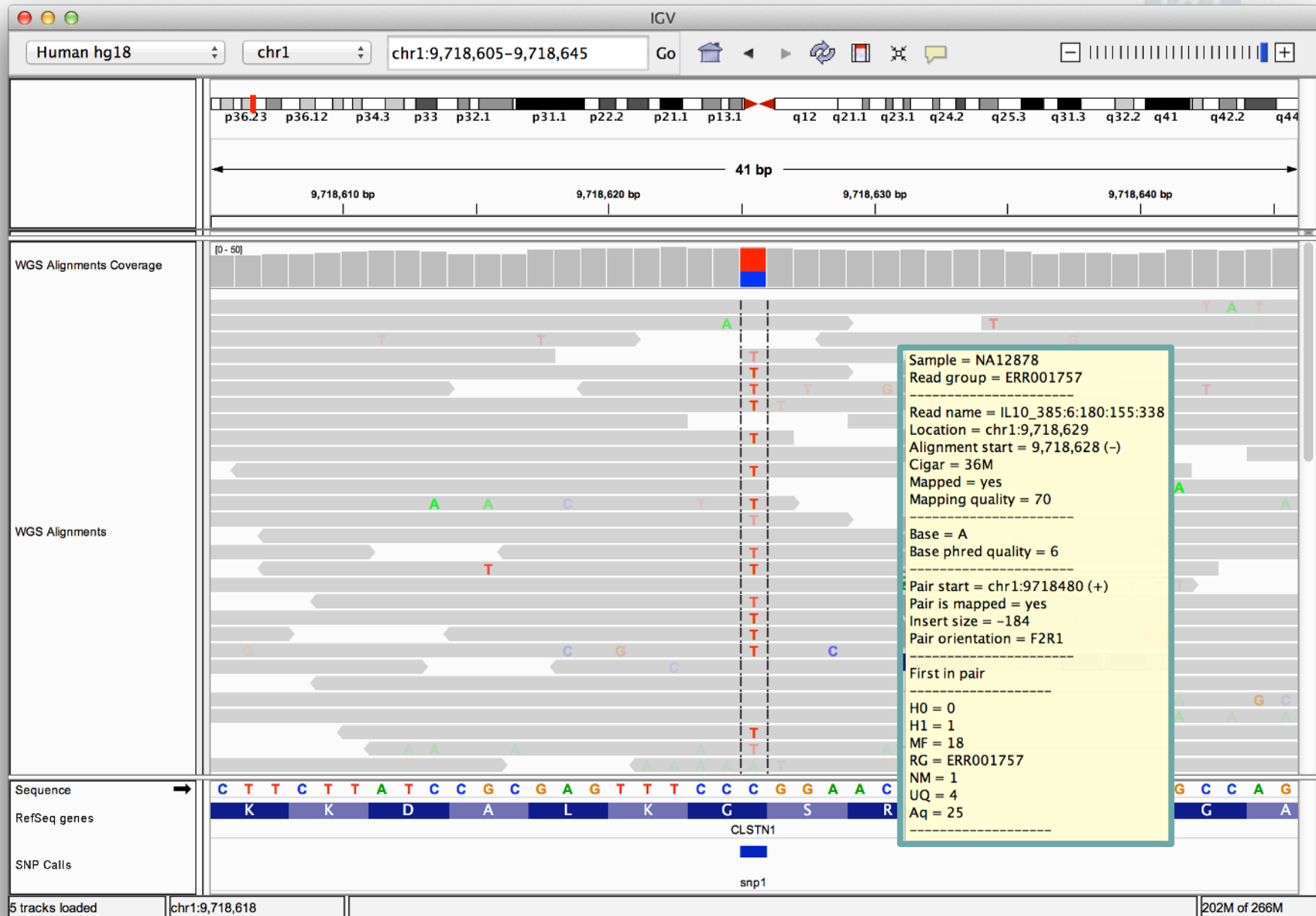
Viewing SNPs



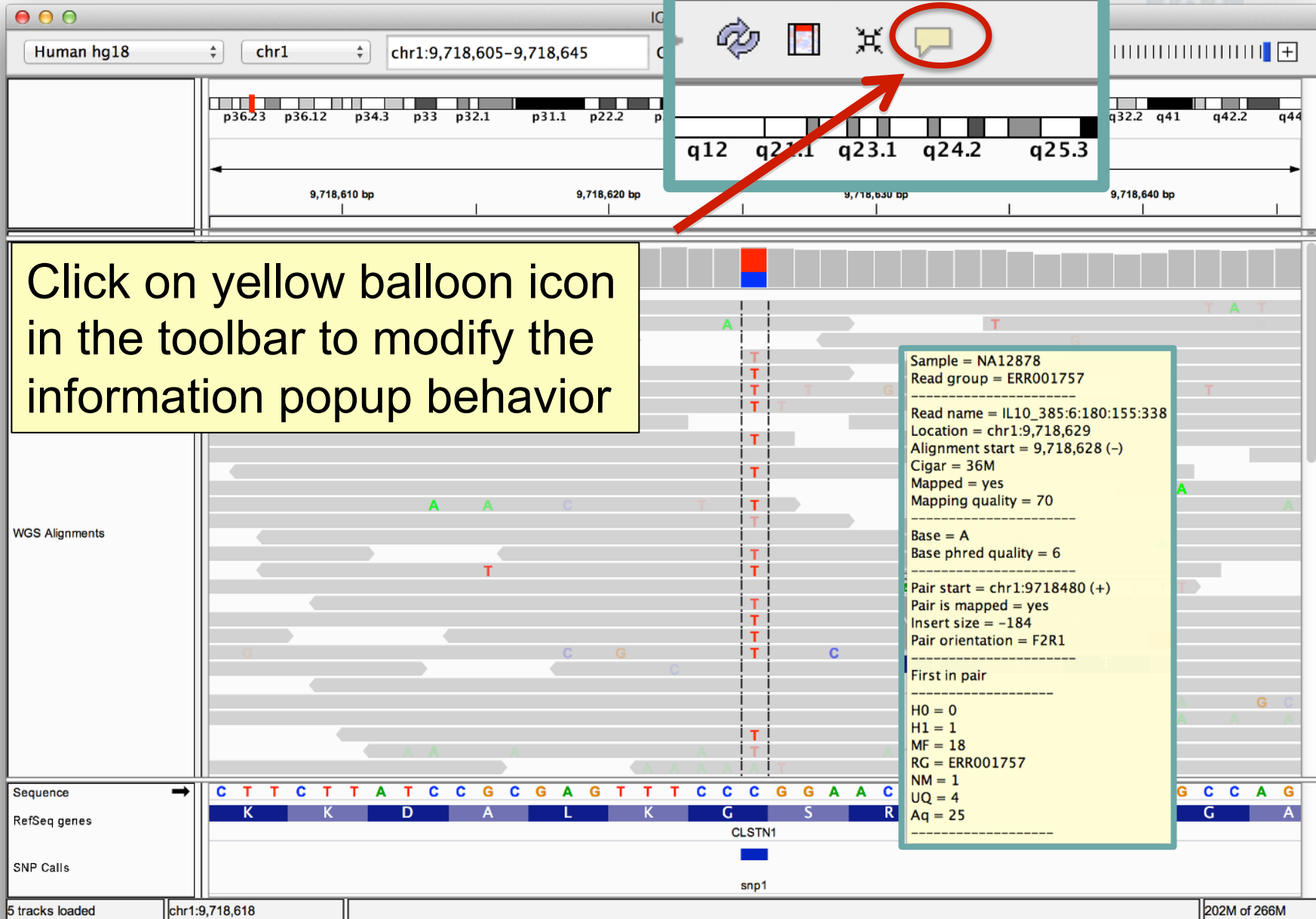
Viewing SNPs



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

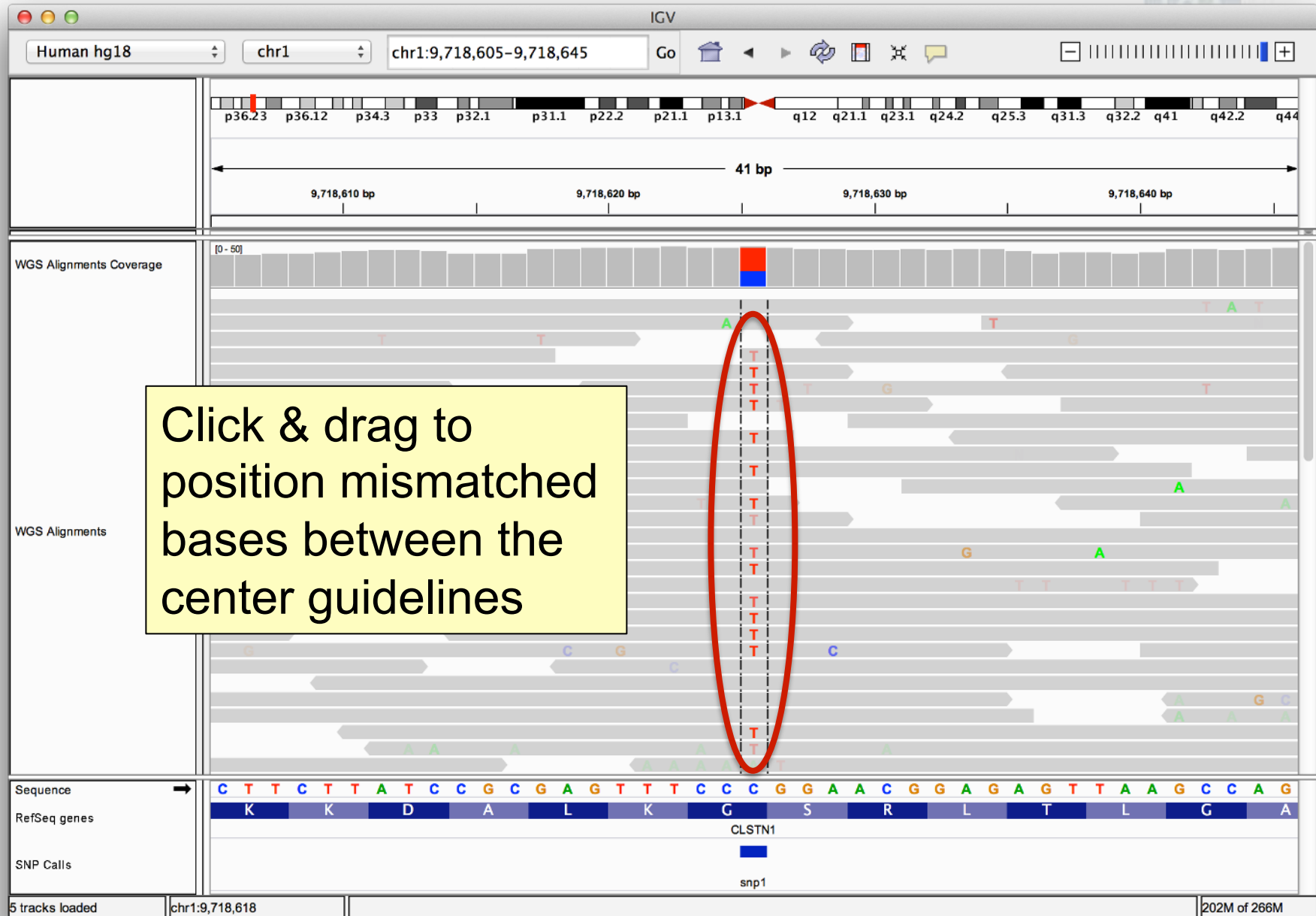
Sequence → C T T C T T A T C C G C G A G T T T C C C G G A A C
RefSeq genes K K D A L K G S R
SNP Calls

CLSTN1
snp1

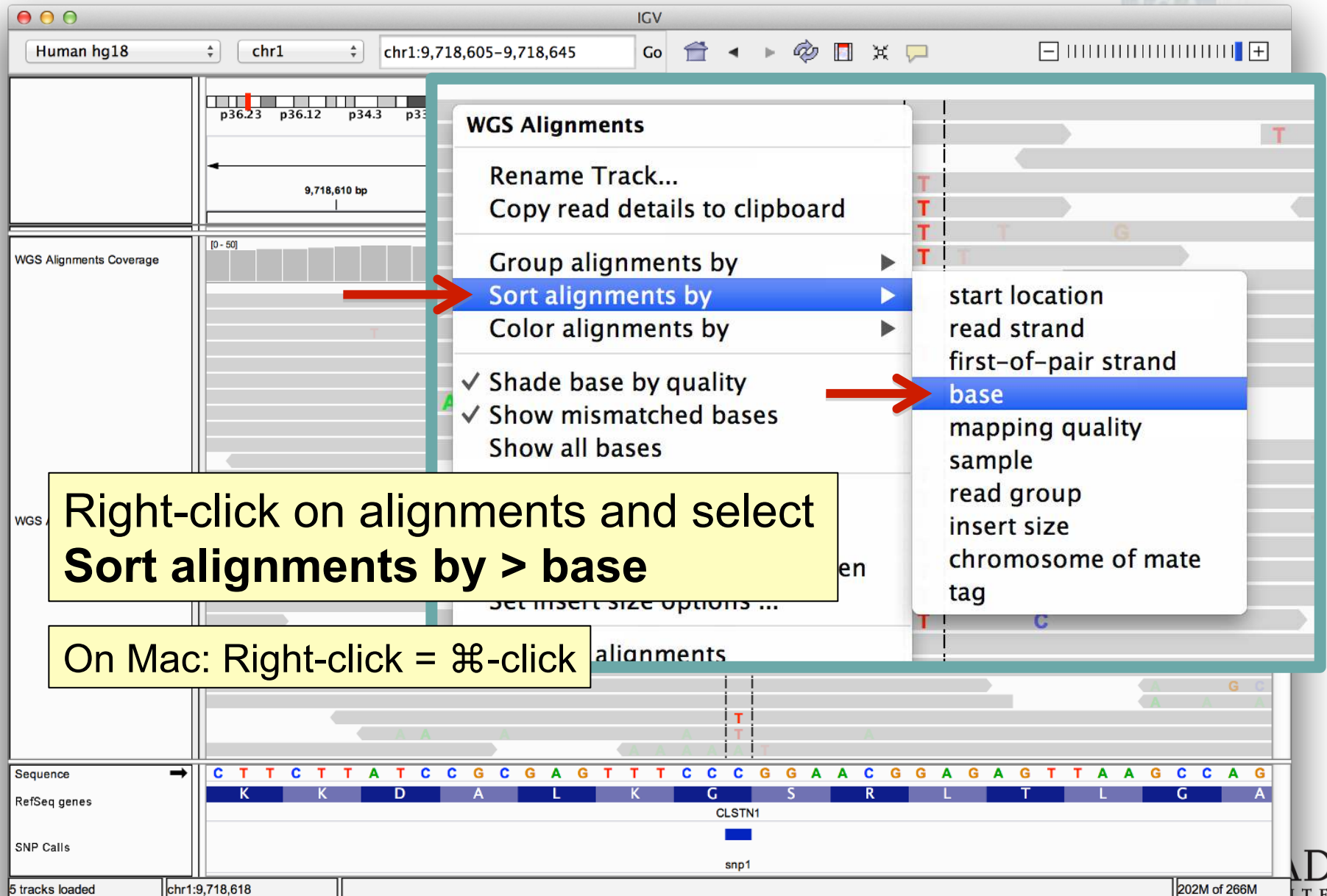
Sample = NA12878
Read group = ERR001757
Read name = IL10_385:6:180:155:338
Location = chr1:9,718,629
Alignment start = 9,718,628 (-)
Cigar = 36M
Mapped = yes
Mapping quality = 70
Base = A
Base phred quality = 6
Pair start = chr1:9718480 (+)
Pair is mapped = yes
Insert size = -184
Pair orientation = F2R1
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



Viewing SNPs



The screenshot shows the IGV interface with a track of WGS alignments. A right-click context menu is open over the alignments, showing options like 'Rename Track...', 'Copy read details to clipboard', 'Group alignments by', 'Sort alignments by', and 'Color alignments by'. The 'Sort alignments by' option is selected, and a sub-menu is open showing various sorting criteria. The 'base' option is highlighted in the sub-menu. A red arrow points from the 'Sort alignments by' option to the 'base' option. Another red arrow points from the 'base' option to the 'base' option in the sub-menu. A yellow box with text is overlaid on the screenshot, providing instructions on how to right-click on alignments and select 'Sort alignments by > base'. A second yellow box provides a note for Mac users: 'On Mac: Right-click = ⌘-click'.

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**

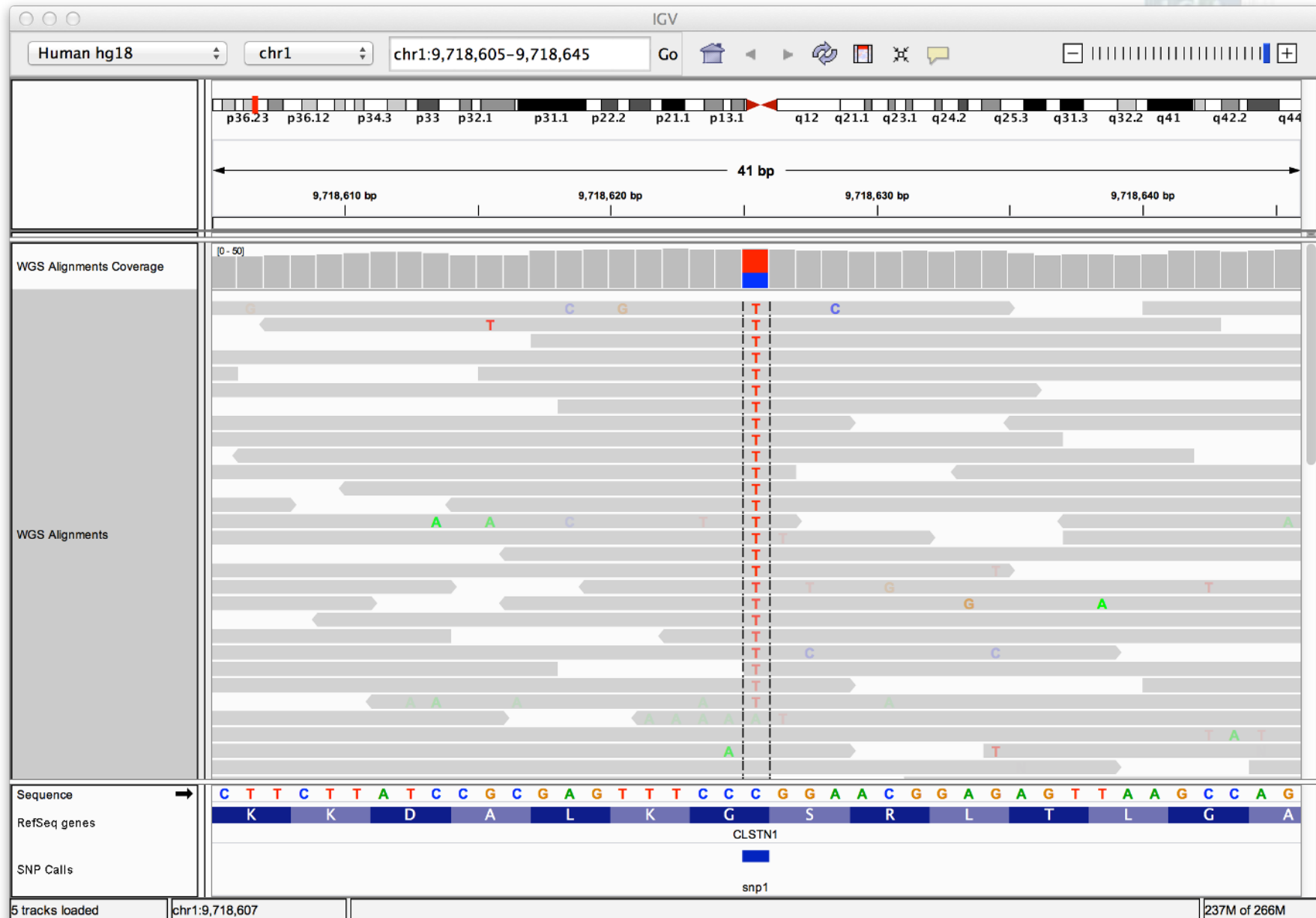
On Mac: Right-click = ⌘-click

Sequence
RefSeq genes
SNP Calls

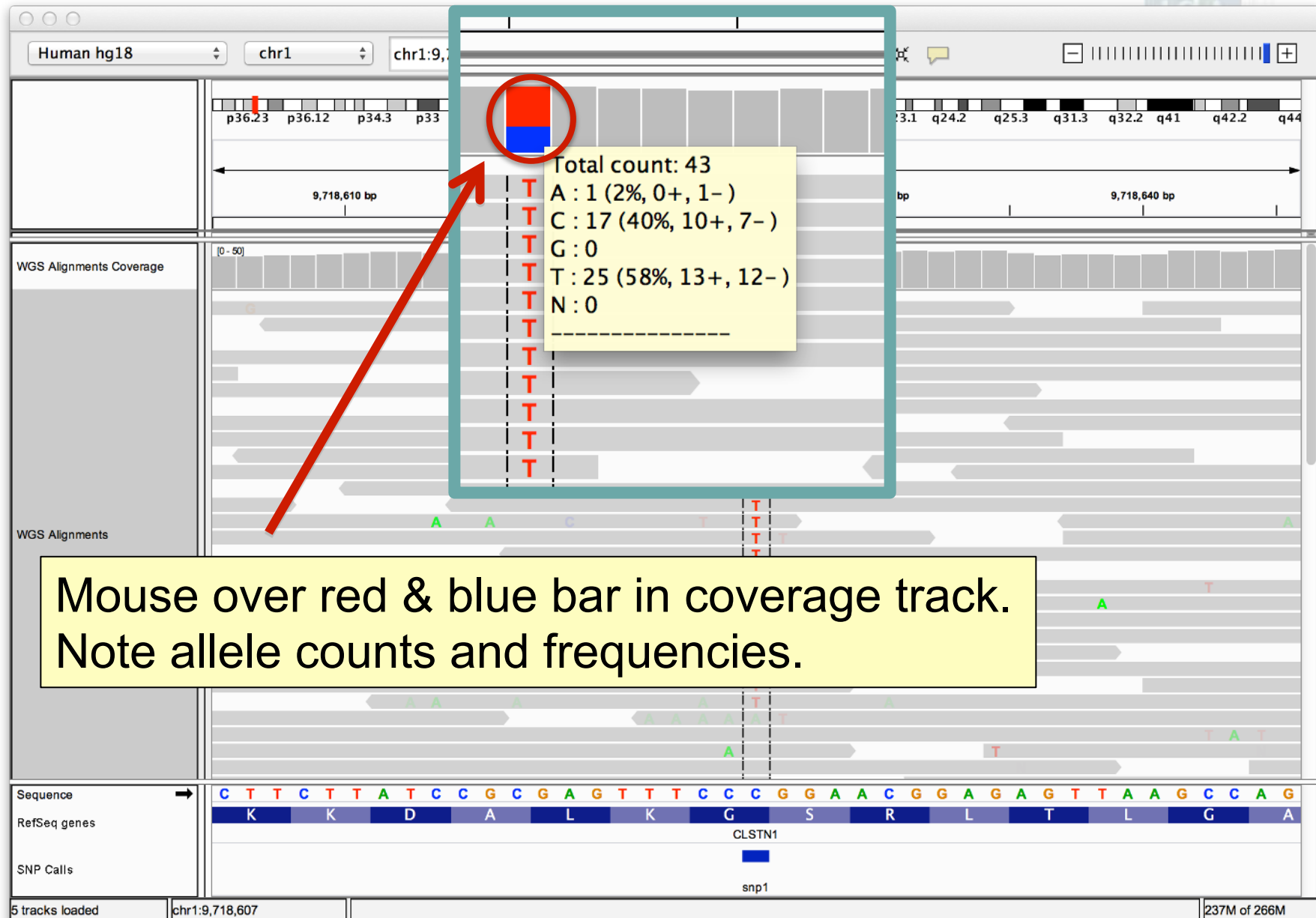
CLSTN1
snp1

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

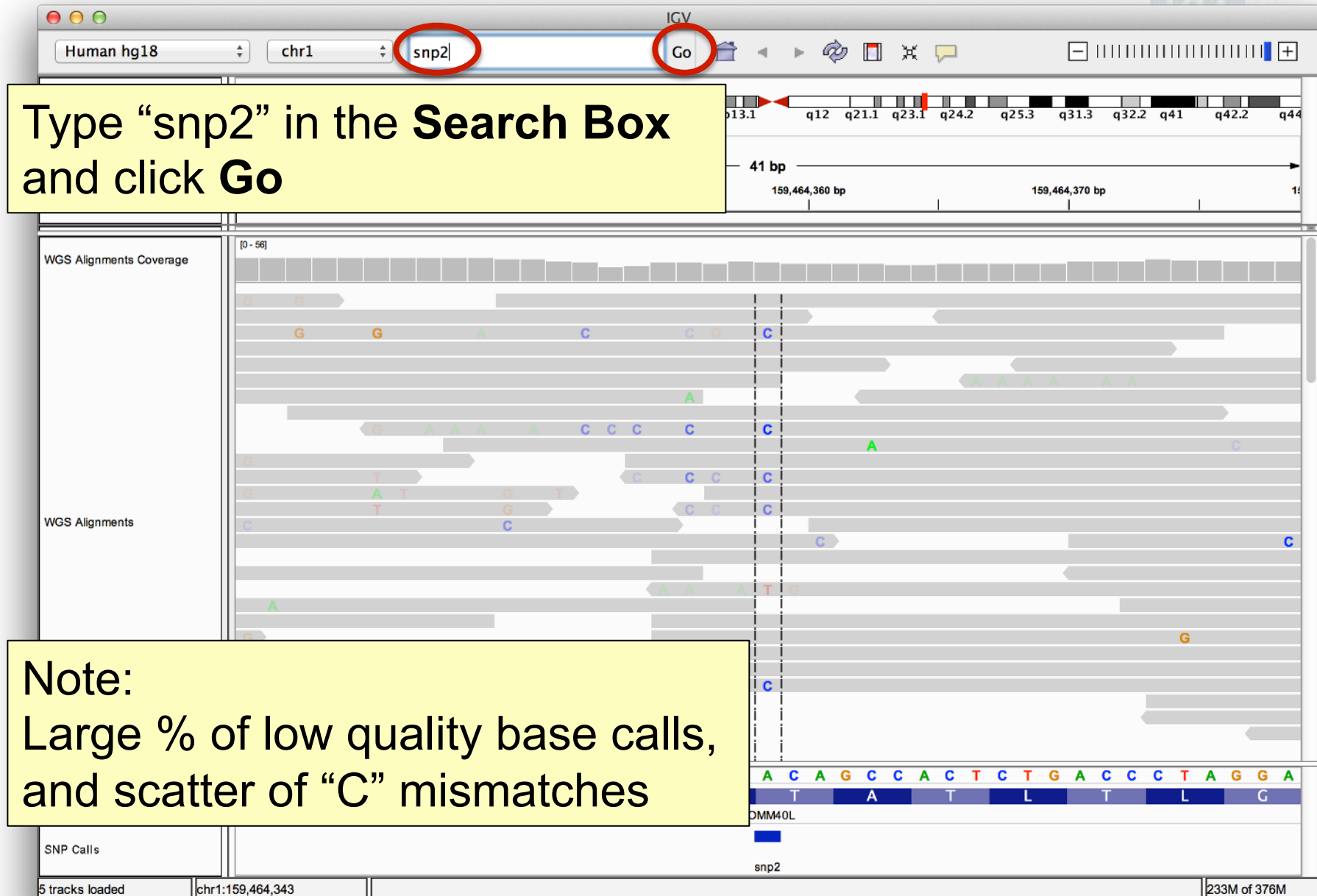
Viewing SNPs



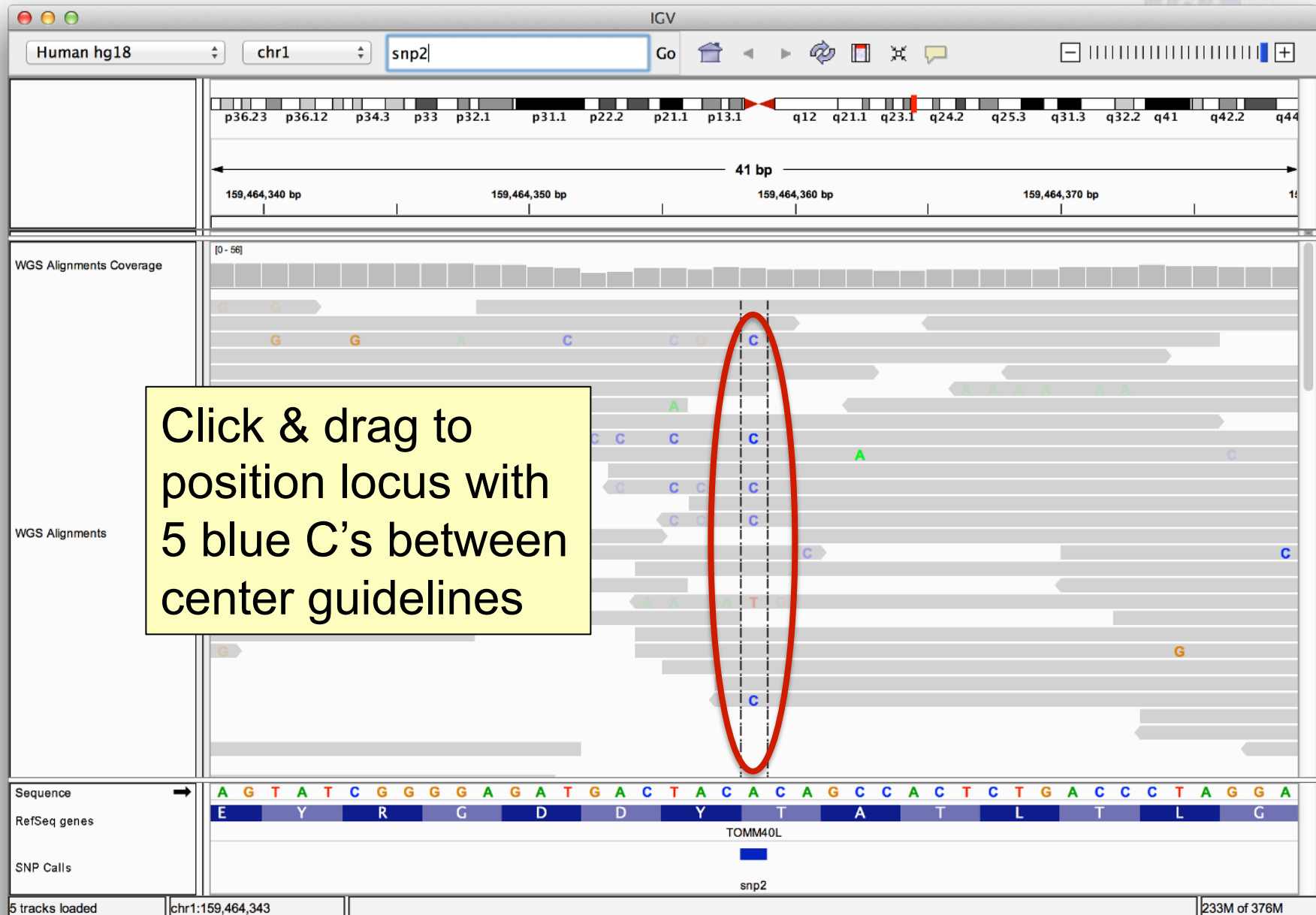
Viewing SNPs



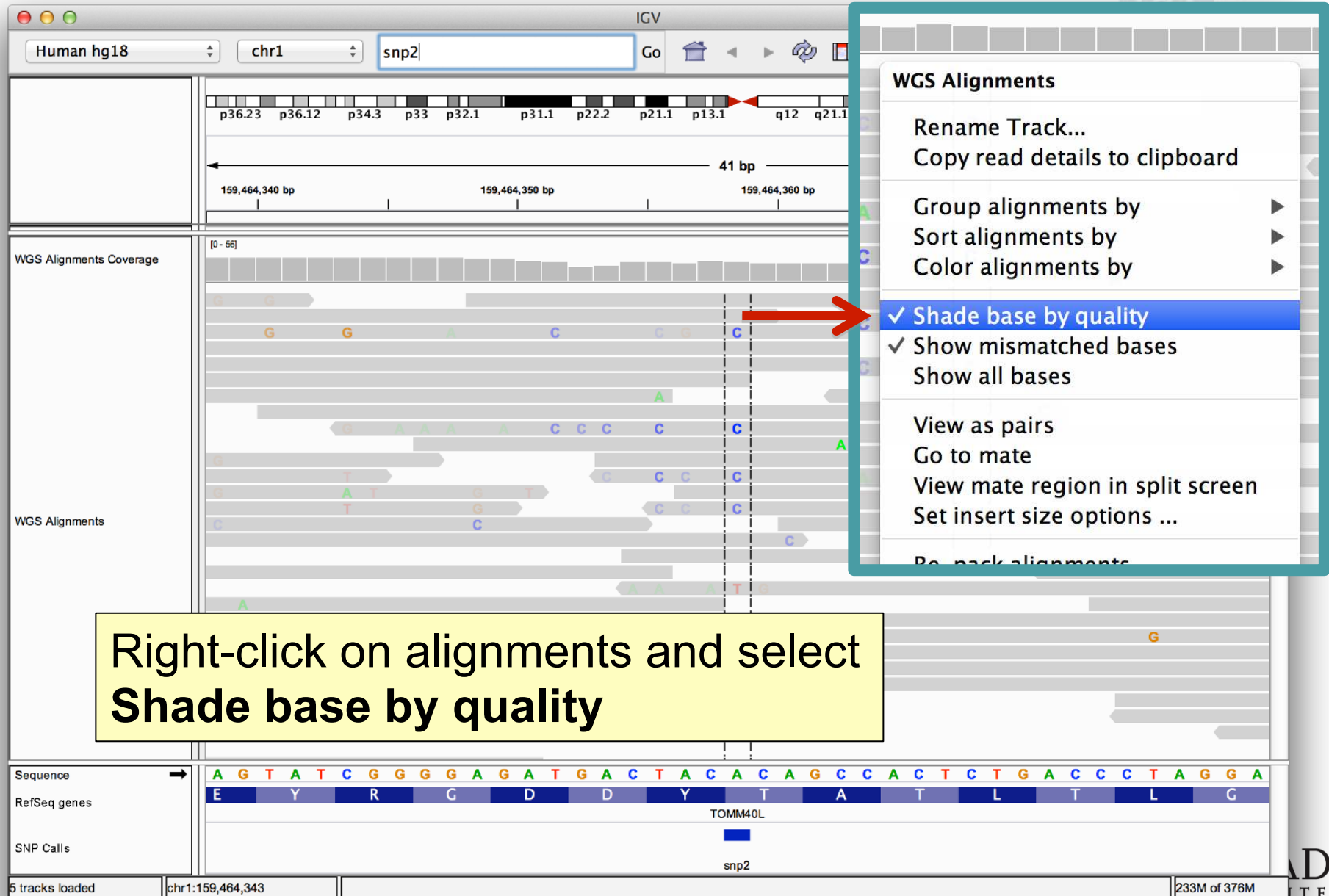
Viewing SNPs



Viewing SNPs



Viewing SNPs



The screenshot shows the IGV interface with the following components:

- Top Bar:** Human hg18, chr1, snp2, Go button, and navigation icons.
- Chromosome Map:** A map of chromosome 1 with a red arrow pointing to the current region.
- WGS Alignments Coverage:** A track showing the coverage of the WGS alignments.
- WGS Alignments:** A track showing individual sequencing reads with colored bases (A, C, G, T) and quality scores.
- Sequence:** A track showing the reference 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:** A track showing the gene TOMM40L.
- SNP Calls:** A track showing the SNP call for snp2.

The right-click context menu for WGS Alignments includes the following options:

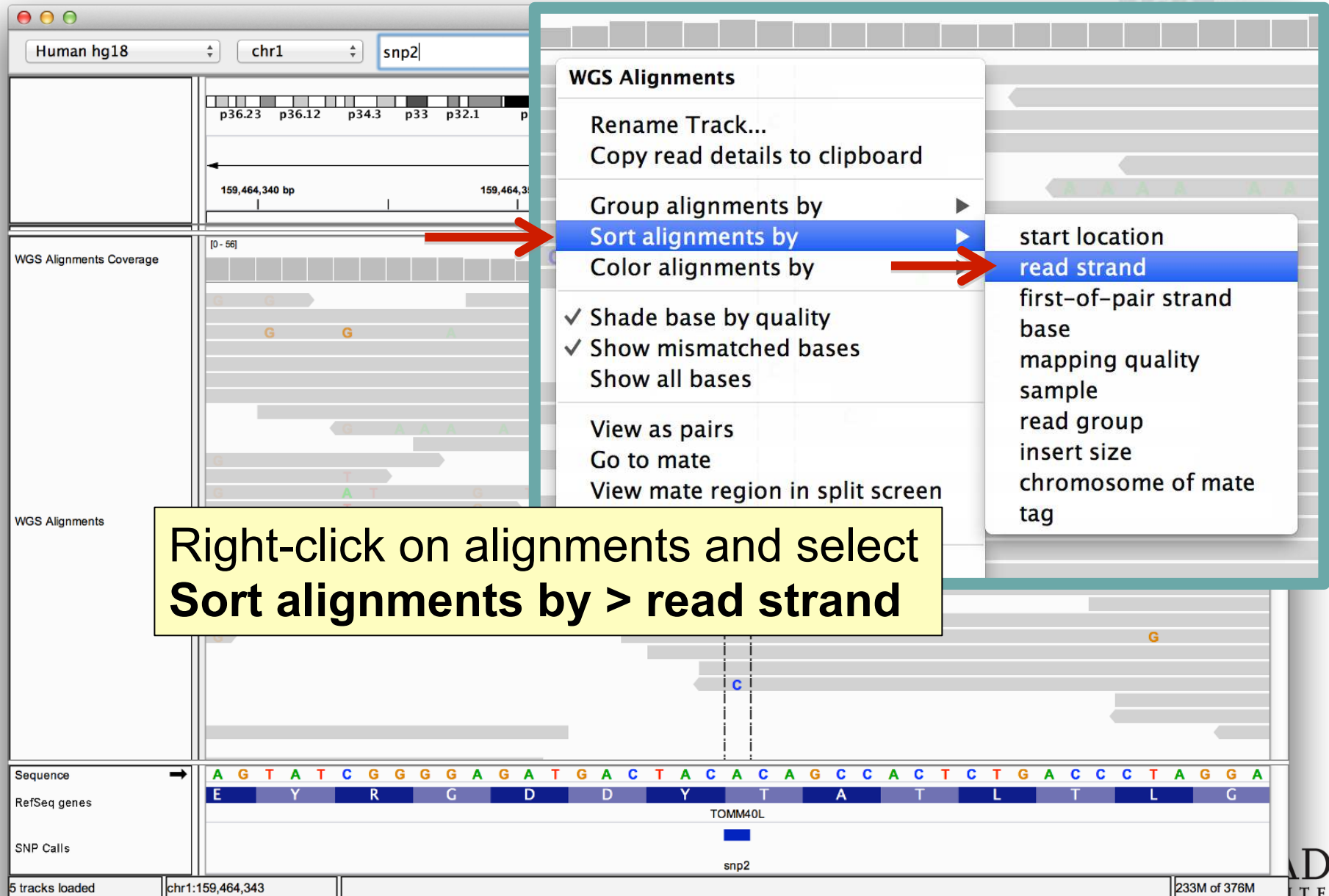
- Rename Track...
- Copy read details to clipboard
- Group alignments by
- Sort alignments by
- Color alignments by
- ✓ Shade base by quality**
- ✓ Show mismatched bases**
- Show all bases
- View as pairs
- Go to mate
- View mate region in split screen
- Set insert size options ...
- Re-pack alignments

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

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

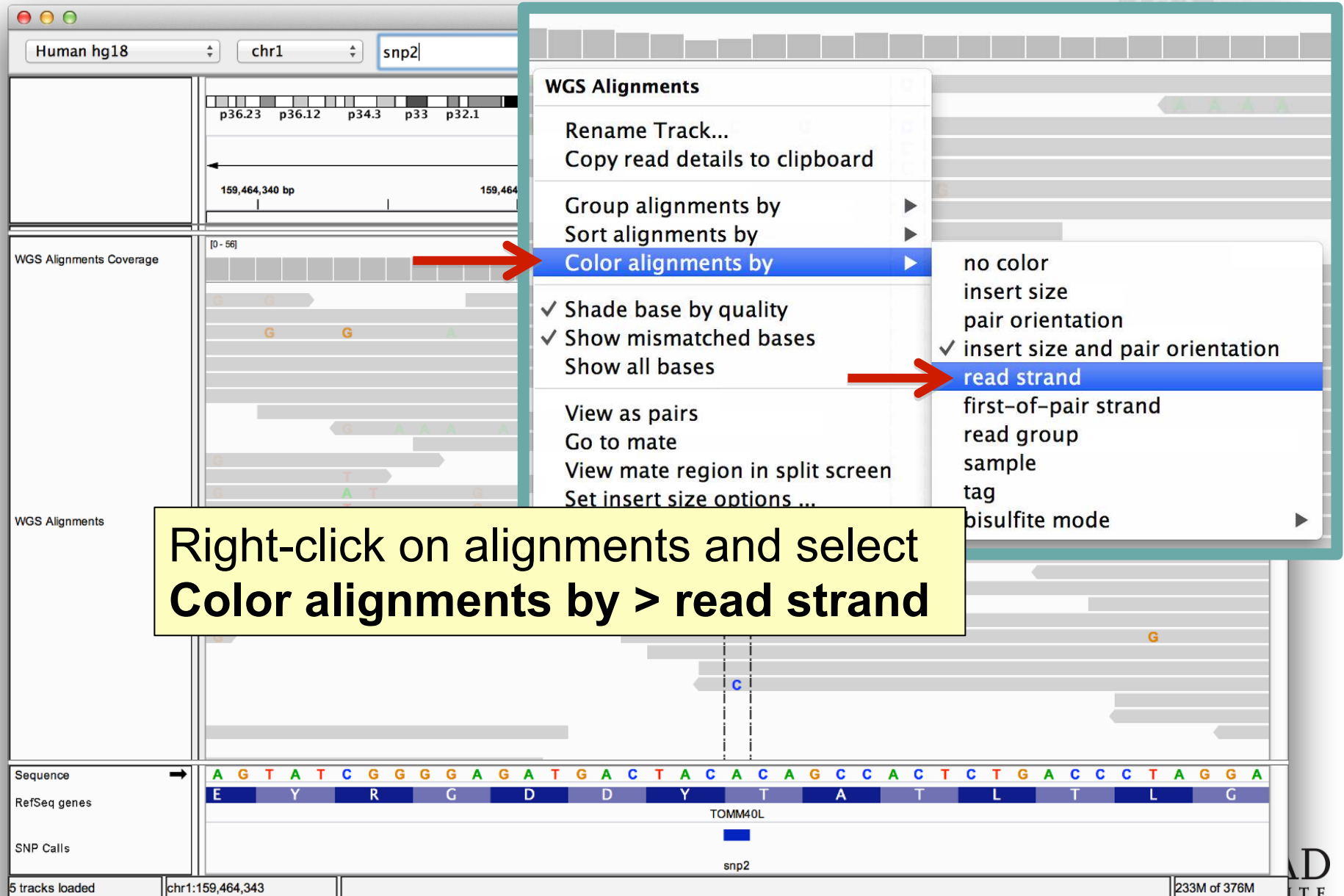
start location
read strand
first-of-pair strand
base
mapping quality
sample
read group
insert size
chromosome of mate
tag

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 Coverage

WGS Alignments

Sequence RefSeq genes SNP Calls

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

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

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 ...

Viewing SNPs



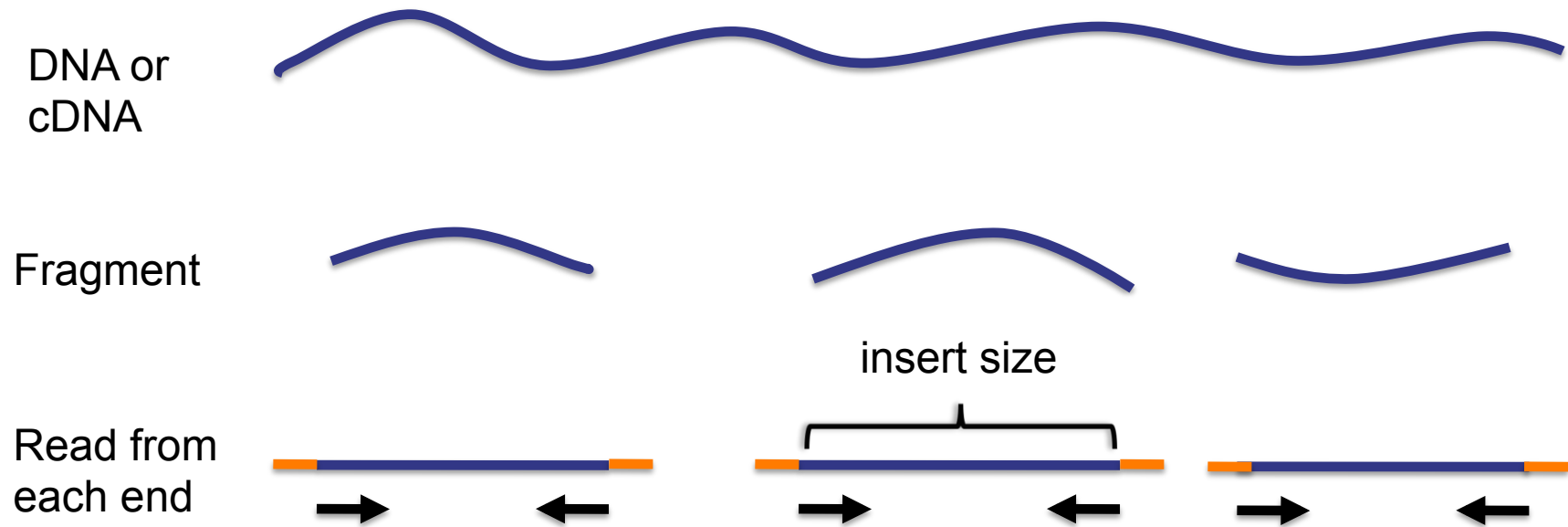
Viewing Structural Events

Structural events

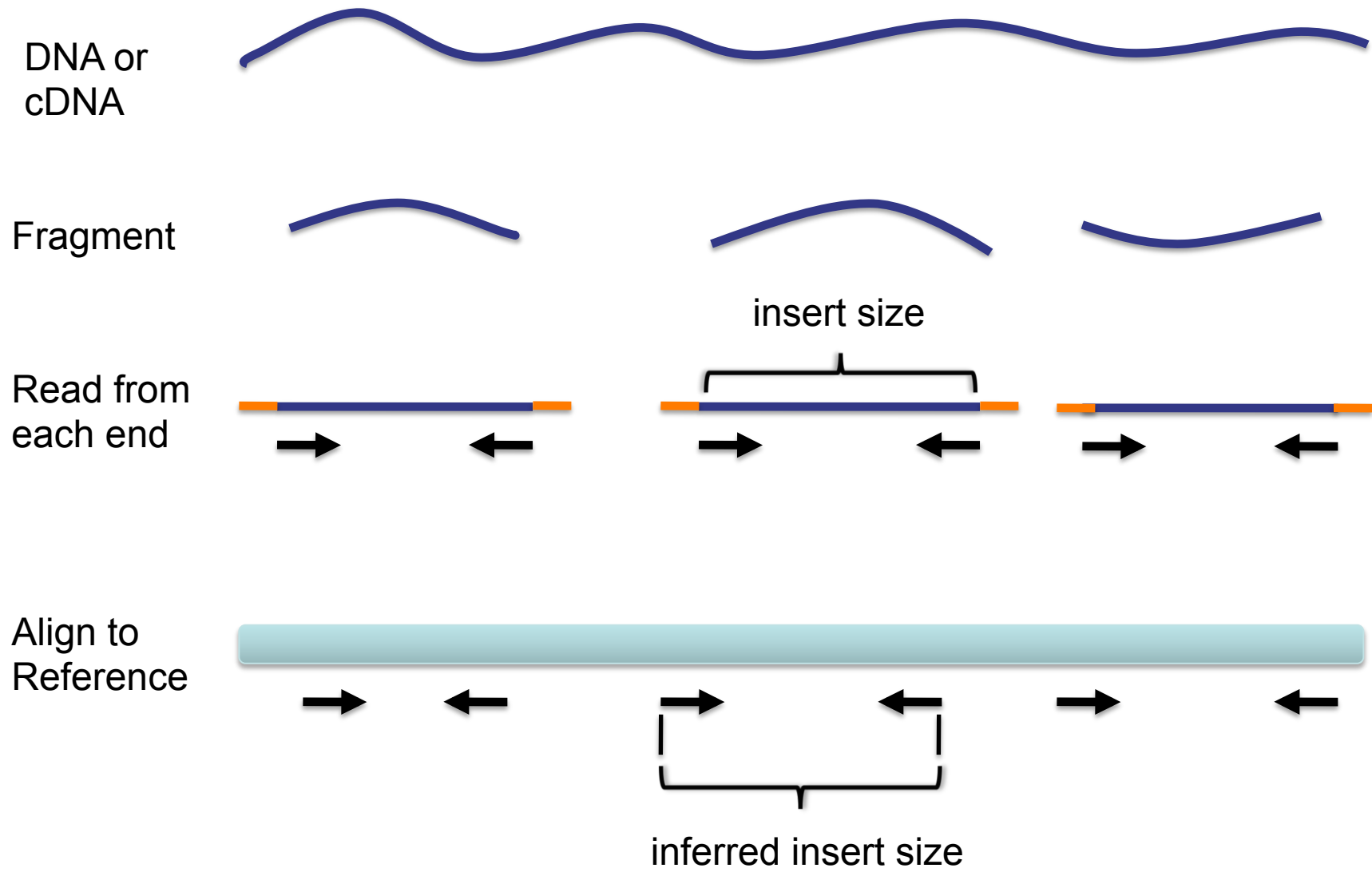


- Paired reads can yield evidence for genomic “structural events”, such as deletions, translocations, and inversions.
- Alignment coloring options help highlight these events based on:
 - Inferred insert size (template length)
 - Pair orientation (relative strand of pair)

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

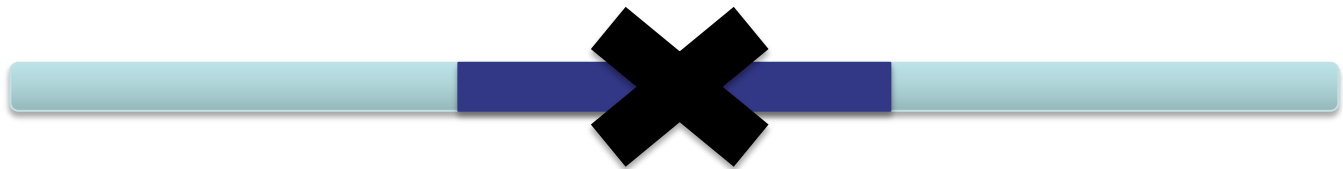


Deletion

Reference
Genome



Subject



Deletion



Reference
Genome



Subject



Deletion

Reference
Genome



Subject



Deletion

Reference
Genome



Subject

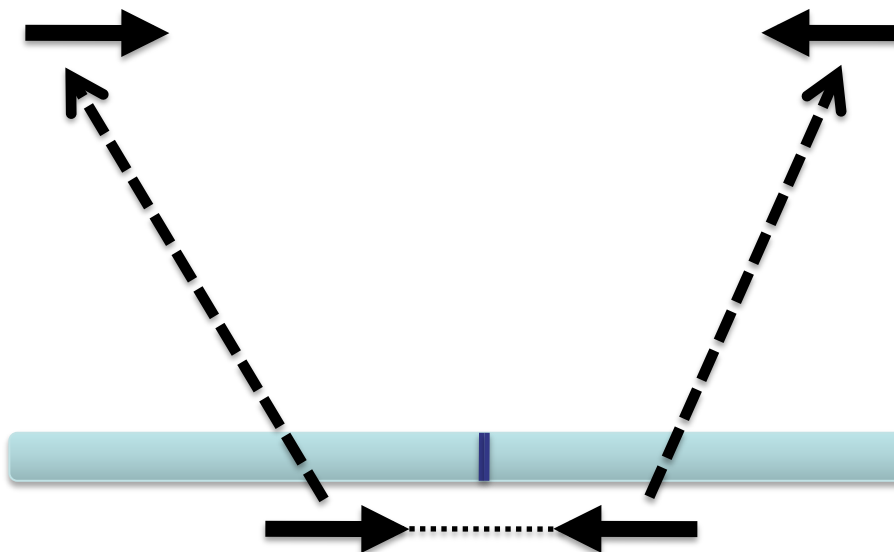


Deletion

Reference
Genome



Subject



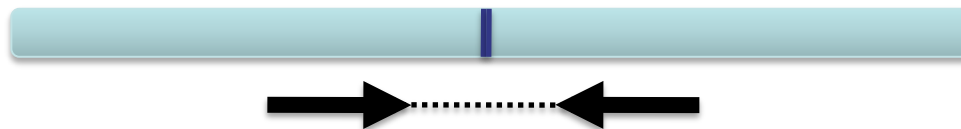
Deletion

Reference
Genome



inferred insert size

Subject



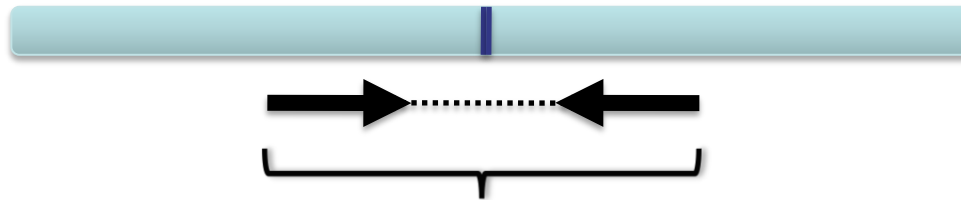
Deletion

Reference
Genome



inferred insert size

Subject



expected insert size

Deletion

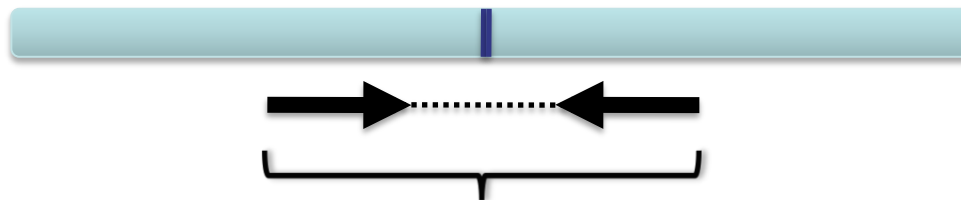
Inferred insert size is $>$ expected value

Reference
Genome



inferred insert size

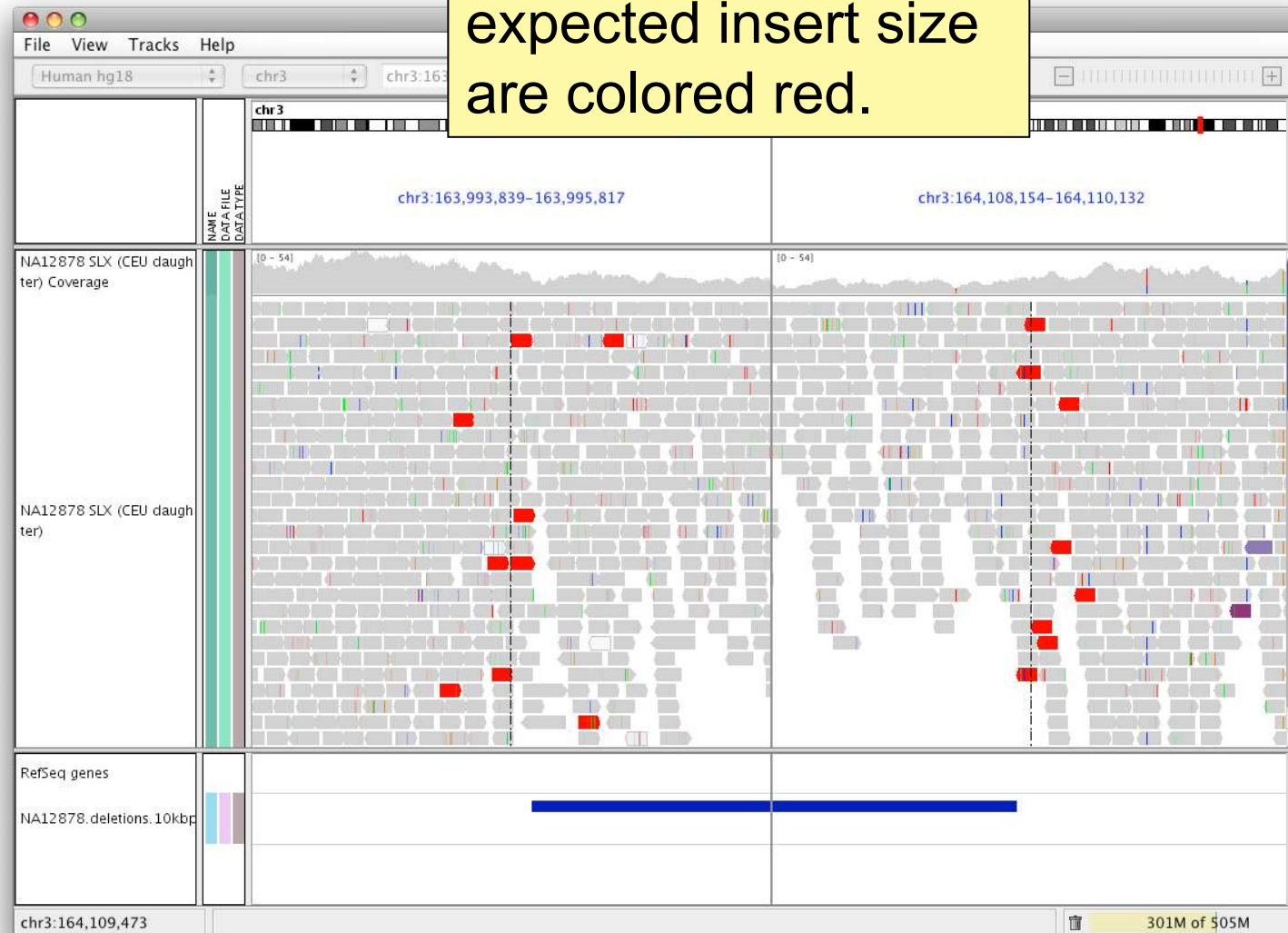
Subject



expected insert size

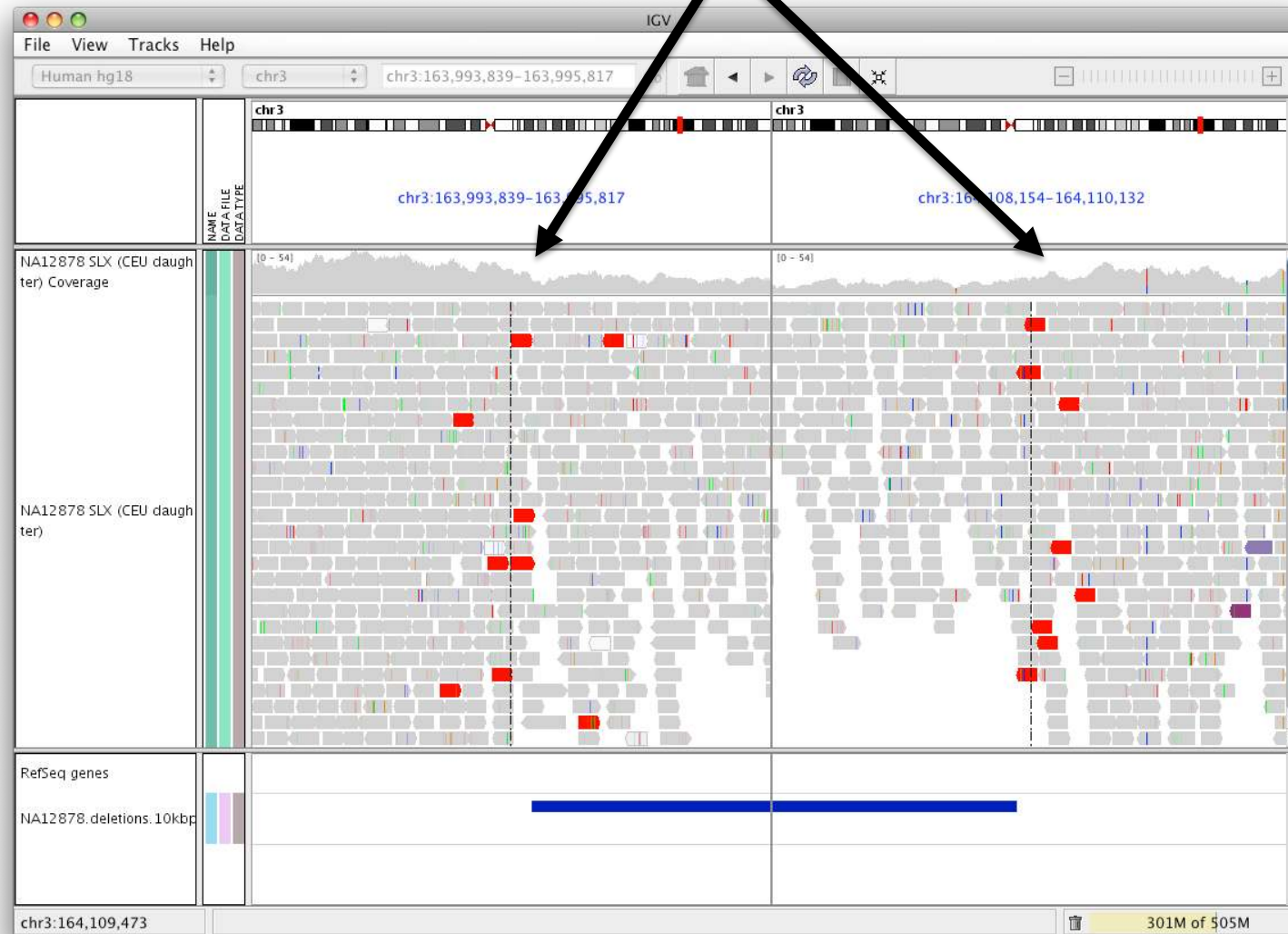
Deletion

Pairs with larger than expected insert size are colored red.



Deletion

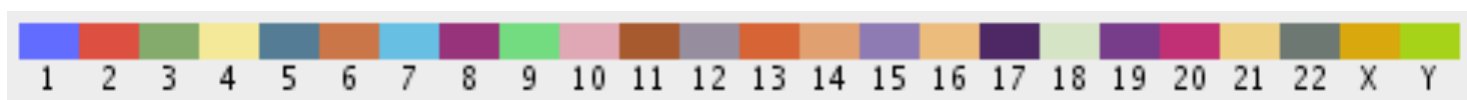
Note drop in coverage



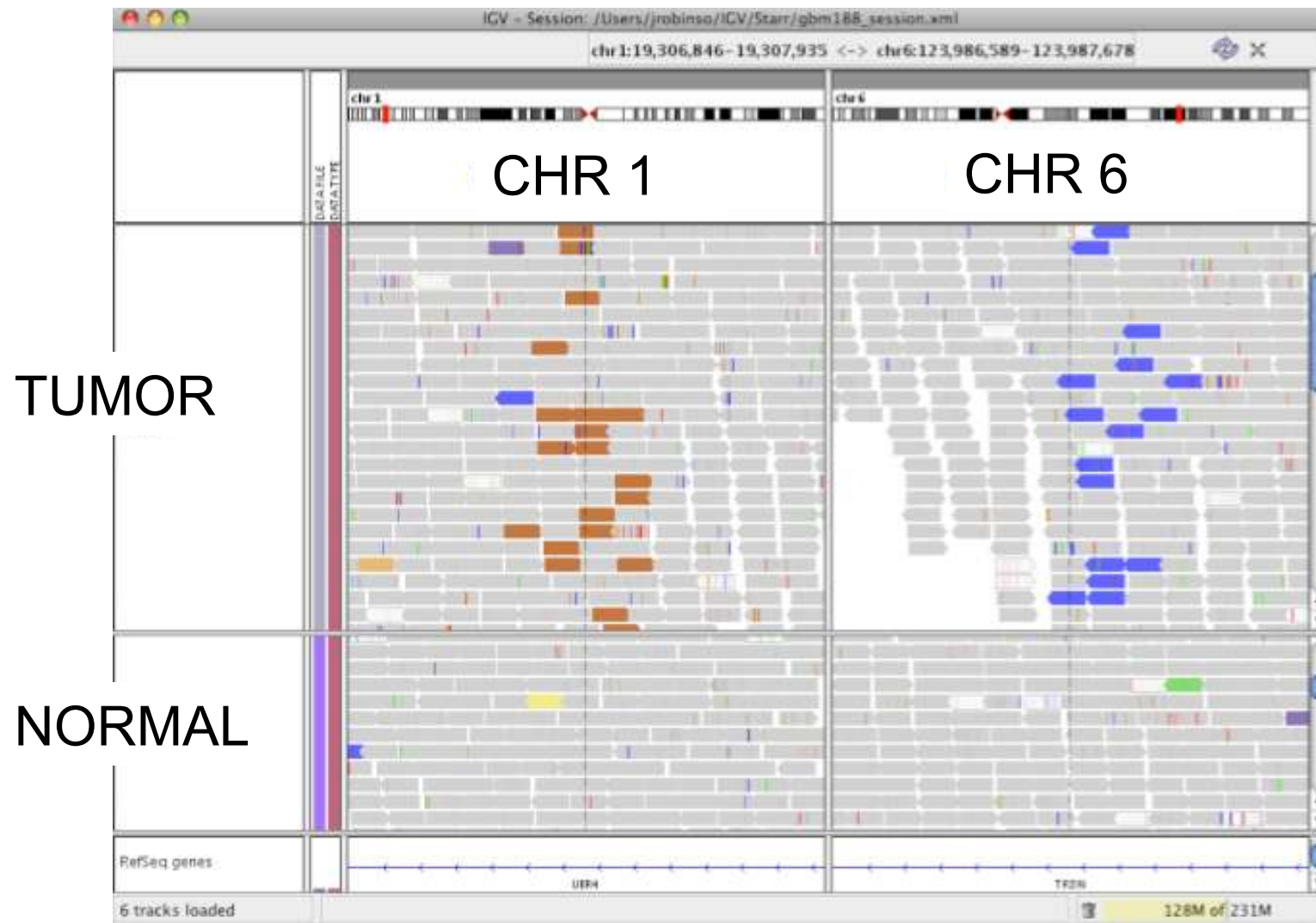
Insert size color scheme

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

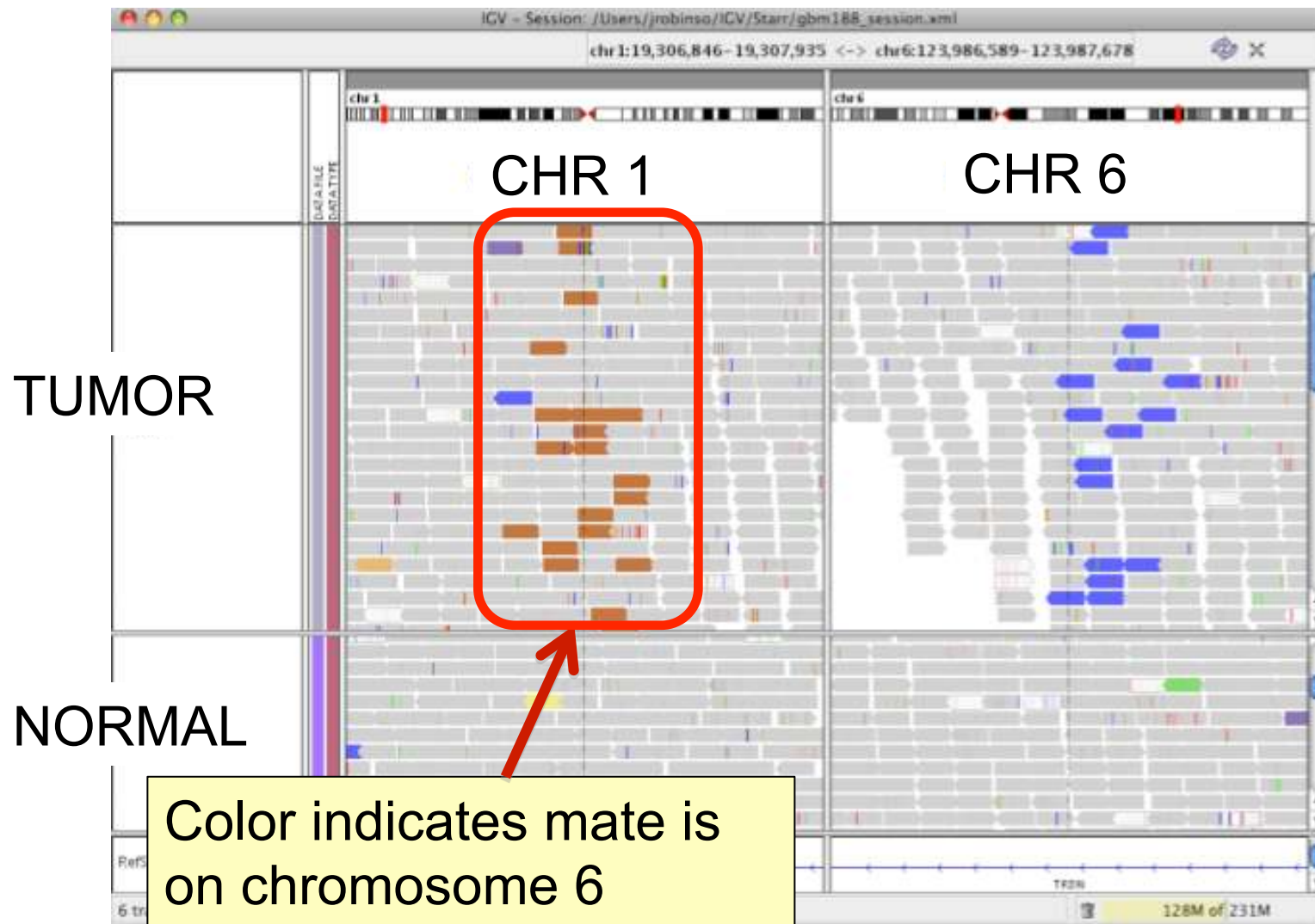
Each end colored by chromosome of its mate



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

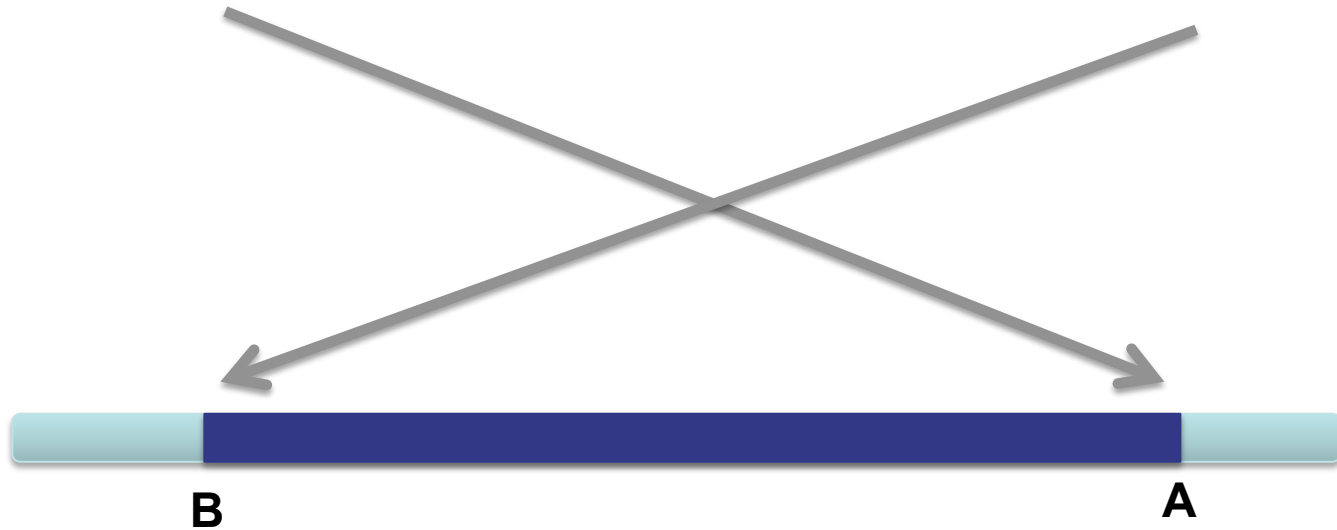


Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



Inversion

Reference
Genome



Subject



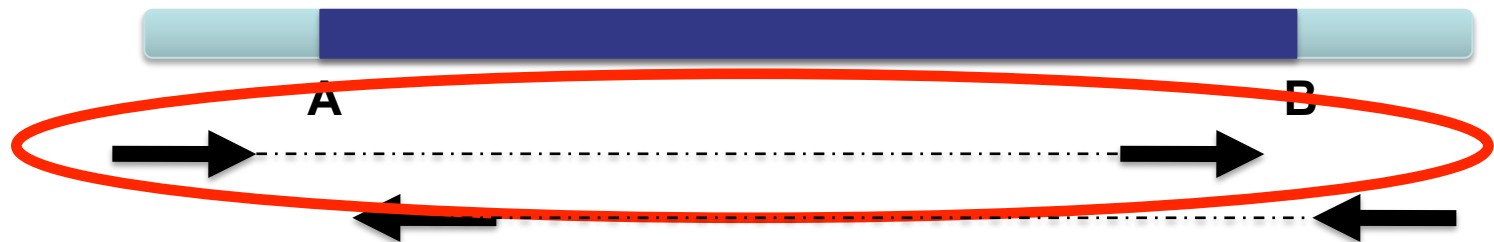
Inversion

Reference
Genome



Inversion

Reference
Genome



Anomaly –
Expected pair orientation is
inward facing ( )

Inversion

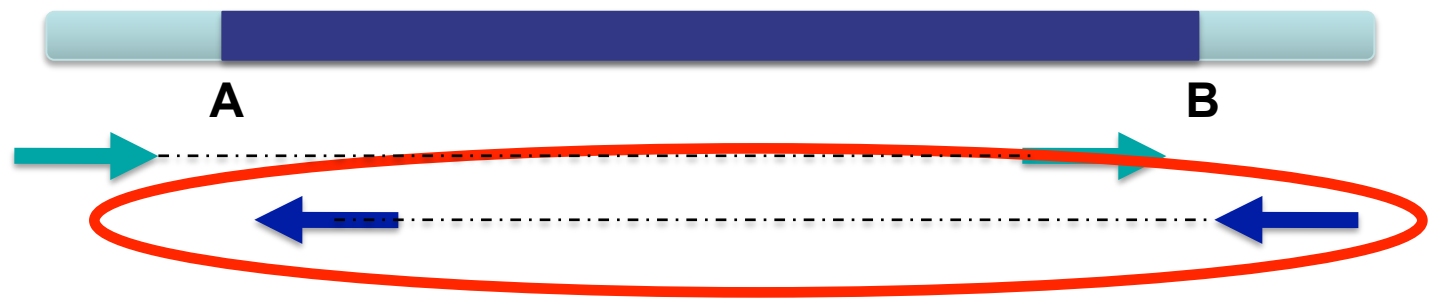
Reference
Genome



“Left” side pair

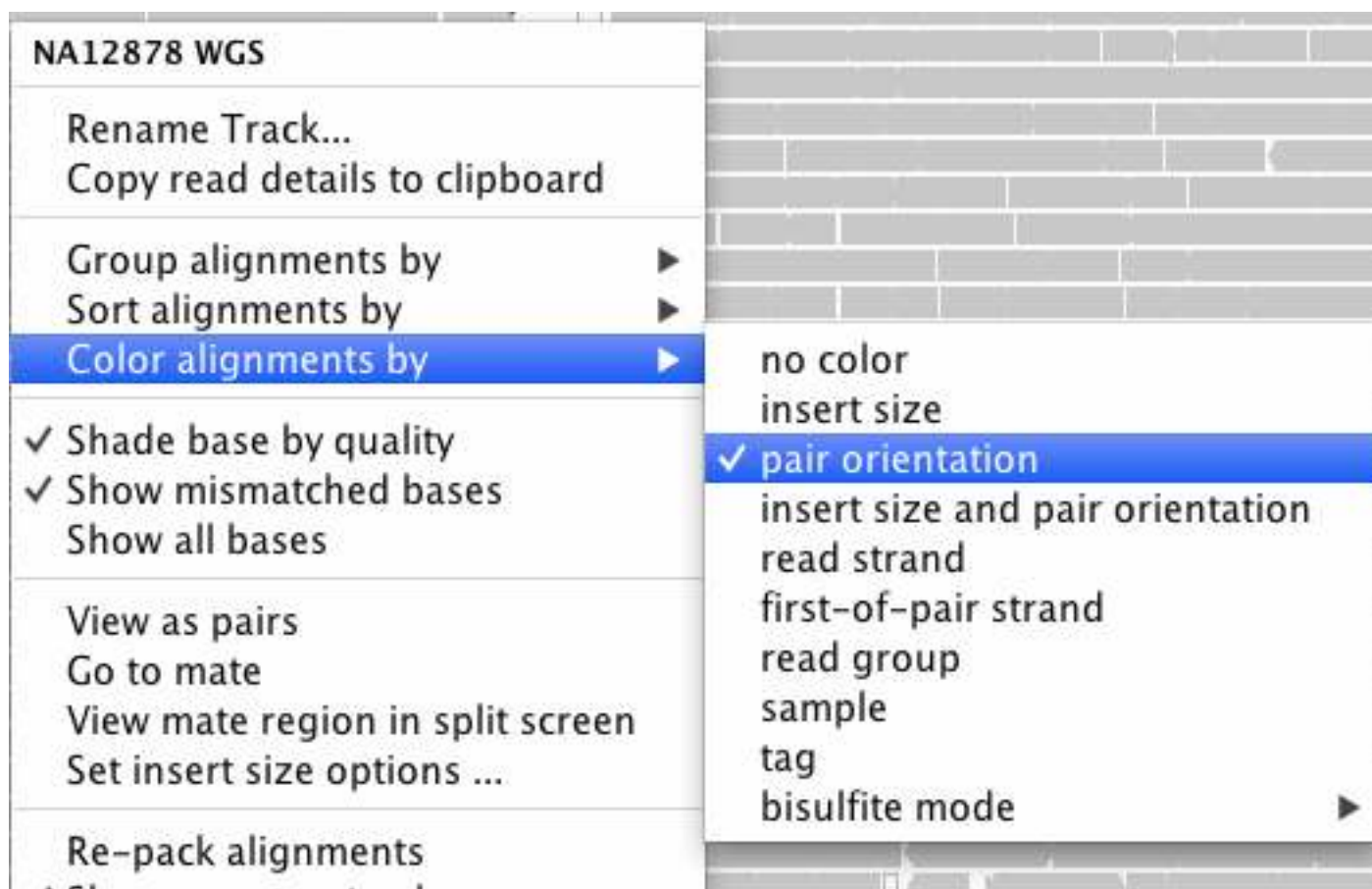
Inversion

Reference
Genome

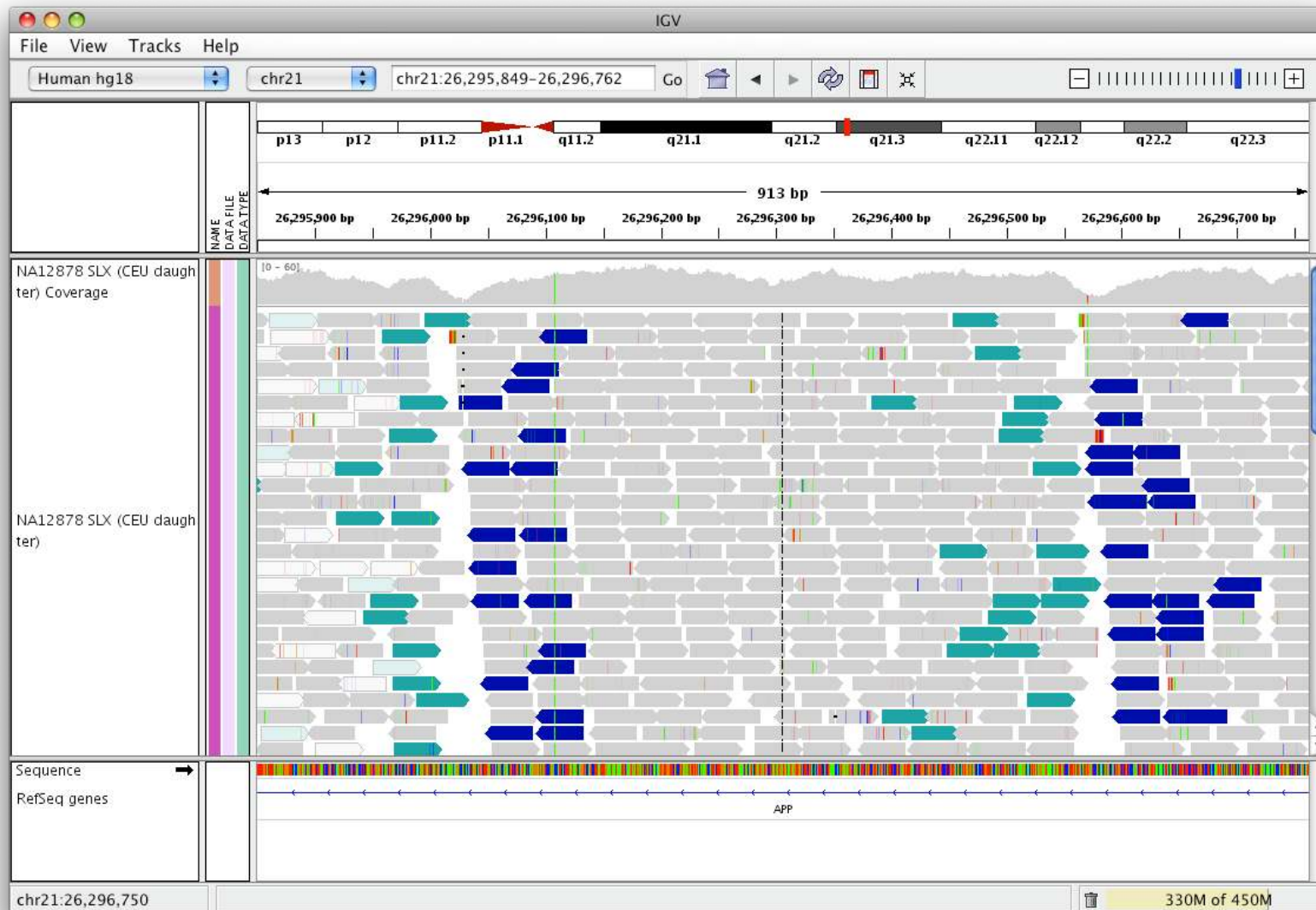


“Right” side pair

Color by pair orientation

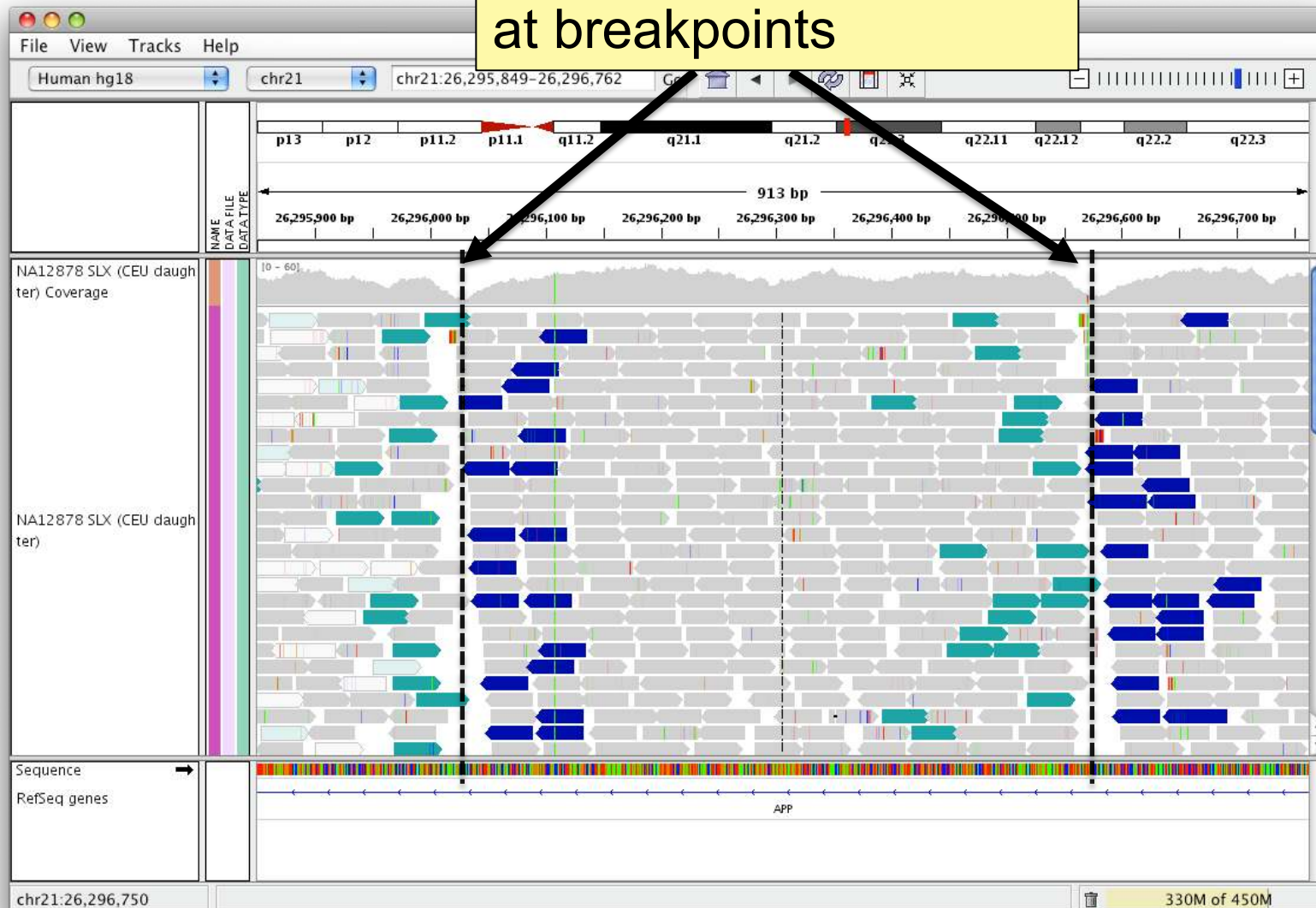


Inversion

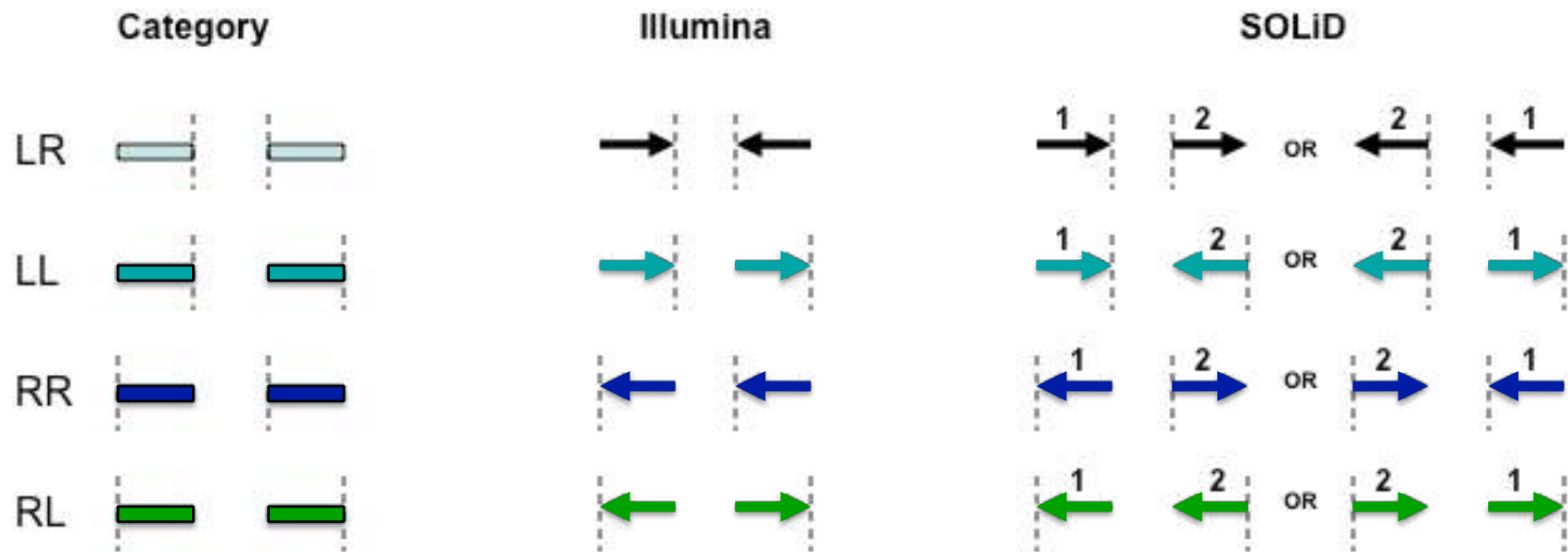


Inversion

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

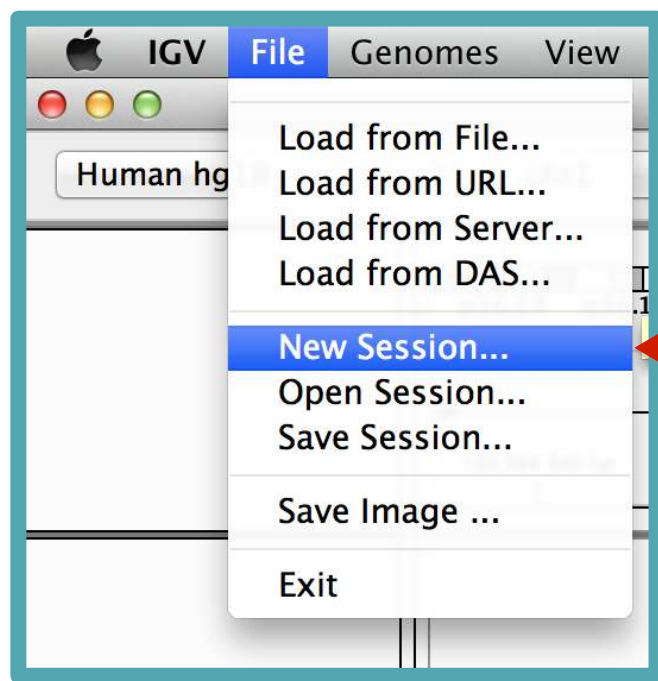
RNA-Seq



Hands-on exercise

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

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



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

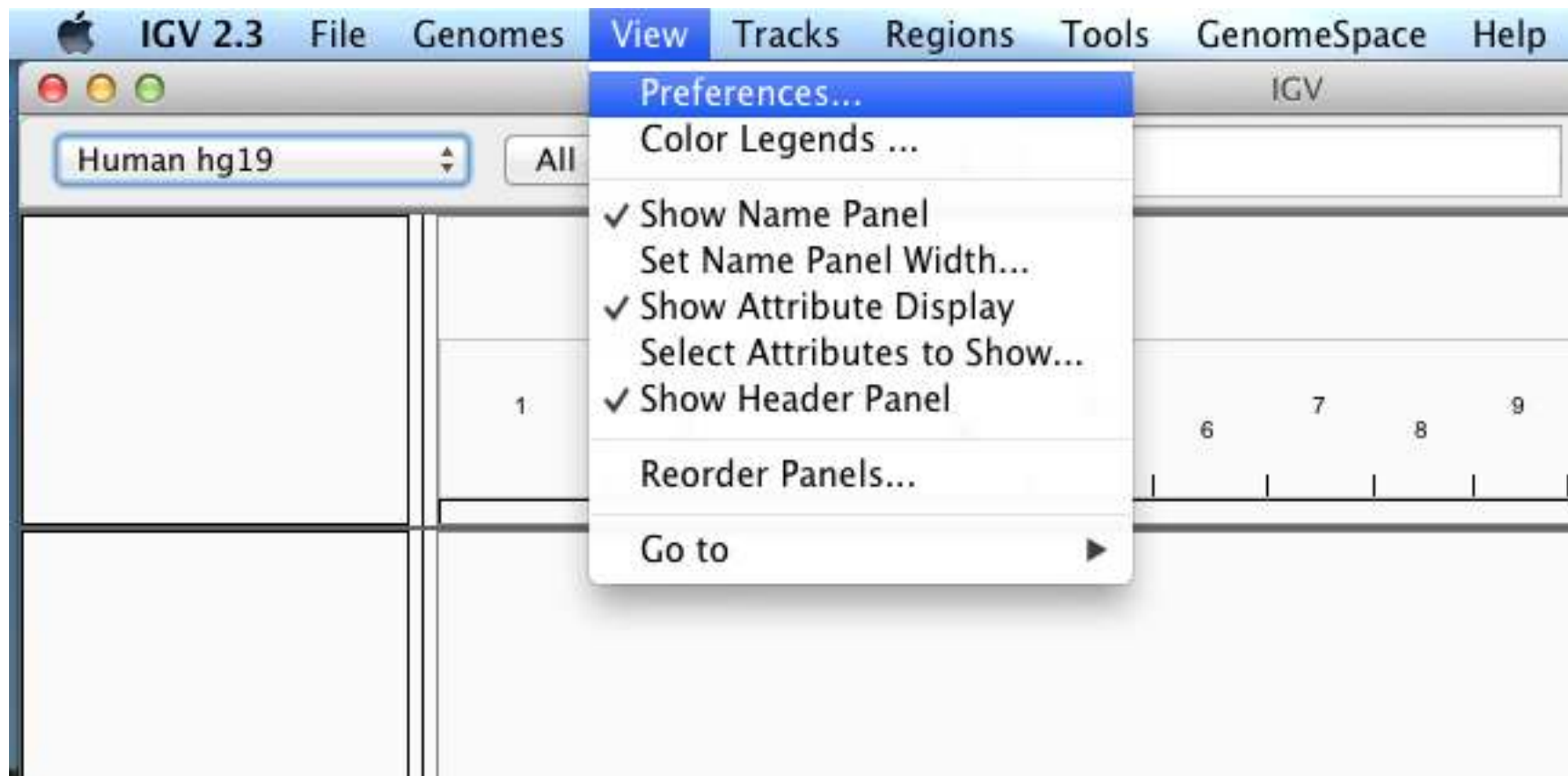
RNA-Seq Setup



- Step 1: Tune settings for RNA.

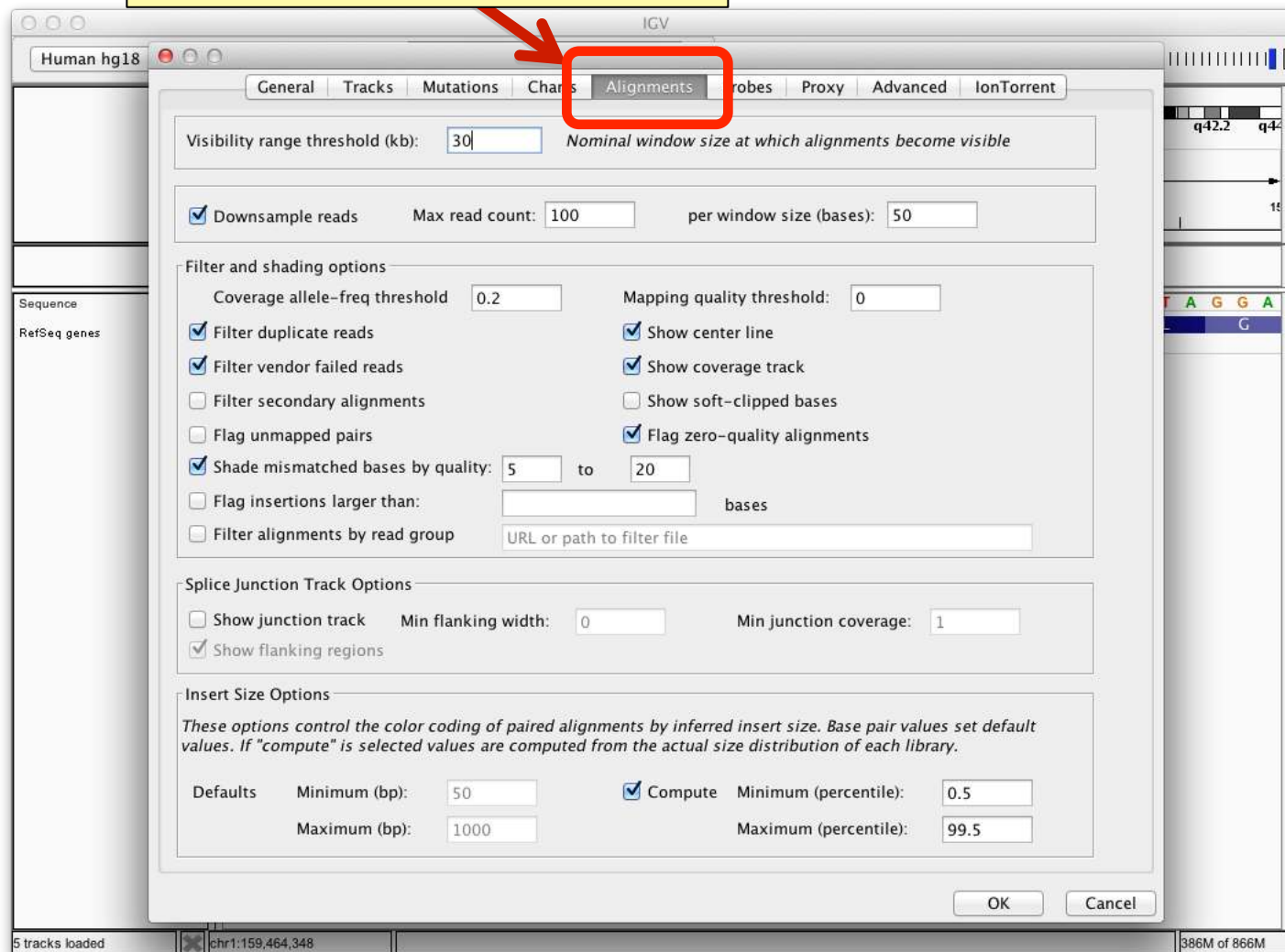
RNA-seq alignments

Select **View > Preferences...**

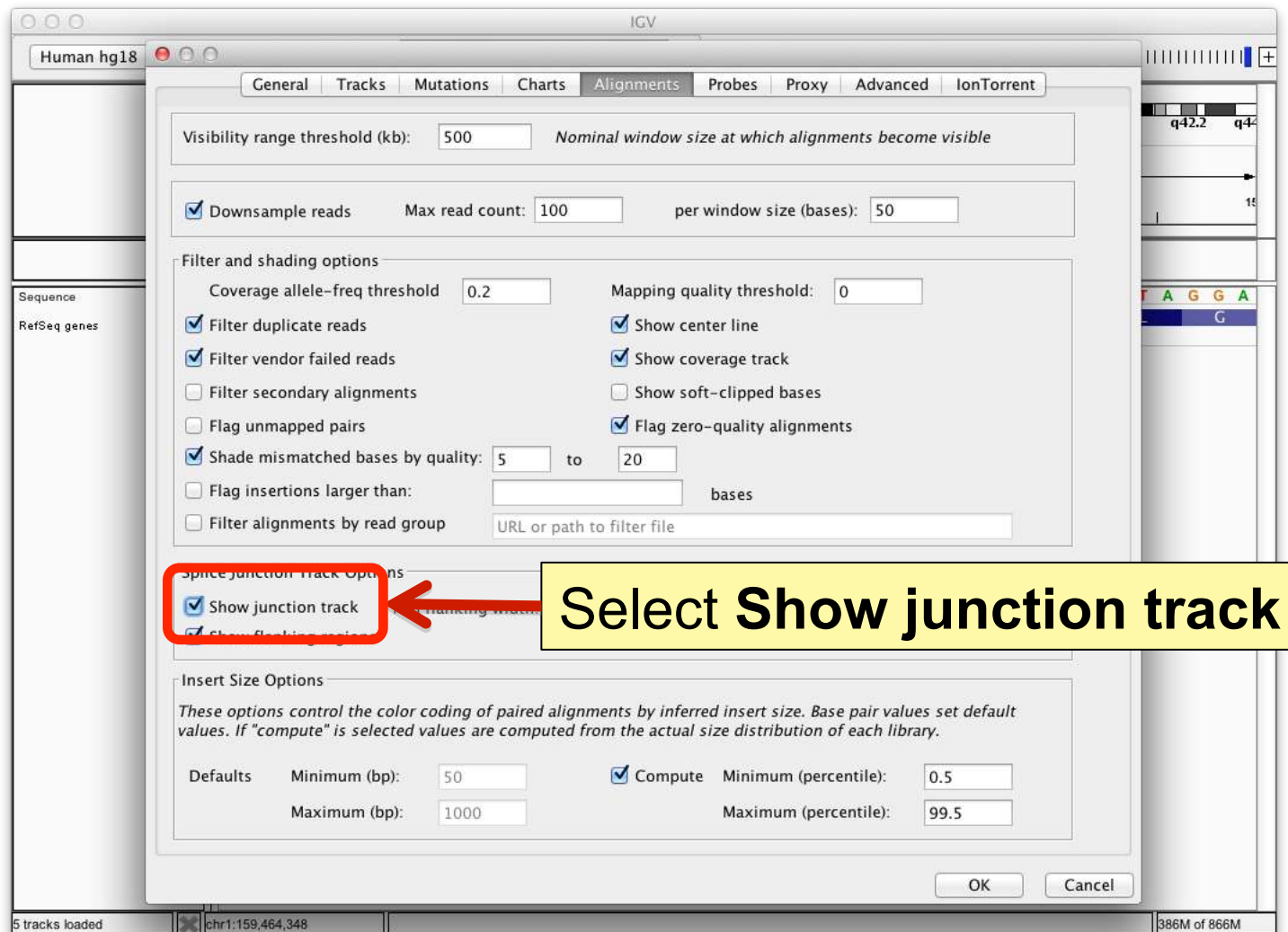


RNA-seq alignments

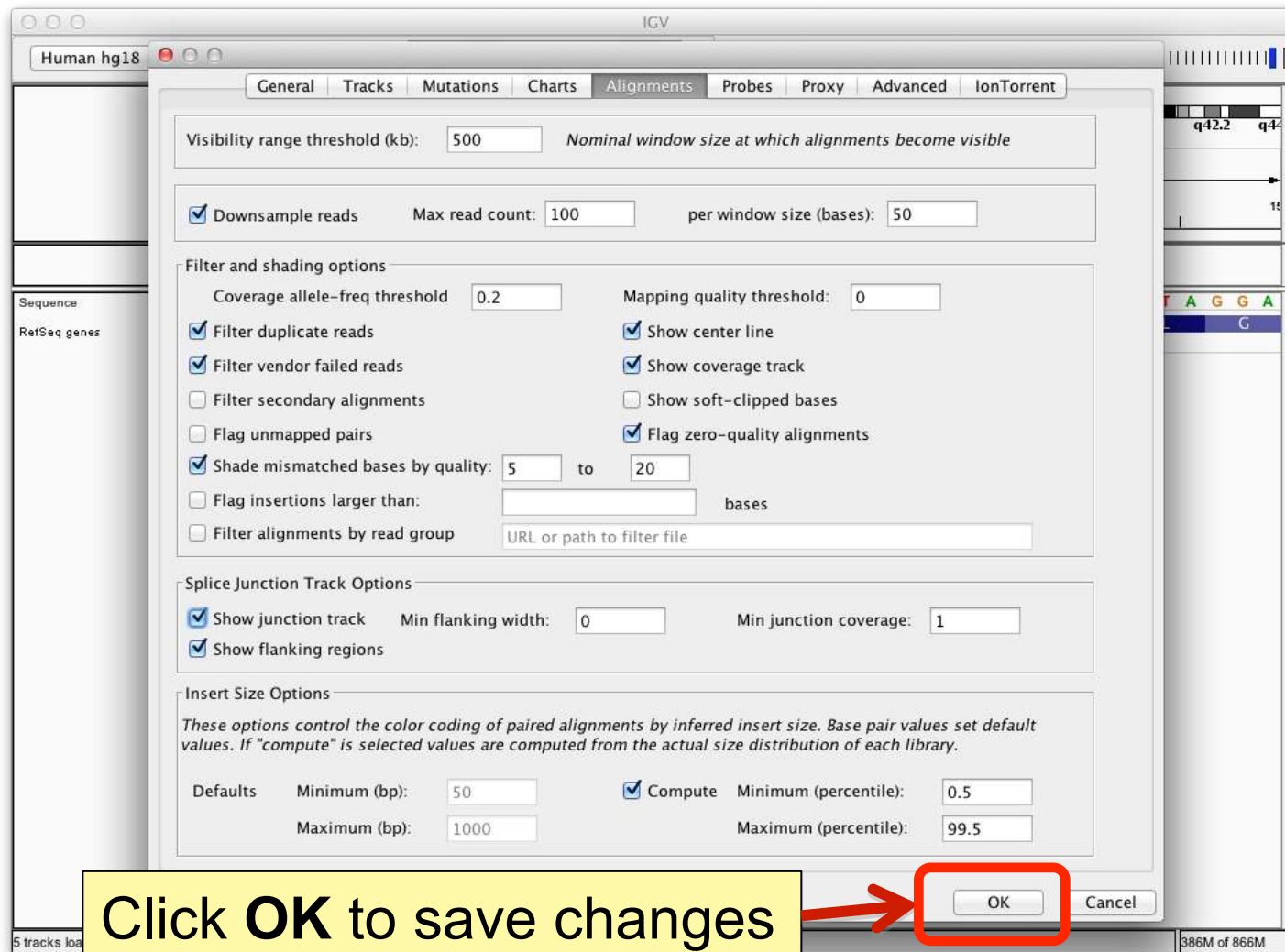
Click **Alignments** tab



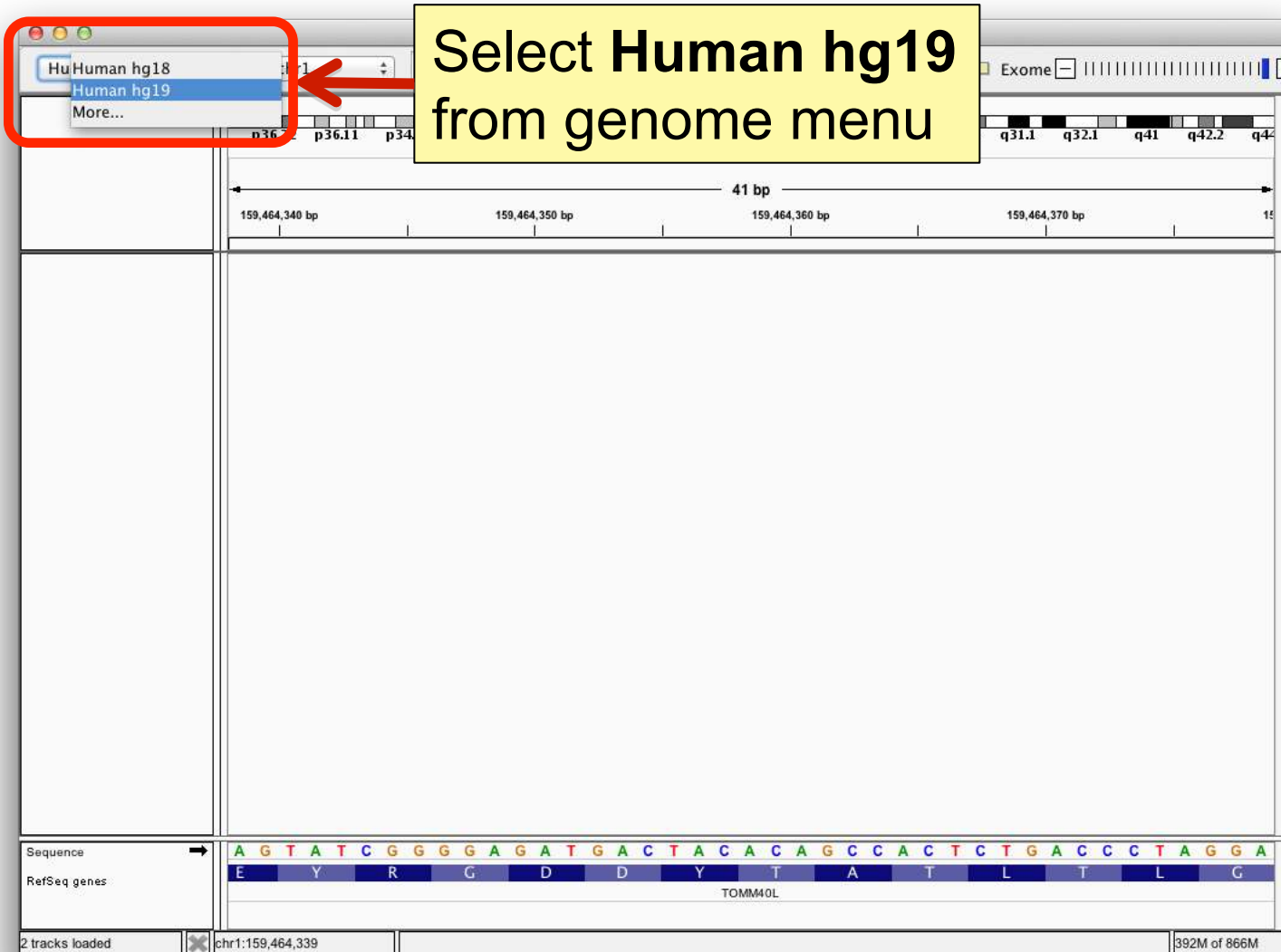
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments



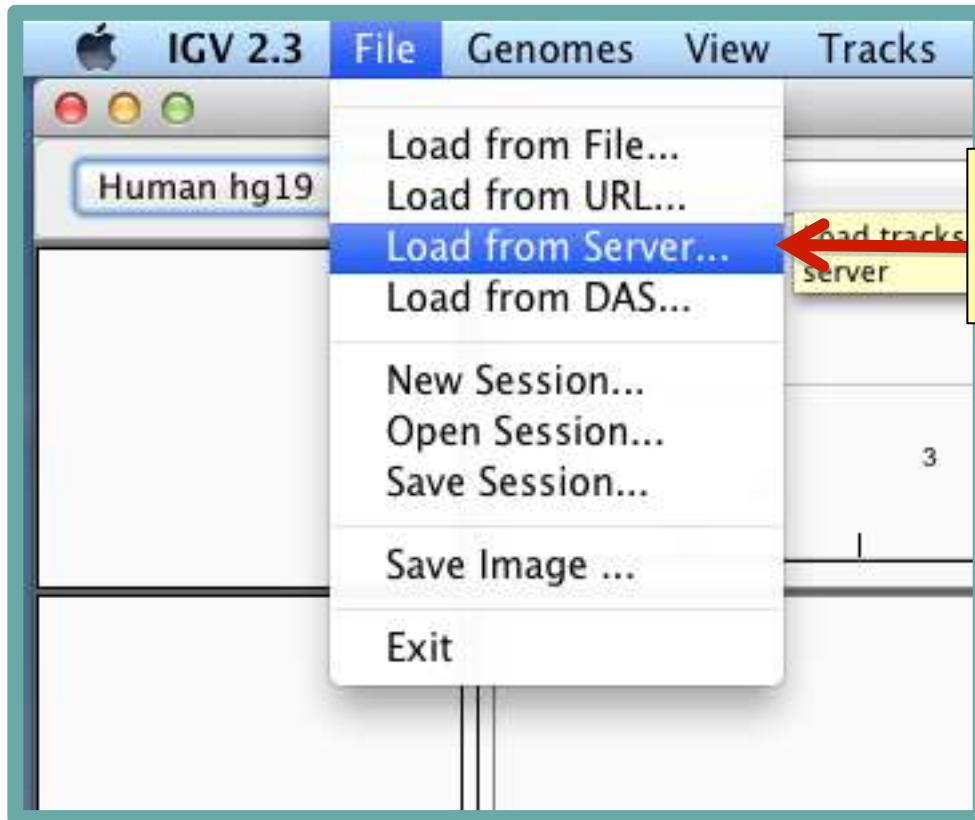
The screenshot shows the IGV interface with a red box around the genome menu in the top-left corner. The menu lists "HuHuman hg18", "Human hg19" (which is highlighted), and "More...". A red arrow points from a yellow text box to the "Human hg19" option. The yellow box contains the text "Select Human hg19 from genome menu". The main view shows a genomic track with a 41 bp scale bar and coordinates from 159,464,340 bp to 159,464,370 bp. The bottom track shows the 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" and the RefSeq gene "TOMM40L".

Select Human hg19
from genome menu

Sequence
RefSeq genes

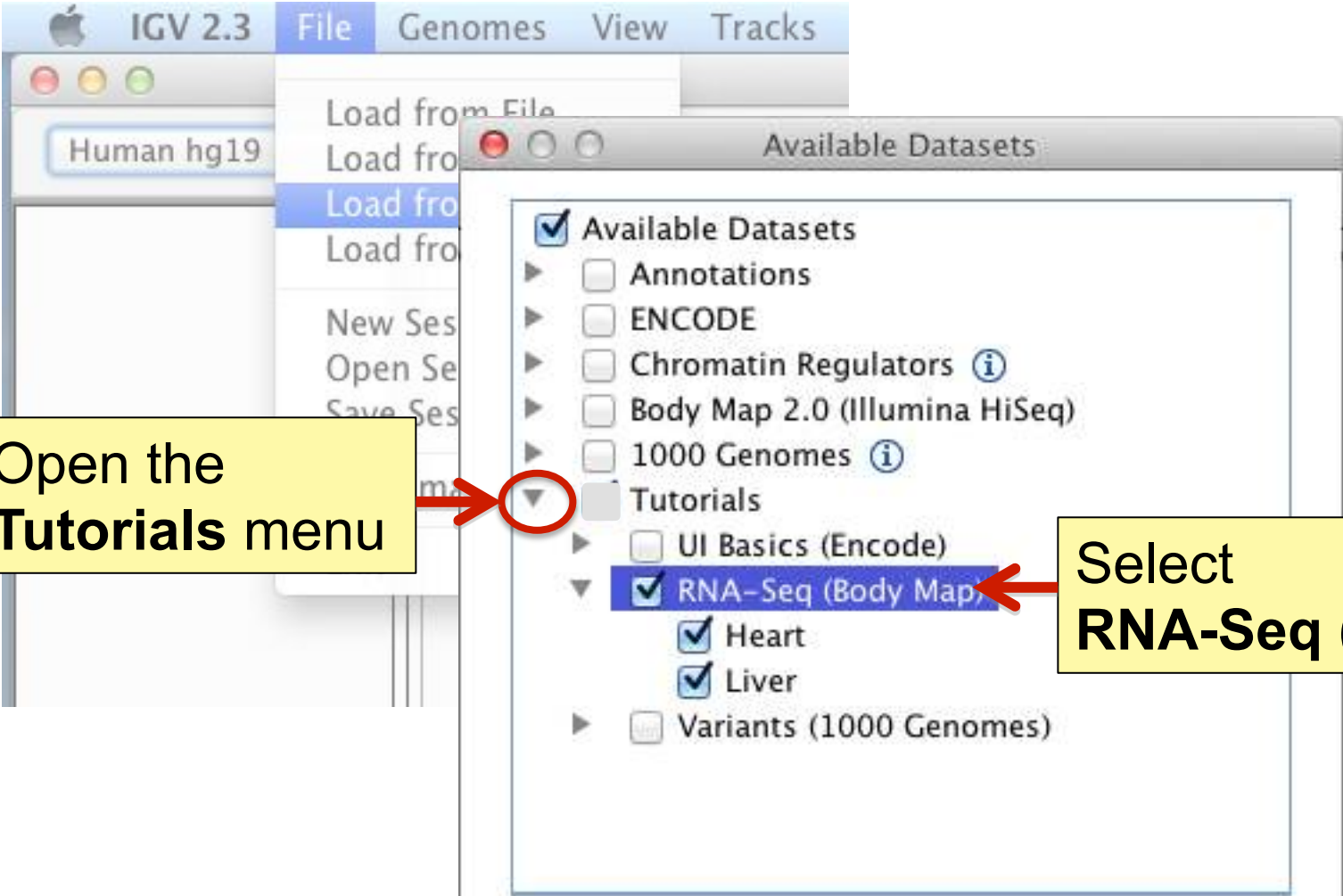
2 tracks loaded | chr1:159,464,339 | 392M of 866M

RNA-seq alignments



Select:
File > Load from Server...

RNA-seq alignments



IGV 2.3 File Genomes View Tracks

Human hg19

Load from File
Load from
Load from
Load from

New Ses
Open Se
Save Ses

Available Datasets

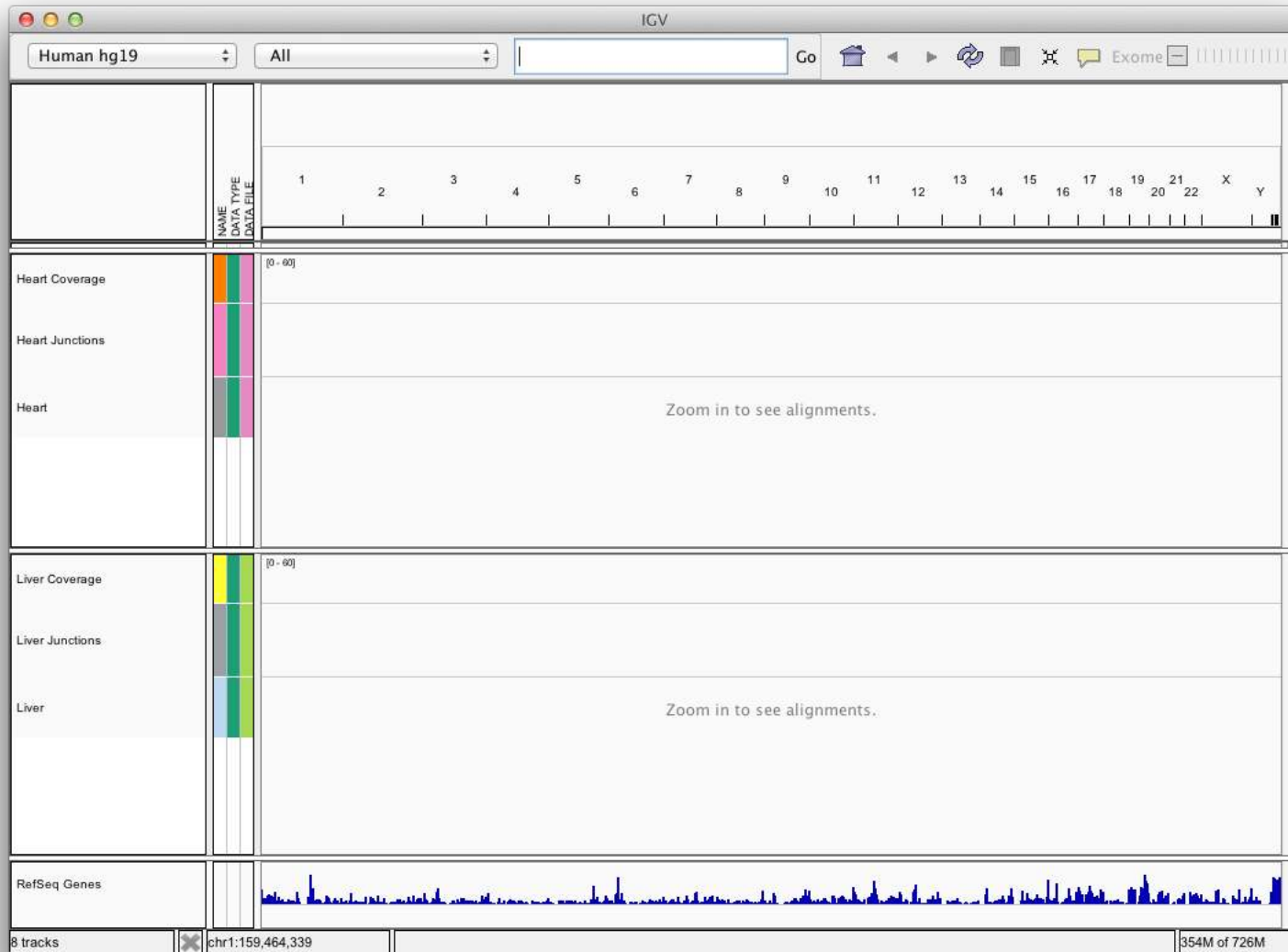
- ☒ Available Datasets
- ☐ Annotations
- ☐ ENCODE
- ☐ Chromatin Regulators ⓘ
- ☐ Body Map 2.0 (Illumina HiSeq)
- ☐ 1000 Genomes ⓘ
- ☒ Tutorials
 - ☐ UI Basics (Encode)
 - ☒ RNA-Seq (Body Map)
 - ☒ Heart
 - ☒ Liver
 - ☐ Variants (1000 Genomes)

Cancel OK

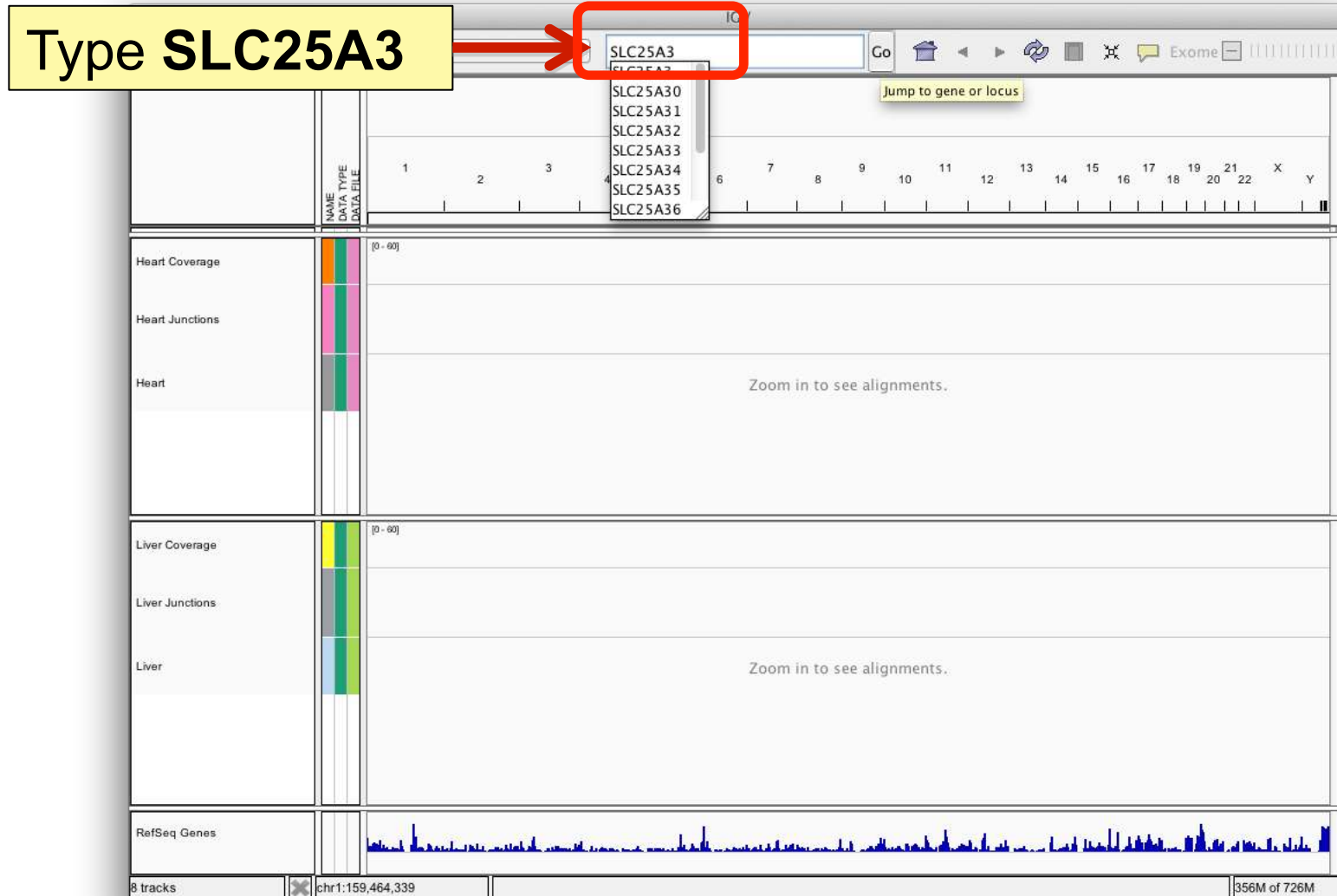
Open the Tutorials menu

Select RNA-Seq (Body Map)

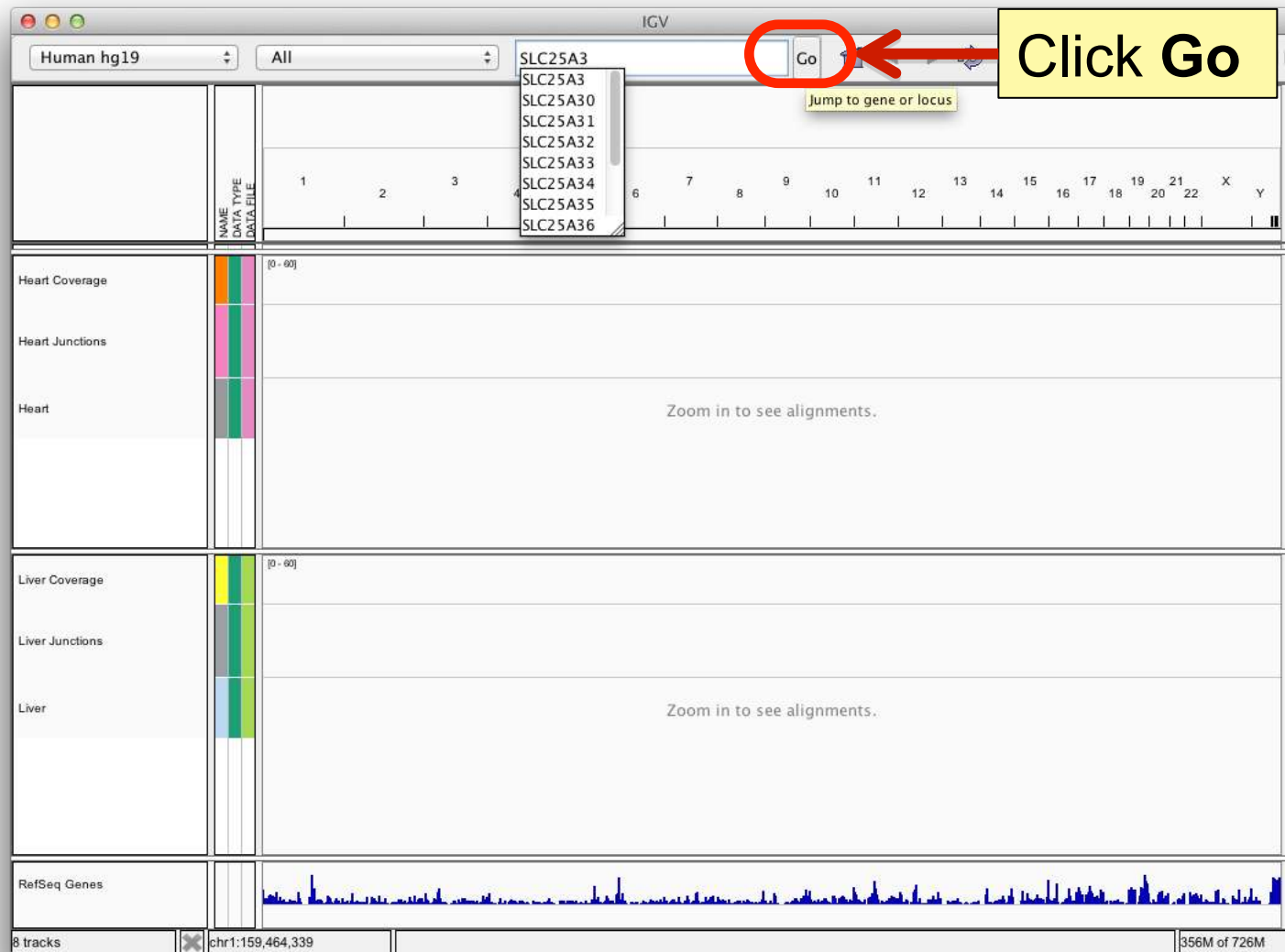
RNA-seq alignments



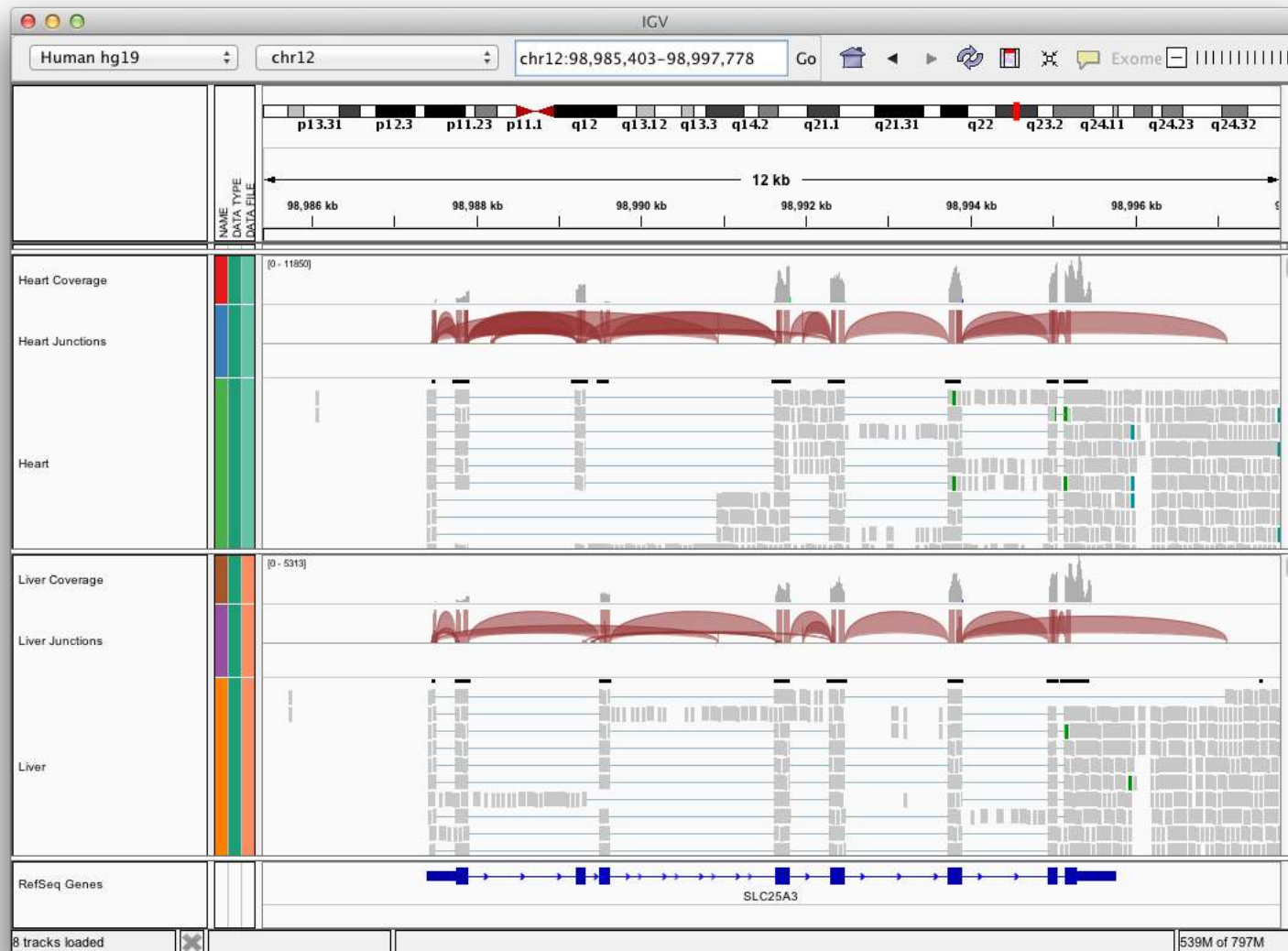
RNA-seq alignments



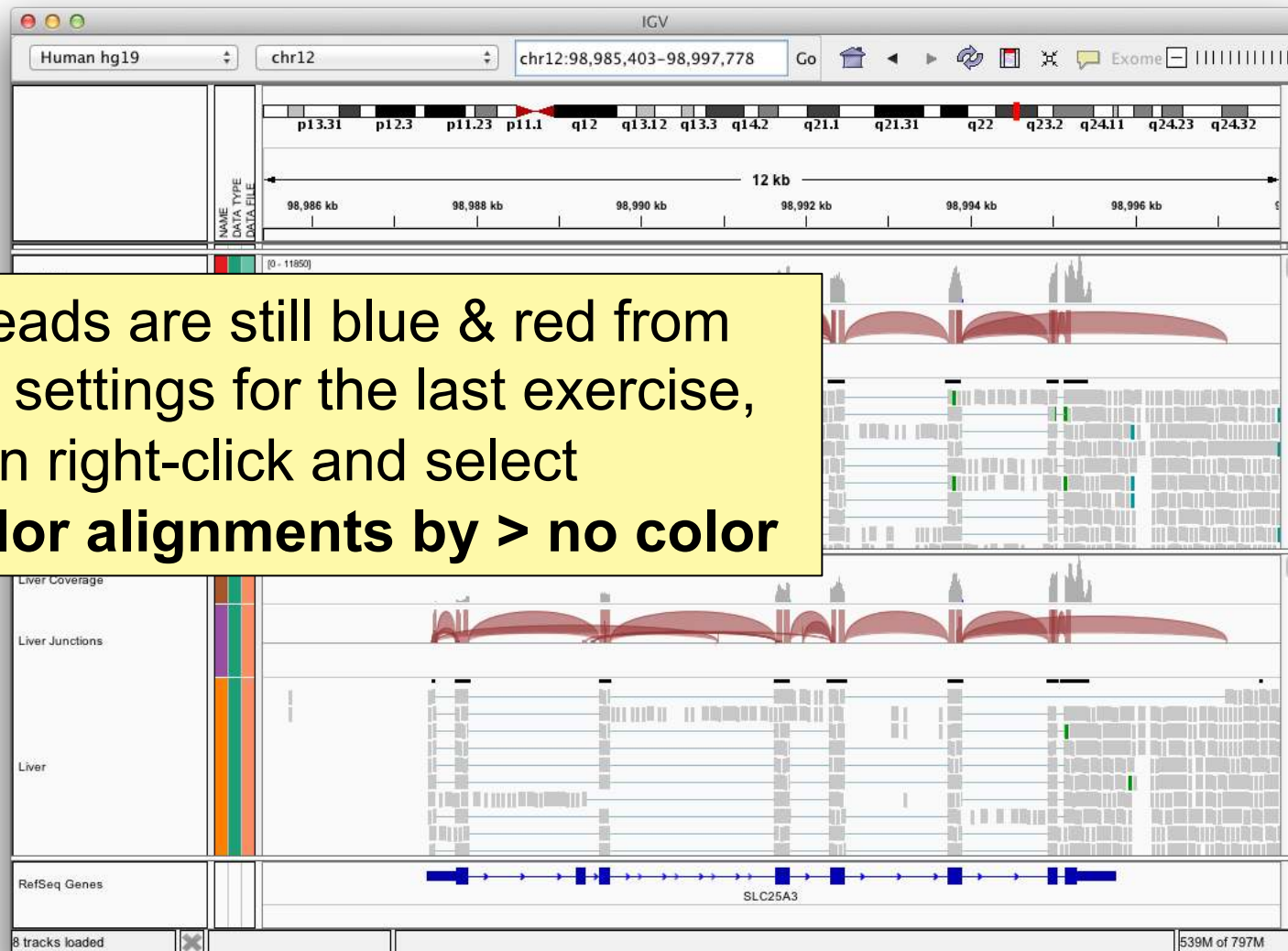
RNA-seq alignments



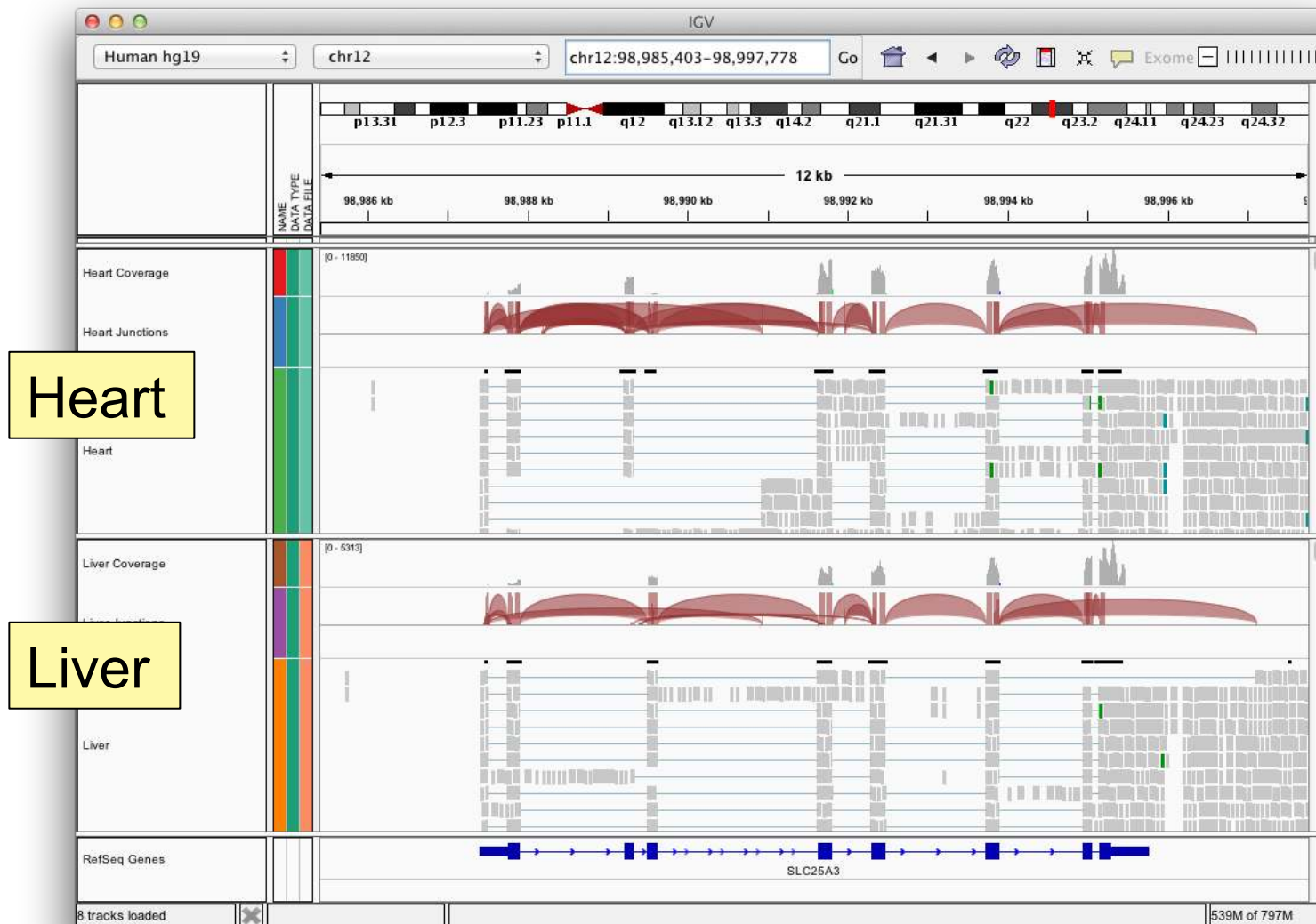
RNA-seq alignments



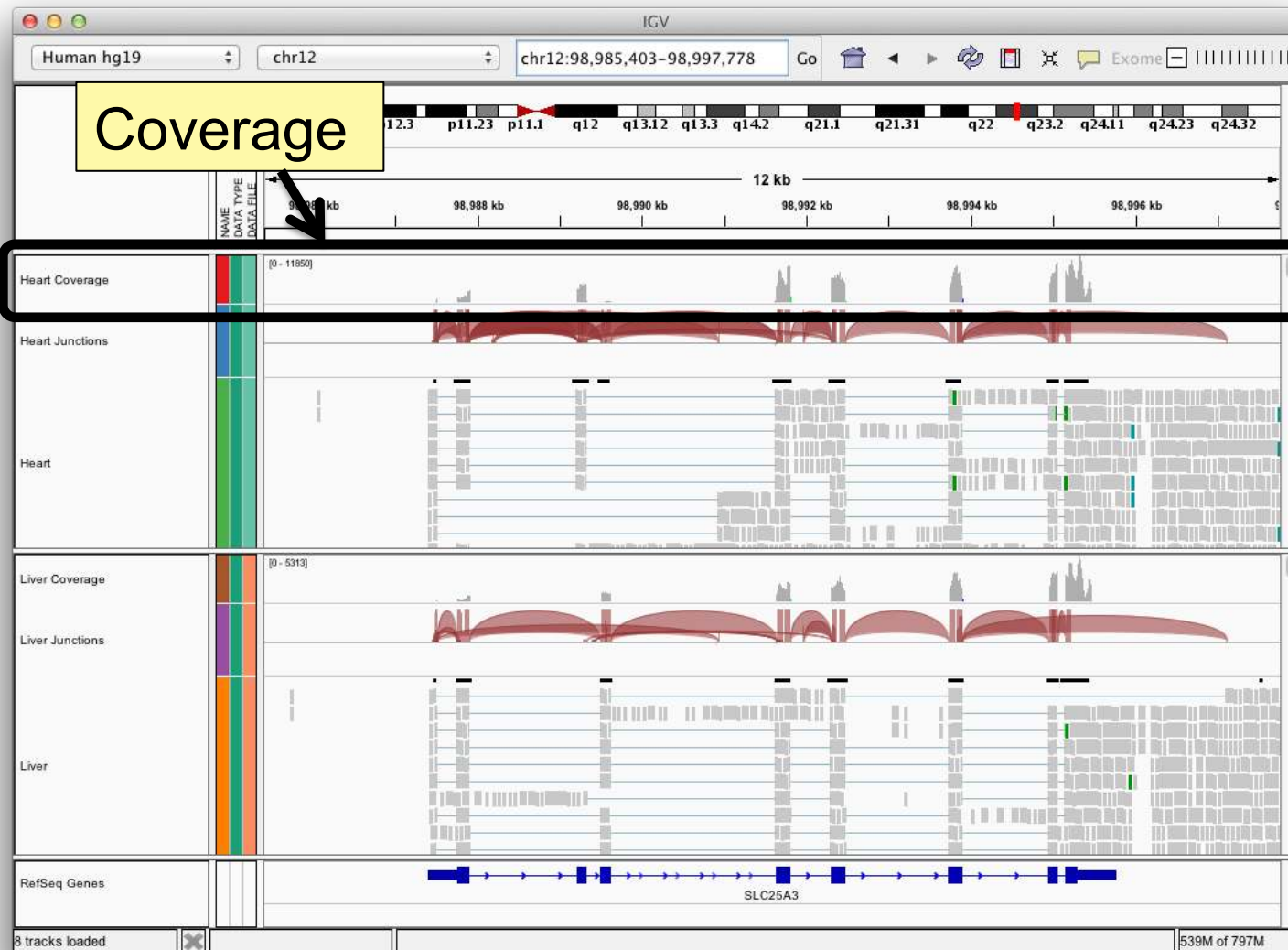
RNA-seq alignments



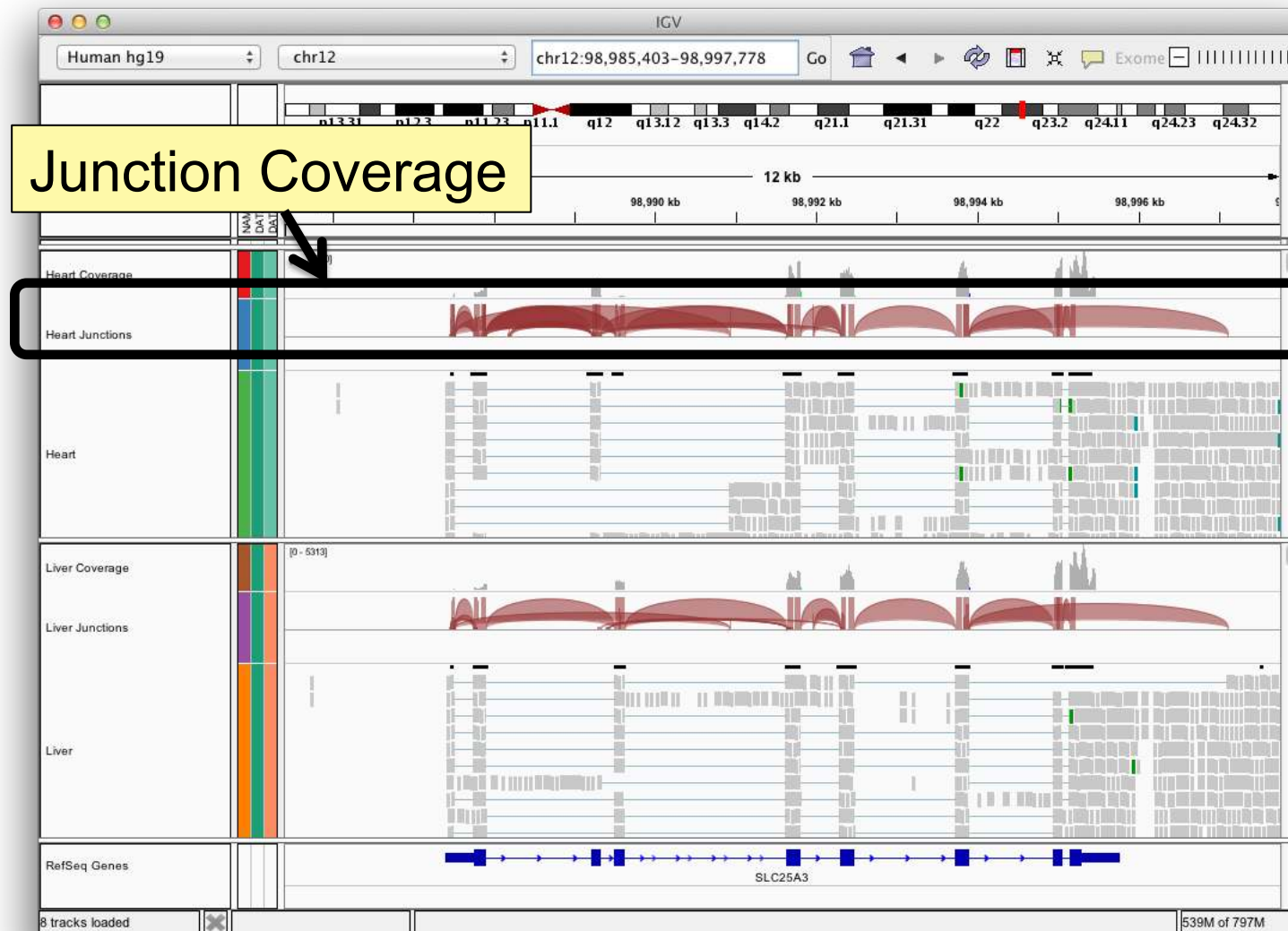
RNA-seq alignments



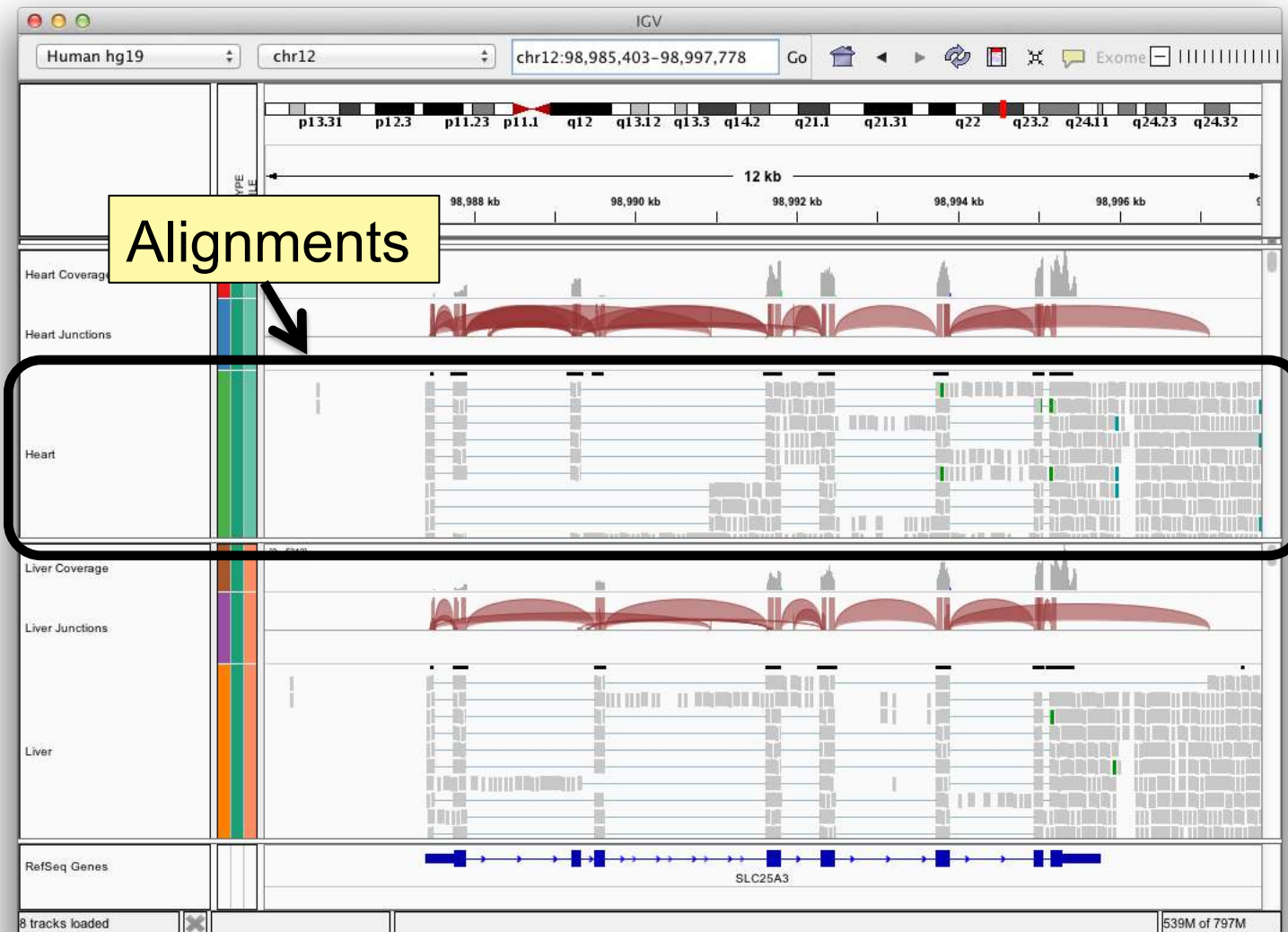
RNA-seq alignments



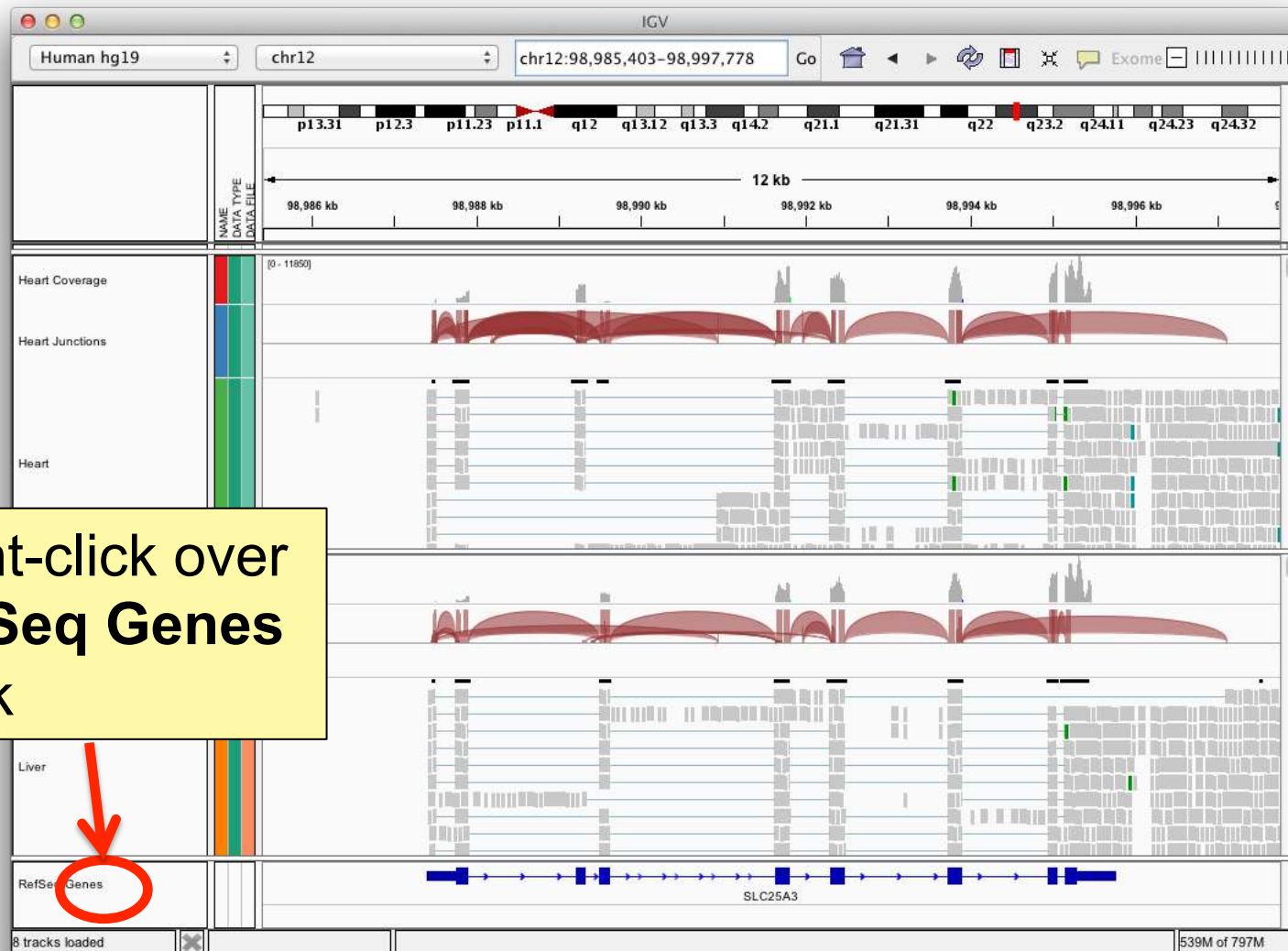
RNA-seq alignments



RNA-seq alignments

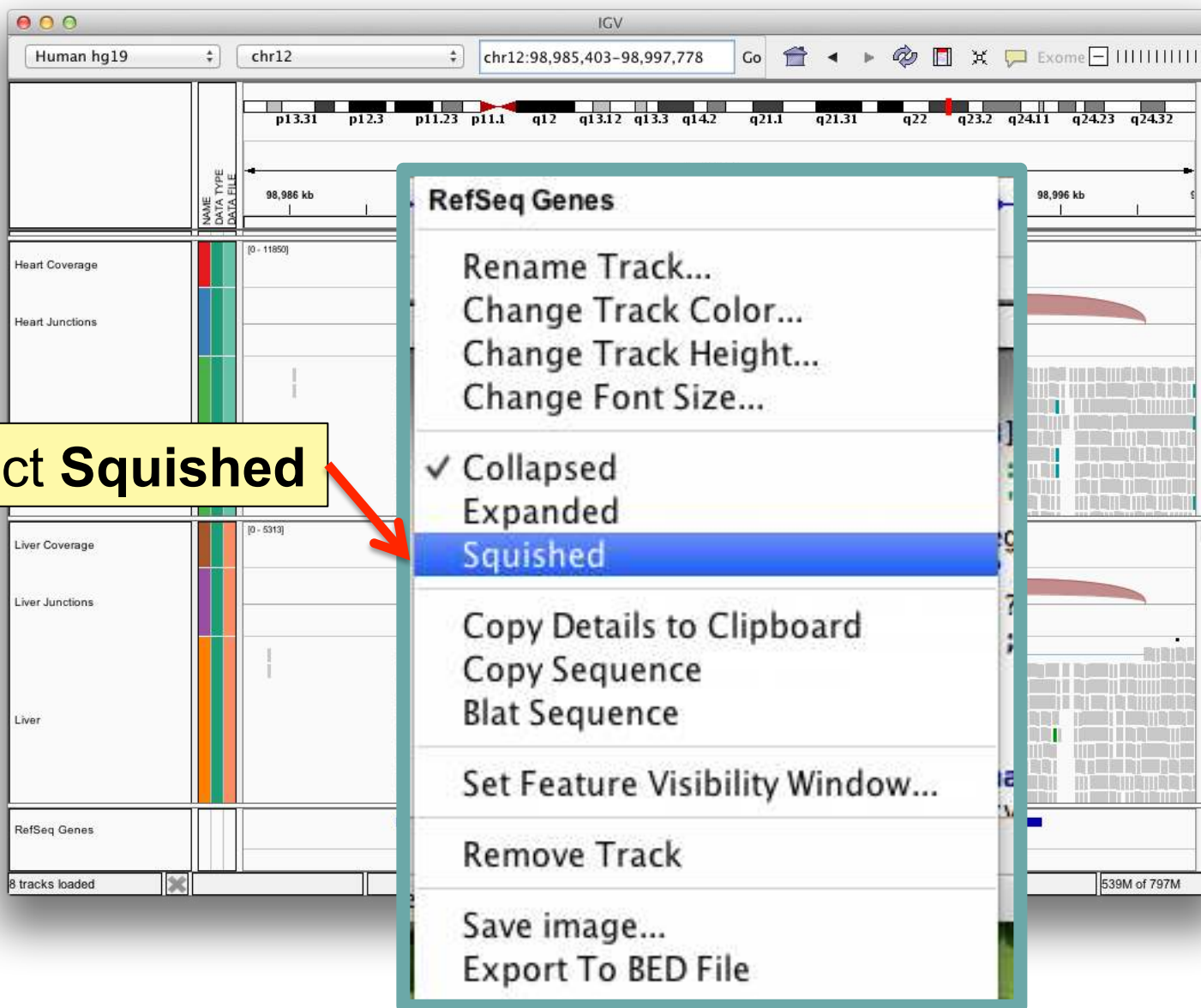


RNA-seq alignments



Right-click over
RefSeq Genes
track

RNA-seq alignments



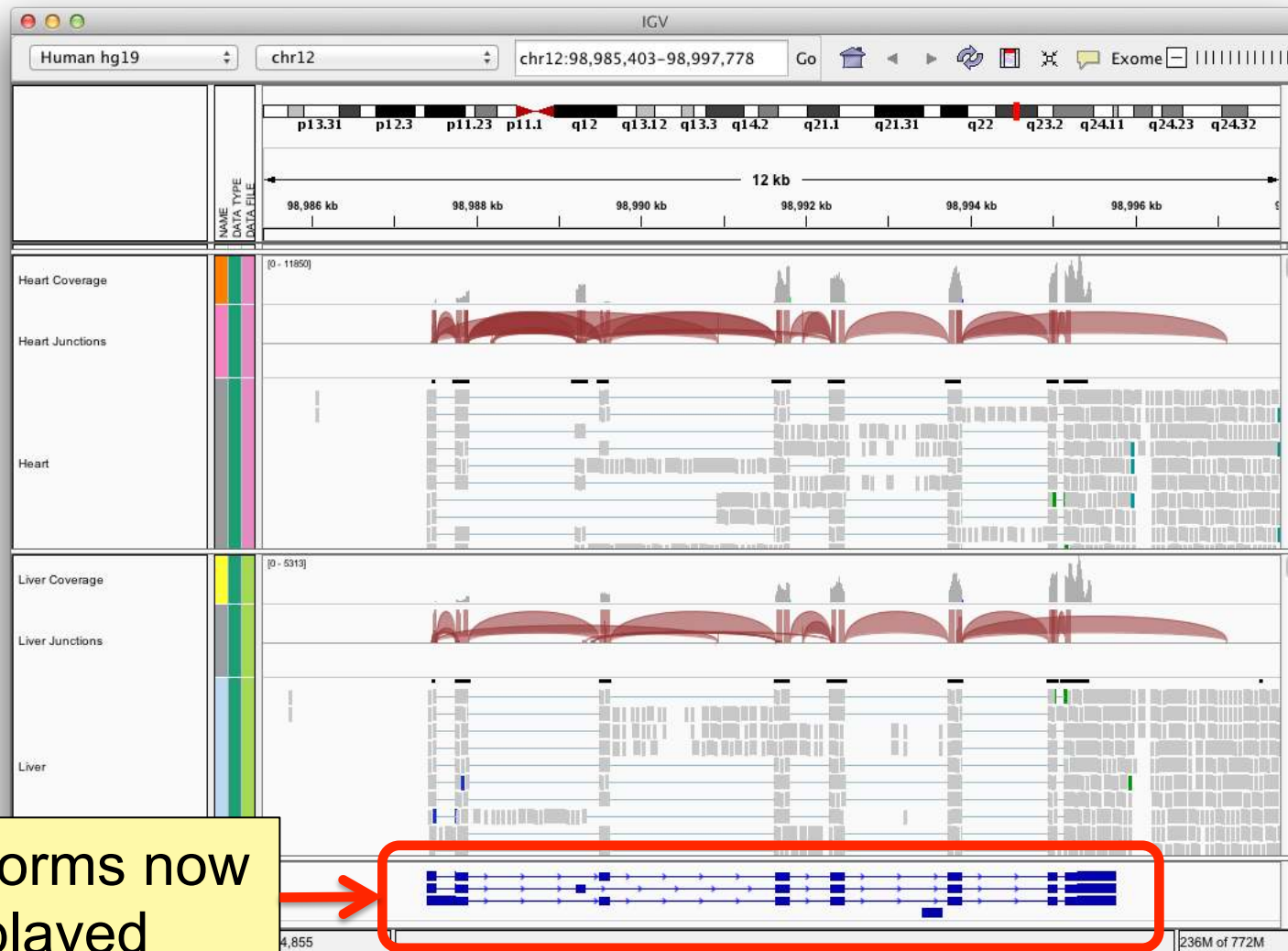
Human hg19 chr12 chr12:98,985,403-98,997,778 Go

RefSeq Genes

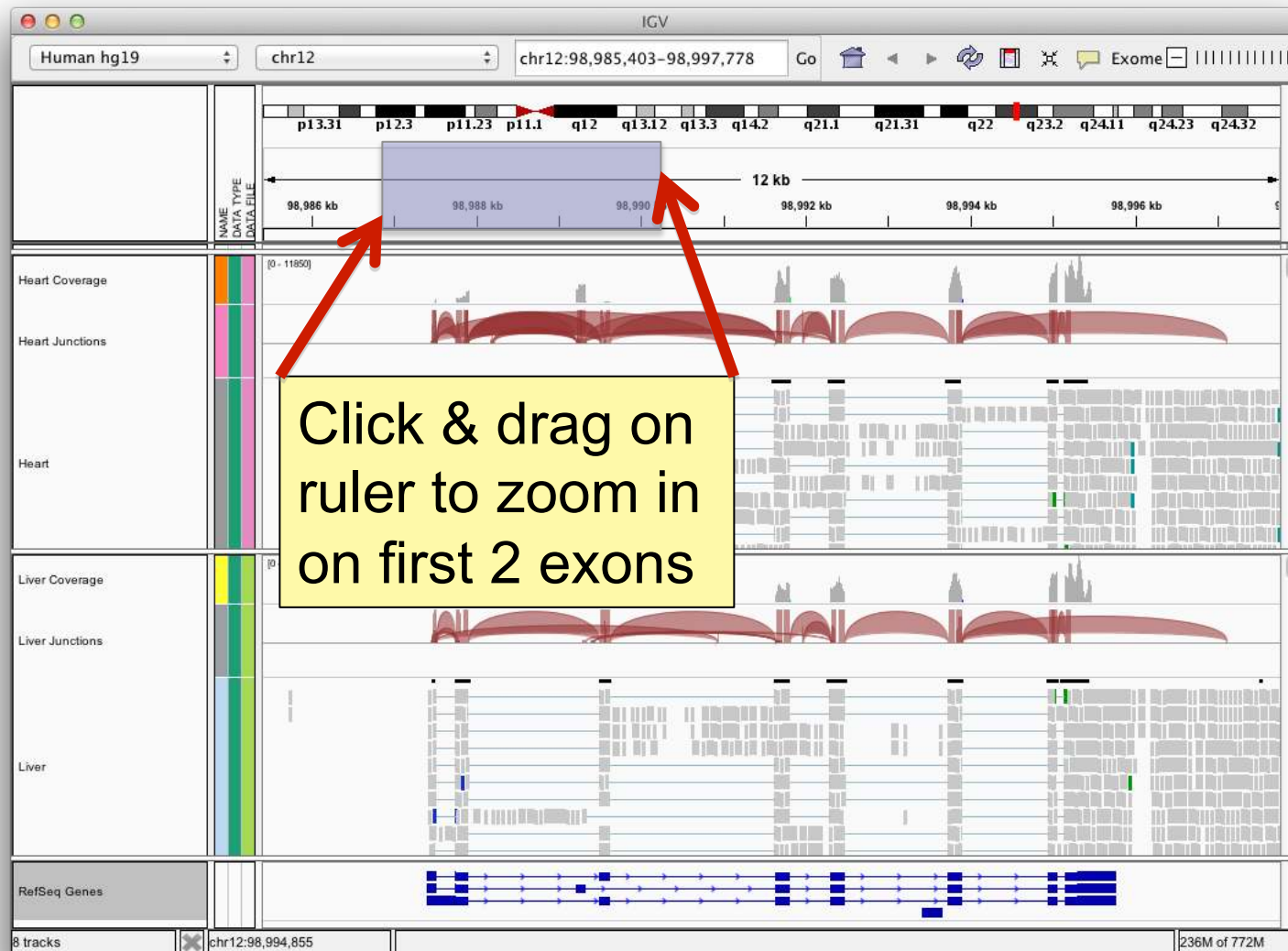
- Rename Track...
- Change Track Color...
- Change Track Height...
- Change Font Size...
- ✓ Collapsed
- Expanded
- Squished**
- Copy Details to Clipboard
- Copy Sequence
- Blat Sequence
- Set Feature Visibility Window...
- Remove Track
- Save image...
- Export To BED File

Select Squished

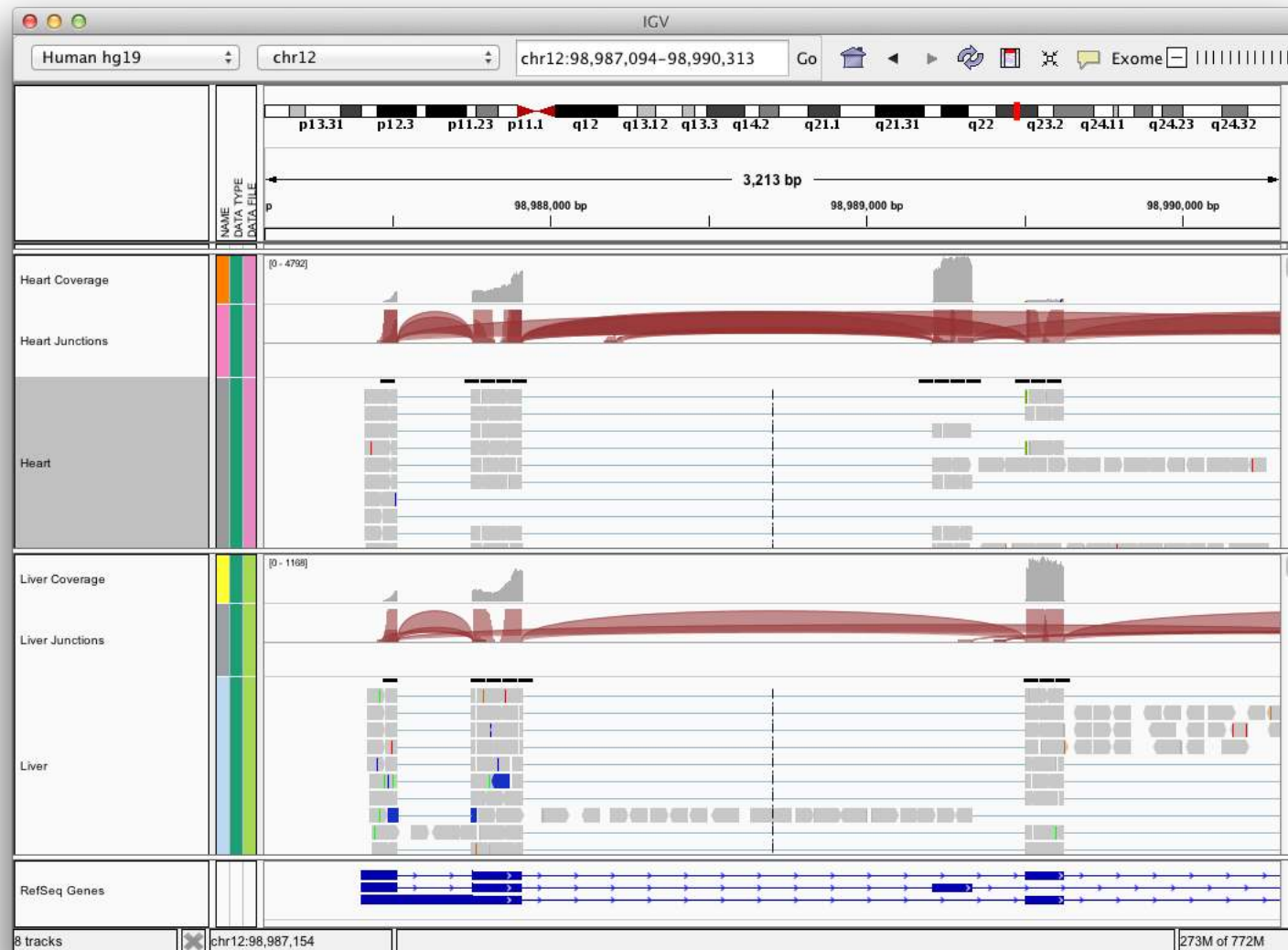
RNA-seq alignments



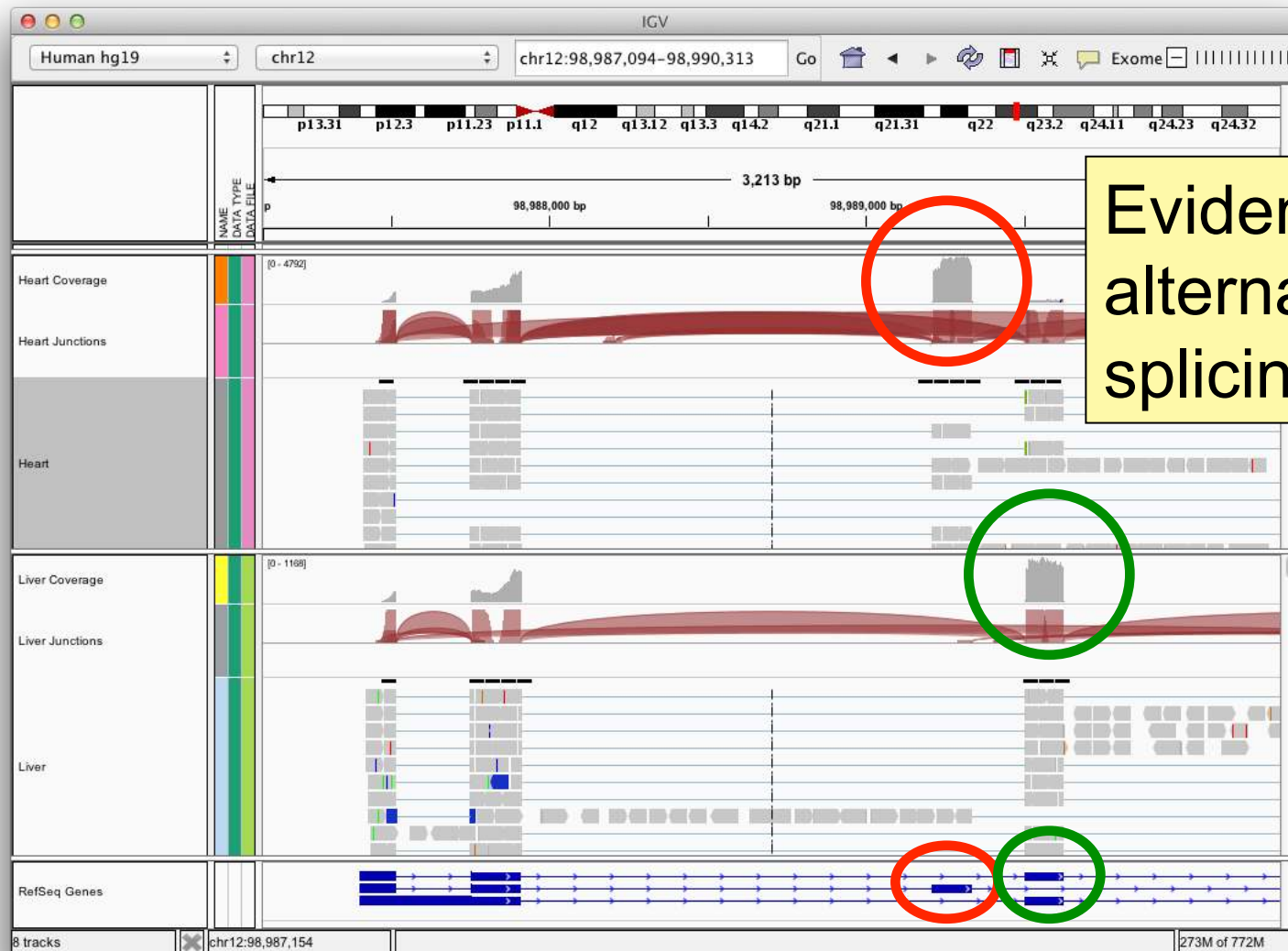
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments



Evidence of
alternative
splicing

Sashimi plot

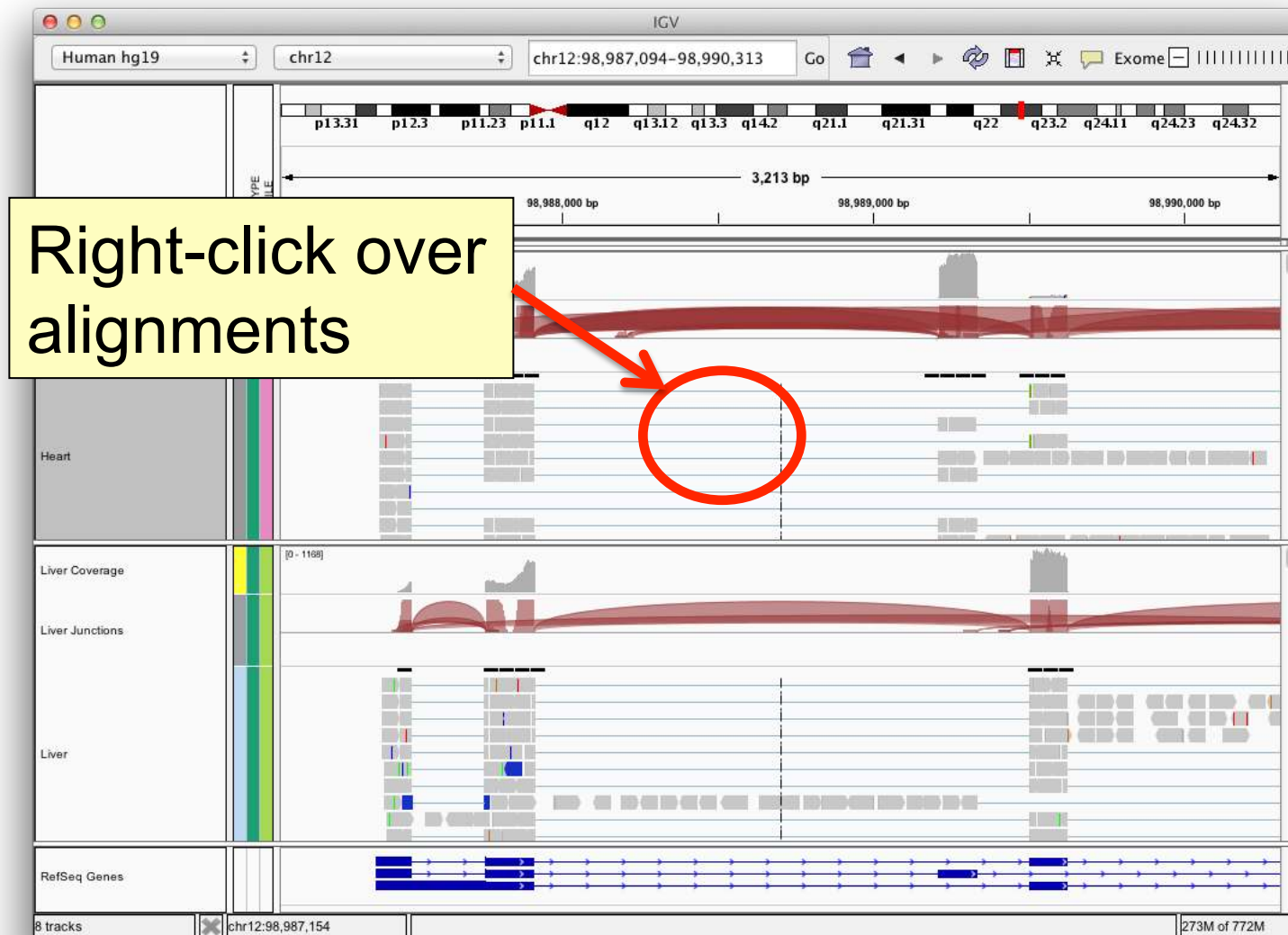


Viewing RNA splicing with Sashimi Plots

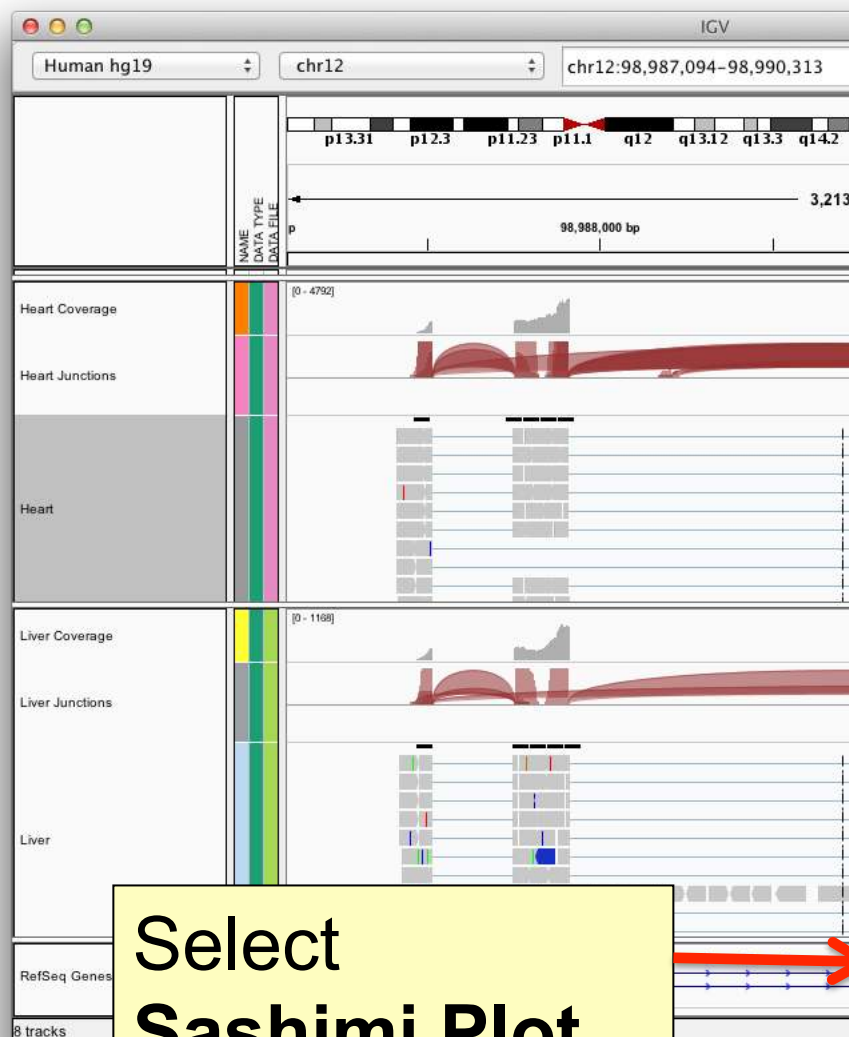
Reference: Katz Y, Wang ET, Silterra J, Schwartz S, Wong B, Mesirov JP, Airolidi EM, Burge, CB.

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

RNA-seq alignments



RNA-seq alignments



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

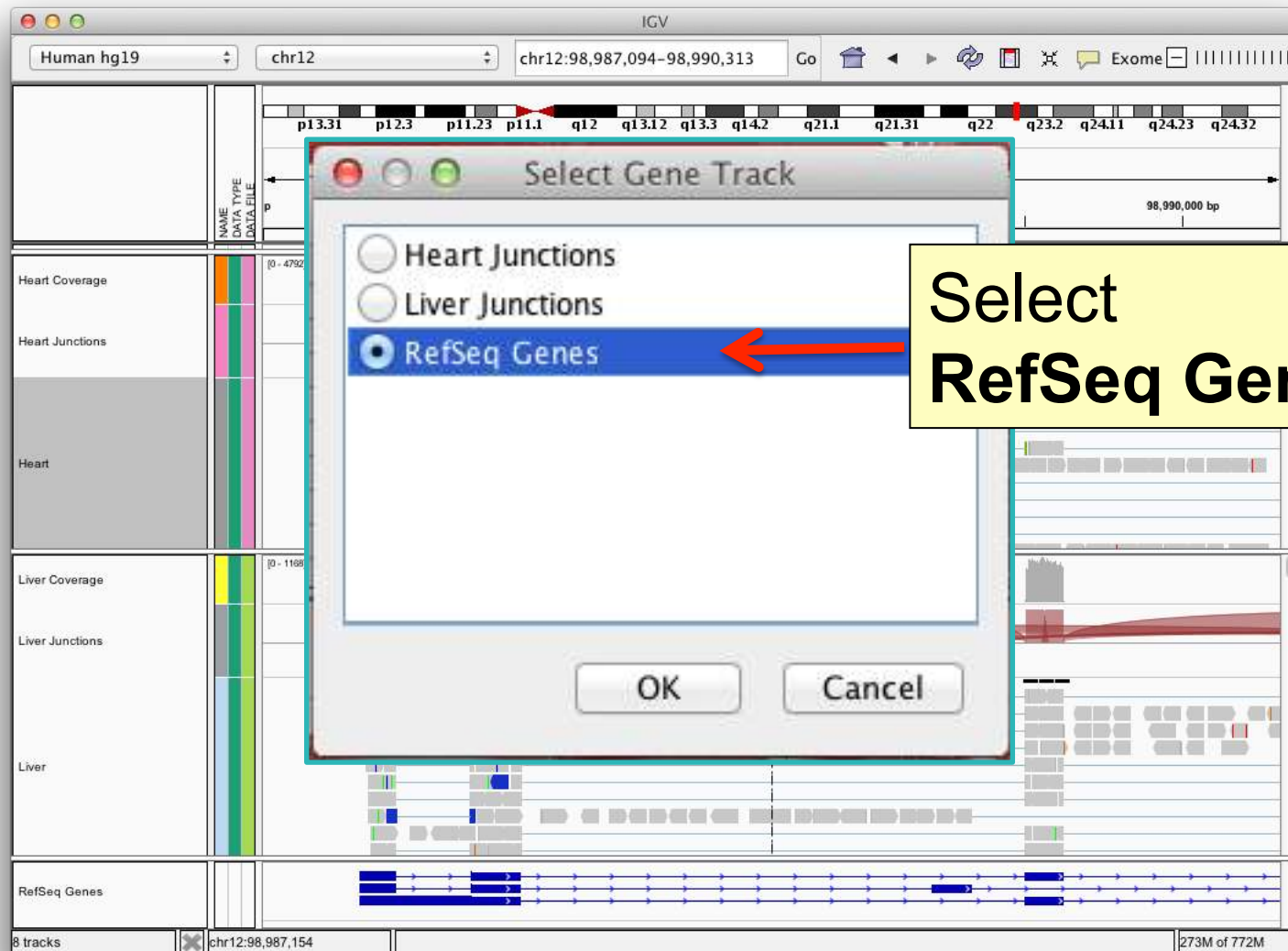
Remove Track

Save image...

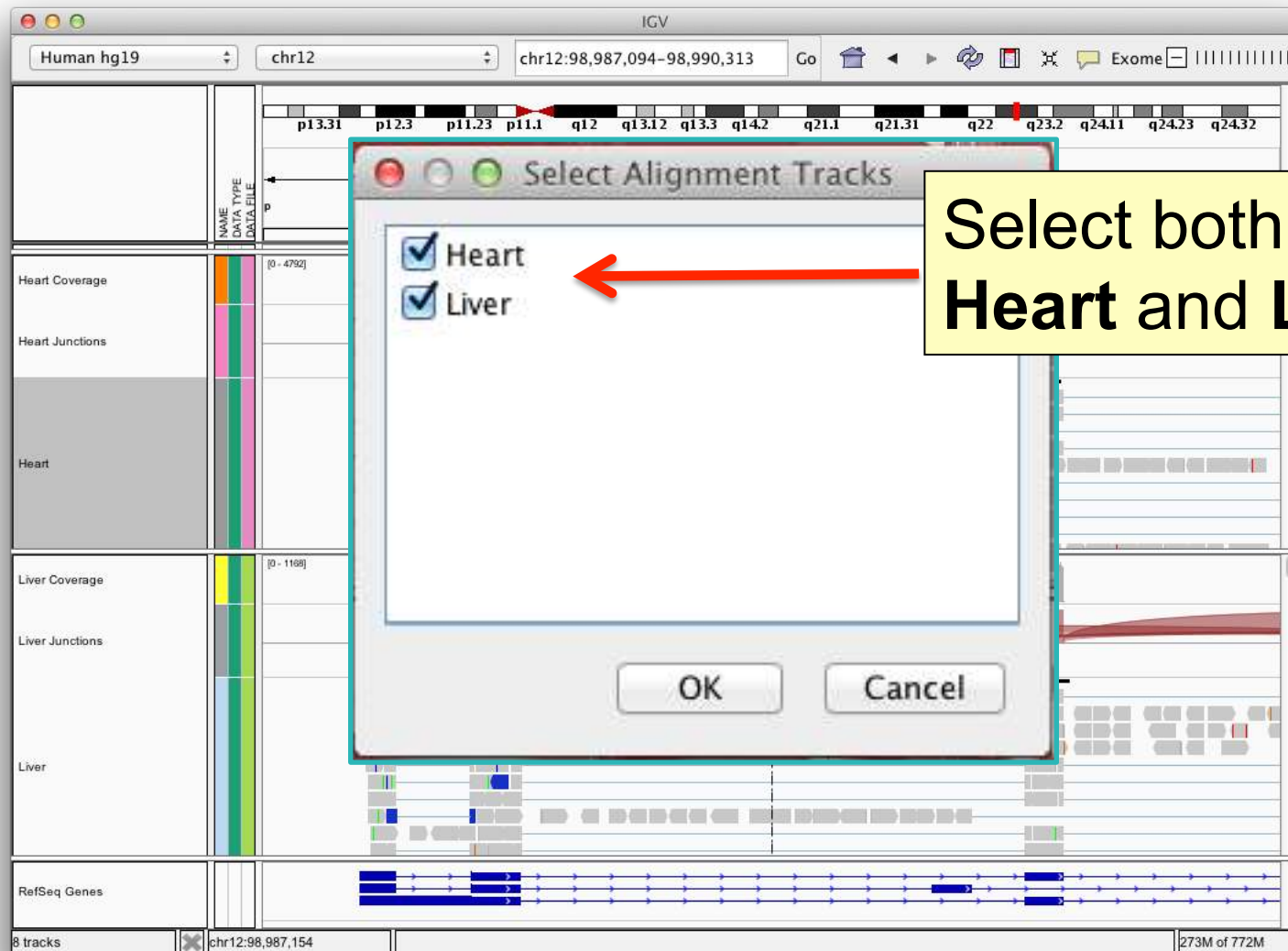


Integrative
Genomics
Viewer

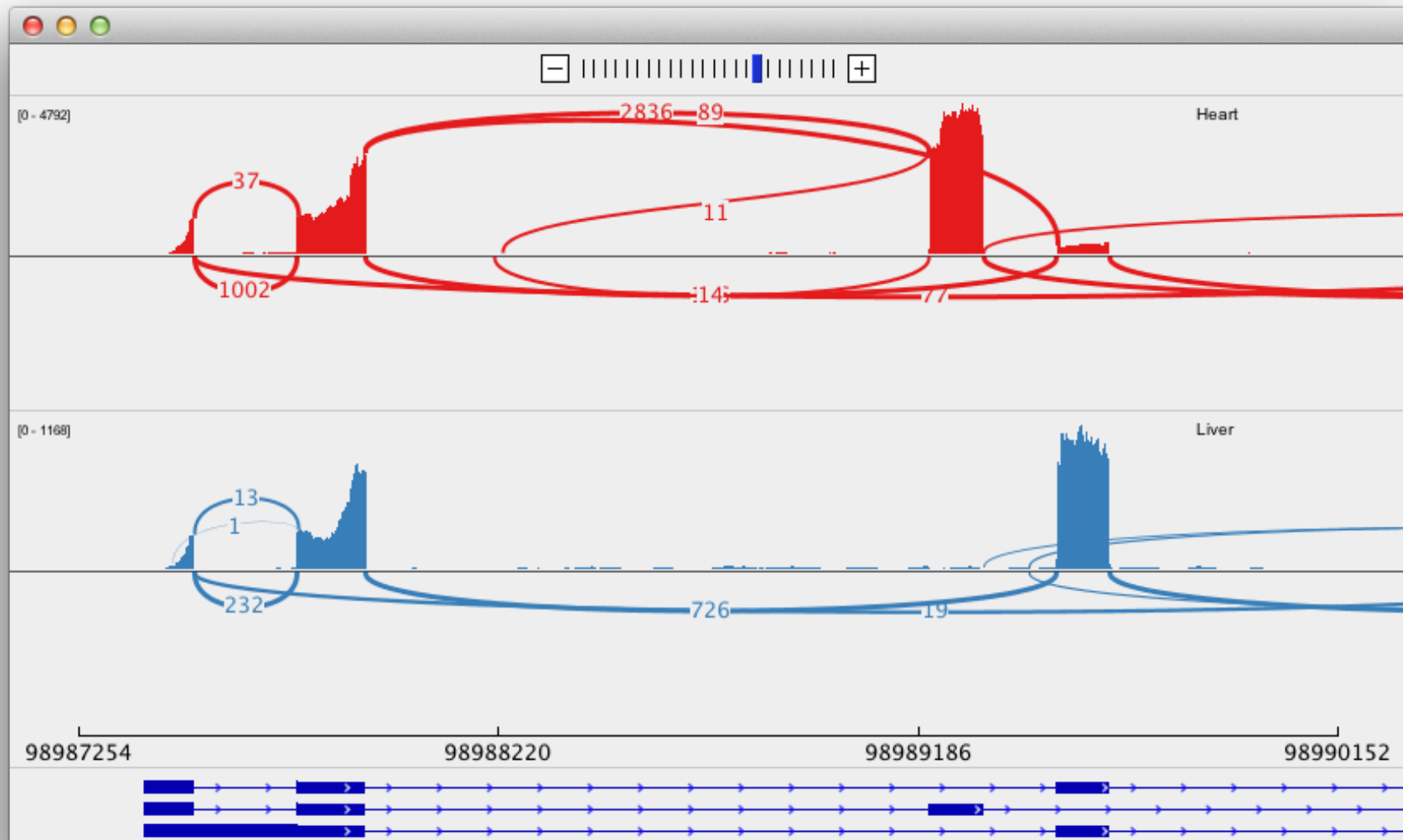
RNA-seq alignments



RNA-seq alignments



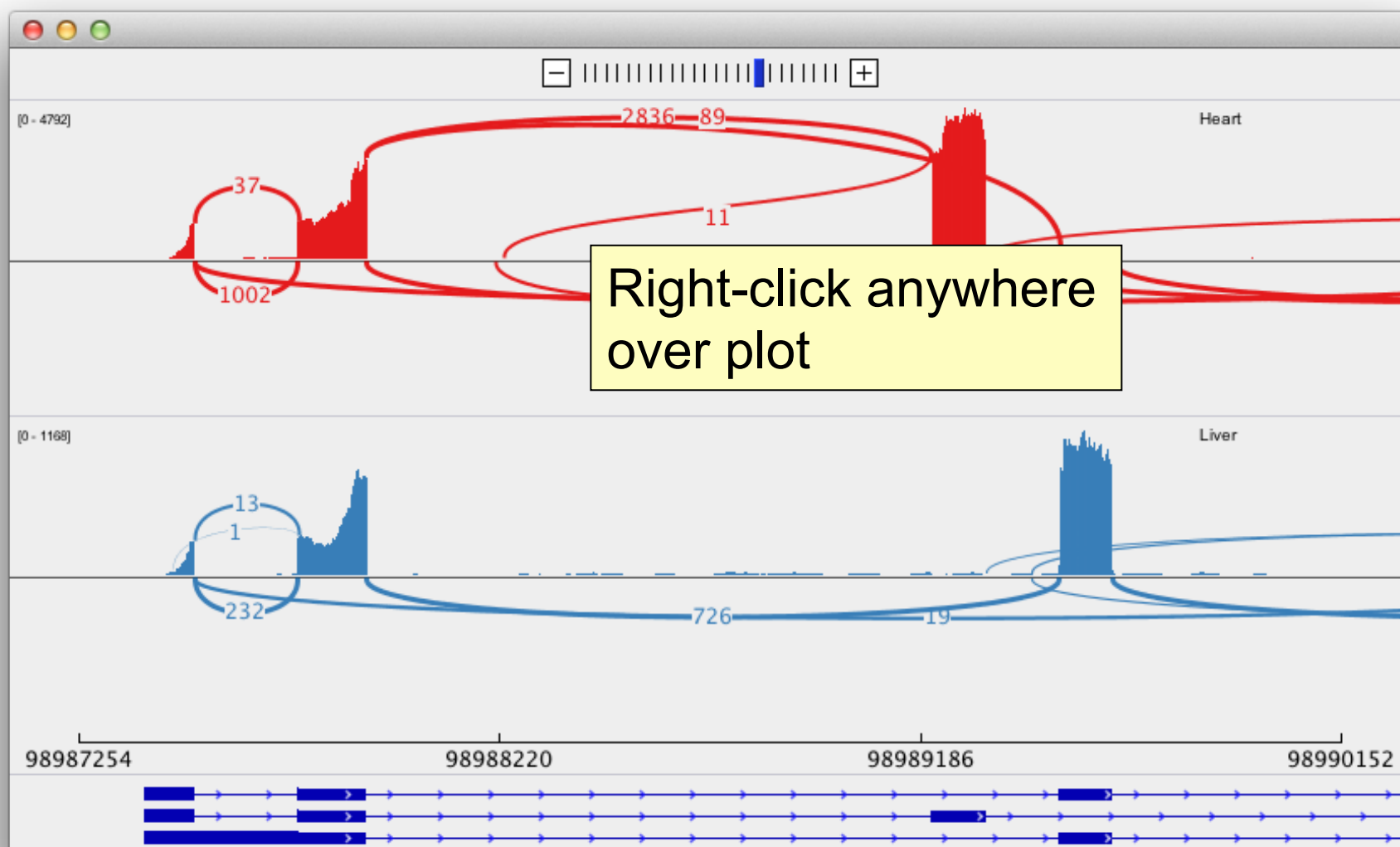
RNA-seq alignments



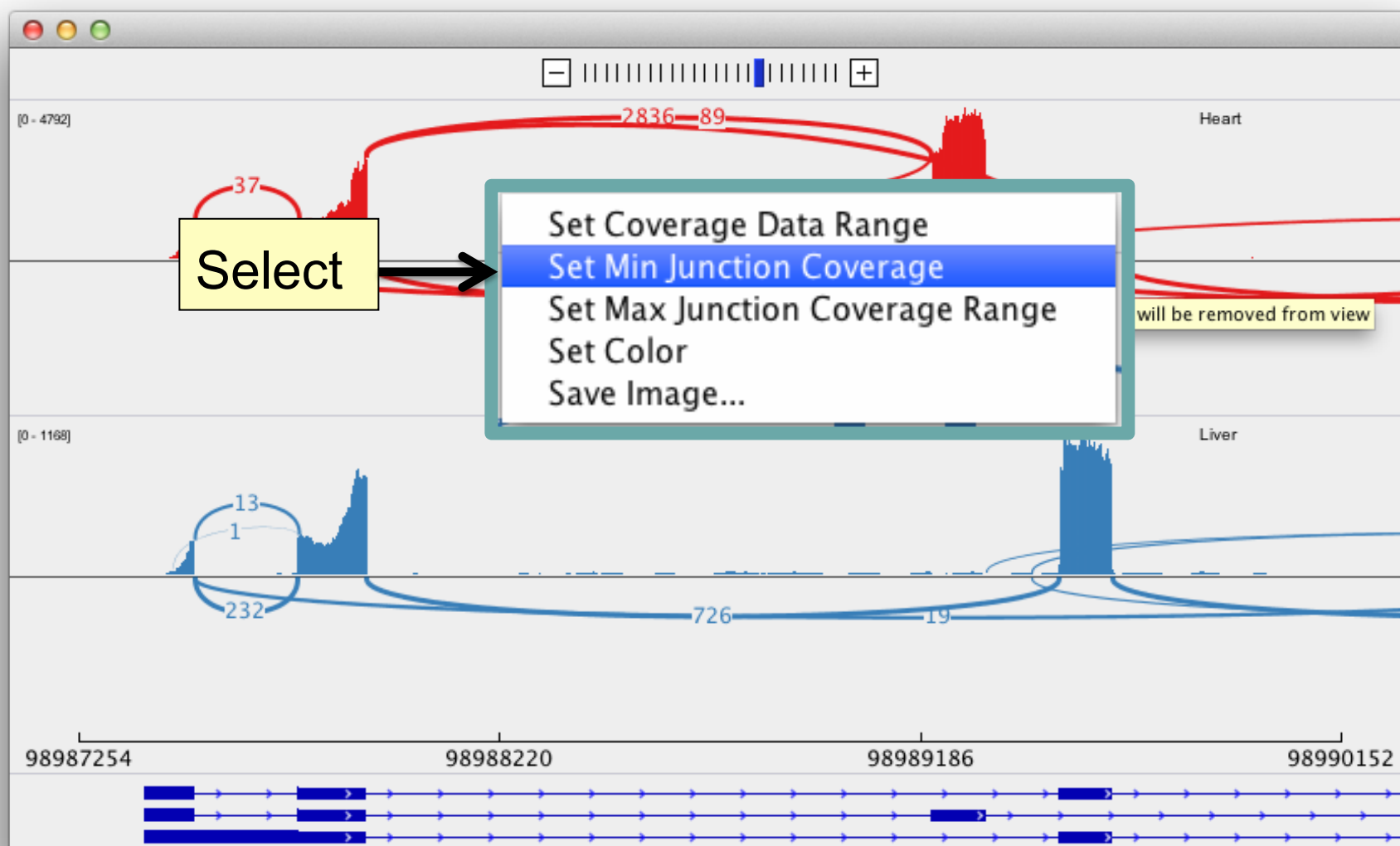
RNA-seq alignments



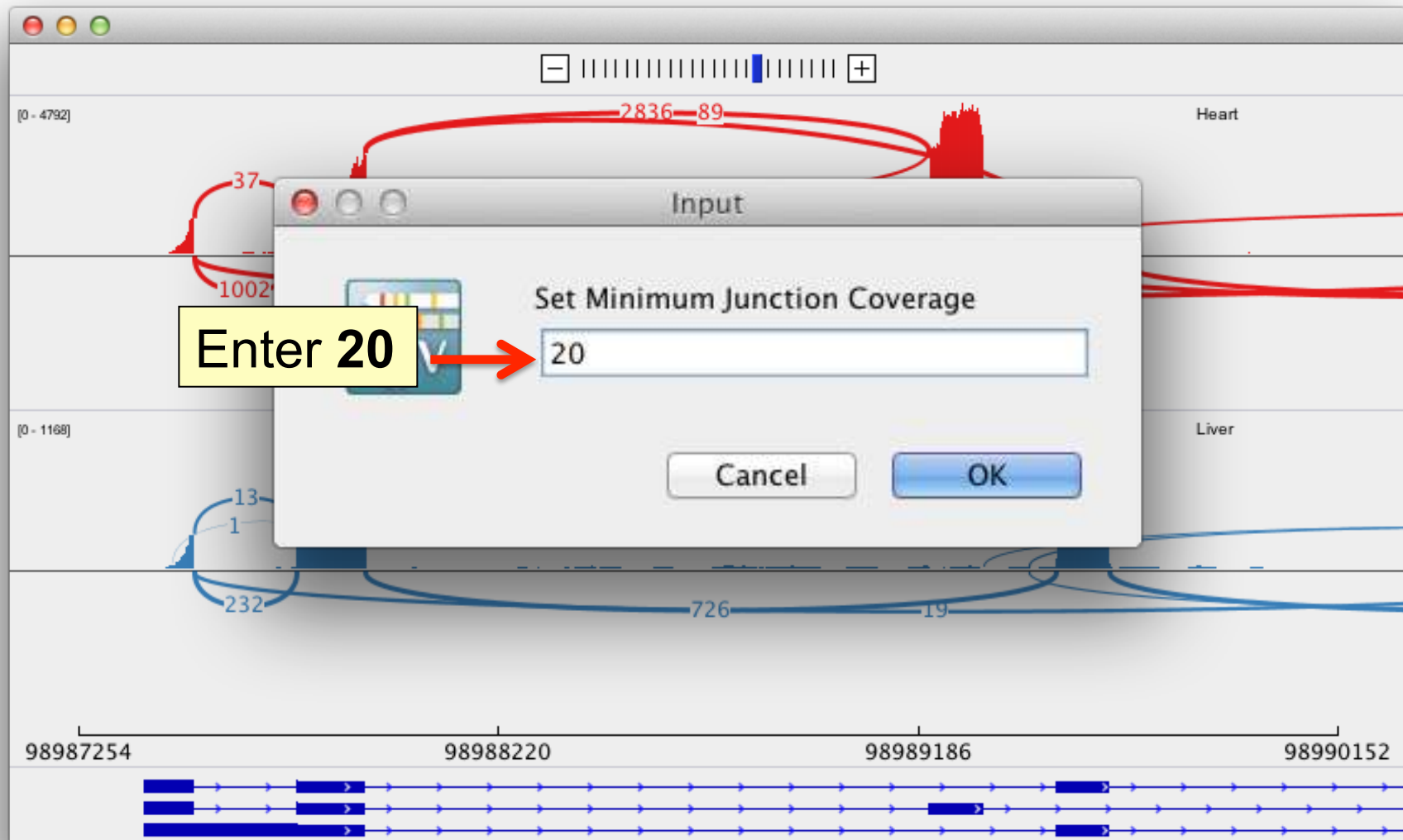
RNA-seq alignments



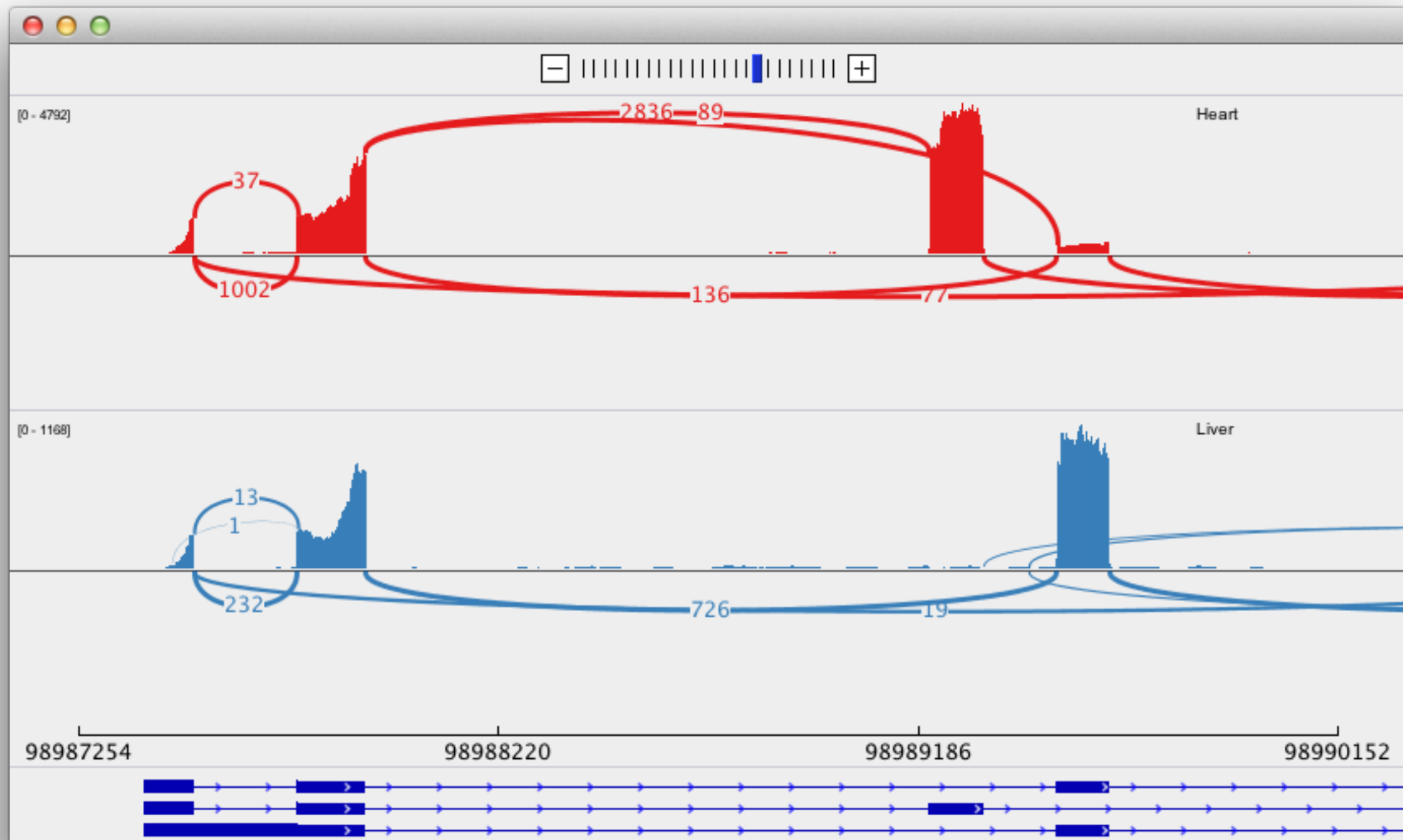
RNA-seq alignments



RNA-seq alignments



RNA-seq alignments



igvtools

igvtools

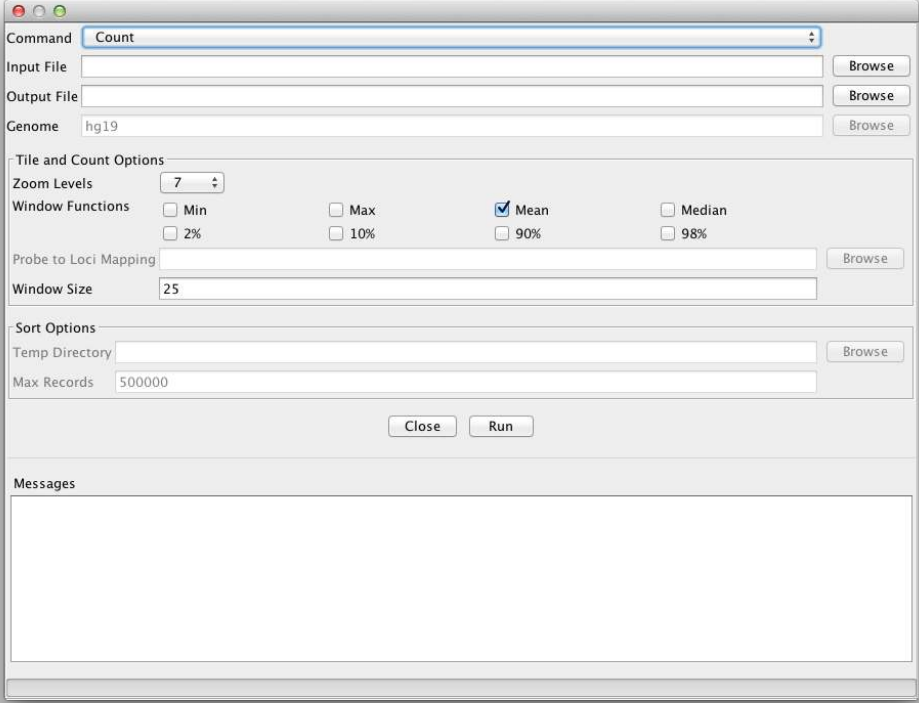


A set of utilities for preparing files for efficient display.

toTDF	• Converts sorted data file to a binary tiled data file (TDF). • Supported file formats: .wig, .cn, .snp, .igv, .gct
count	• Computes average alignment or feature density over a specified window size across the genome. • Supported file formats: .sam, .bam, .aligned, .sorted.txt, .bed
sort	• Sorts file by genomic start position. • Supported file formats: .cn, .igv, .sam, .aligned, .bed.
index	• Creates an index file for alignment or feature file. • Supported file formats: .sam, .aligned, .sorted.txt, .bed

igvtools

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



The screenshot shows the IGVtools application window. It has a title bar with standard window controls. The main interface is divided into several sections:

- Command:** A dropdown menu set to "Count".
- Input File:** A text field with a "Browse" button.
- Output File:** A text field with a "Browse" button.
- Genome:** A text field set to "hg19" with a "Browse" button.
- Tile and Count Options:**
 - Zoom Levels:** A dropdown menu set to "7".
 - Window Functions:** A grid of checkboxes for "Min", "Max", "Mean" (checked), "Median", "2%", "10%", "90%", and "98%".
 - Probe to Loci Mapping:** A text field with a "Browse" button.
 - Window Size:** A text field set to "25".
- Sort Options:**
 - Temp Directory:** A text field with a "Browse" button.
 - Max Records:** A text field set to "500000".
- Buttons:** "Close" and "Run" buttons.
- Messages:** A large empty text area at the bottom.

igvtools toTDF



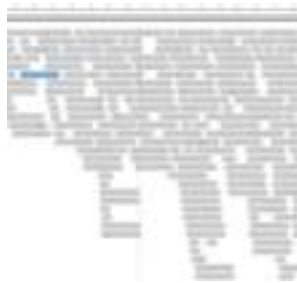
The **toTDF** utility converts large ASCII data files into tiled data format (.tdf) files.

TDF files have the following advantages:

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

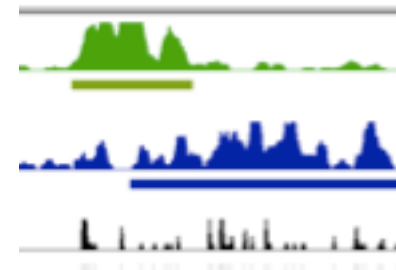
igvtools count

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



Alignments

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



Read Density

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

igvtools sort



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

Example:

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

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

igvtools index



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

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

Example: `igvtools index myFeatures.bed`

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

Computing coverage: igvtools



Hands-on exercise

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

Data source

Illumina BodyMap

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

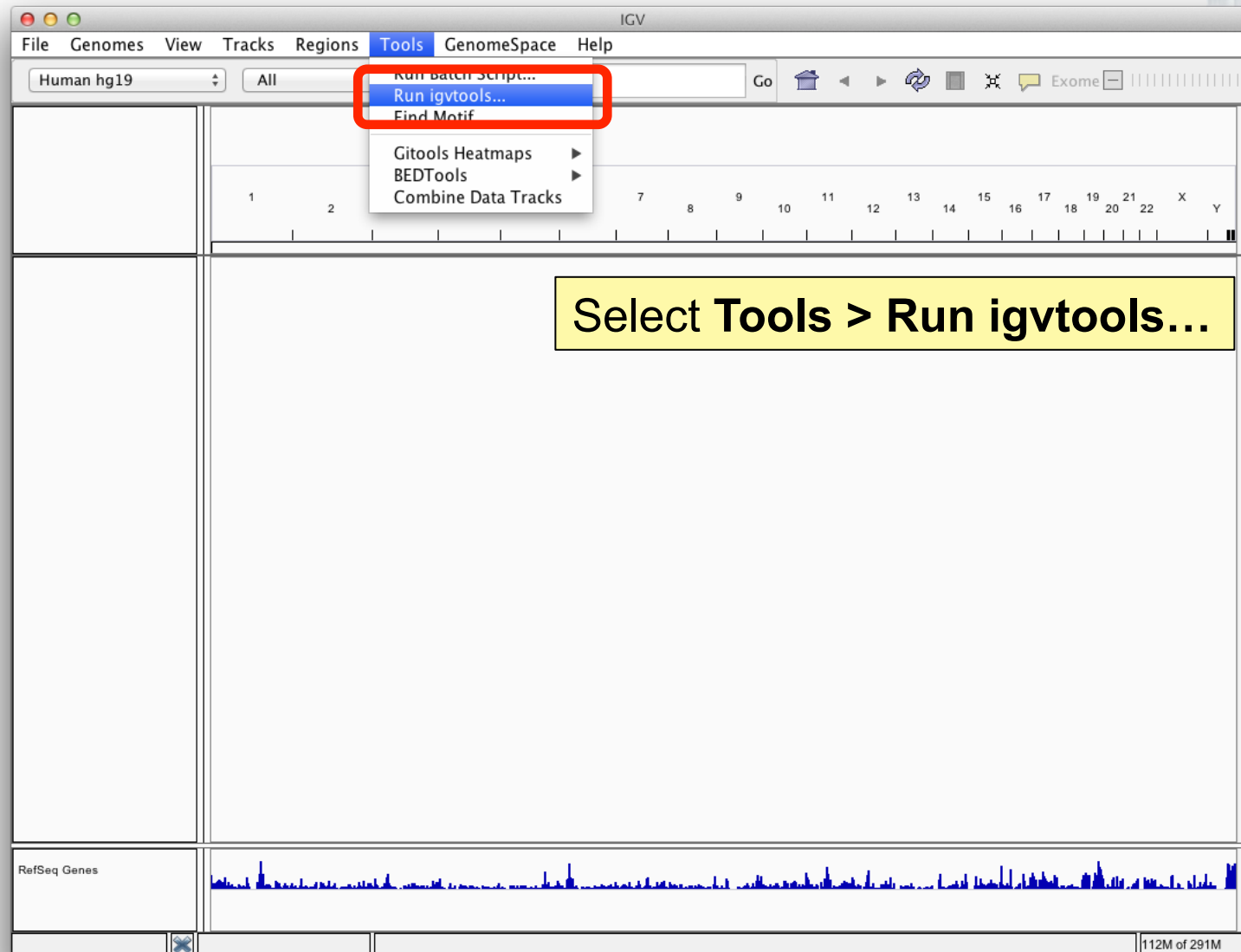
Files included in the zip:

heart.bodyMap.bam

heart.bodyMap.bam.bai

sacCer3.fa (used in next exercise)

Computing coverage: igvtools



Computing coverage: igvtools

A screenshot of the 'Command' dialog box in the IGVtools application. The 'Command' dropdown menu is highlighted with a red rectangle and contains the text 'Count'. A yellow callout box with the text 'Select Count command' points to this dropdown. The dialog box includes fields for 'Input File', 'Output File', and 'Genome' (set to 'hg19'). Below these are 'Tile and Count Options' including 'Zoom Levels' (set to 7), 'Window Functions' (with 'Mean' selected), and 'Probe to Loci Mapping'. There are also 'Sort Options' for 'Temp Directory' and 'Max Records' (set to 500000). At the bottom are 'Close' and 'Run' buttons, and a 'Messages' section at the very bottom.

Command: **Count**

Input File: Browse

Output File: 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

Close Run

Messages

Computing coverage: igvtools

A screenshot of the igvtools application window. The 'Input File' field is highlighted with a red rectangle and contains the path '/Users/jrobinso/files/heart.bodyMap.bam'. The 'Browse' button next to it is also highlighted with a red rectangle. Below the 'Input File' field, the 'Output File' field contains the path '/Users/jrobinso/files/heart.bodyMap.bam.tdf'. 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, 2%, 10%). The 'Sort Options' section includes a 'Temp Directory' field and a 'Max Records' field set to '500000'. At the bottom, there are 'Close' and 'Run' buttons. A 'Messages' section is at the very bottom.

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 ☐ 2% ☐ 10%

Probe to Loci Mapping

Window Size 25

Sort Options

Temp Directory Browse

Max Records 500000

Close Run

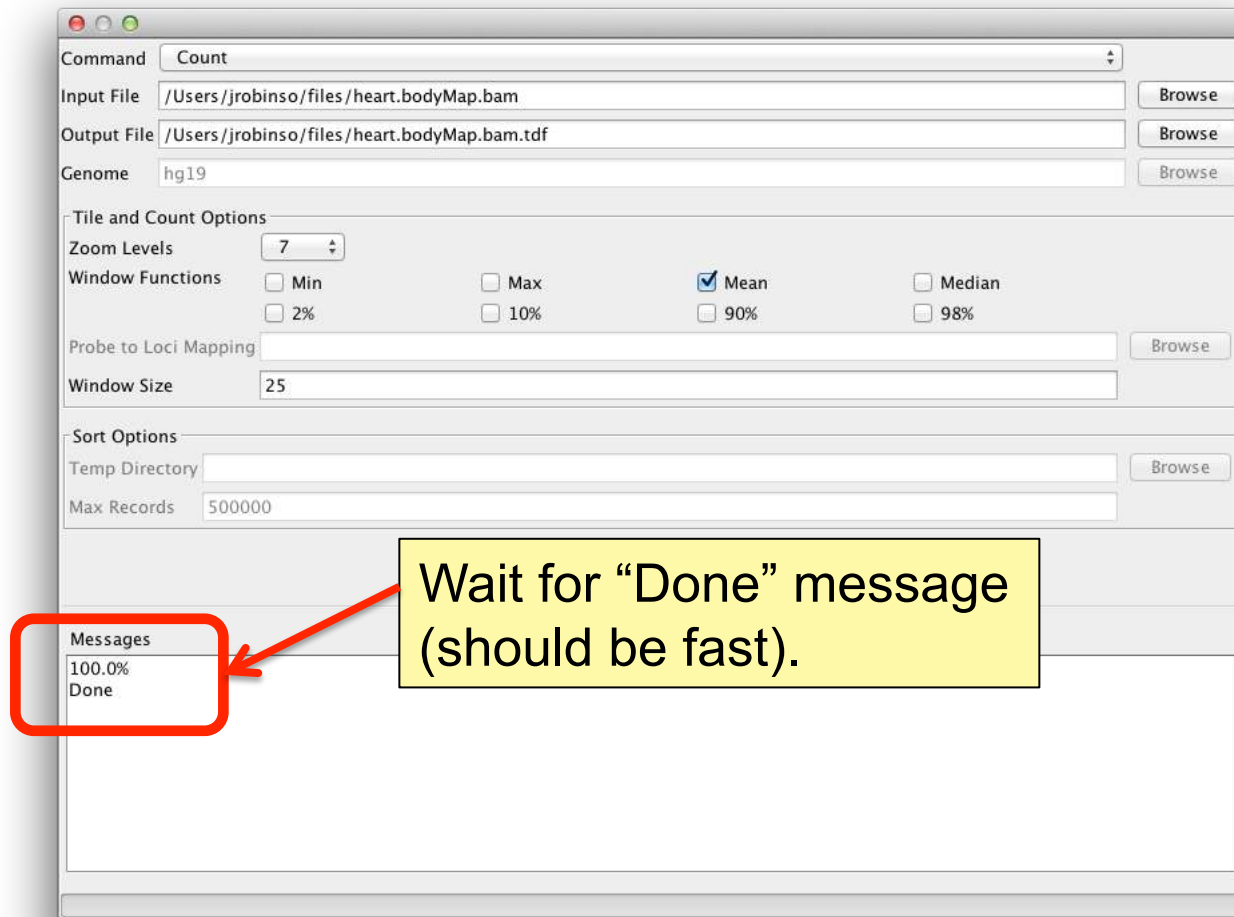
Messages

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

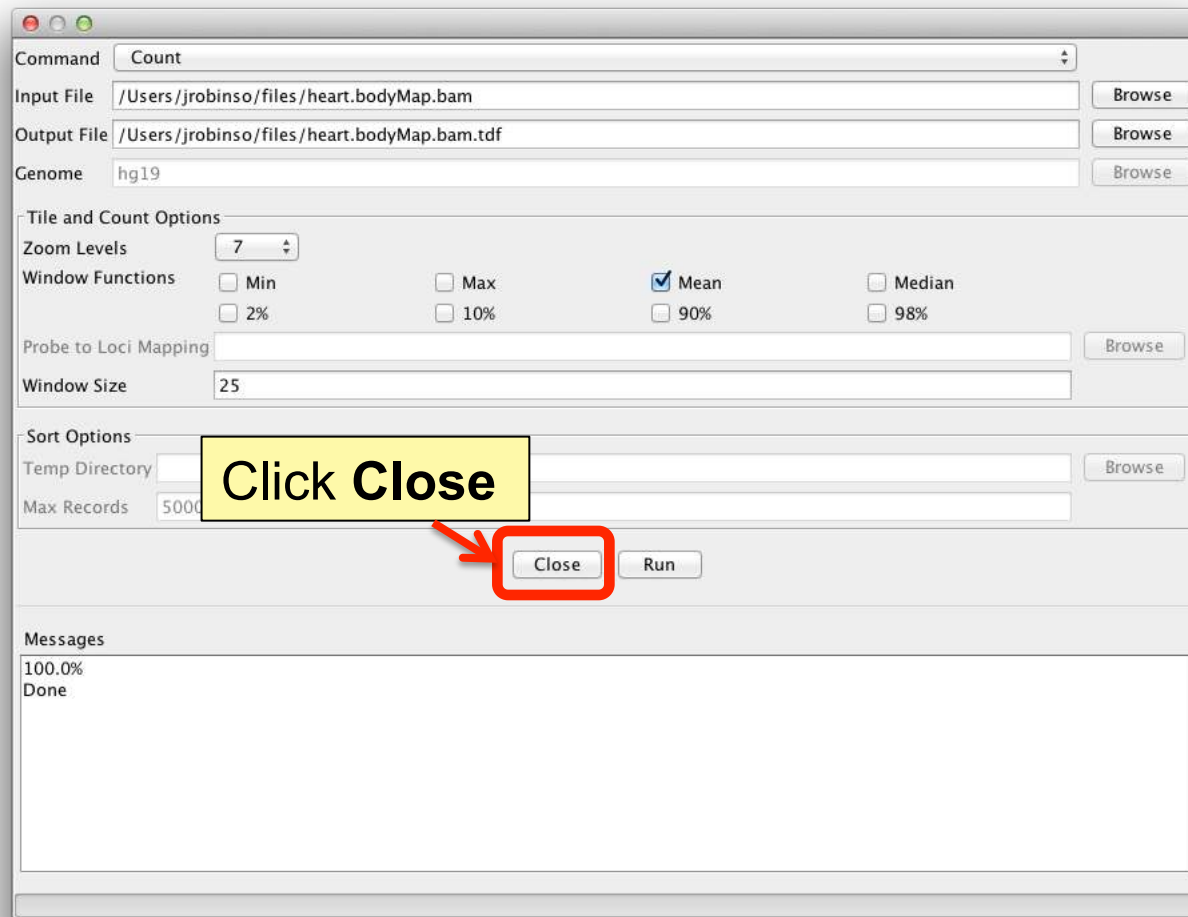
Computing coverage: igvtools

A screenshot of the 'Count' command window in the igvtools application. The window has a title bar with standard macOS window controls. It contains several input fields and buttons. The 'Command' dropdown is set to 'Count'. The 'Input File' field contains '/Users/jrobinso/files/heart.bodyMap.bam' and has a 'Browse' button to its right. The 'Output File' field contains '/Users/jrobinso/files/heart.bodyMap.bam.tdf' and also has a 'Browse' button. The 'Genome' field is set to 'hg19' with a 'Browse' button. Below these is a section titled 'Tile and Count Options' containing a 'Zoom Levels' spinner set to 7, and 'Window Functions' with checkboxes for 'Min', 'Max', 'Mean' (which is checked), 'Median', '2%', '10%', '90%', and '98%'. There is also a 'Probe to Loci Mapping' field with a 'Browse' button and a 'Window Size' field set to 25. A 'Sort Options' section follows with a 'Temp Directory' field and a 'Browse' button, and a 'Max Records' field set to 500000. At the bottom of the main configuration area are 'Close' and 'Run' buttons. A red rectangle highlights the 'Run' button, and a red arrow points from a yellow box containing the text 'Click Run' to this button. Below the configuration area is a 'Messages' section with a large empty text area.

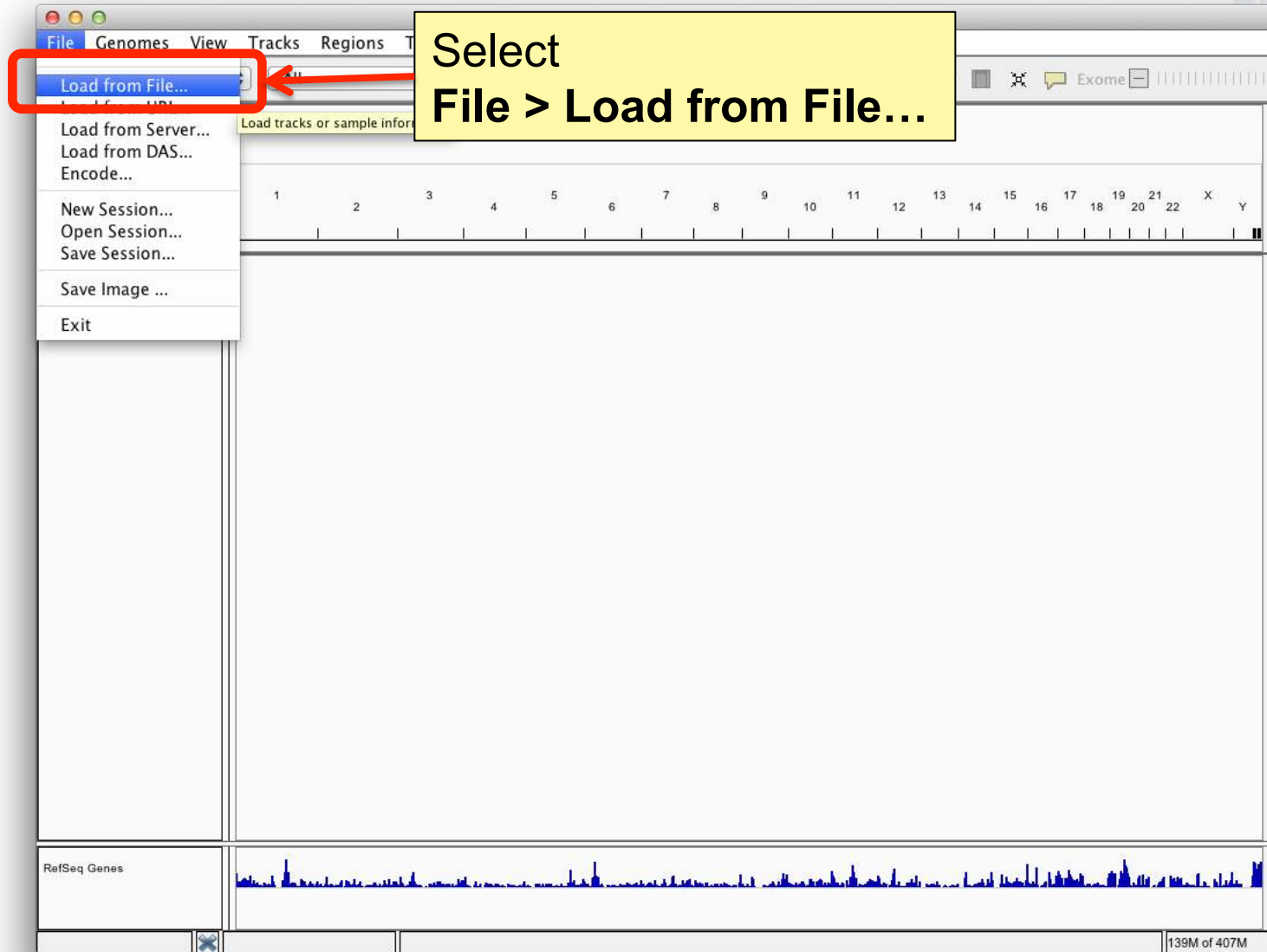
Computing coverage: igvtools



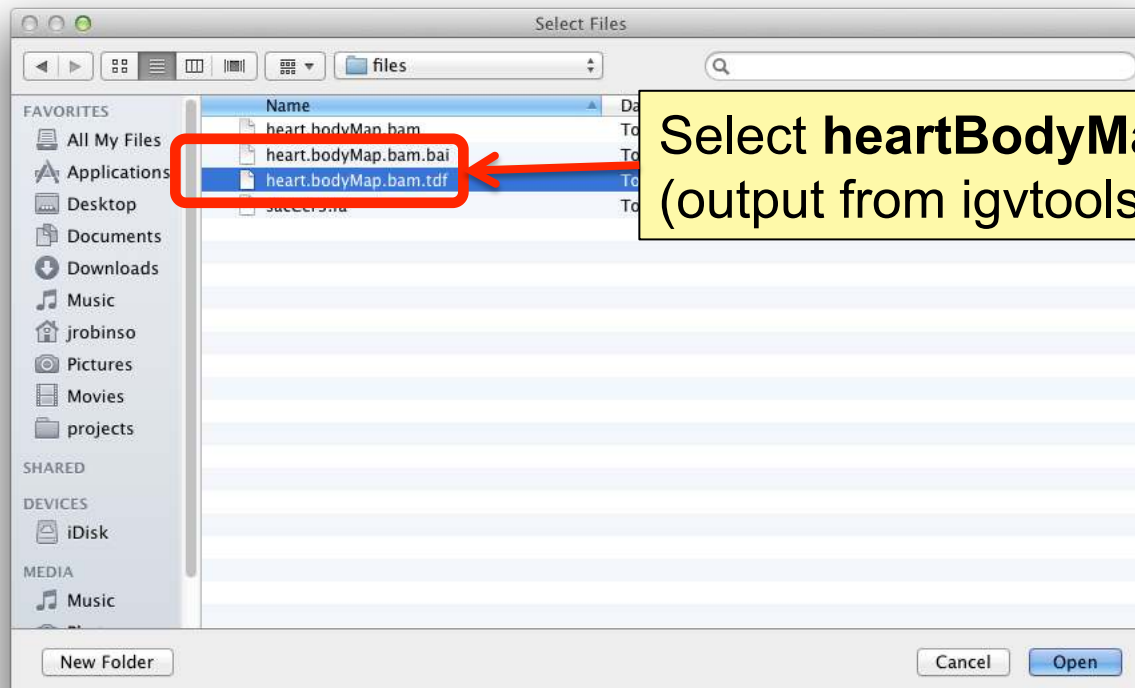
Computing coverage: igvtools



Computing coverage: igvtools

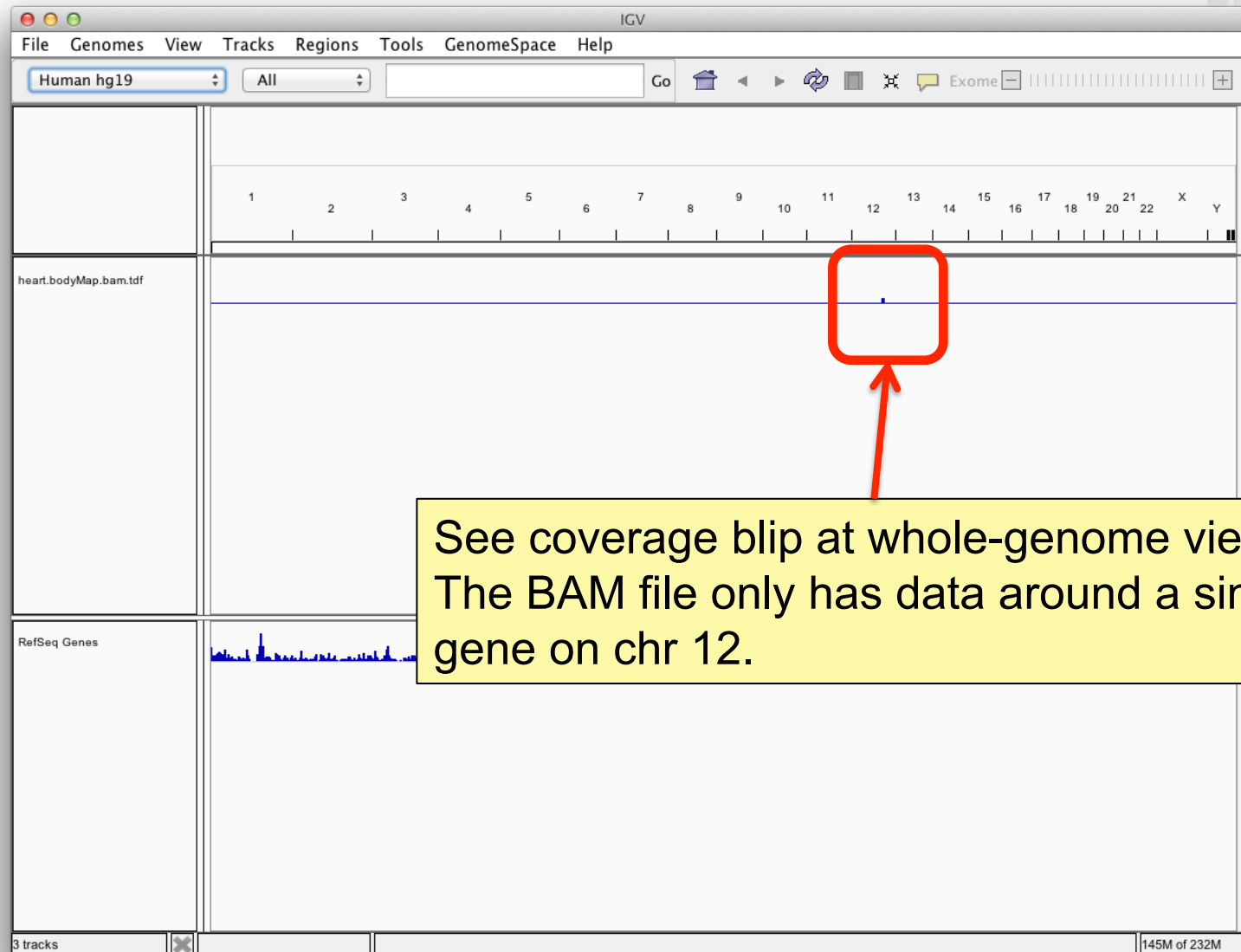


Computing coverage: igvtools



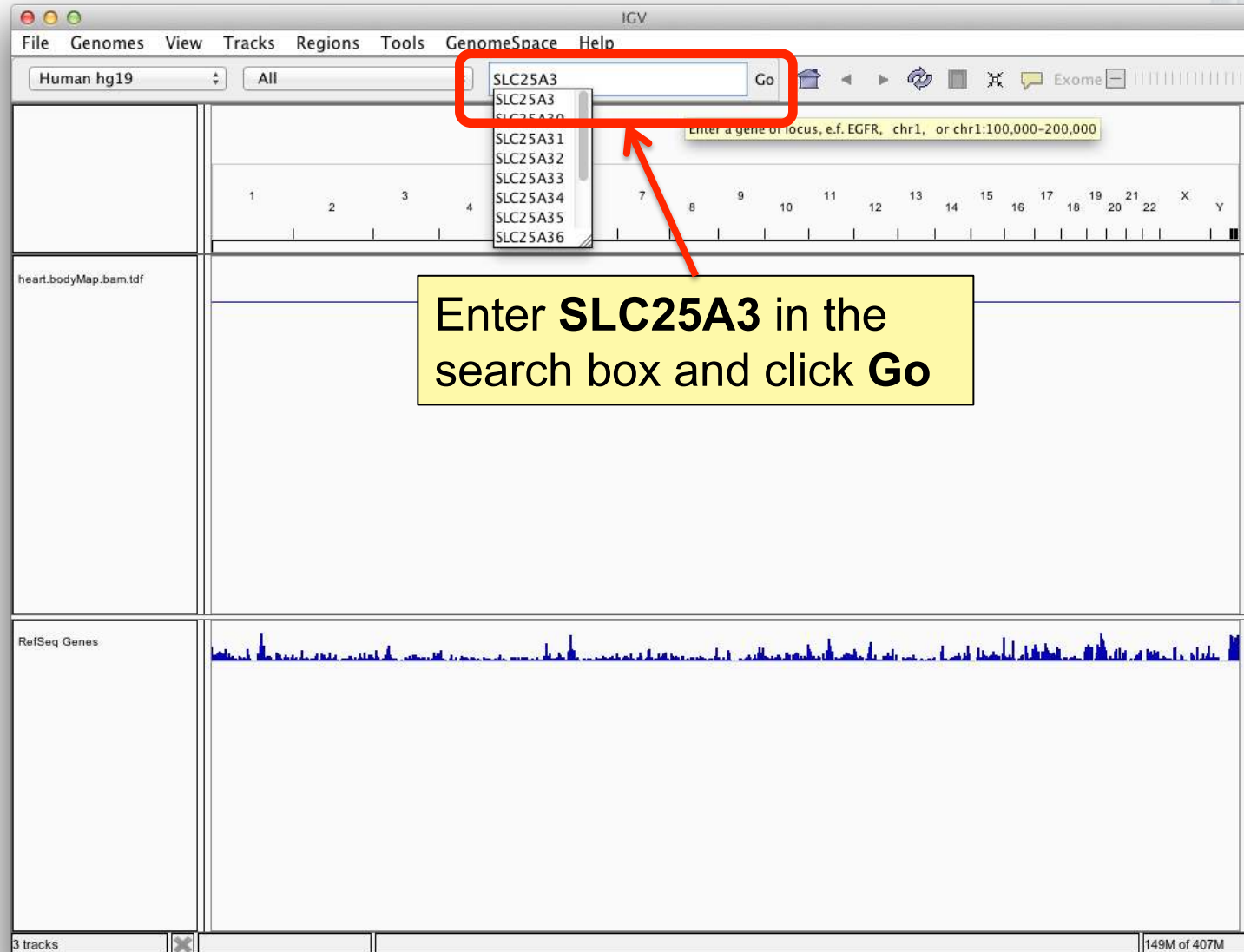
Select **heartBodyMap.bam.tdf**
(output from igvtools)

Computing coverage: igvtools

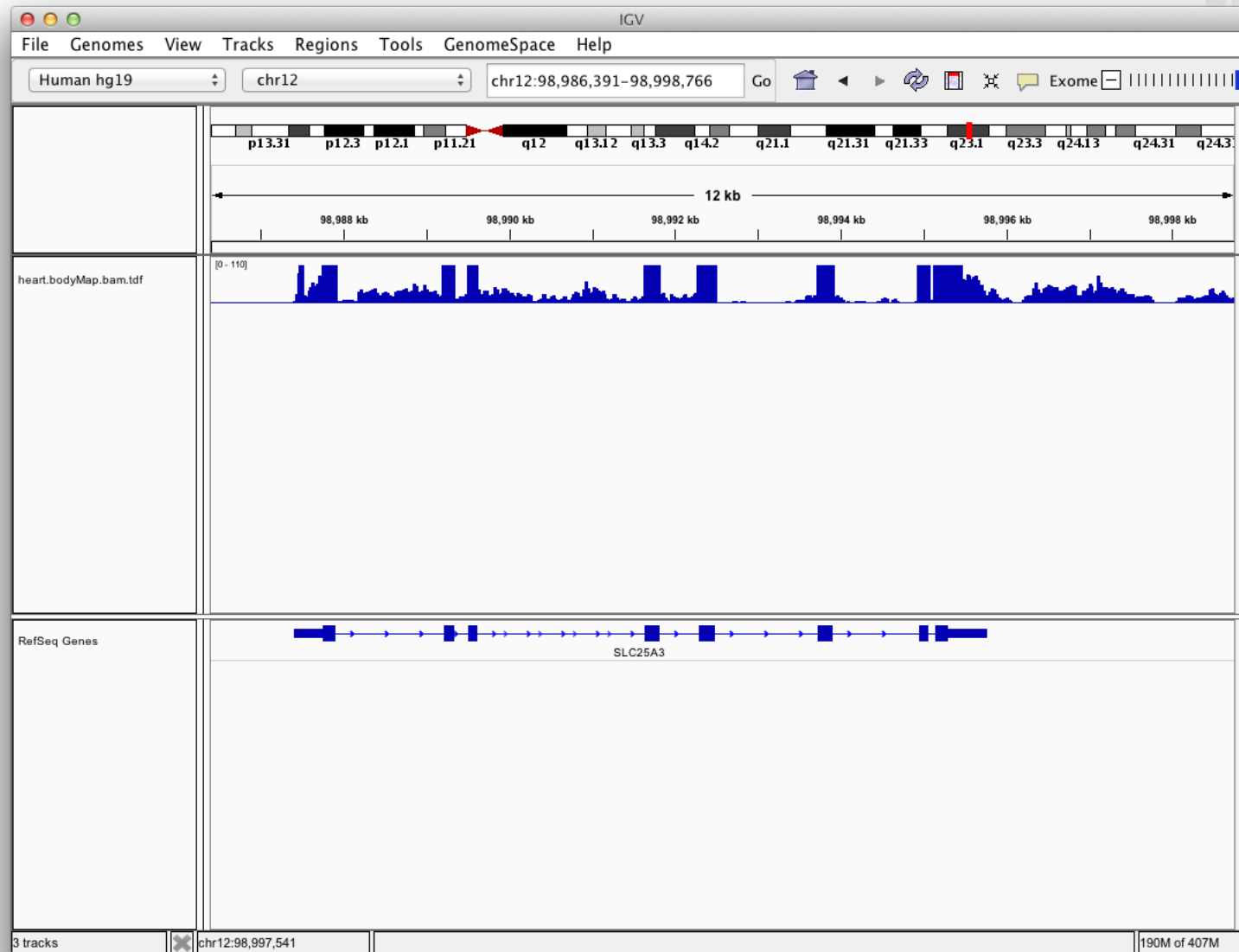


See coverage blip at whole-genome view.
The BAM file only has data around a single gene on chr 12.

Computing coverage: igvtools



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.

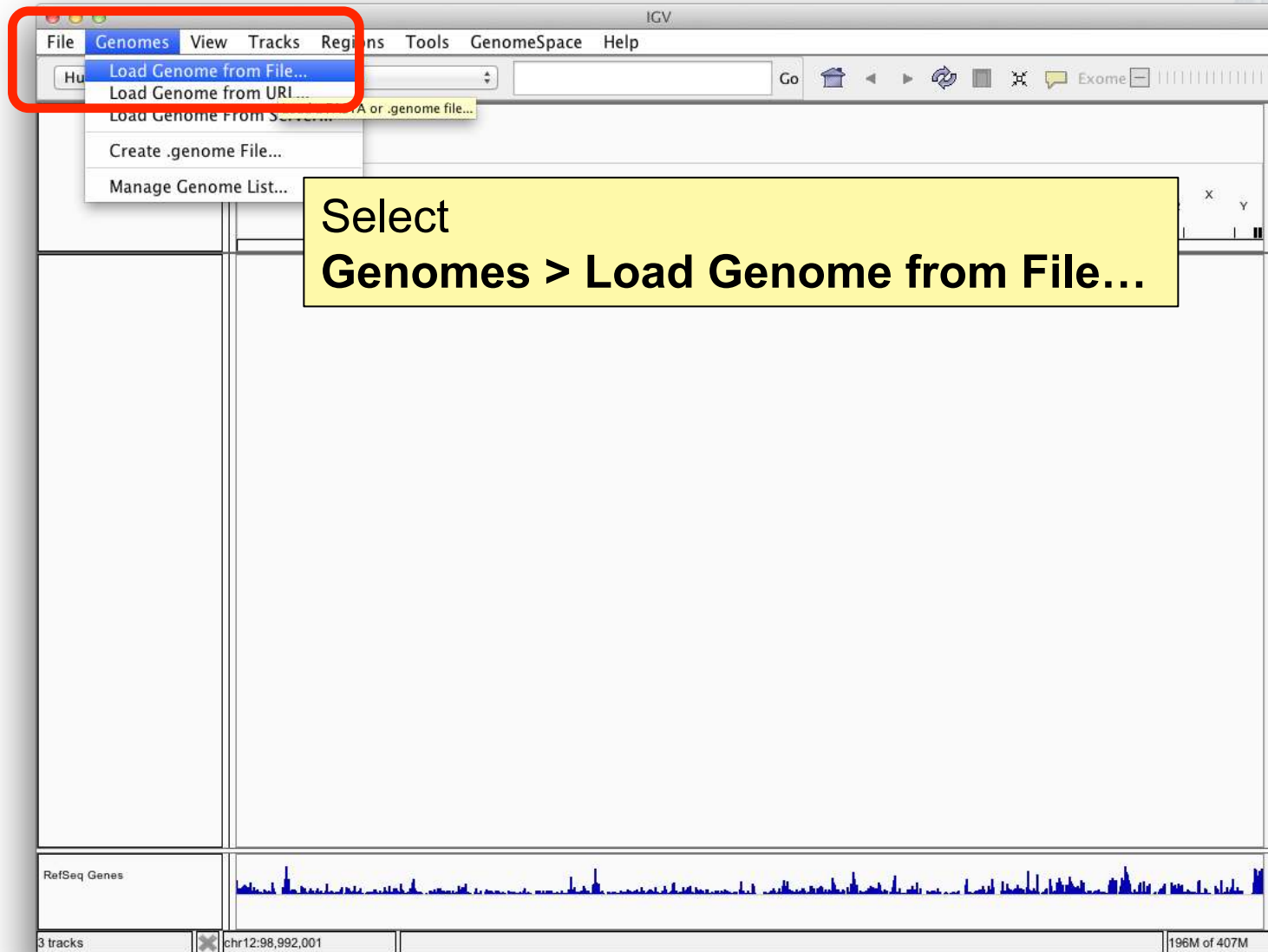
Optionally, package genome annotations with the sequence.

Loading a genome

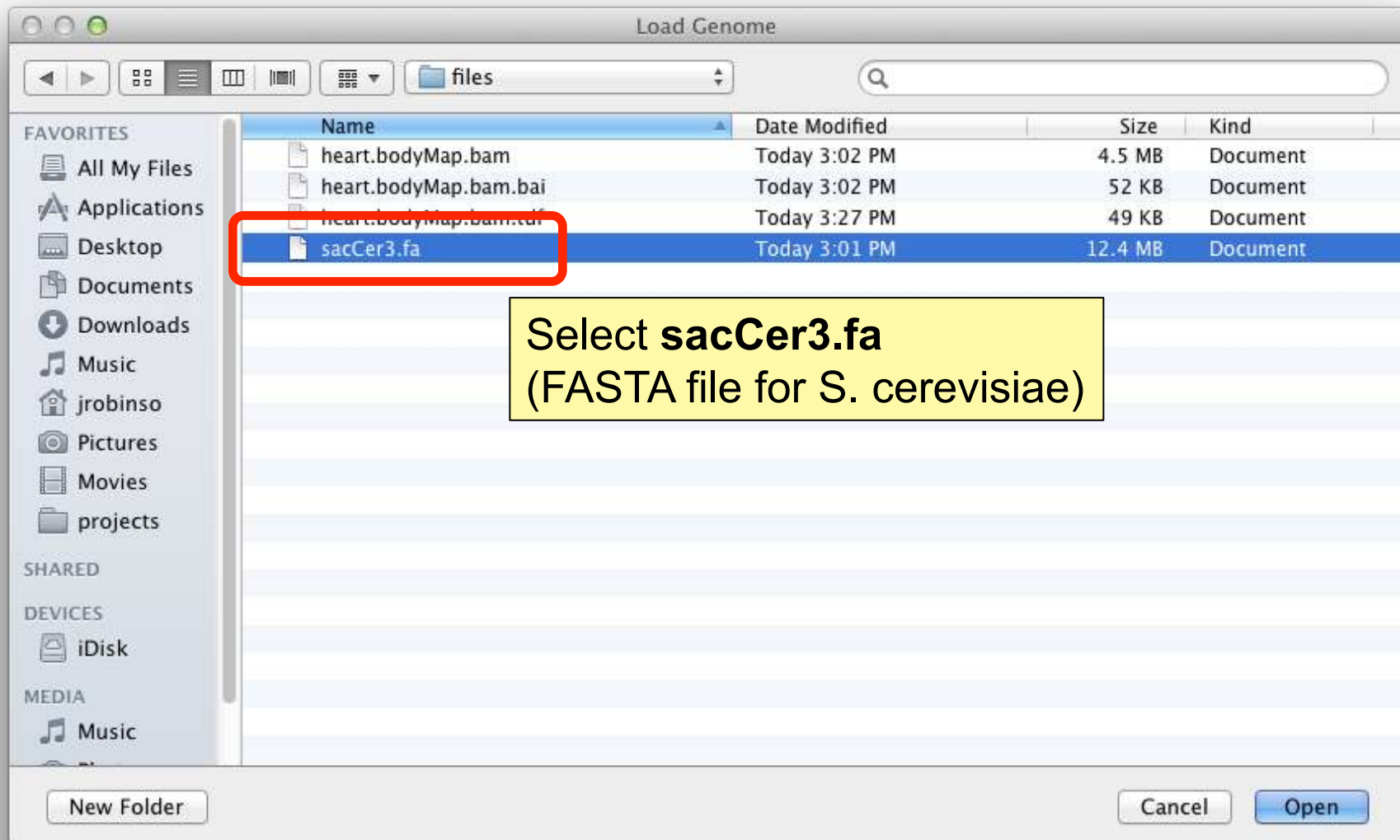


Hands-on exercise

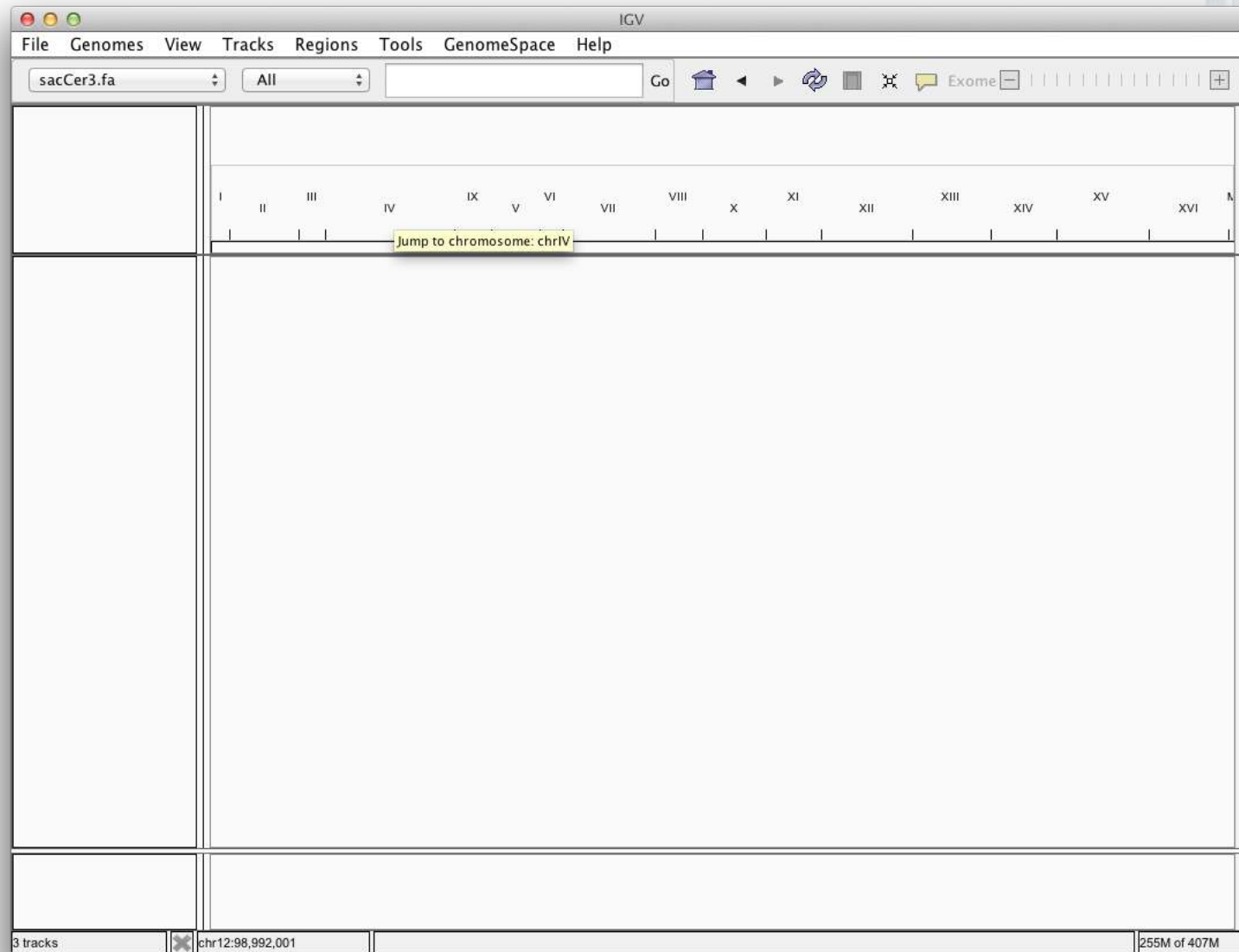
Loading a genome



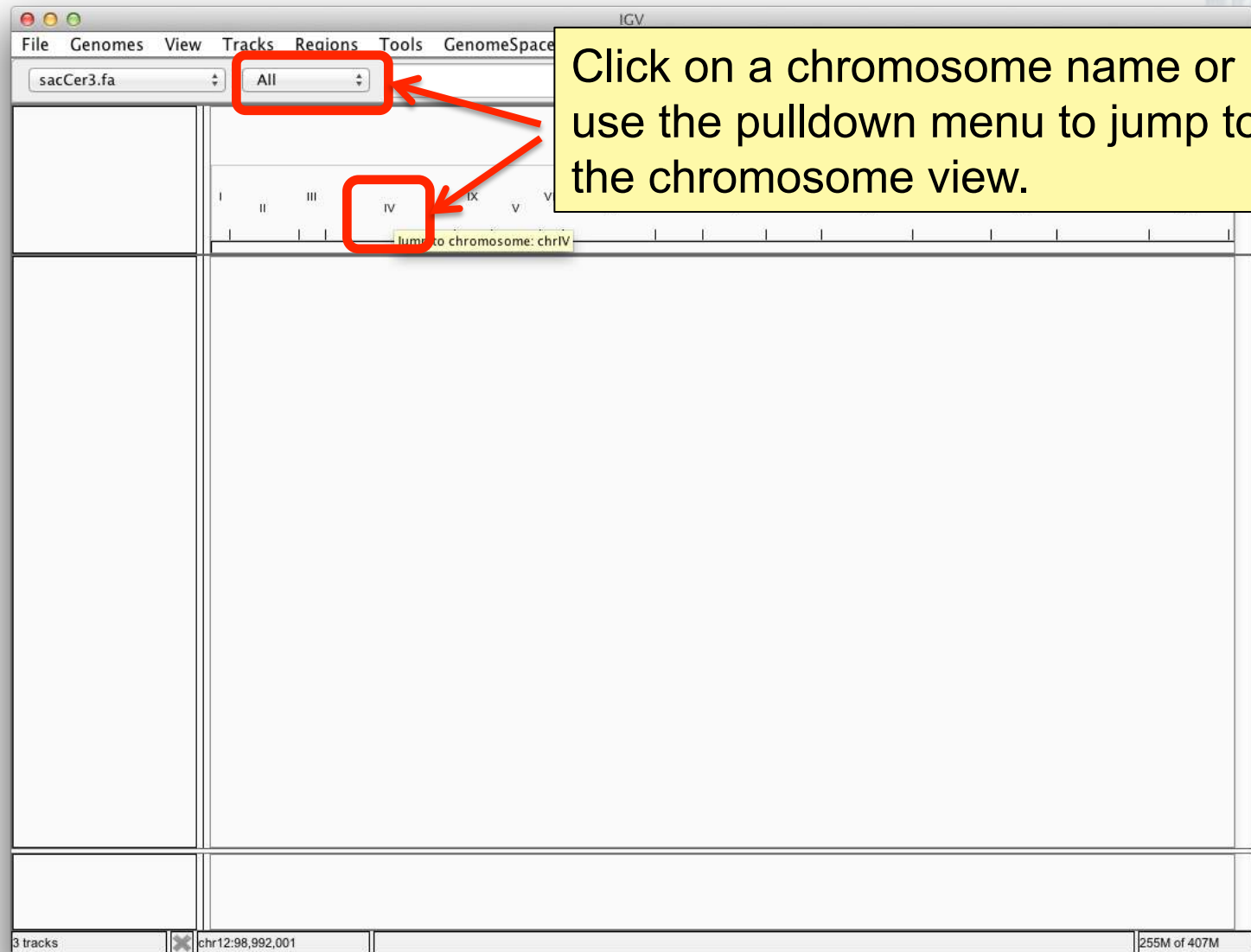
Loading a genome



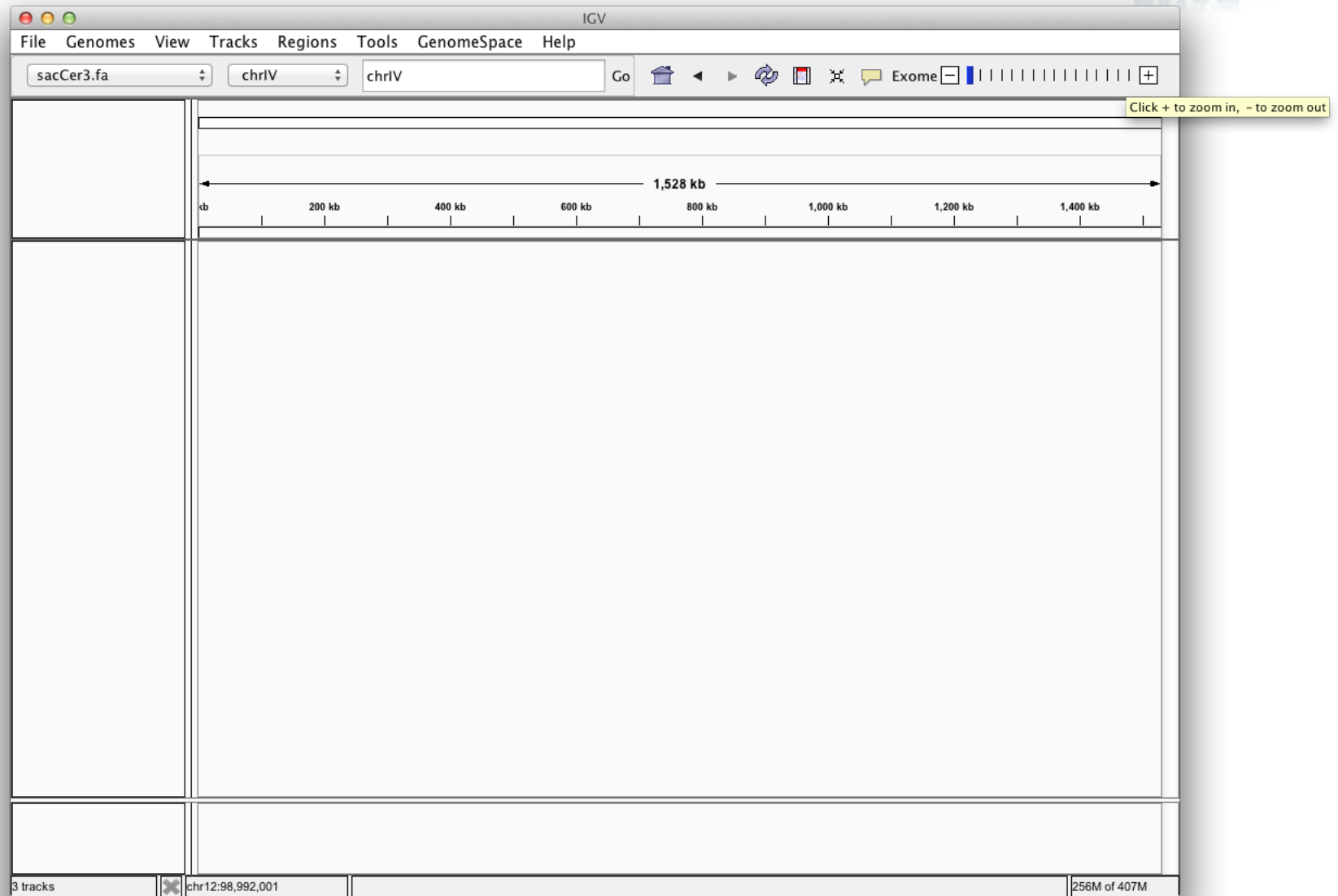
Loading a genome



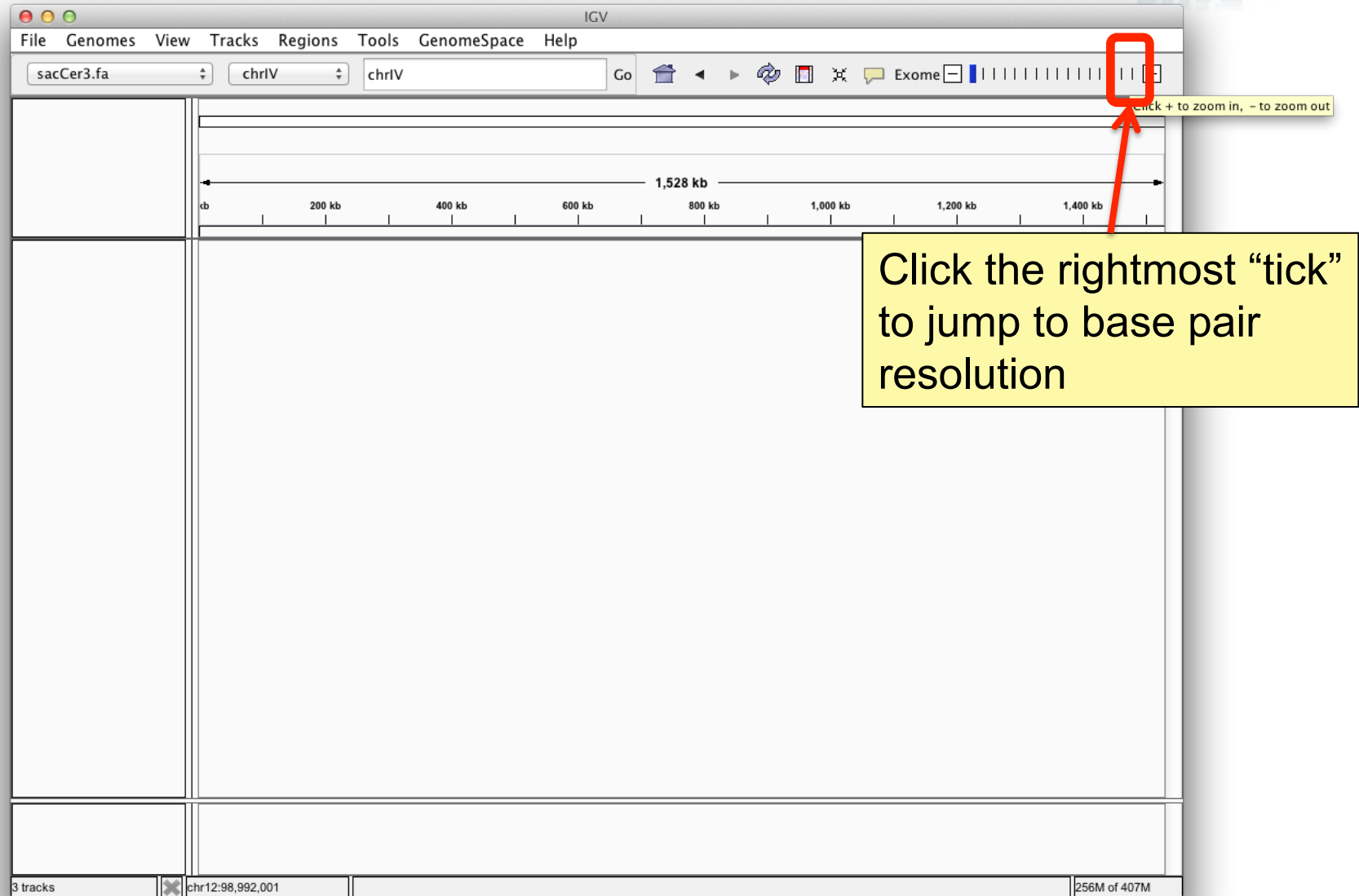
Loading a genome



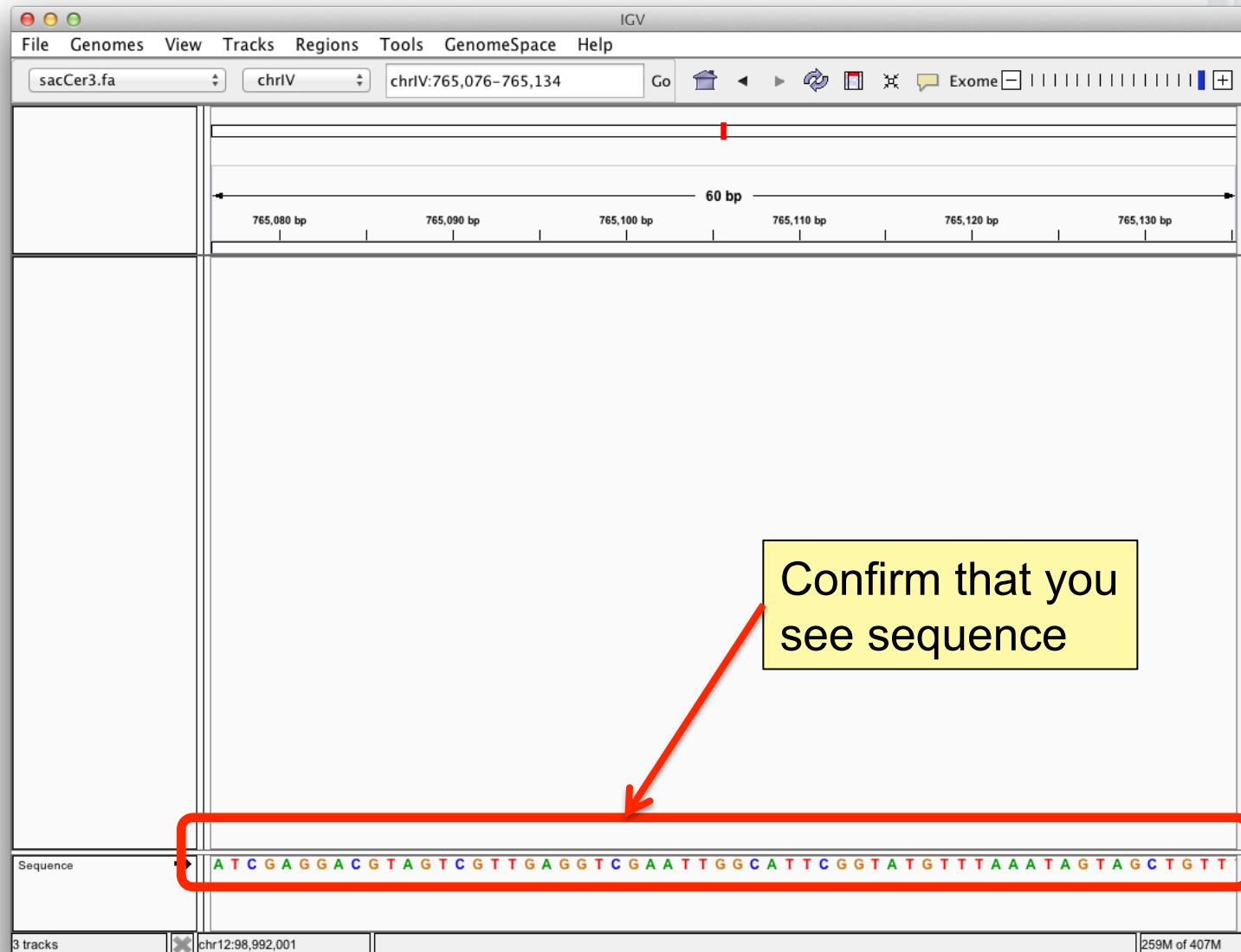
Loading a genome



Loading a genome



Loading a genome



Confirm that you see sequence

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).