

Introduction to the Integrative Genomics Viewer (IGV)

Jim Robinson & Helga Thorvaldsdóttir
Broad Institute of MIT and Harvard

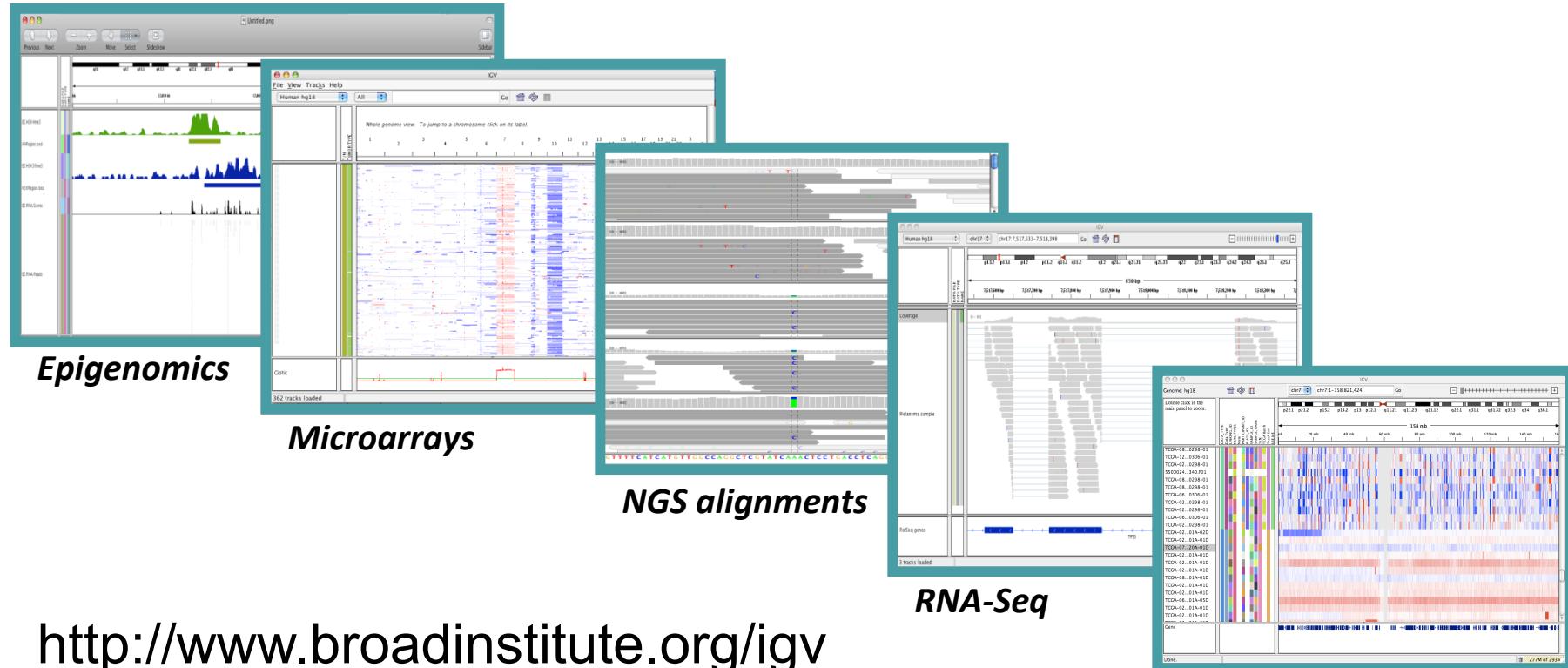
BroadE Workshop, June 2014



Integrative Genomics Viewer (IGV)



Desktop application for the interactive visual exploration of integrated genomic datasets



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

>85,000 registrations

mRNA, CNV, Seq



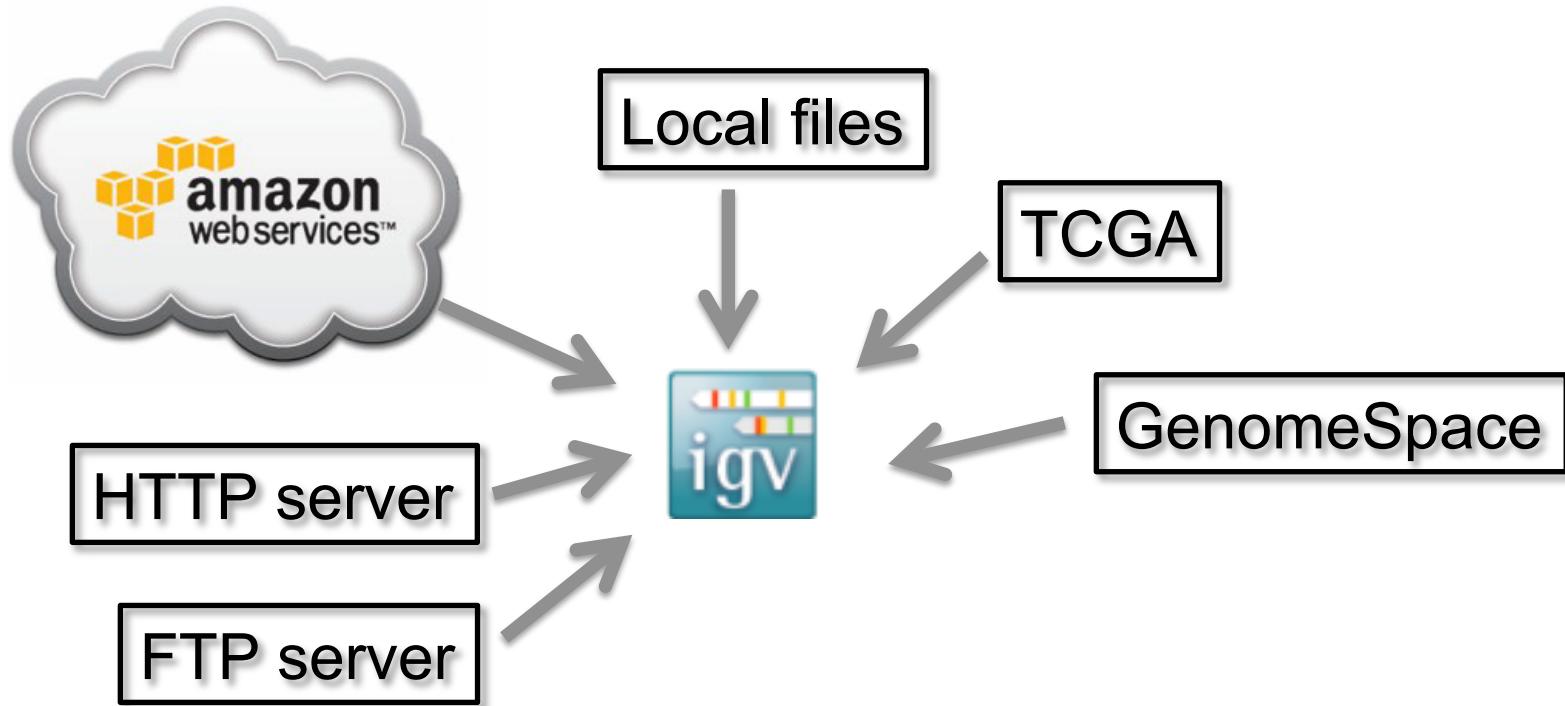
Features



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>

A screenshot of a web browser displaying the "Home | Integrative Genomics Viewer" page at https://www.broadinstitute.org/igv/. The page has a dark teal header with the IGV logo and a search bar. The main content area features a large image of the IGV software interface showing multiple tracks of genomic data. To the left of the main content is a sidebar with a navigation menu:

- Home
- Downloads** (circled with a red arrow)
- Documents
- Hosted Genomes
- FAQ
- IGV User Guide
- File Formats
- Release Notes
- Credits
- Contact

Below the menu is a search bar and links to "Broad Home" and "Cancer Program". The footer contains the "BROAD INSTITUTE" logo and the text "© 2012 Broad Institute".

Launch IGV



Integrative
Genomics
Viewer
ALGML

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

Search website

[Broad Home](#)
[Cancer Program](#)

BROAD INSTITUTE
© 2013 Broad Institute

INSTITUTE

Launch IGV



Integrative
Genomics
Viewer
IGV

Downloads | Integrative Genomics Viewer

www.broadinstitute.org/software/igv/download

Home Downloads

Downloads

Integrative Genomics Viewer (Version 2.3)

IGV can either be downloaded on to the local machine, or launched via Java Webstart.

Downloads

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

Download Mac App (button circled in red)

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

Download Binary Distribution

Java Webstart

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

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

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

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

Launch (button circled in red)
Launch with 750 MB
Maximum usable memory for Windows OS with 32-bit Java.

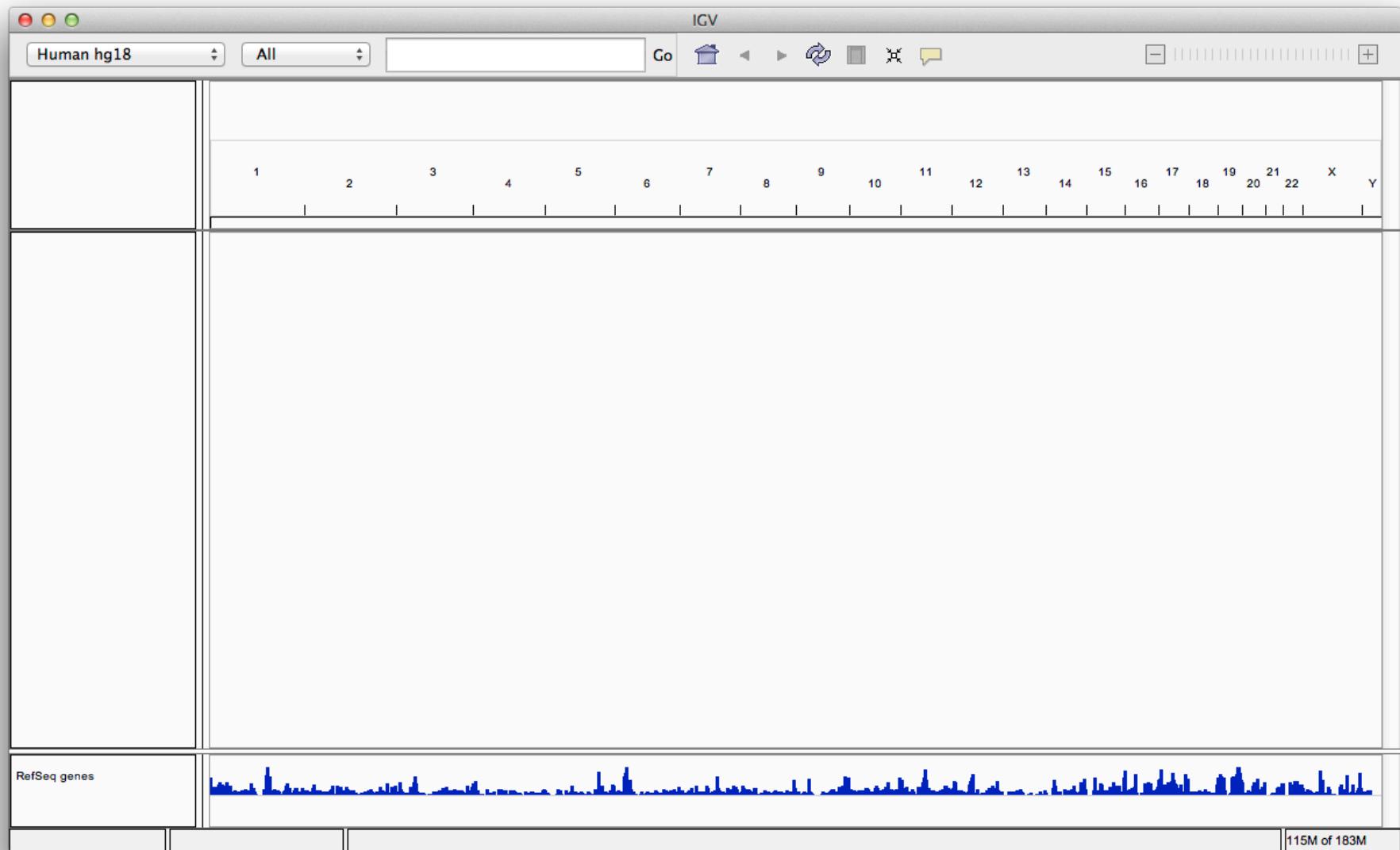
Launch
Launch with 1.2 GB

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

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



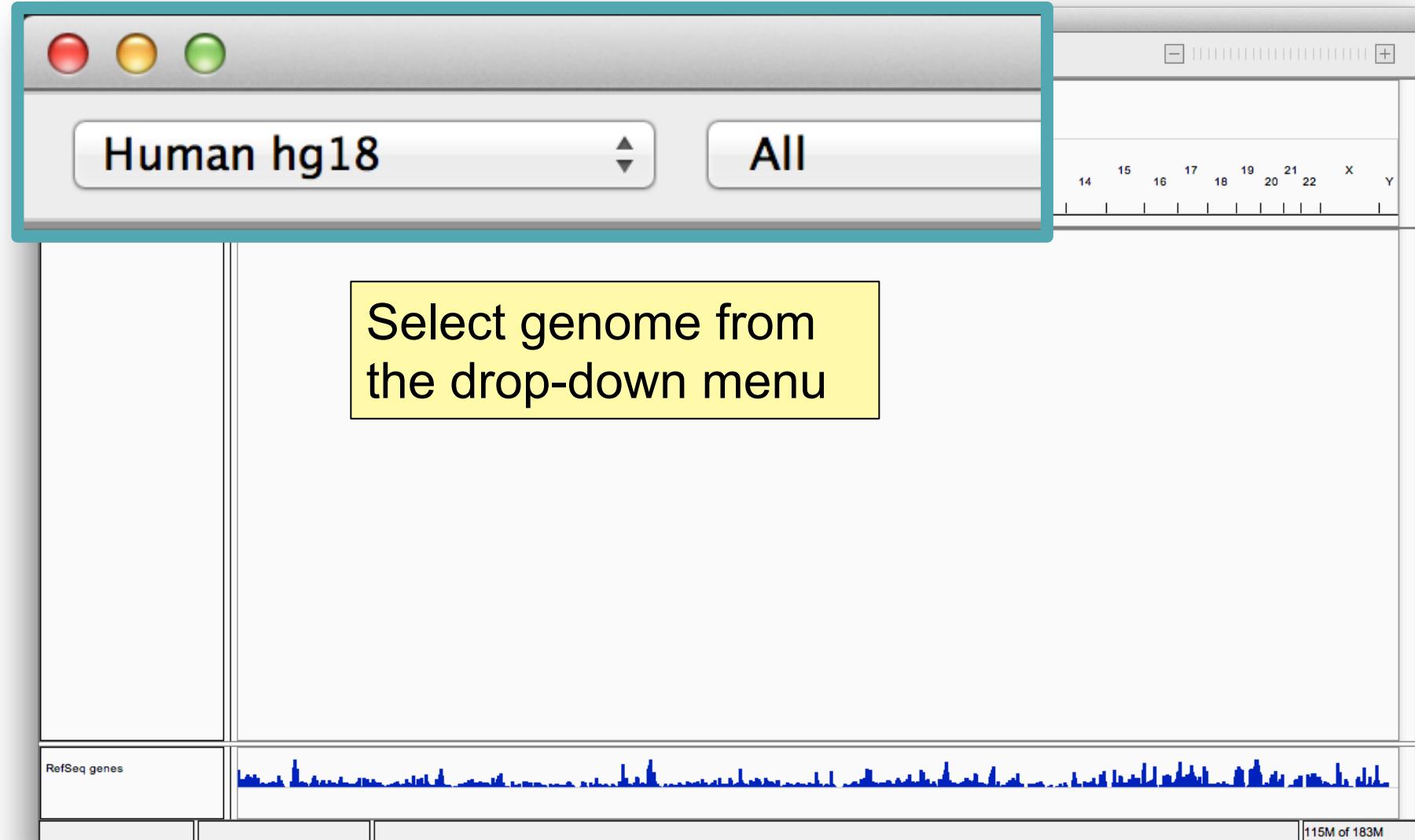
Integrative
Genomics
Viewer



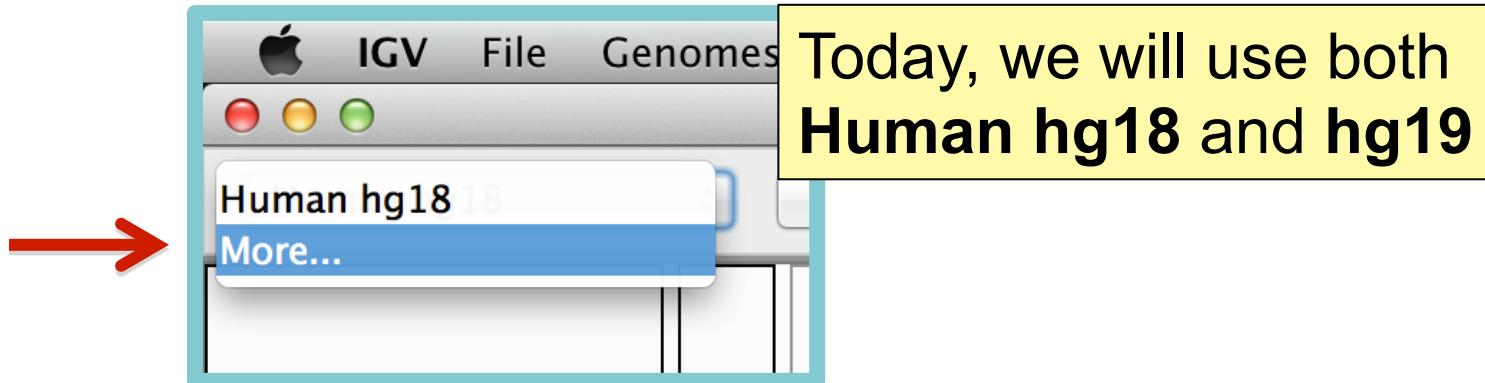
Select the reference genome



Integrative
Genomics
Viewer
ALGVIEWER

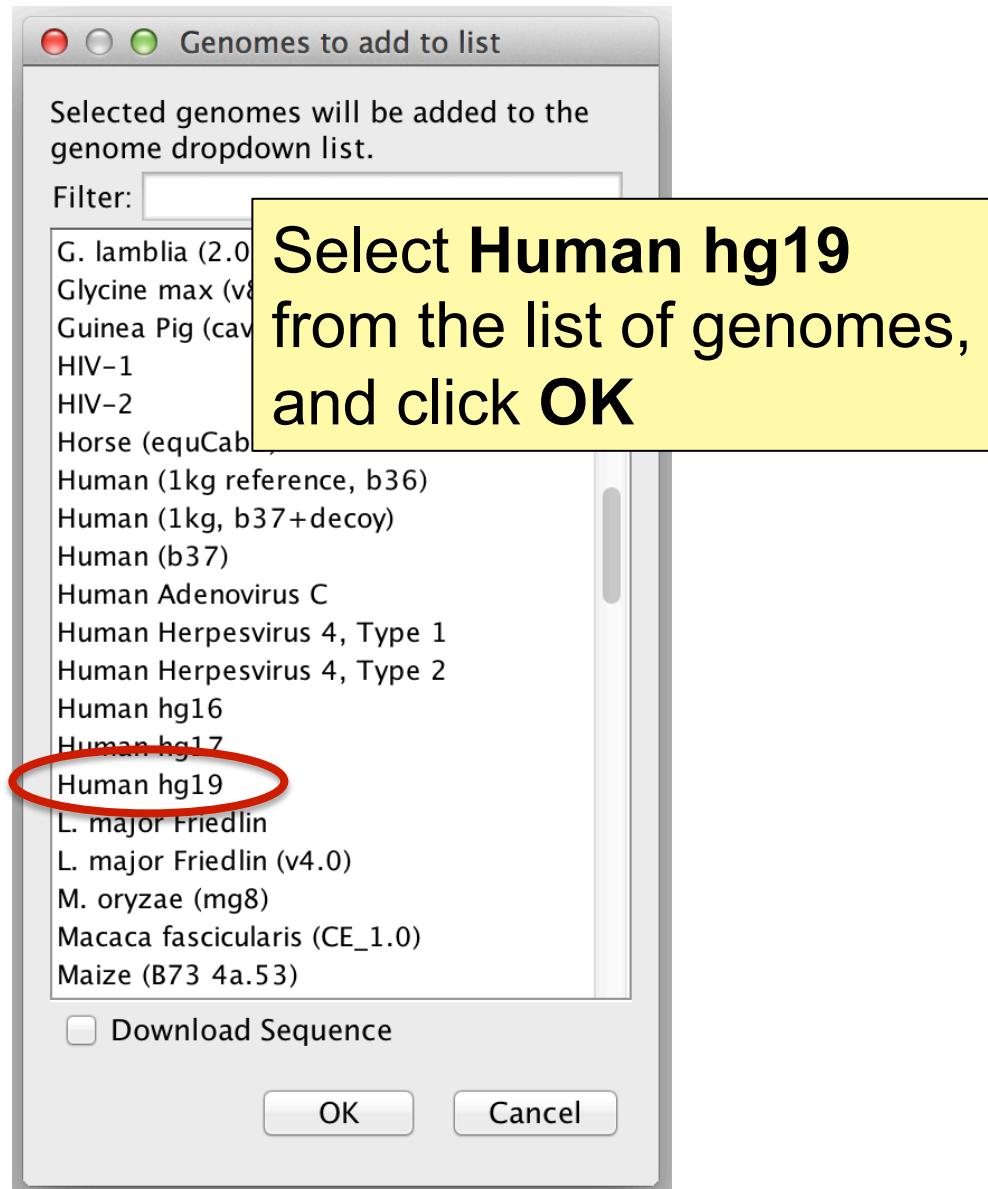


Select the reference genome

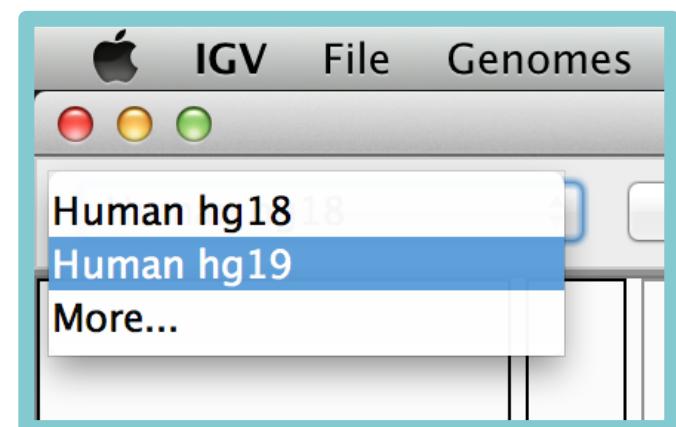
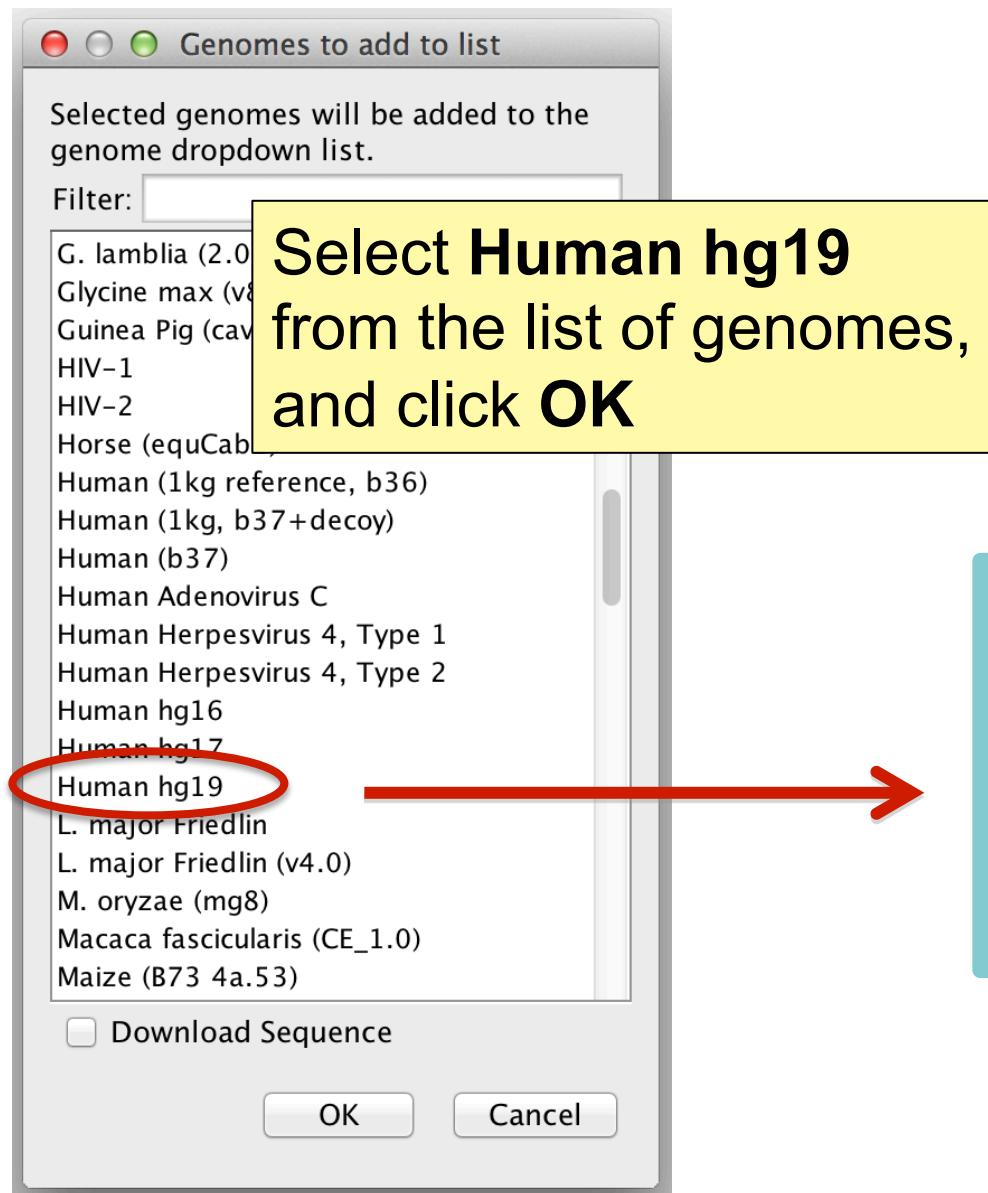


If **Human hg19** is not in the menu,
then click on **More...**

Select the reference genome



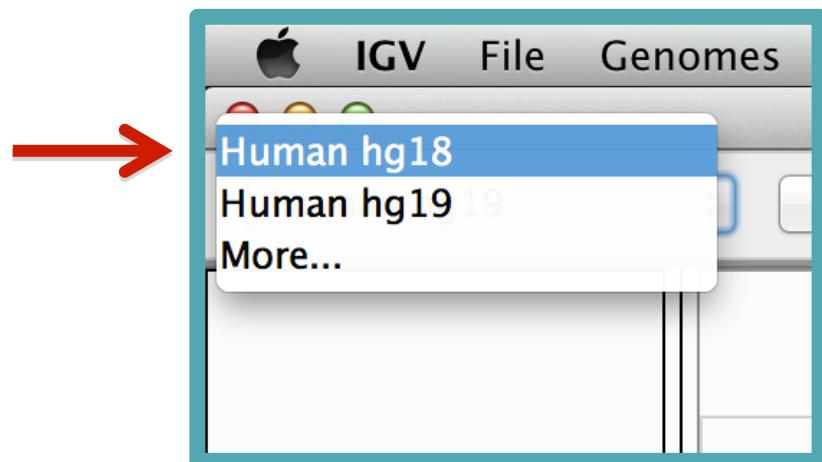
Select the reference genome



Select the reference genome



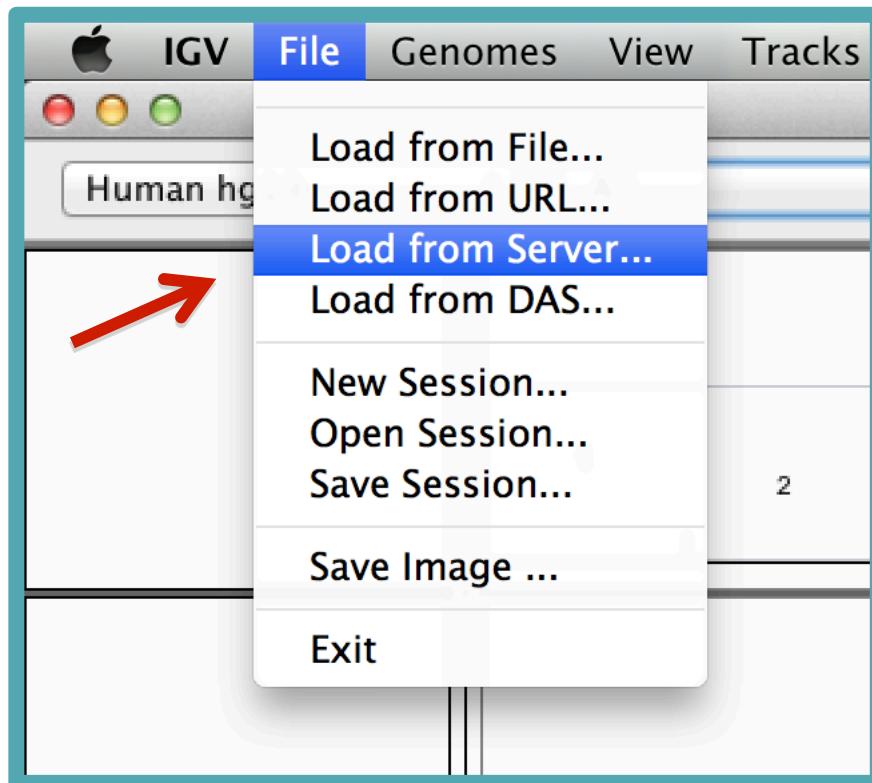
Select Human hg18



Load data



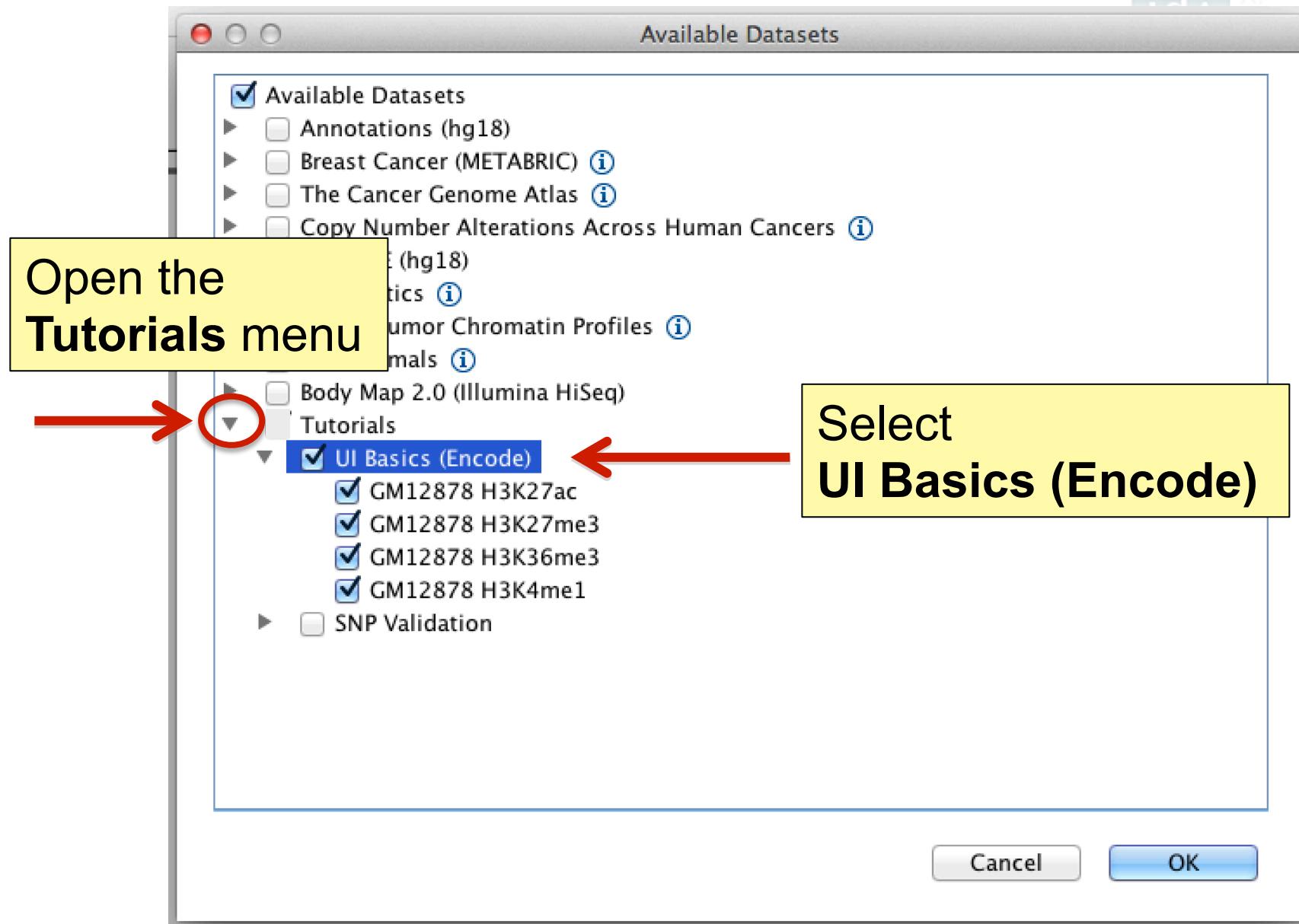
Select **File > Load from Server...**



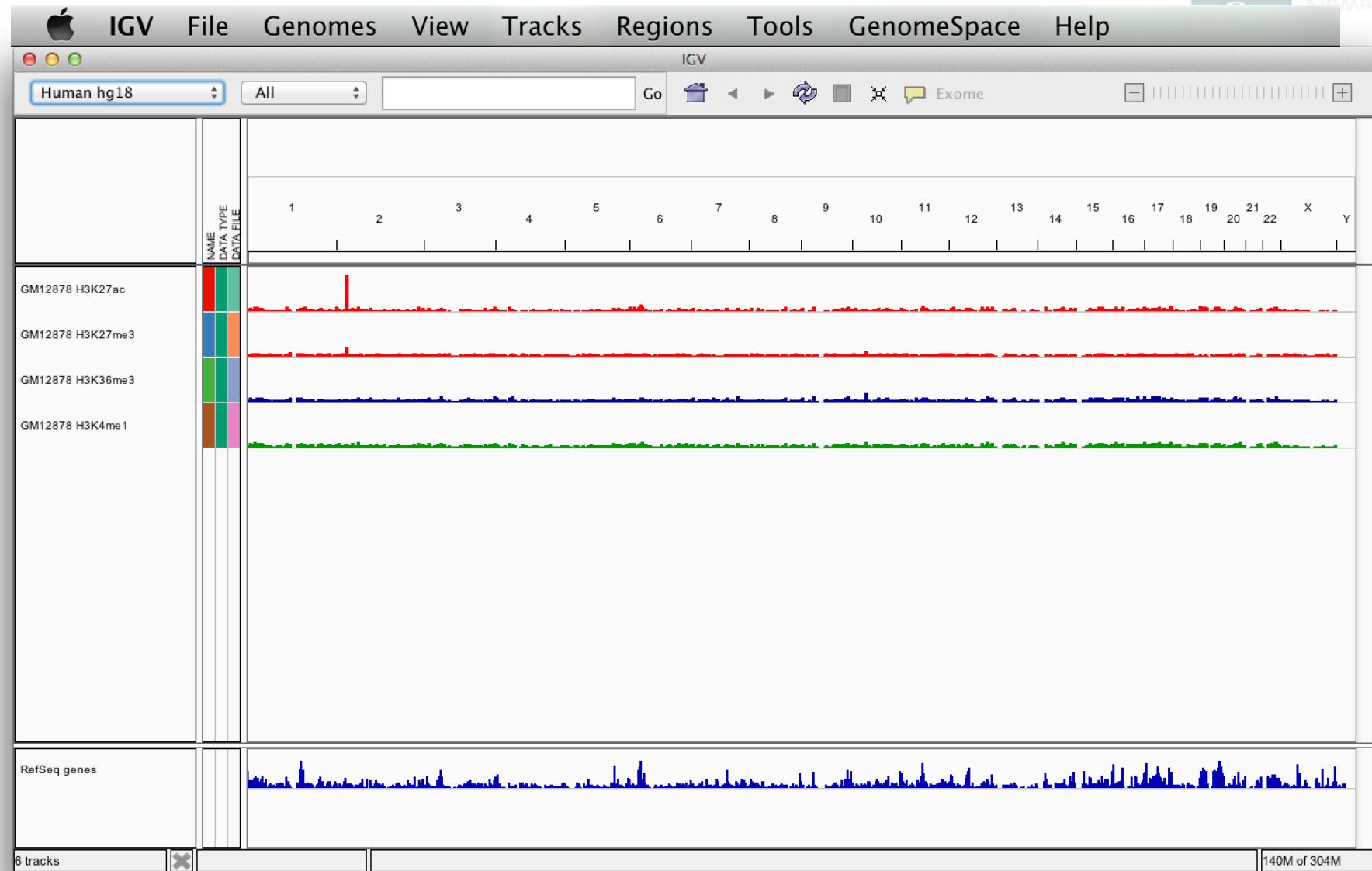
Load data



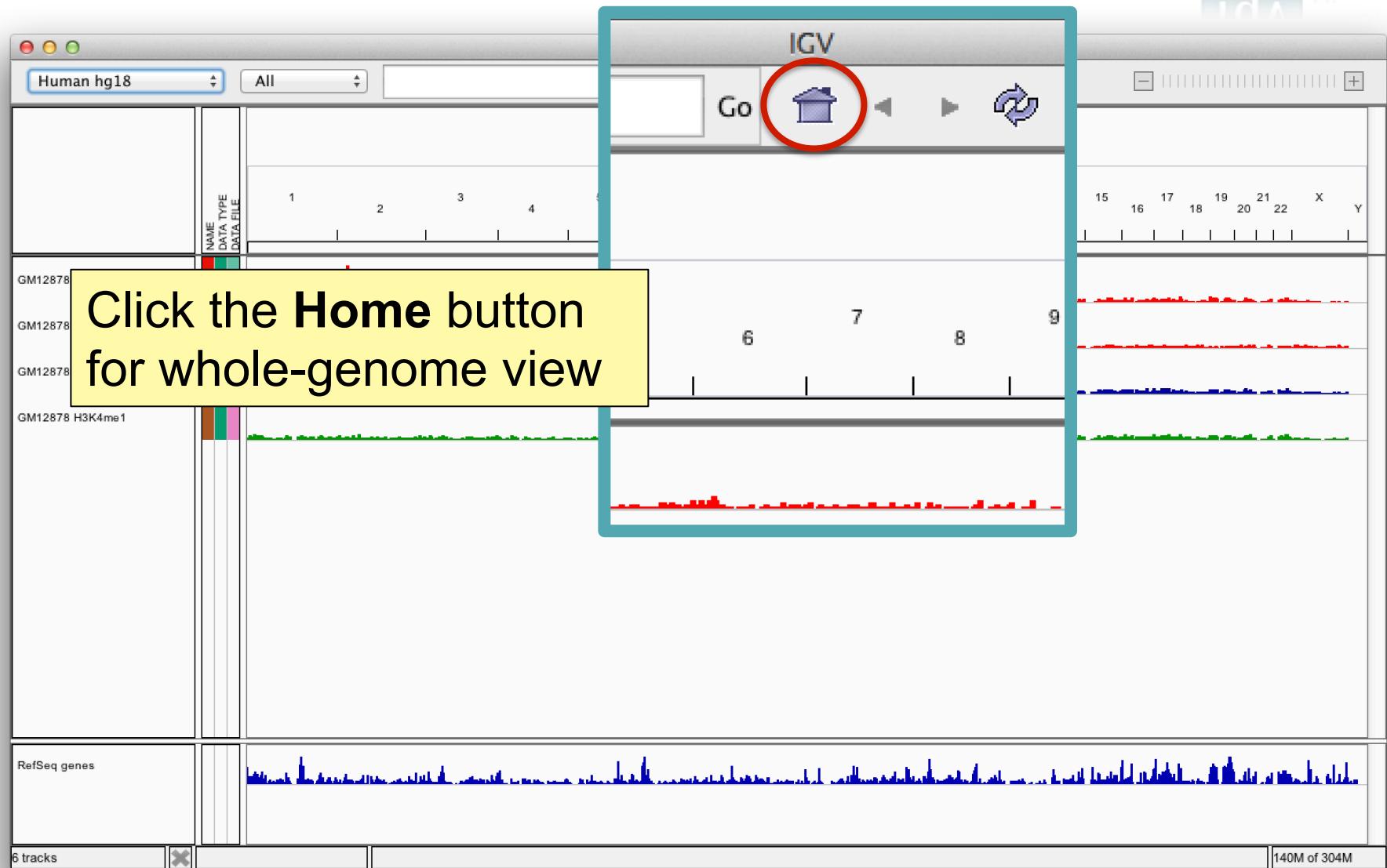
Integrative
Genomics
Viewer
ALGEME



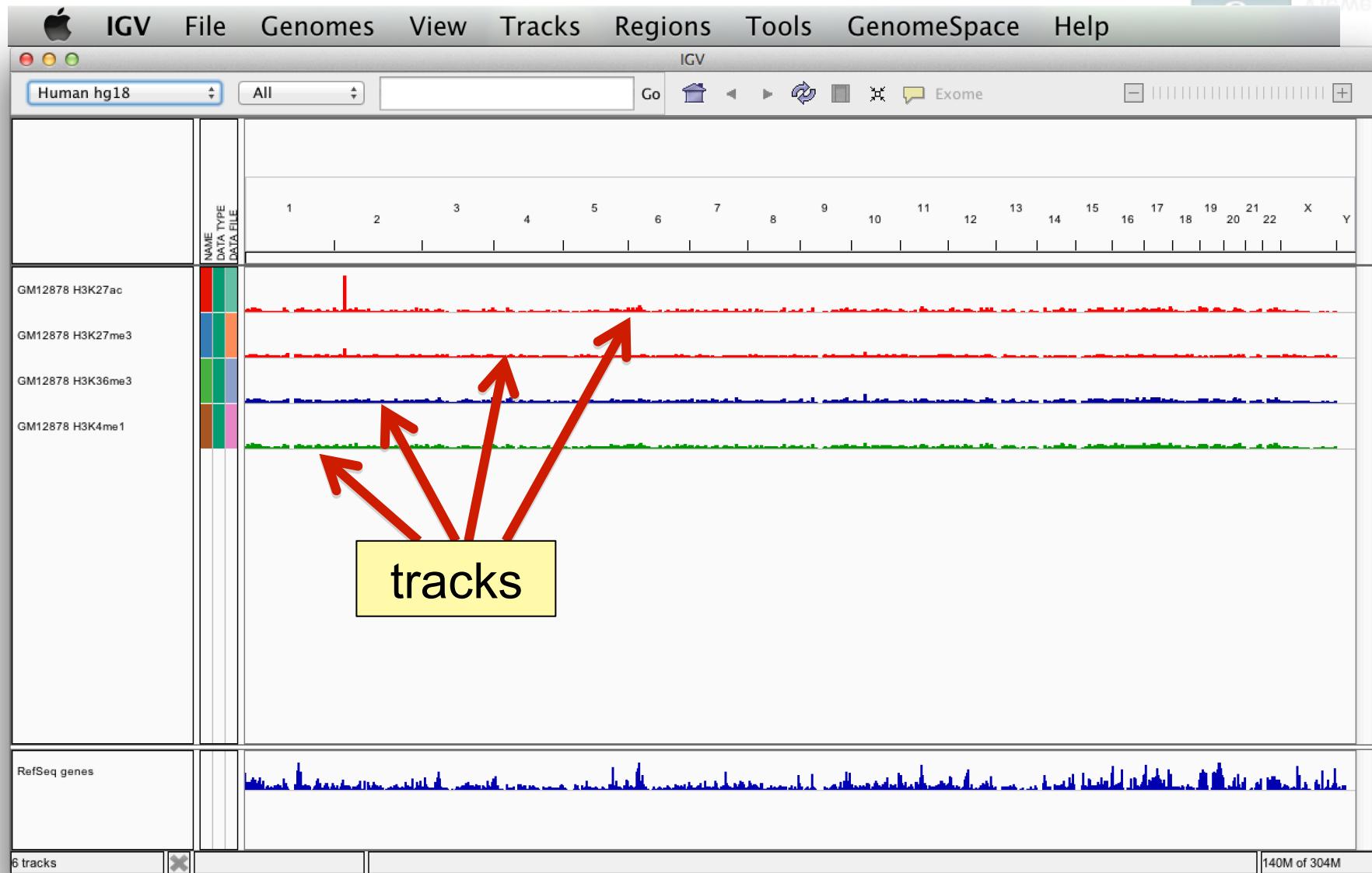
Screen layout



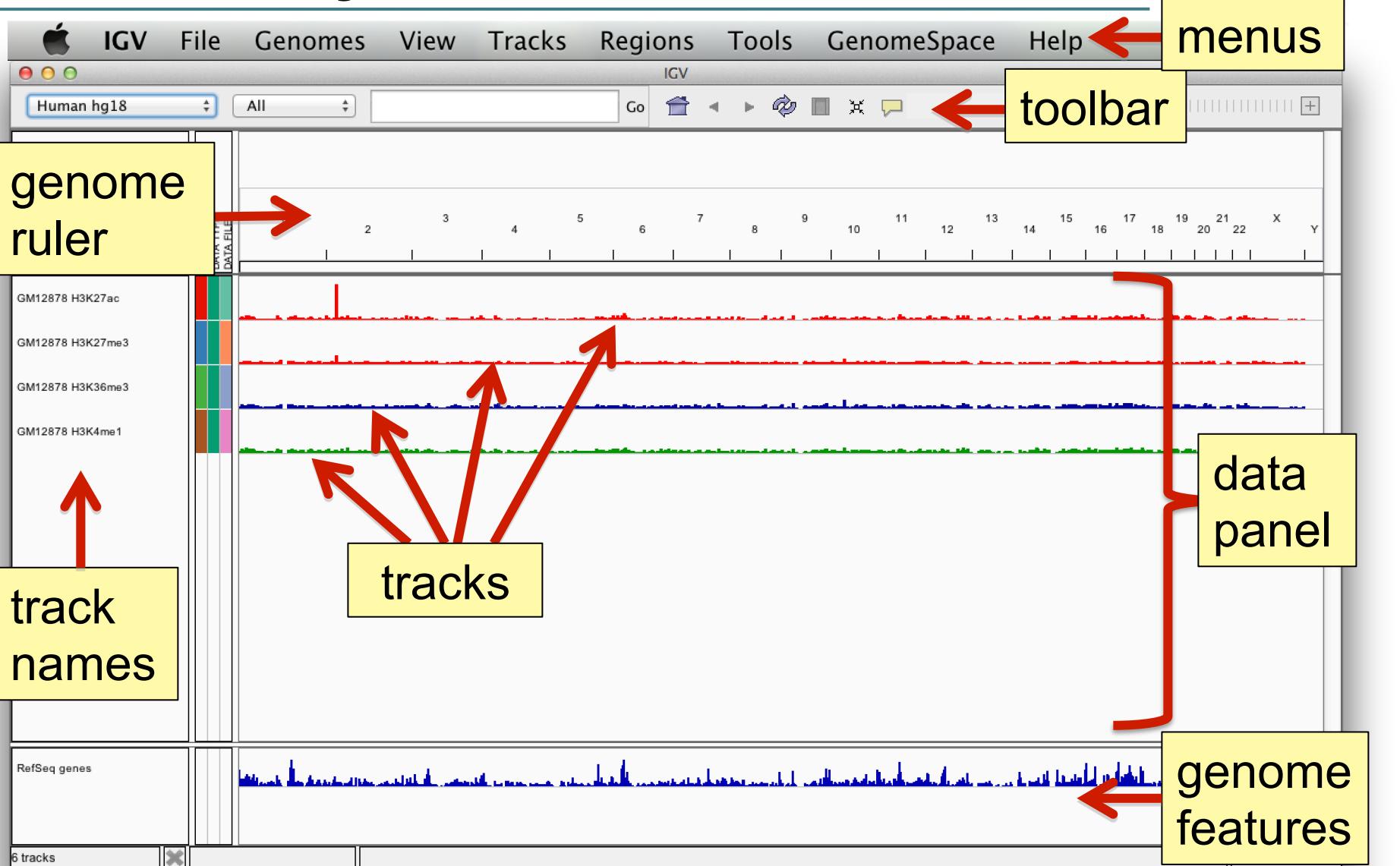
Screen layout



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

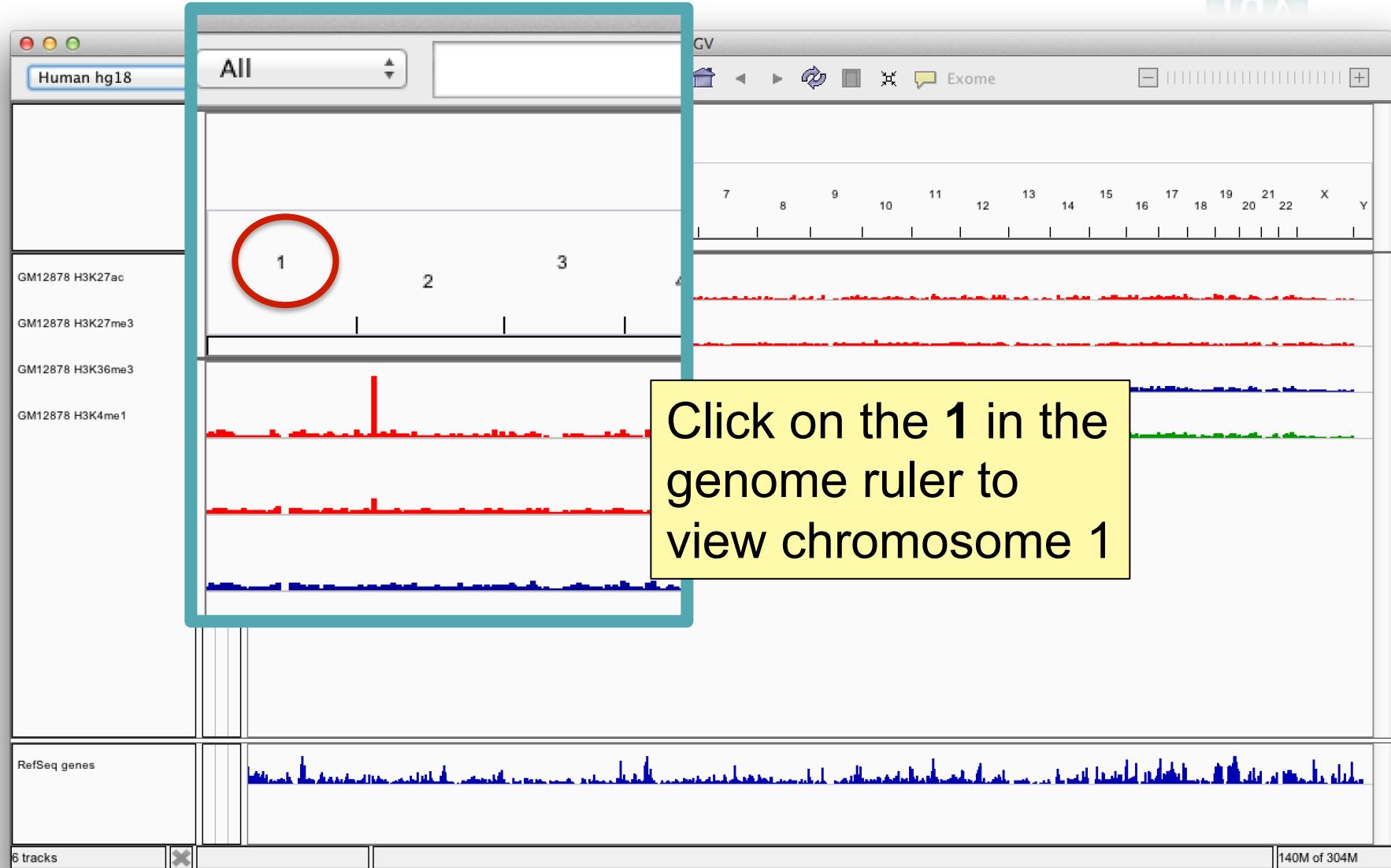


- The **file format** defines the track type.
- The **track type** determines the display options
- IGV supports many different file formats.

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

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

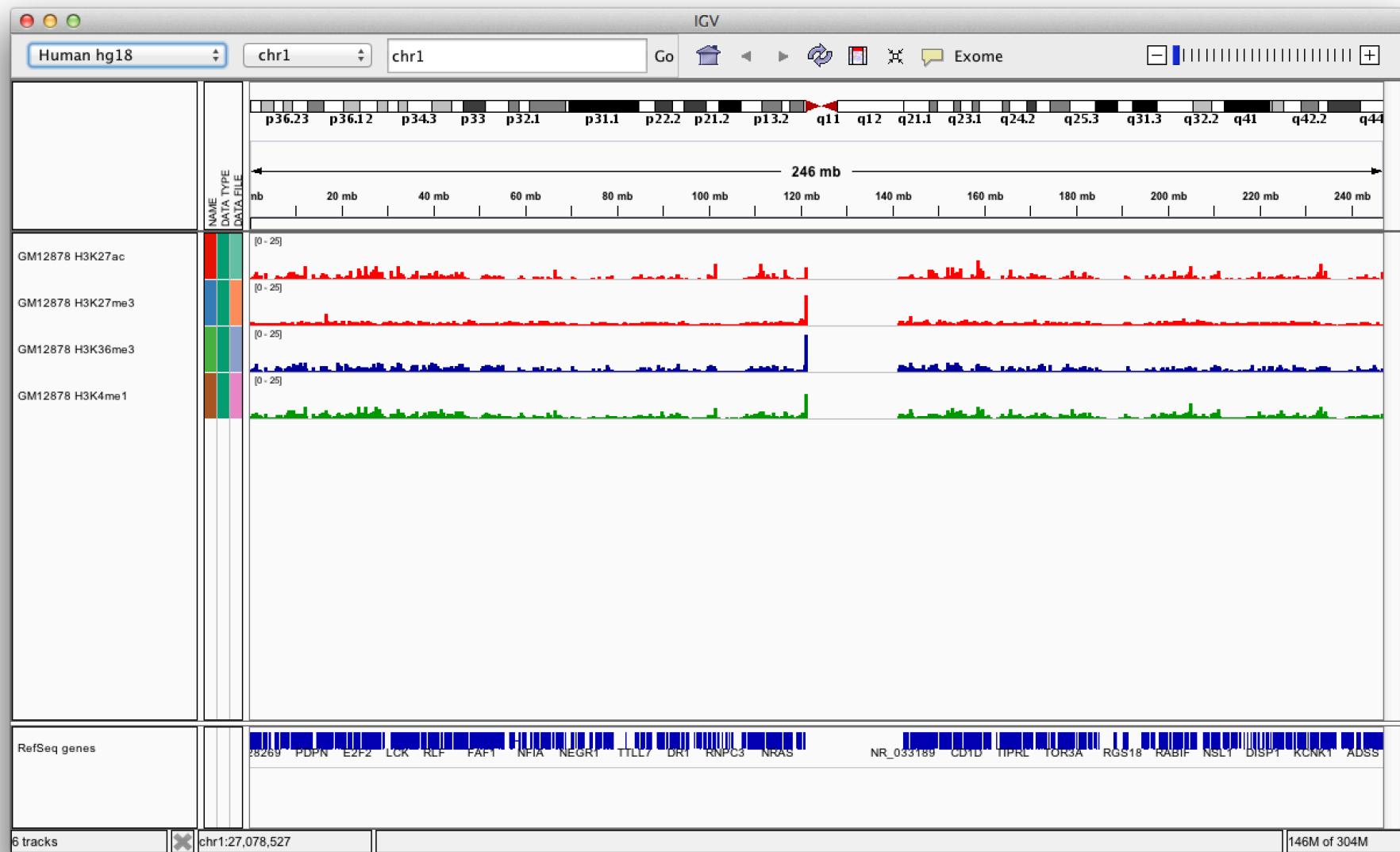
Navigate



Navigate



Integrative
Genomics
Viewer



Navigate



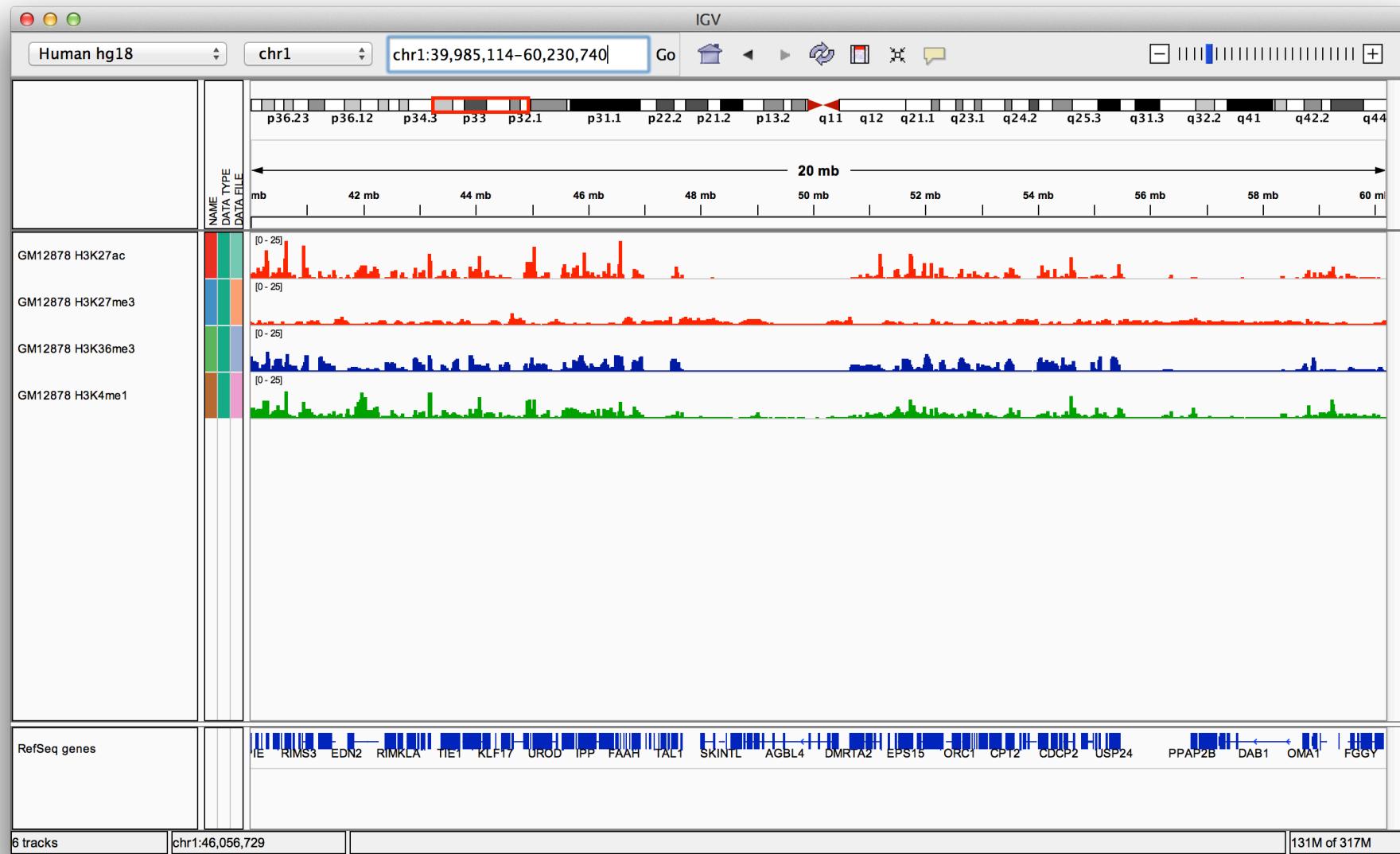
Integrative
Genomics
Viewer



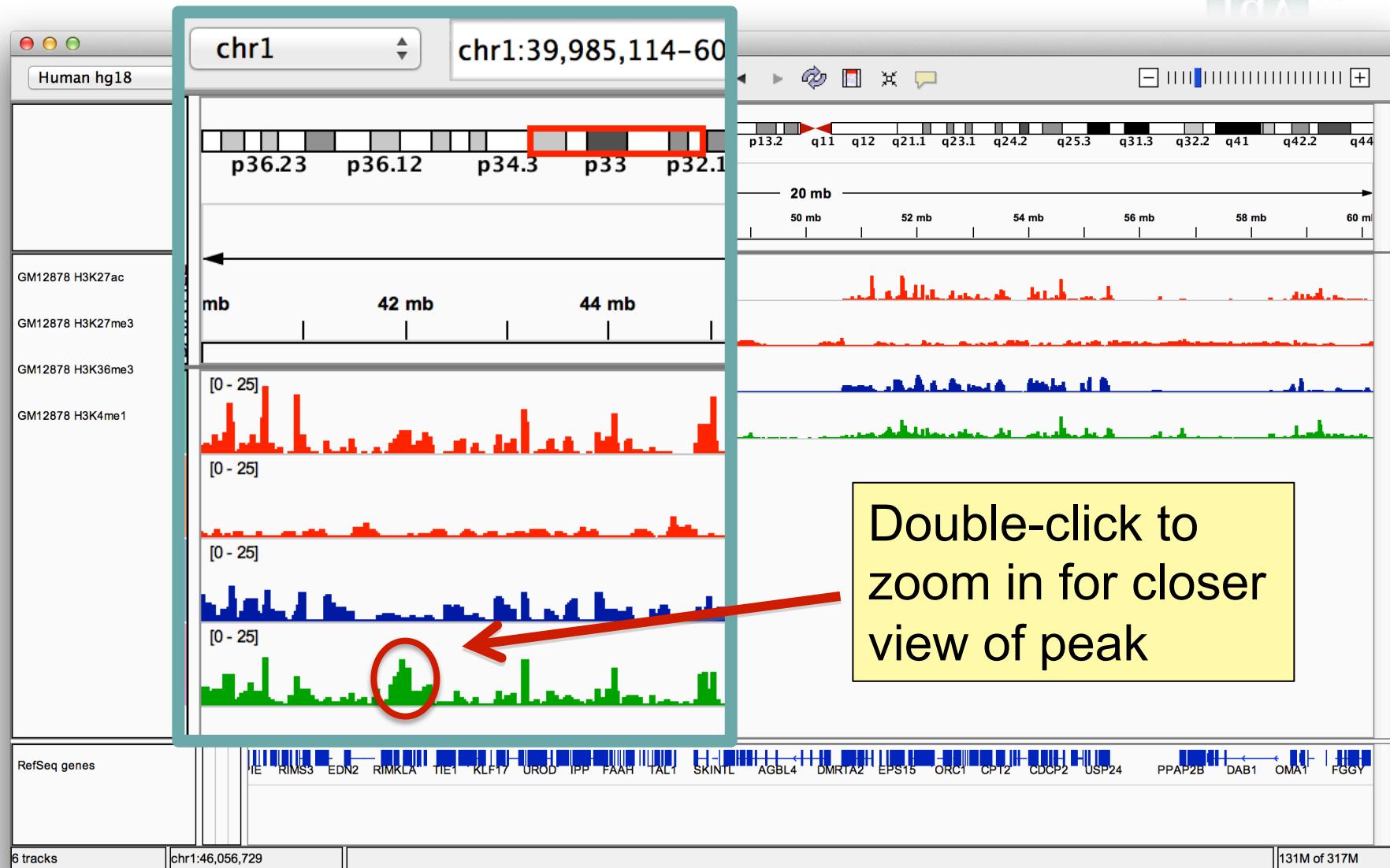
Navigate



Integrative
Genomics
Viewer
ALMEL



Navigate



Navigate



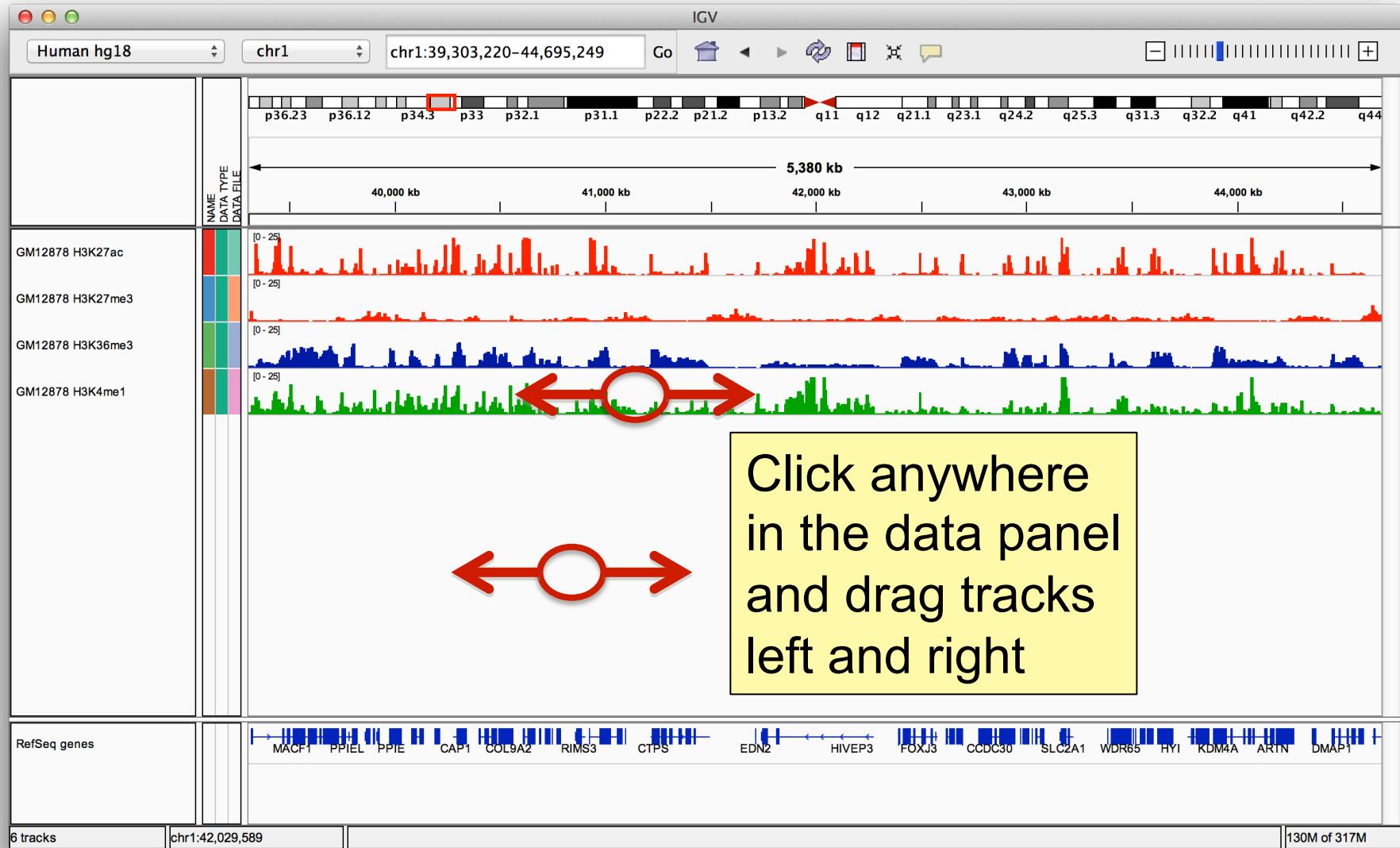
Integrative
Genomics
Viewer



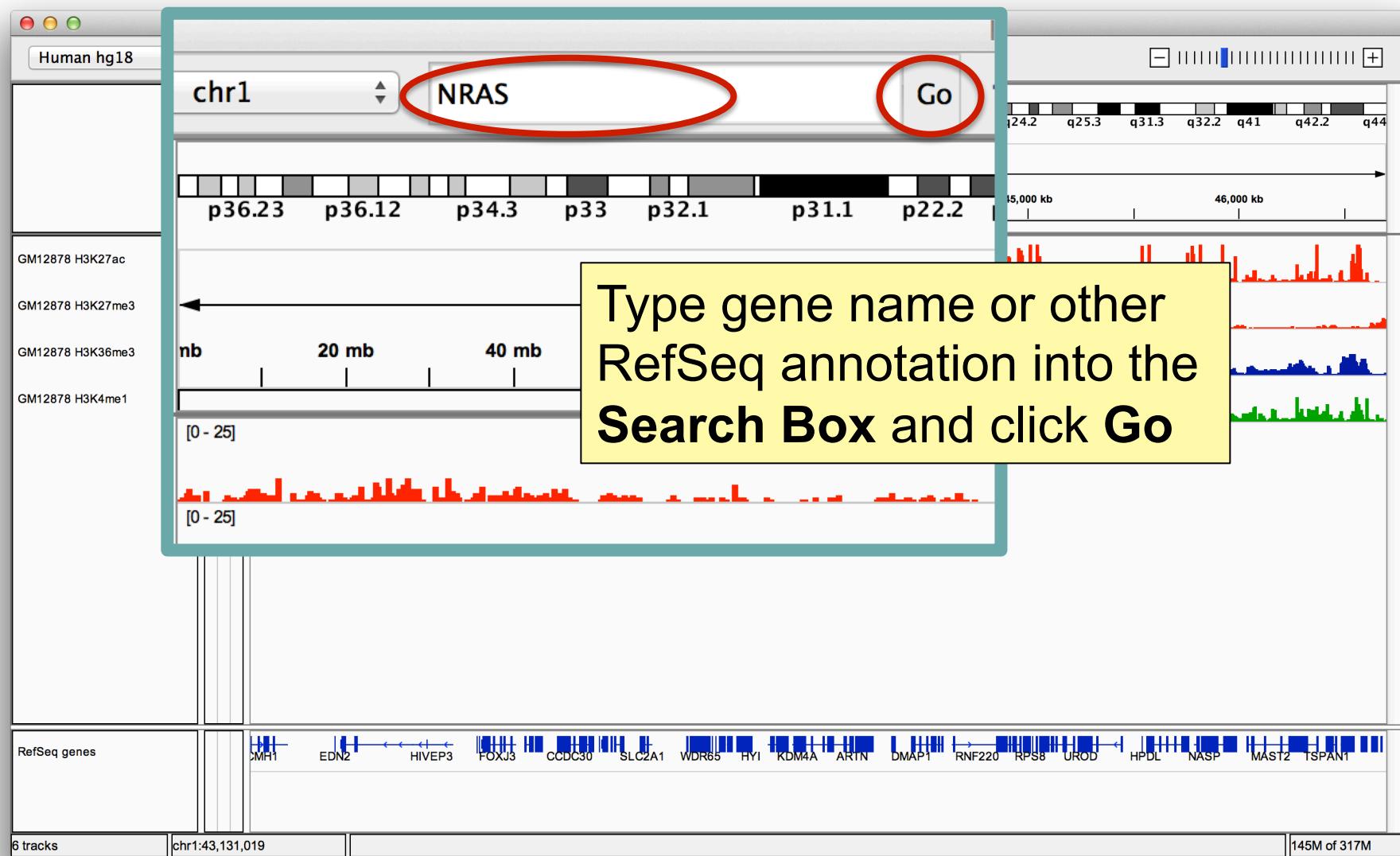
Navigate



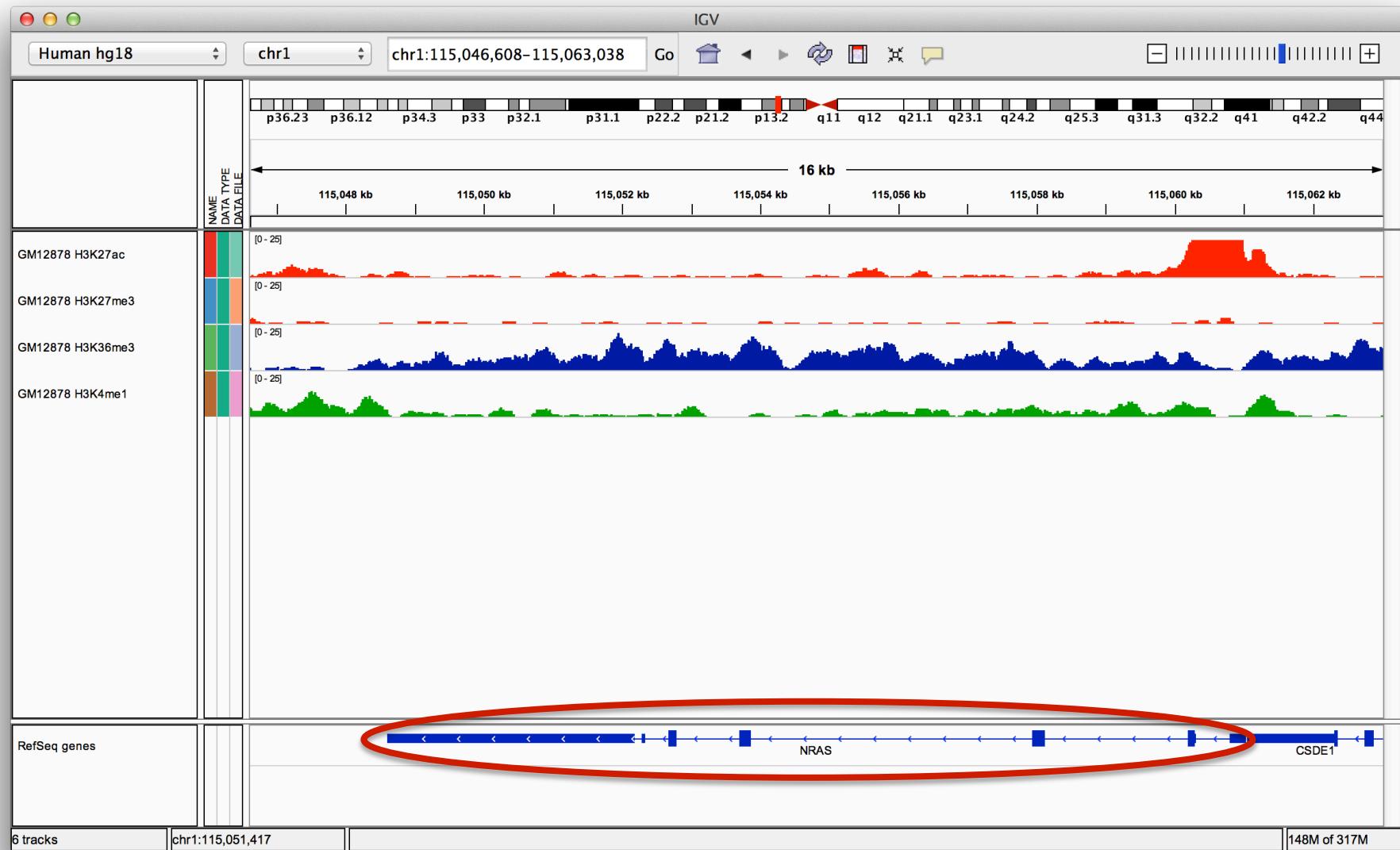
Integrative
Genomics
Viewer



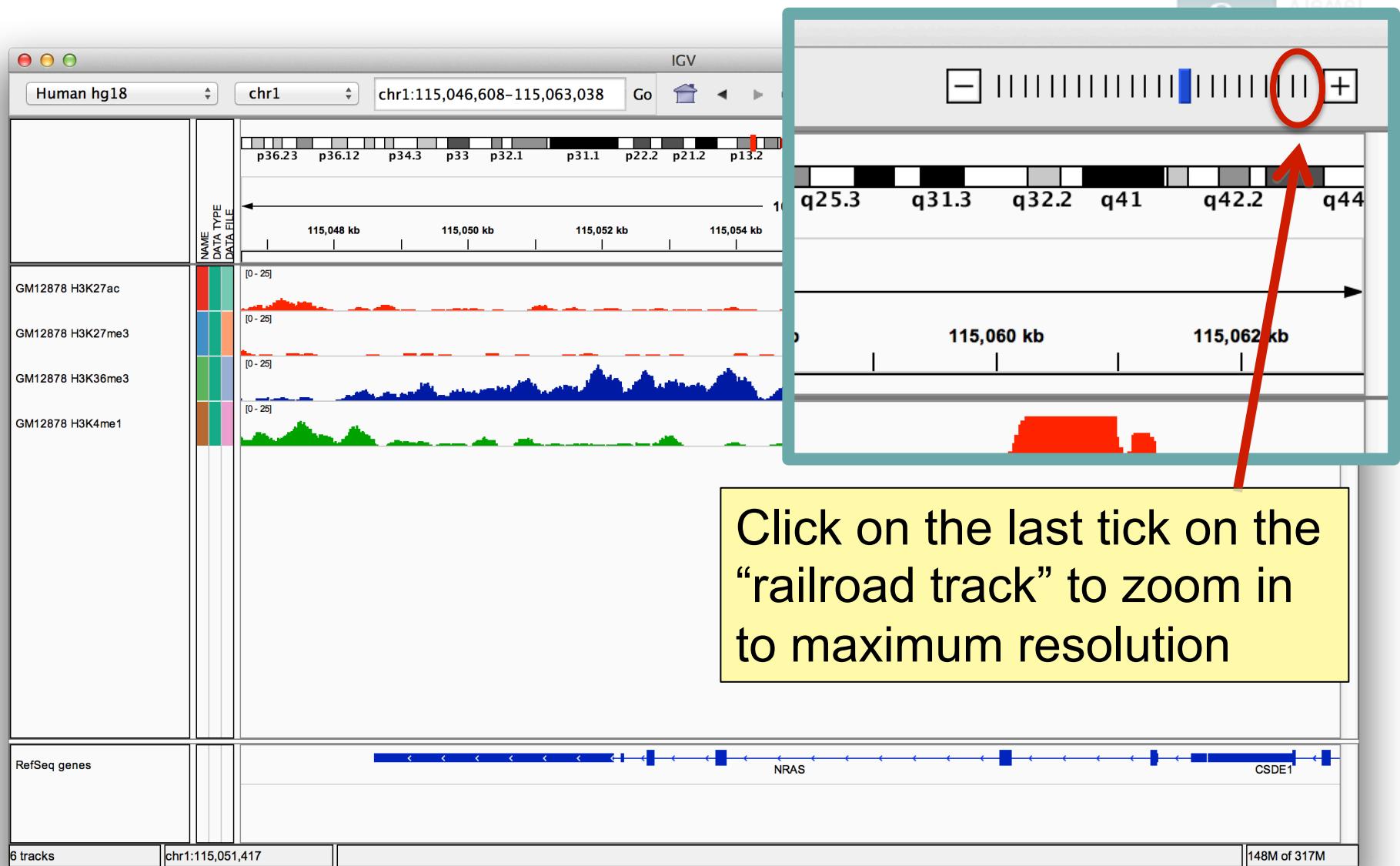
Navigate



Navigate



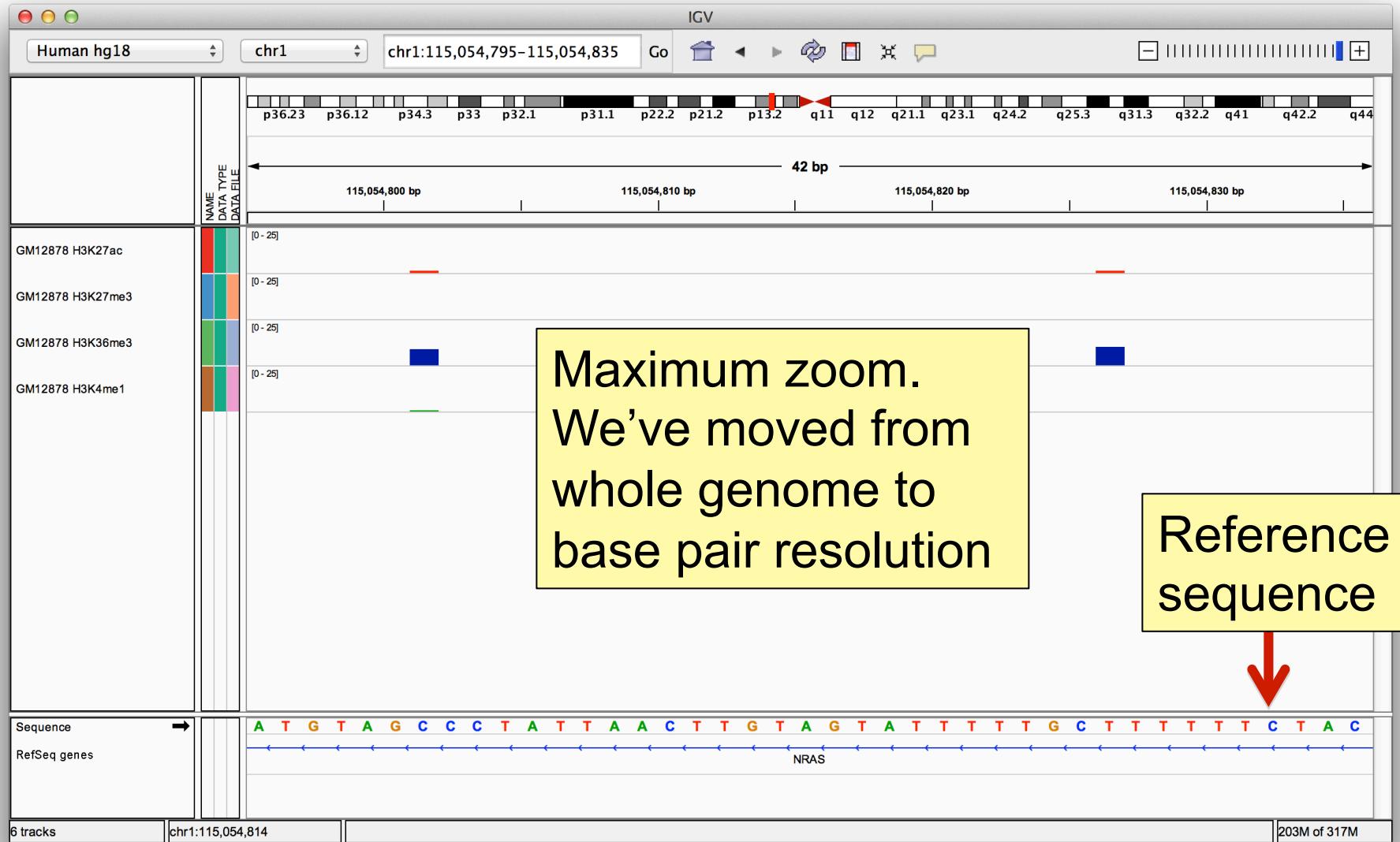
Navigate



Navigate



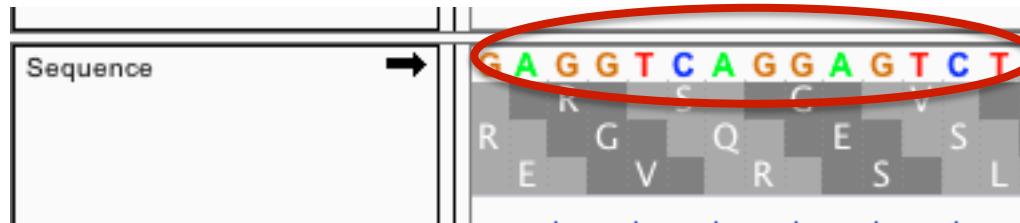
Integrative
Genomics
Viewer



Reference sequence



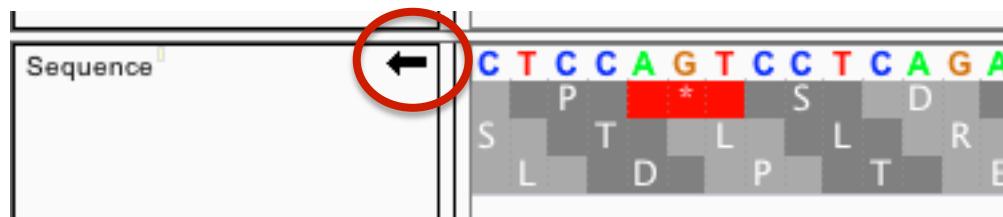
Click anywhere on the sequence to see a 3 frame translation.



By default the sequence for the forward strand is shown.



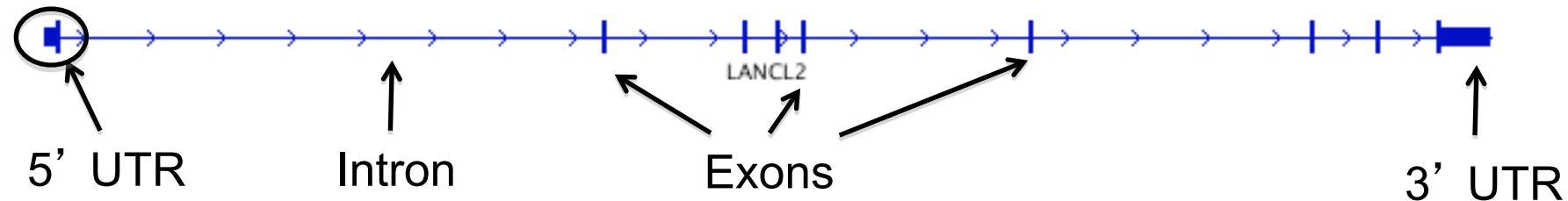
Click the arrow on the left to reverse the strand.



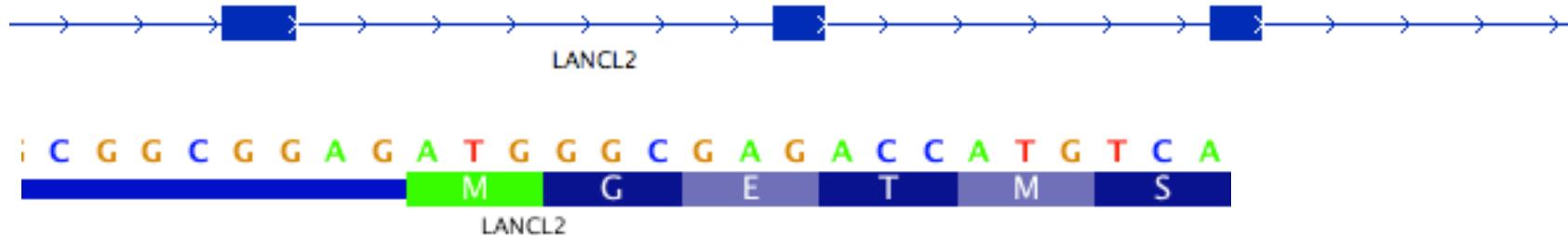
Genome annotation track



UCSC style gene representation



Zoomed in views



Zoomed out views

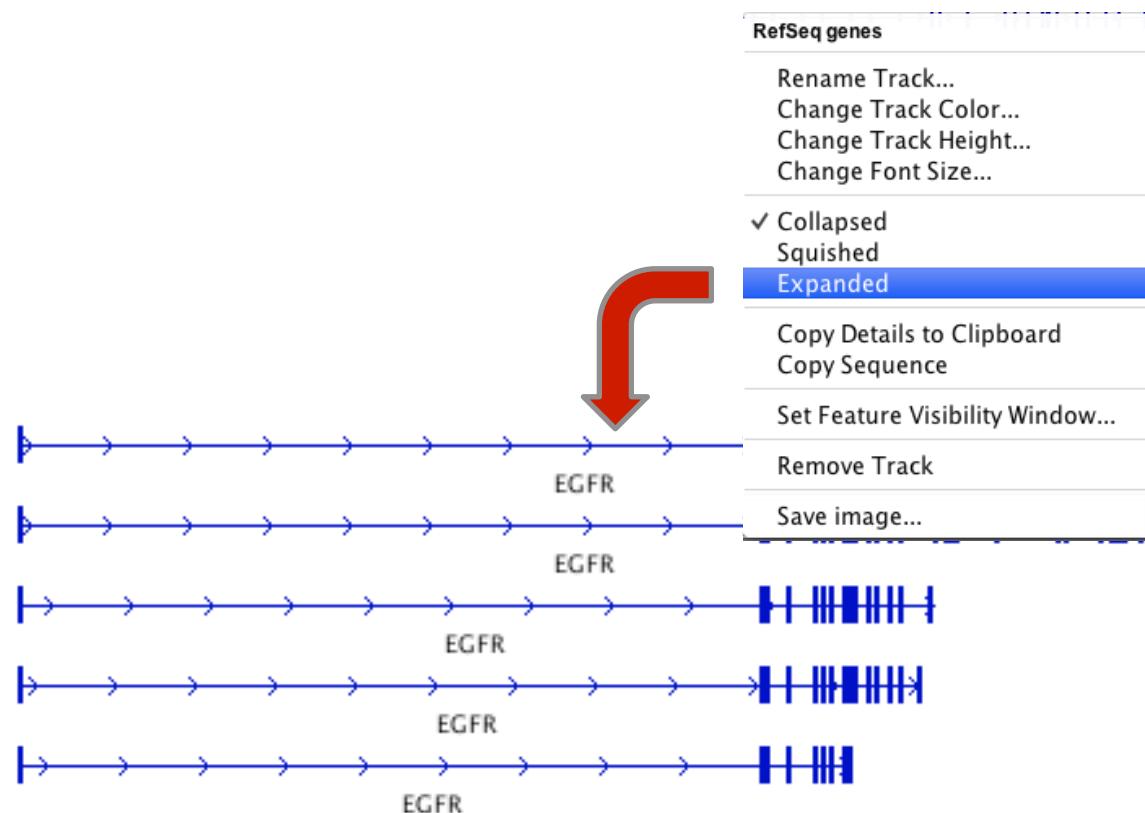


Annotation display mode

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



2. Expand the track using the popup menu

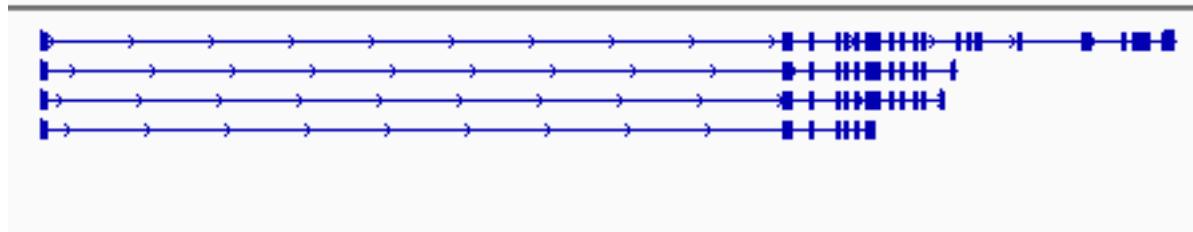
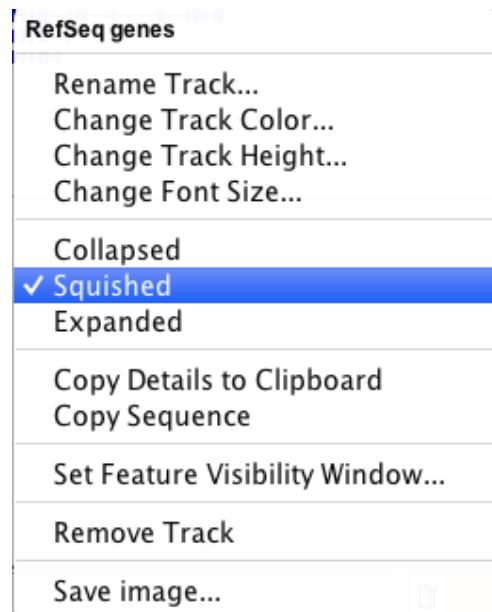




Integrative
Genomics
Viewer
IGV

Annotation display mode

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



Viewing multiple regions



Viewing multiple regions



Integrative
Genomics
Viewer
IGV



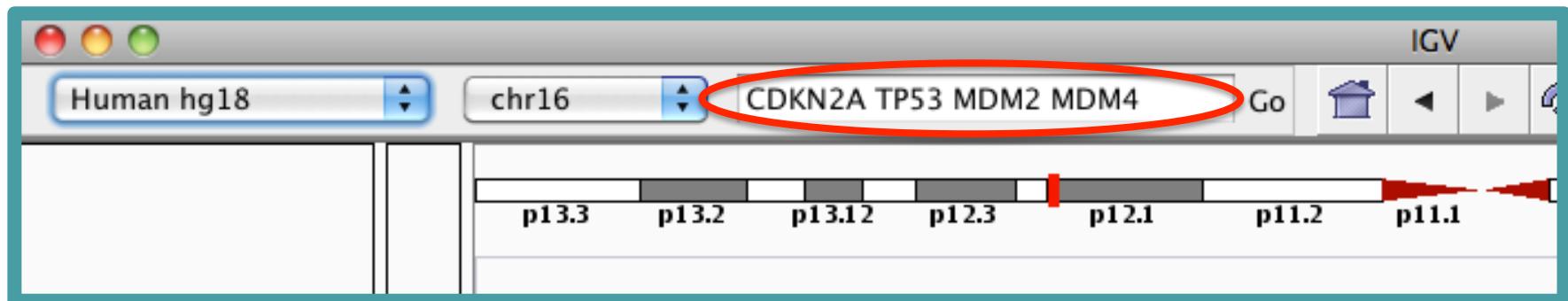
D E

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



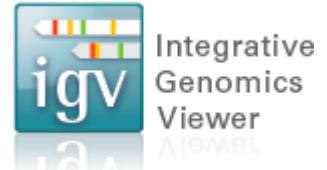
To go back to the standard, single-region view:

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



Viewing Copy Number Data

Data



Data Source:

Verhaak et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98-110. PMCID: 2818769.

Examines glioblastoma subtypes, the associated genomic lesions, and their correlation with treatment outcome.

Data types:

- copy number
- gene expression
- mutation
- sample information

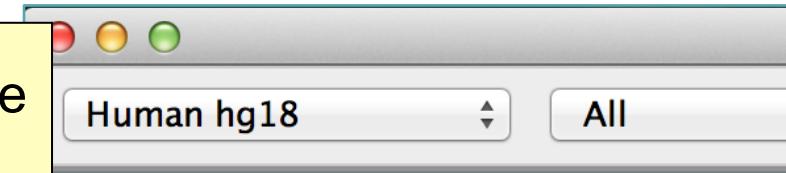
GBM Subtypes – Load data



1.

Select reference genome

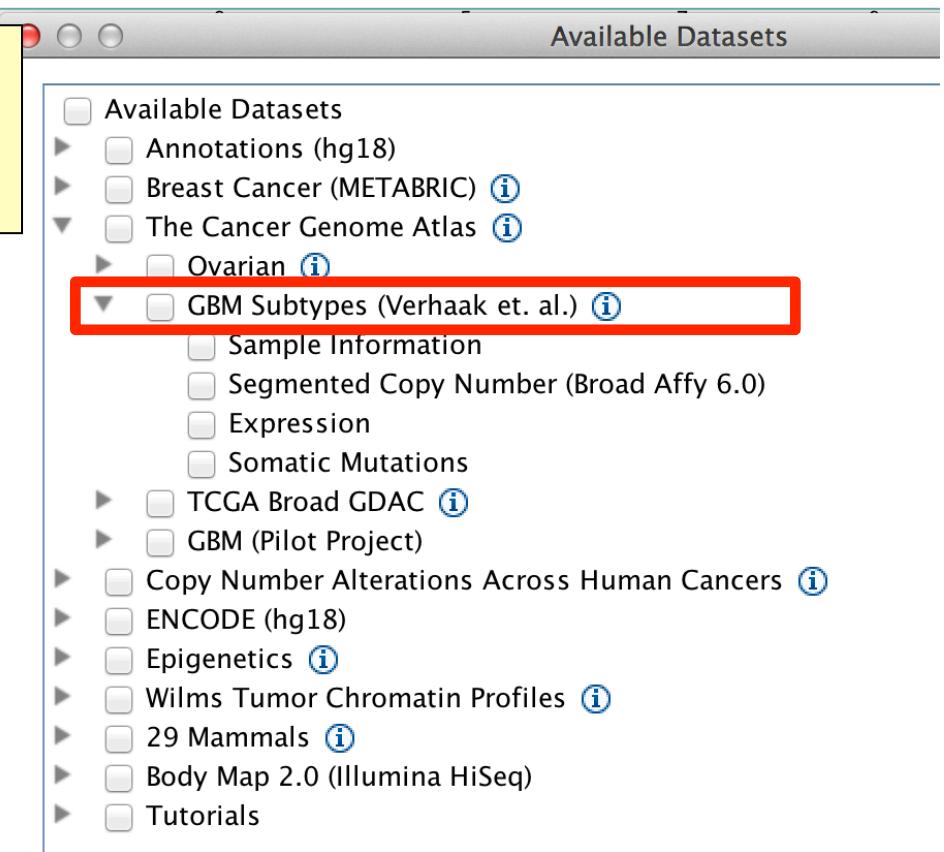
Human hg18



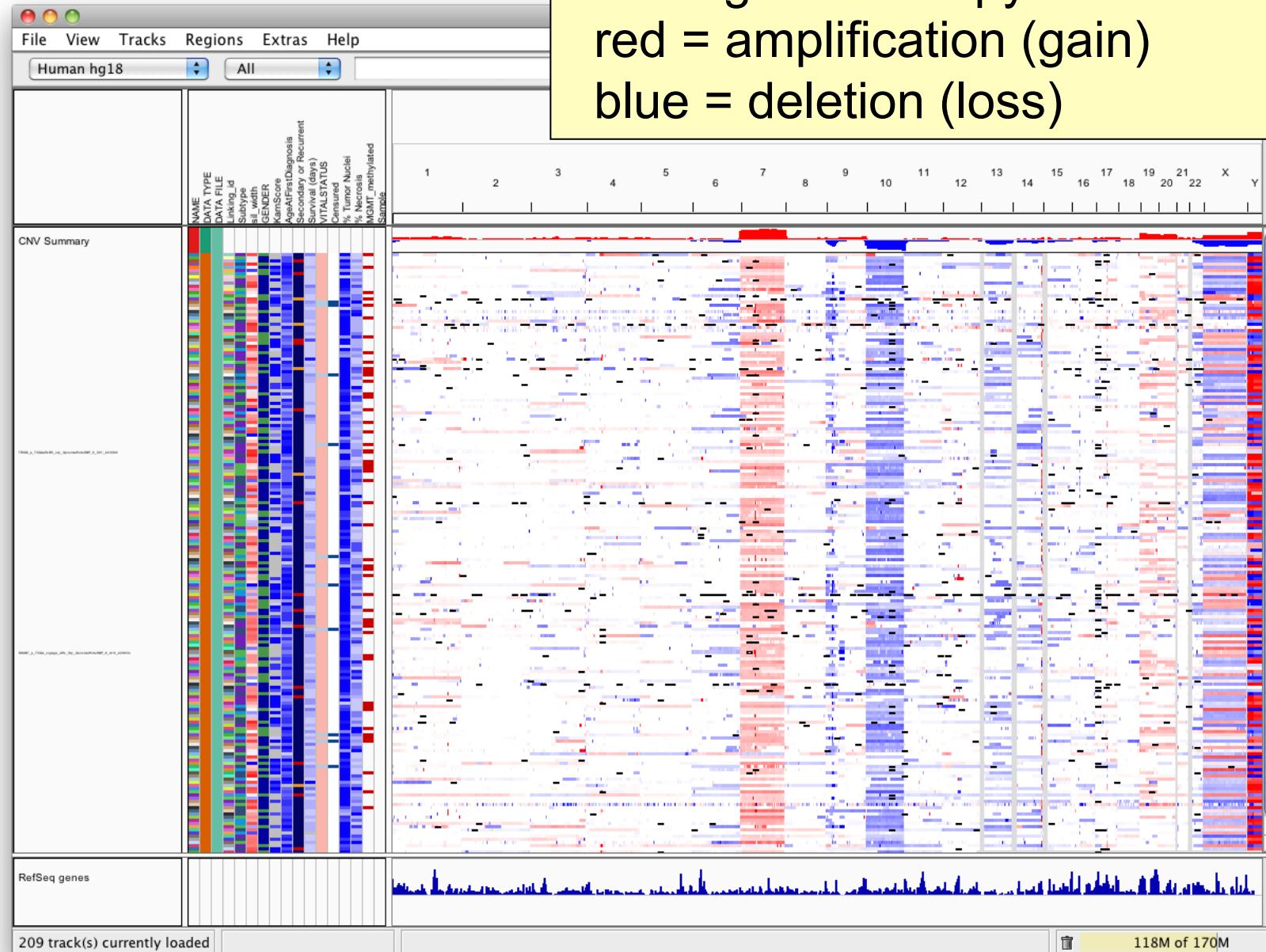
2.

File > Load File from Server

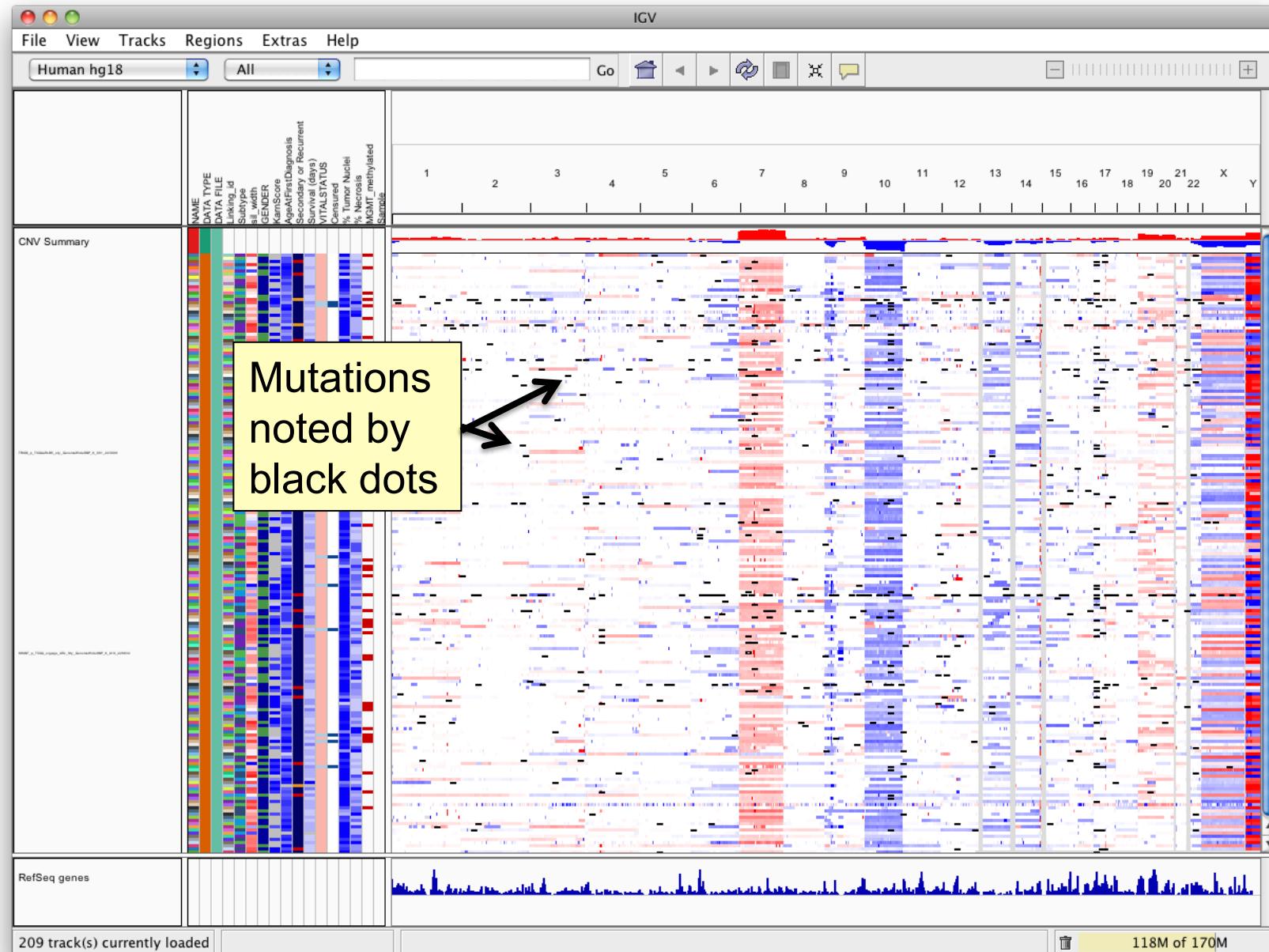
GBM Subtypes (Verhaak et. al)



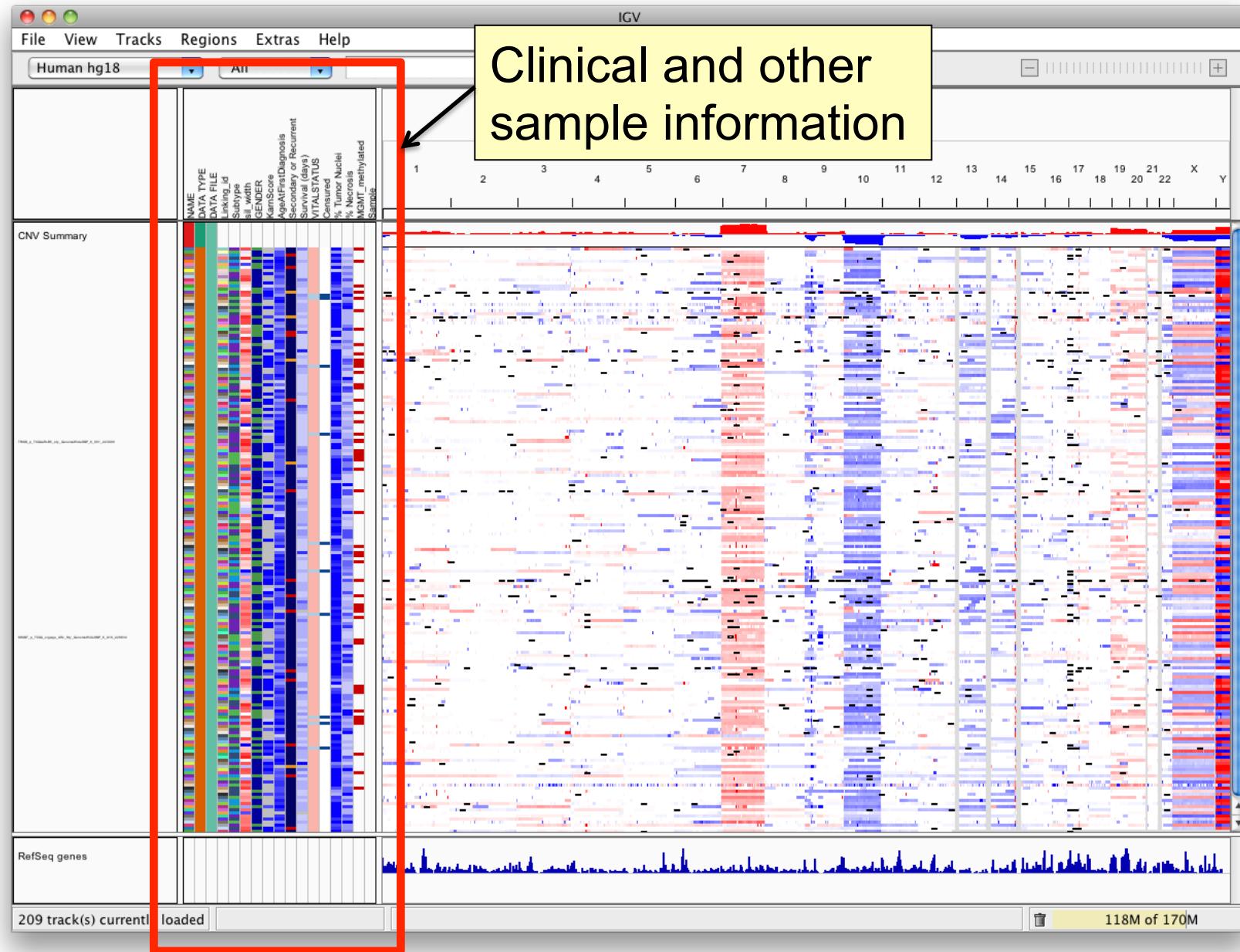
GBM Subtypes



GBM Subtypes



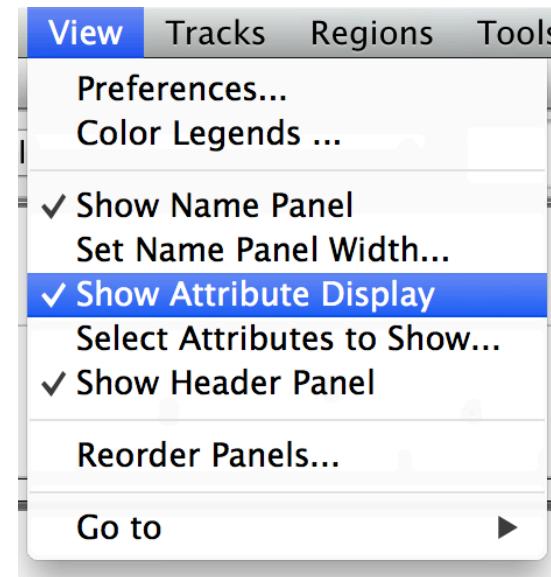
GBM Subtypes



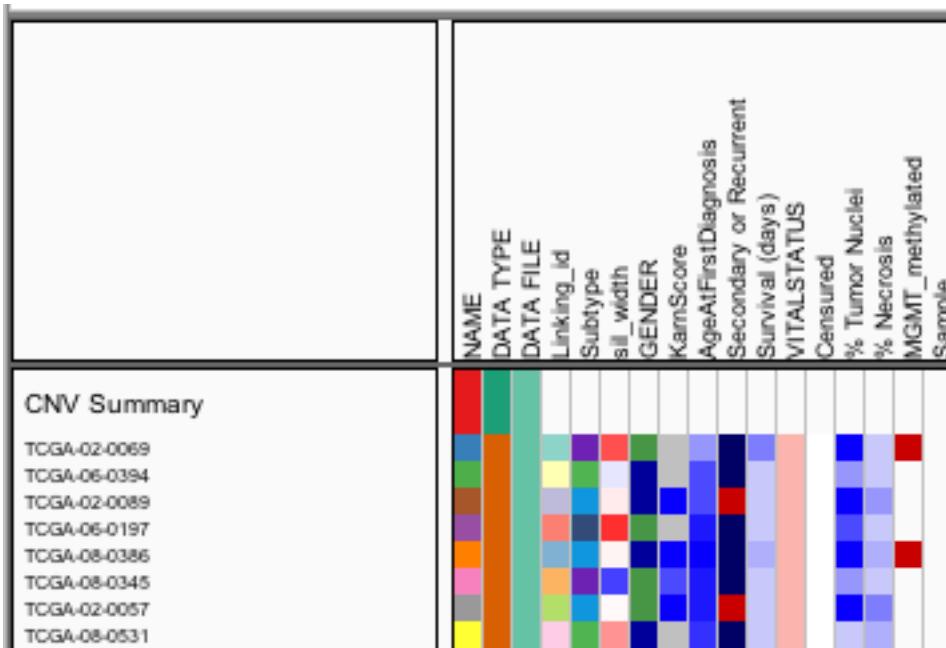
GBM Subtypes



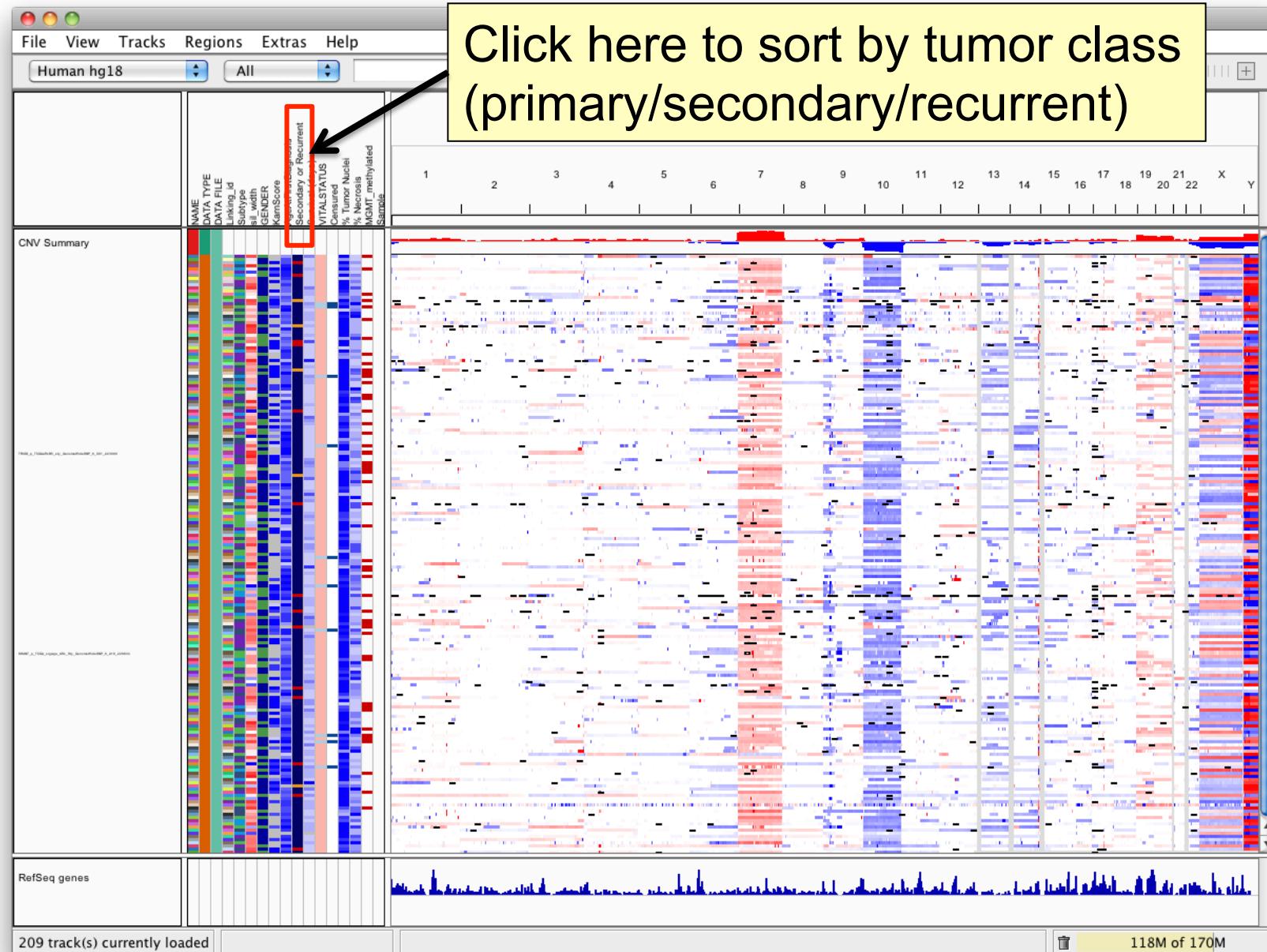
If the attribute panel is missing:
View > Show Attribute Display



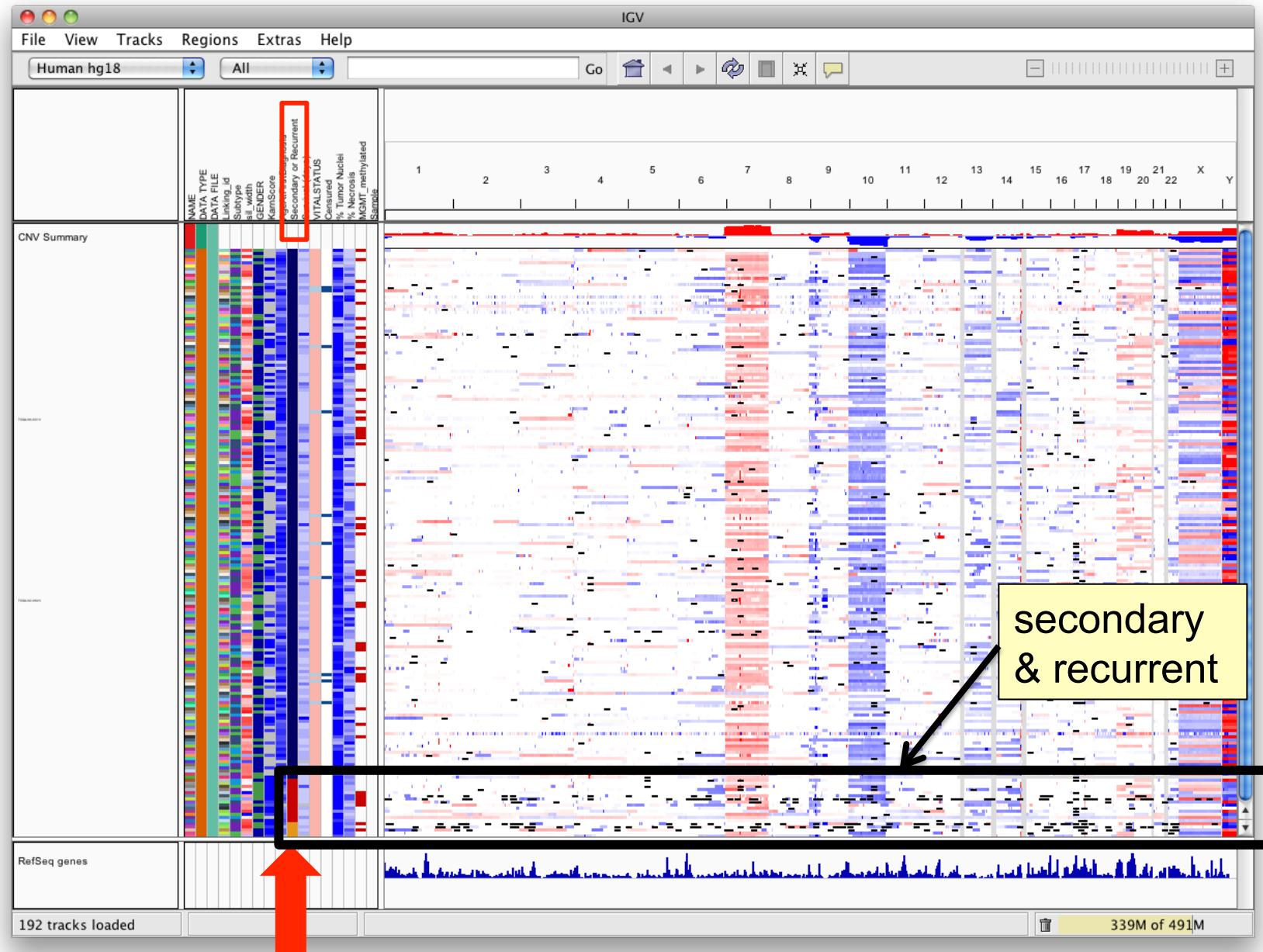
GBM Subtypes



GBM Subtypes



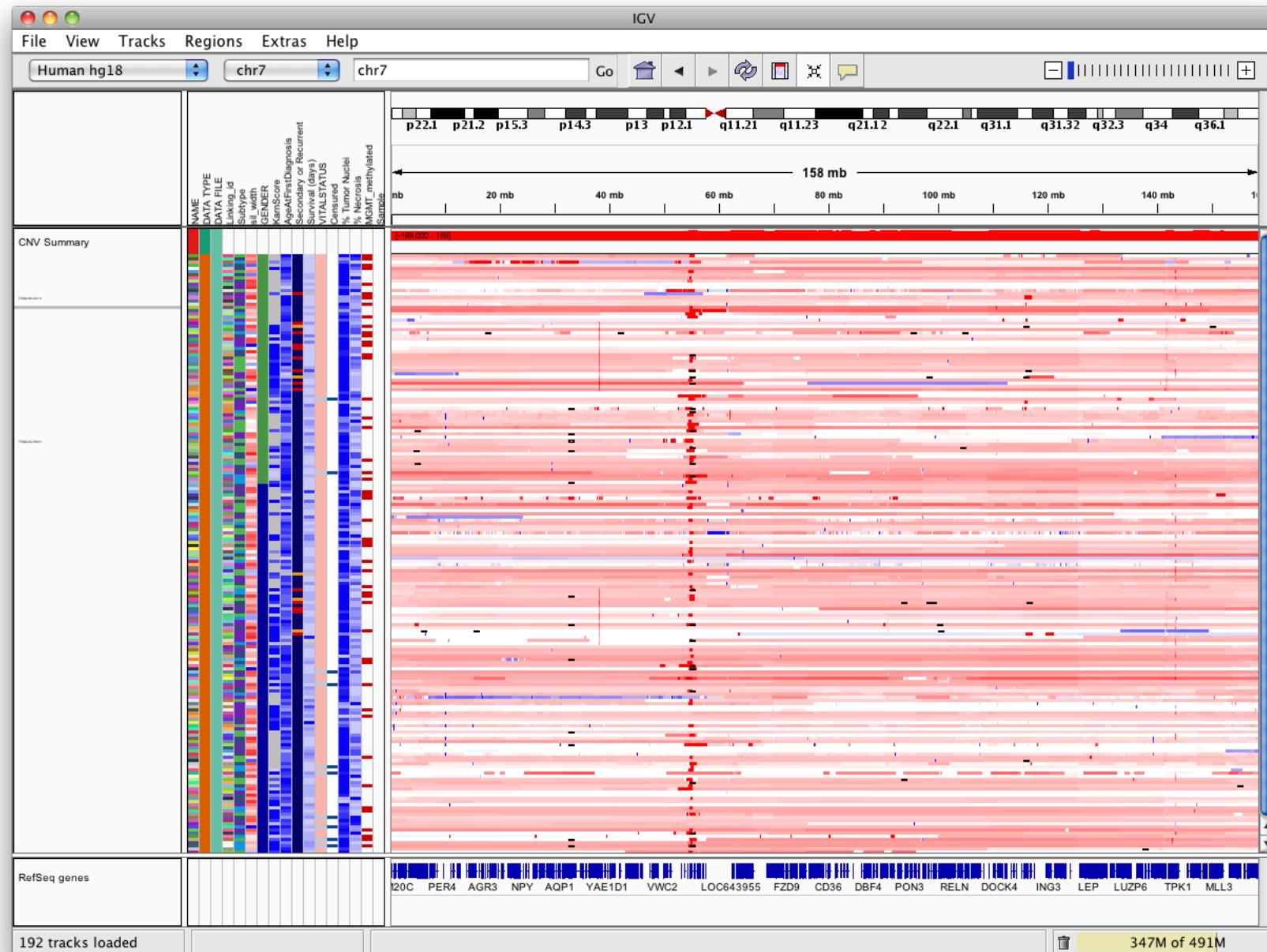
GBM Subtypes – sort by attribute



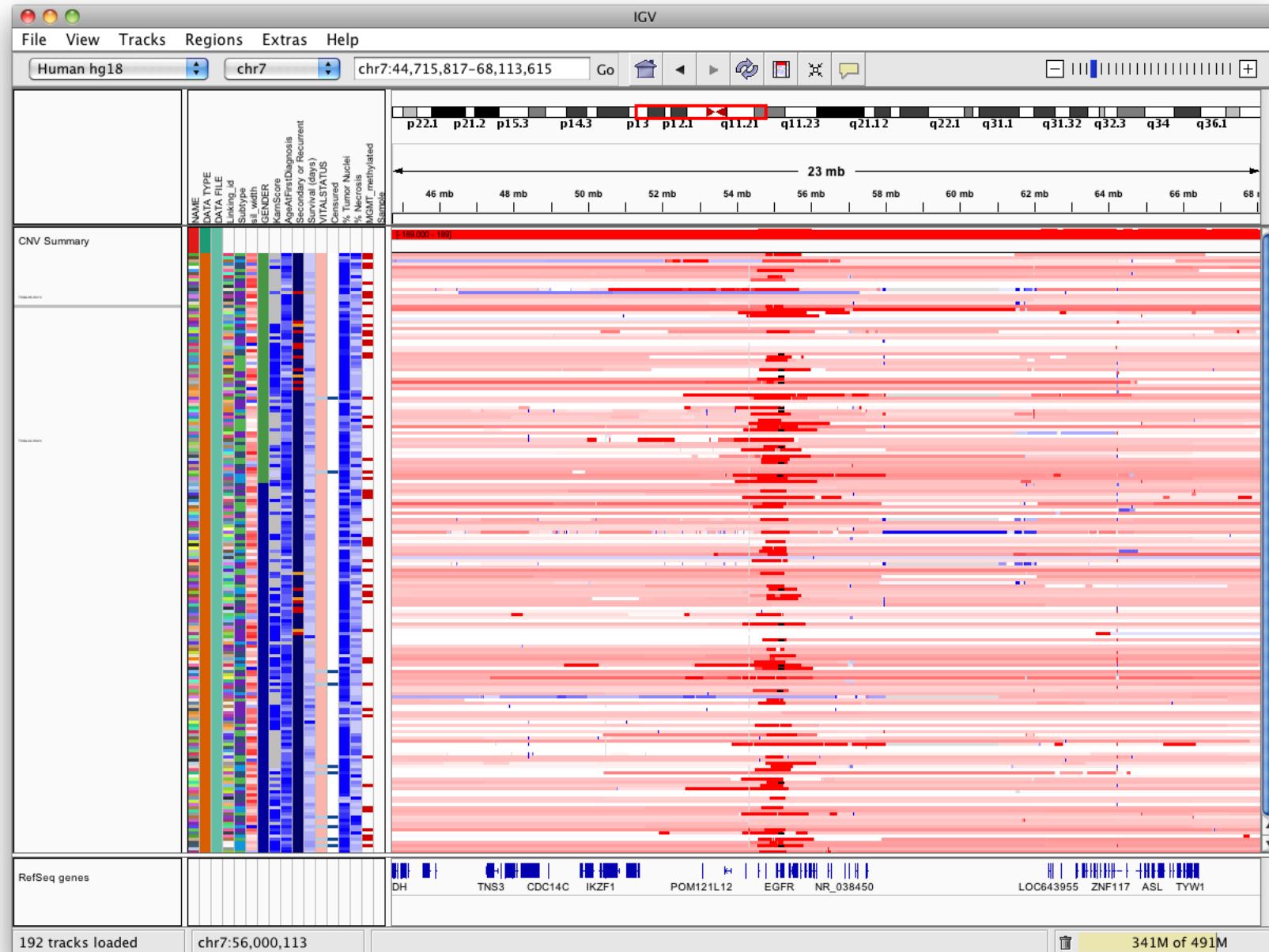
GBM Subtypes



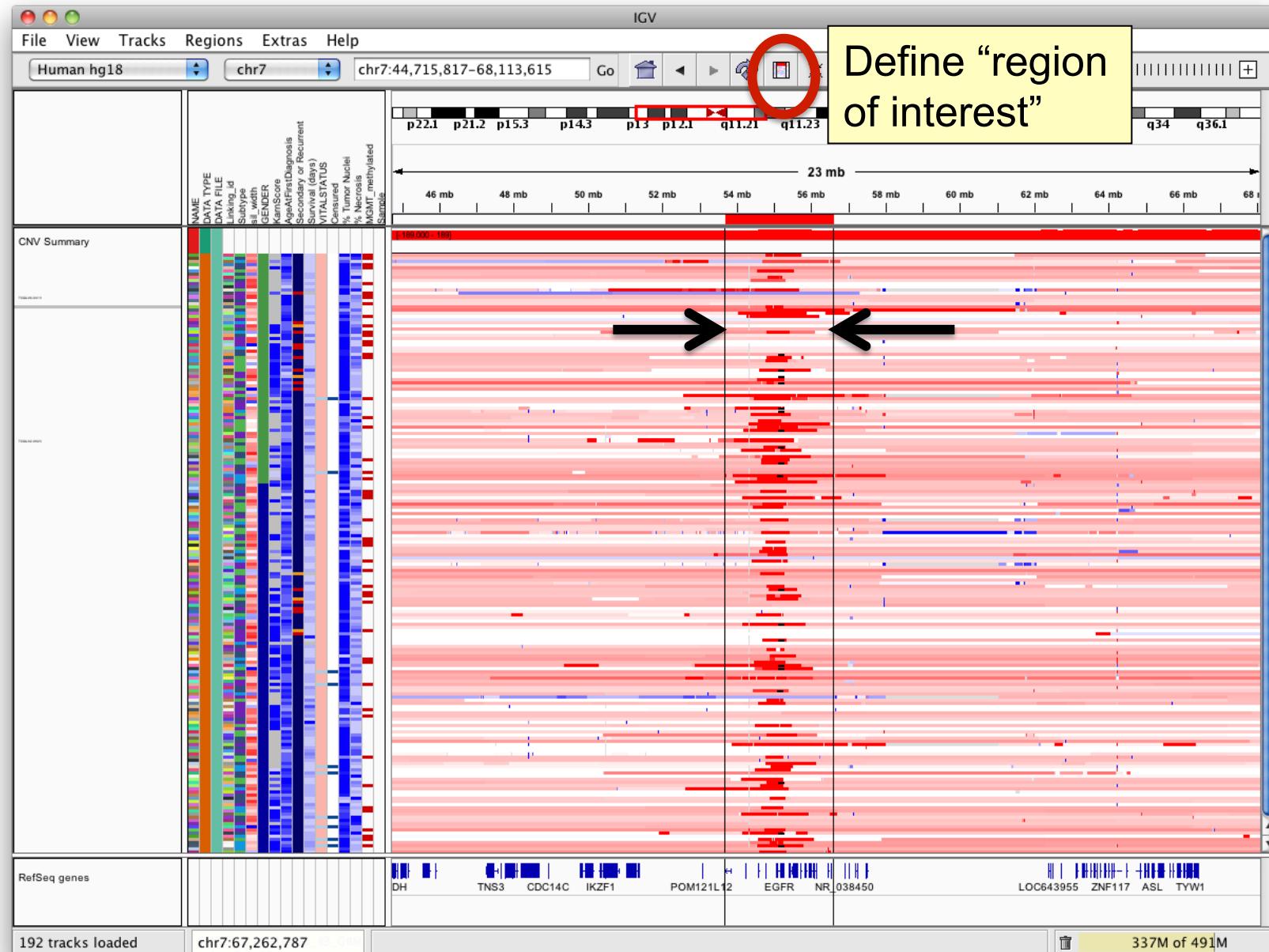
GBM Subtypes



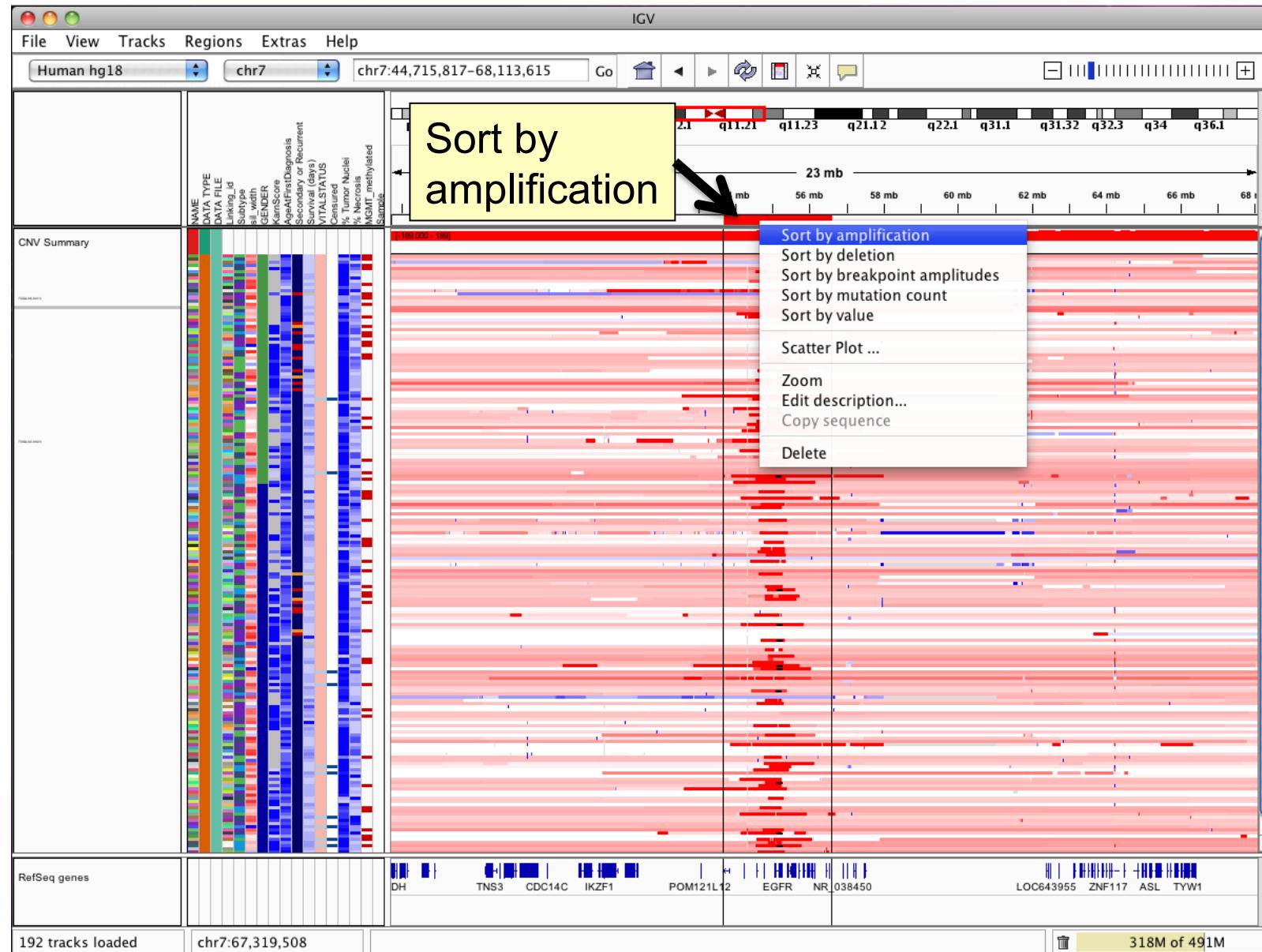
GBM Subtypes



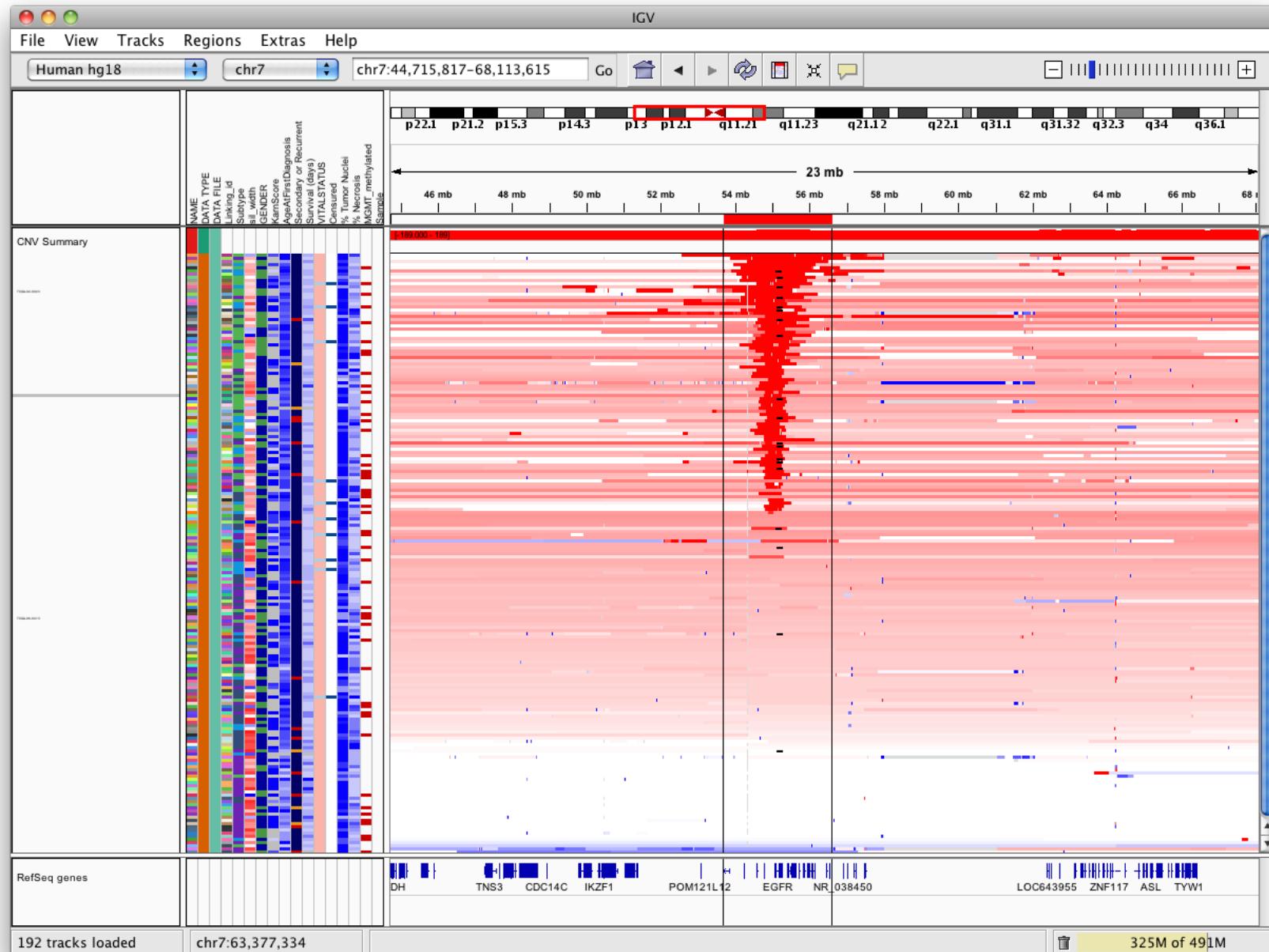
GBM Subtypes



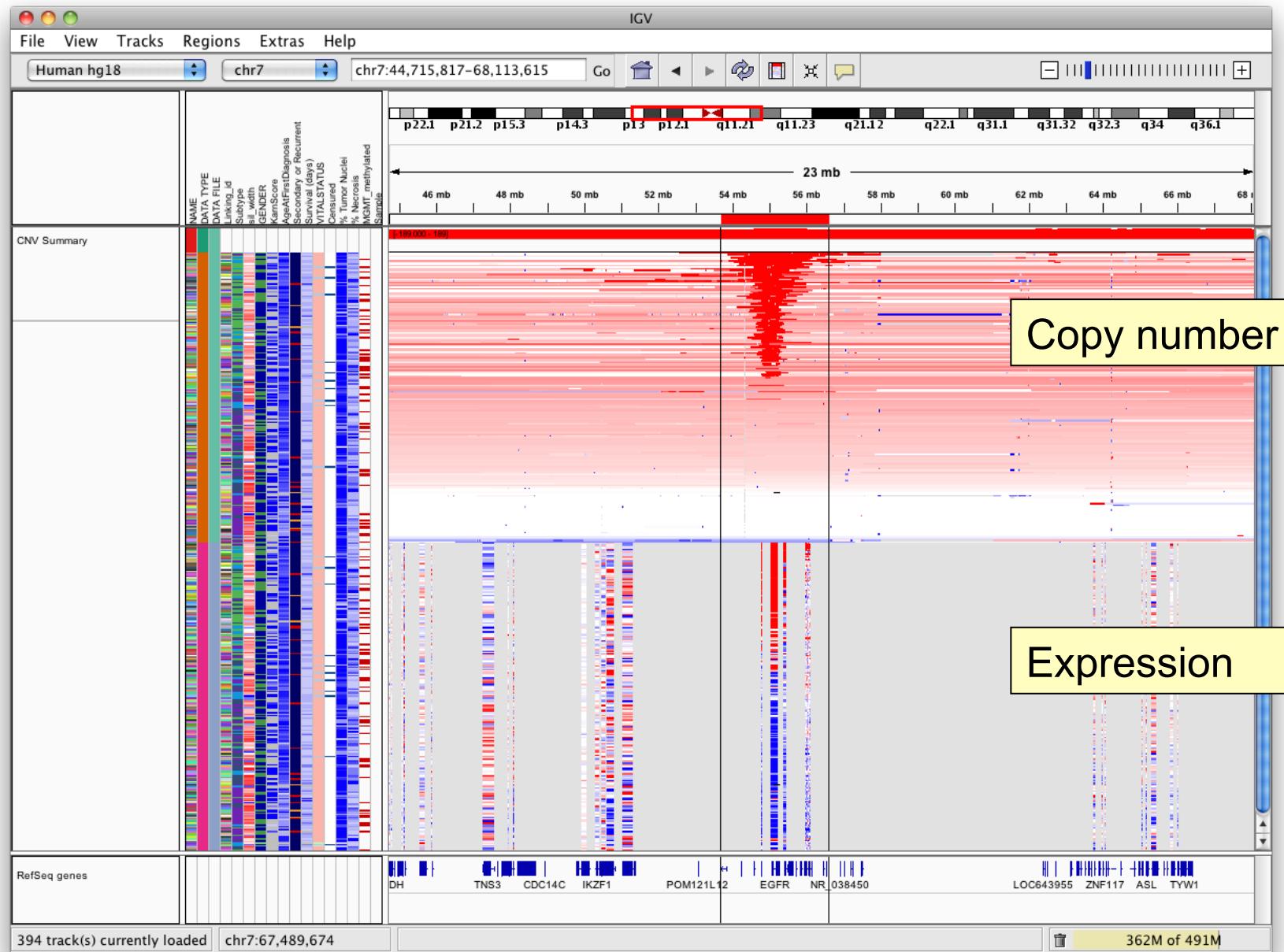
GBM Subtypes – sort by data value



GBM Subtypes – sort by data value



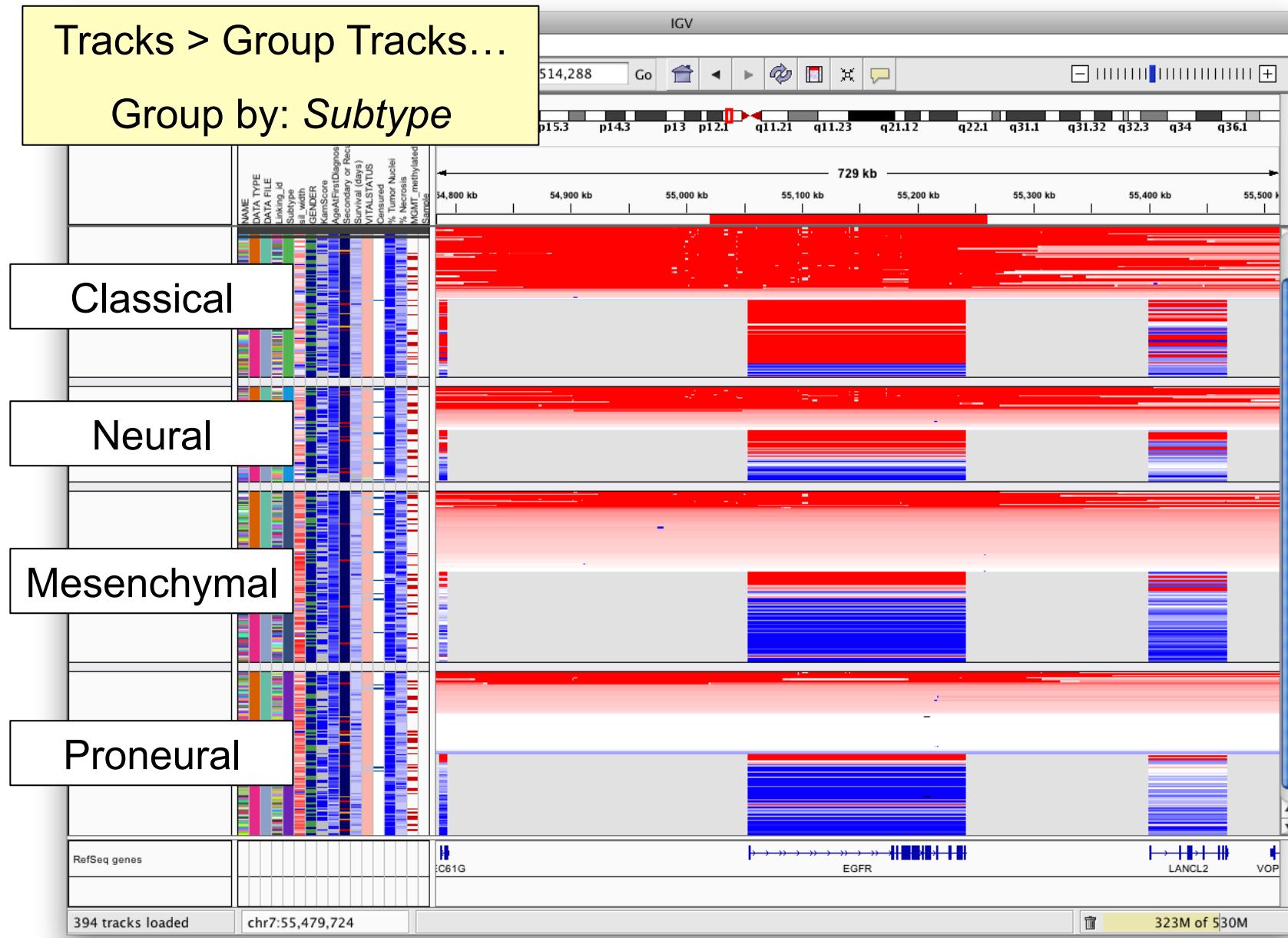
GBM Subtypes



GBM Subtypes - grouping tracks

Tracks > Group Tracks...

Group by: *Subtype*



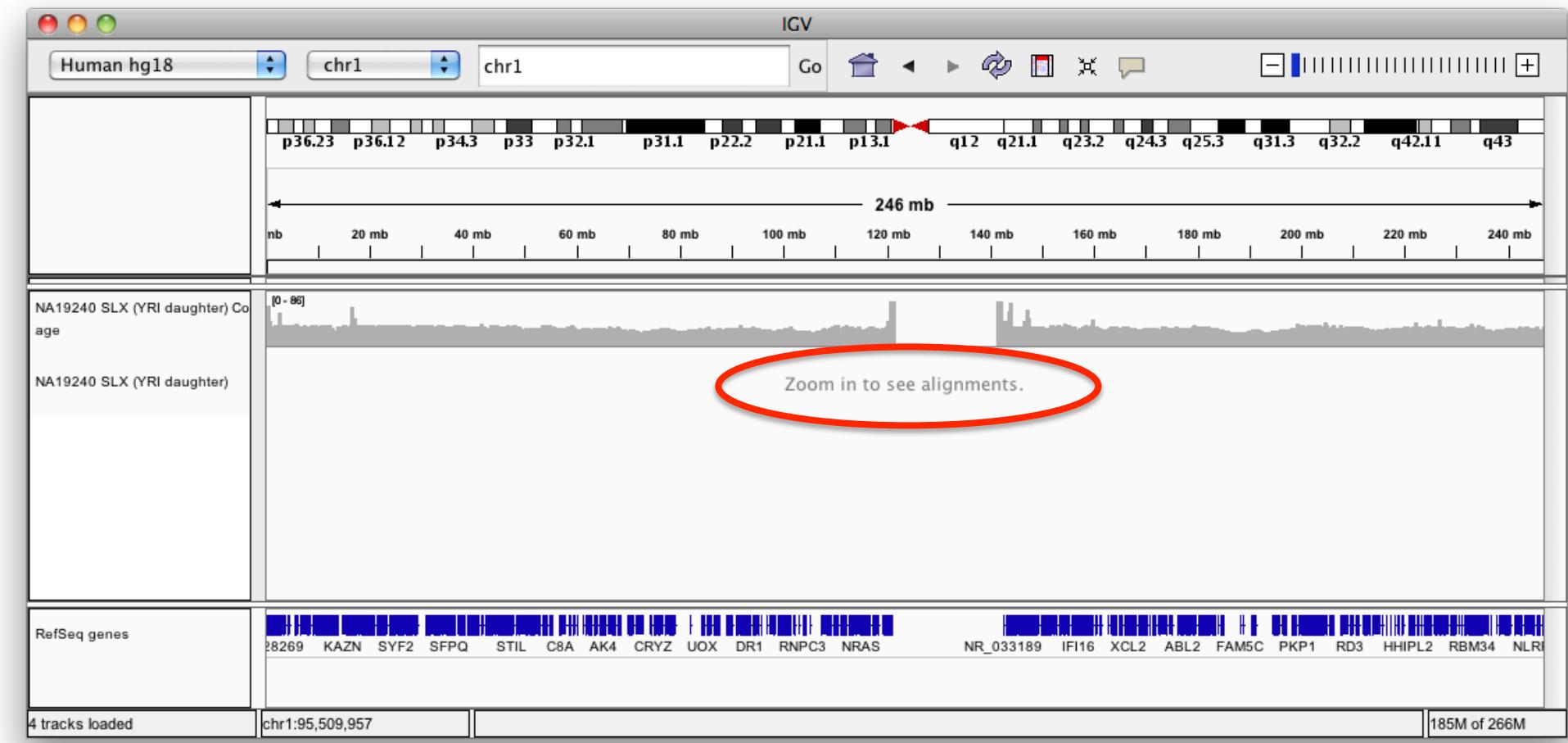
Viewing NGS Data



Integrative
Genomics
Viewer
ALIGATEL

Viewing alignments

Whole chromosome view

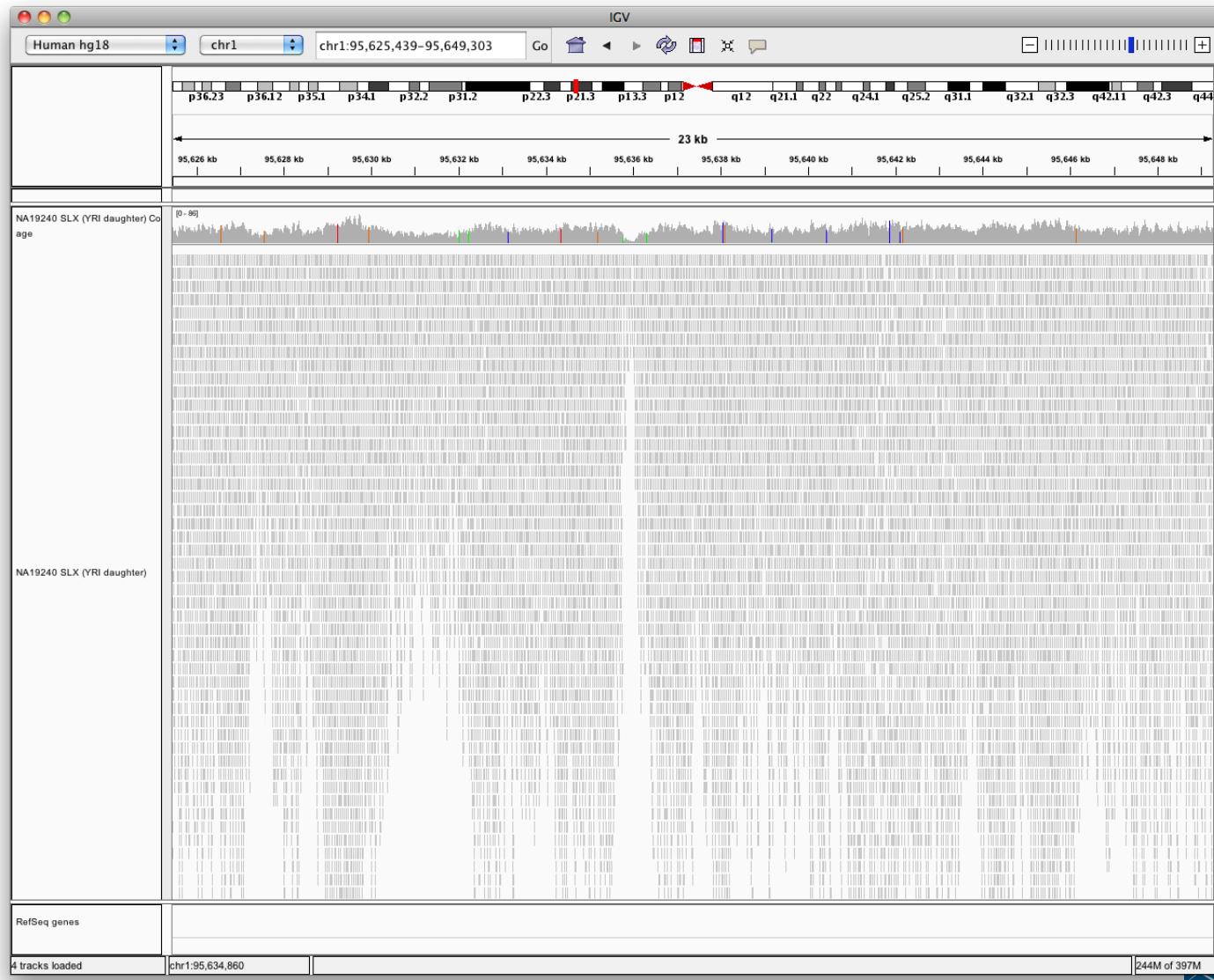




Integrative
Genomics
Viewer
ALIGMENT

Viewing alignments

Zoom in to view alignments



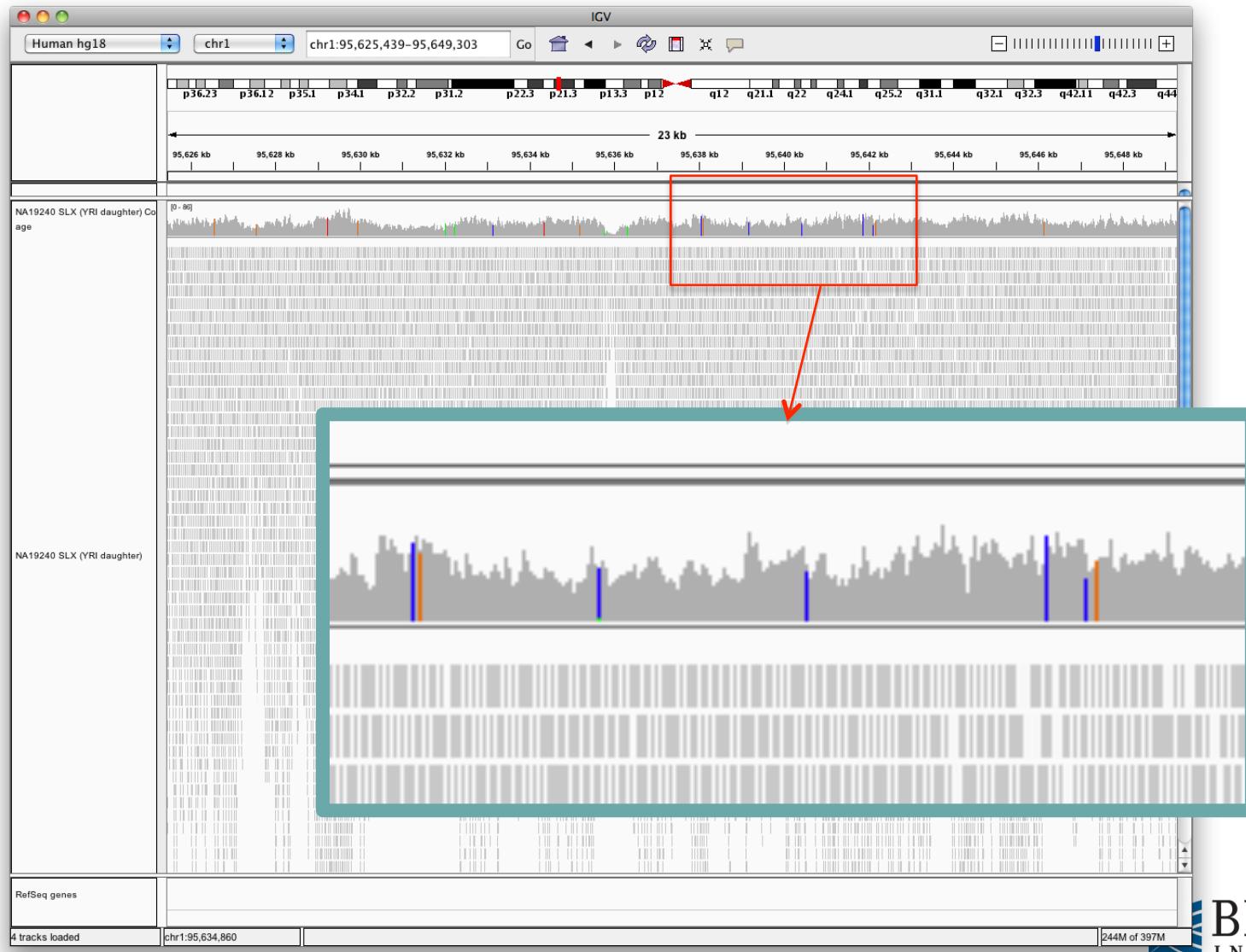
BROAD
INSTITUTE



Integrative
Genomics
Viewer
ALWIGL

Viewing alignments

Coverage track now has more detail



BROAD
INSTITUTE

Viewing alignments



Zoom in to see more detail

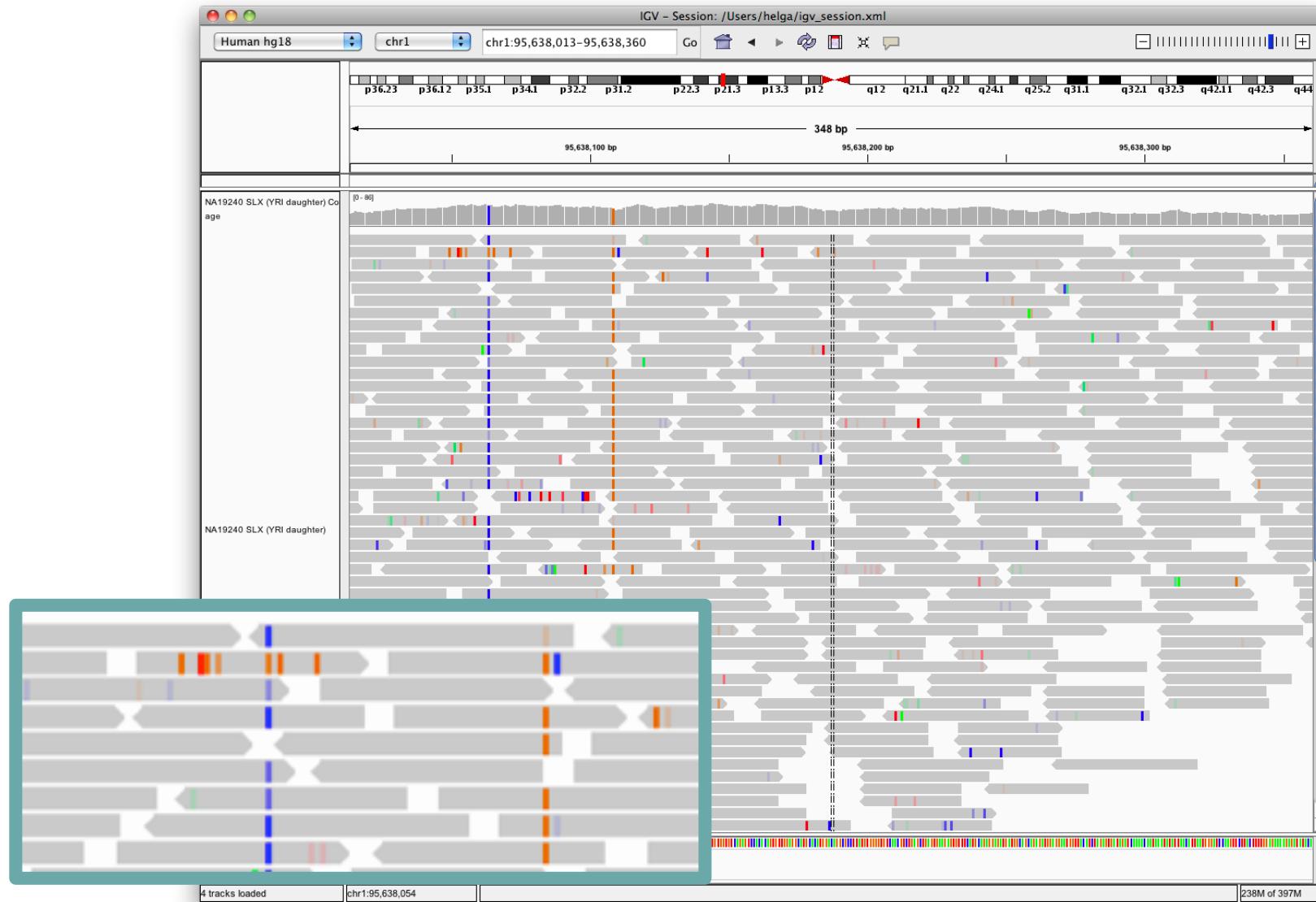


Viewing alignments



Integrative
Genomics
Viewer
ALGGEN

Zoom in to see more detail



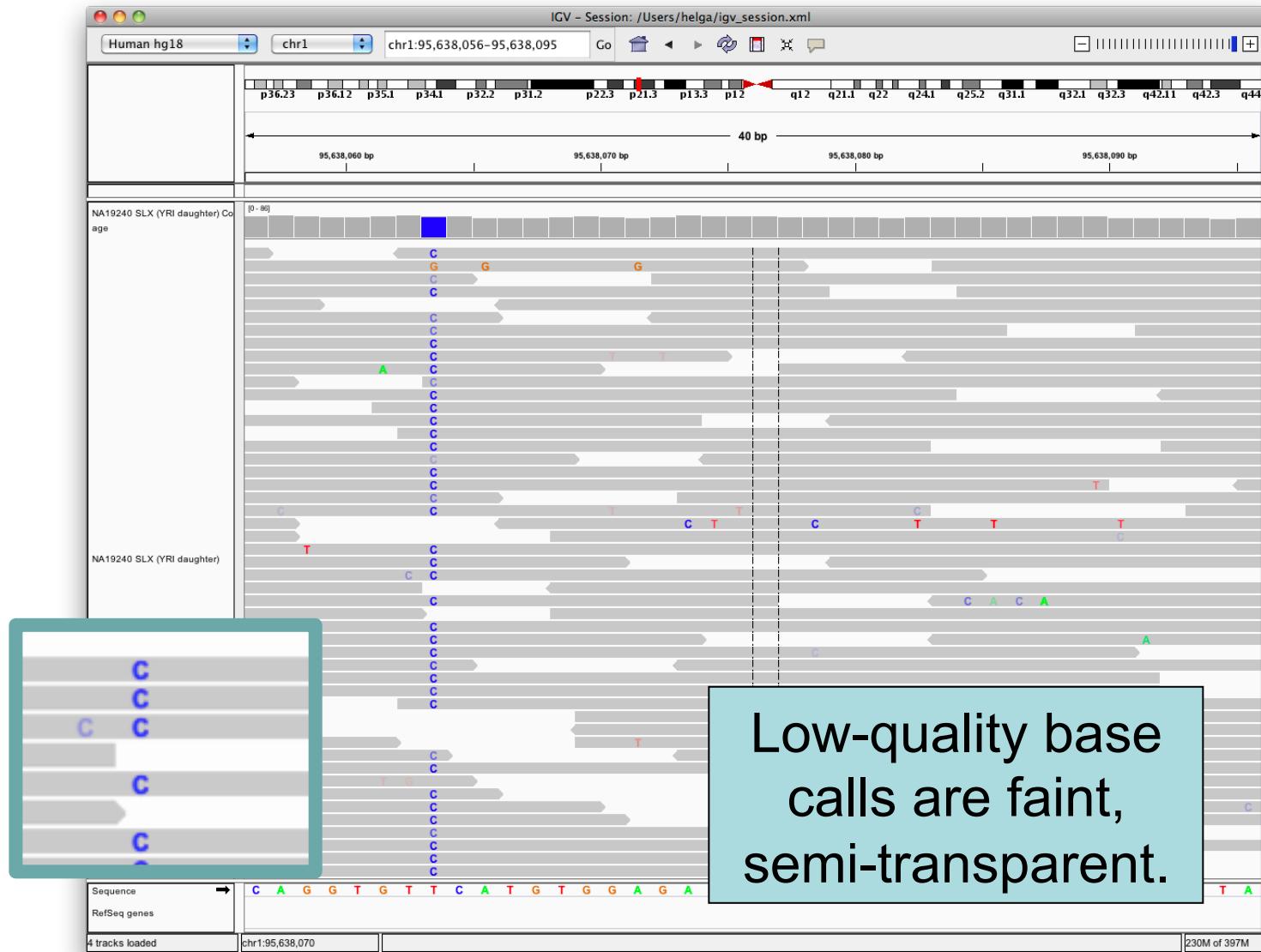
BROAD
INSTITUTE



Integrative
Genomics
Viewer
ALIVE

Viewing alignments

Zoom in to see more detail

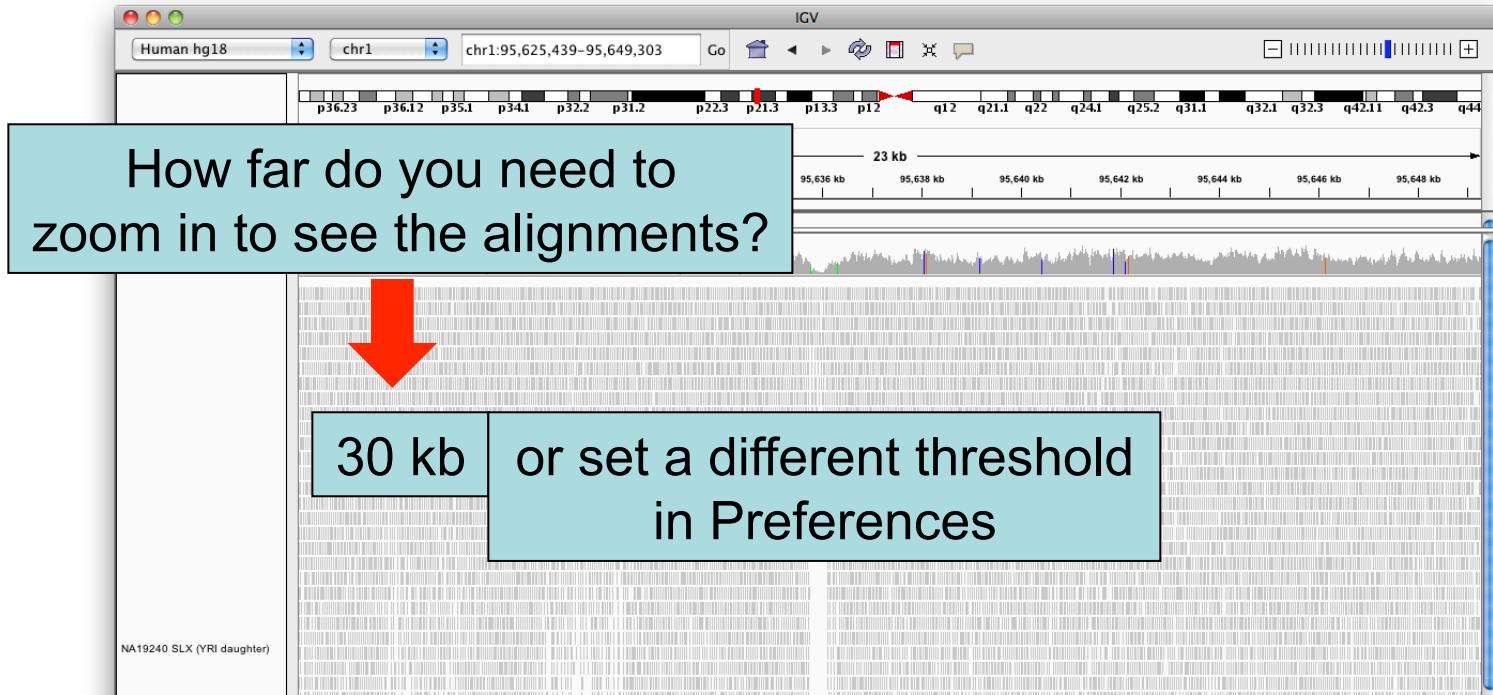




Integrative
Genomics
Viewer
ALGVIEWER

Viewing alignments

Zoom in to view alignments

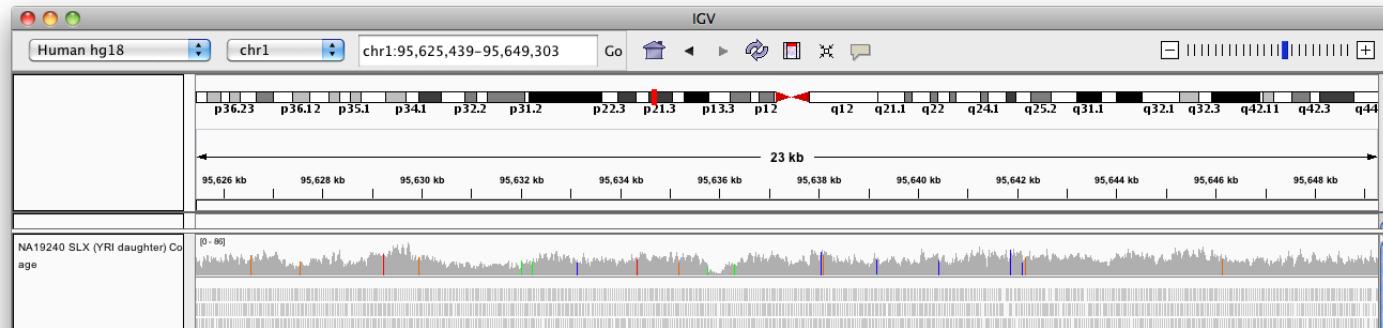


- Higher value → requires more memory
- Low coverage files → ok to use higher value
- Very deep coverage files → use lower value

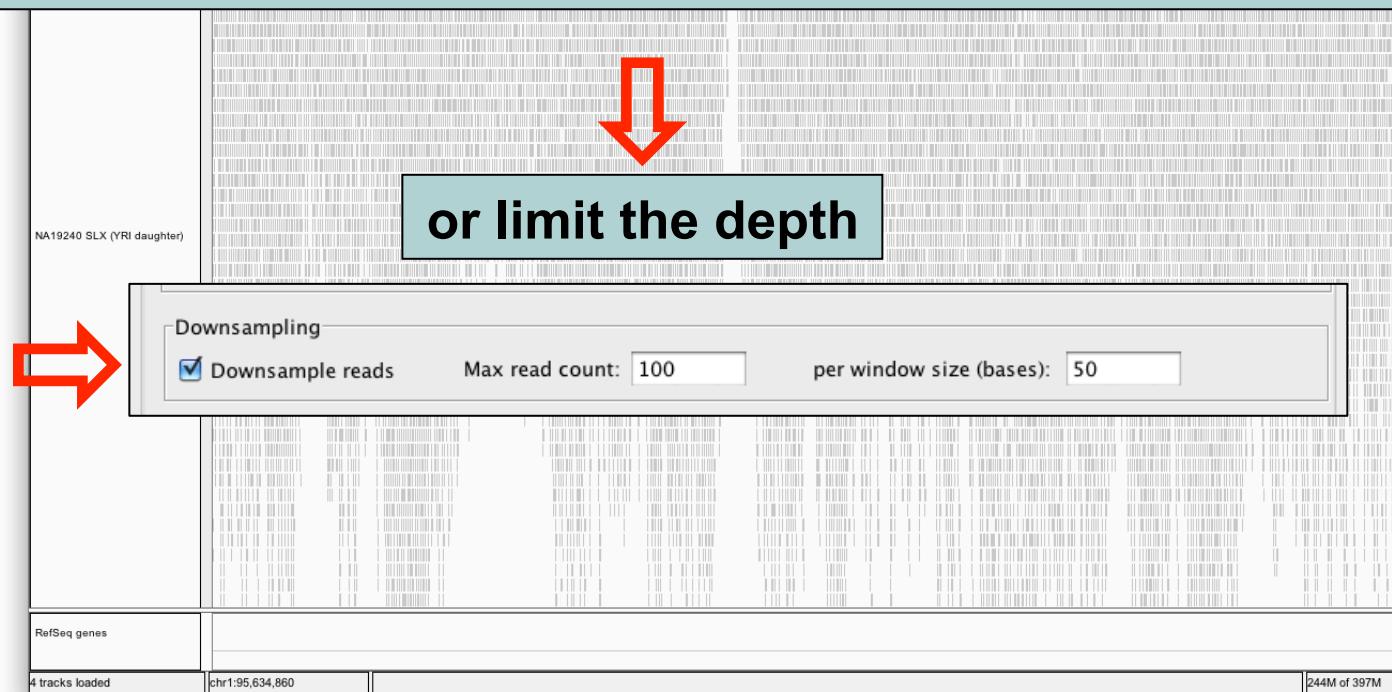
Viewing alignments



Integrative
Genomics
Viewer
ALIGMENT



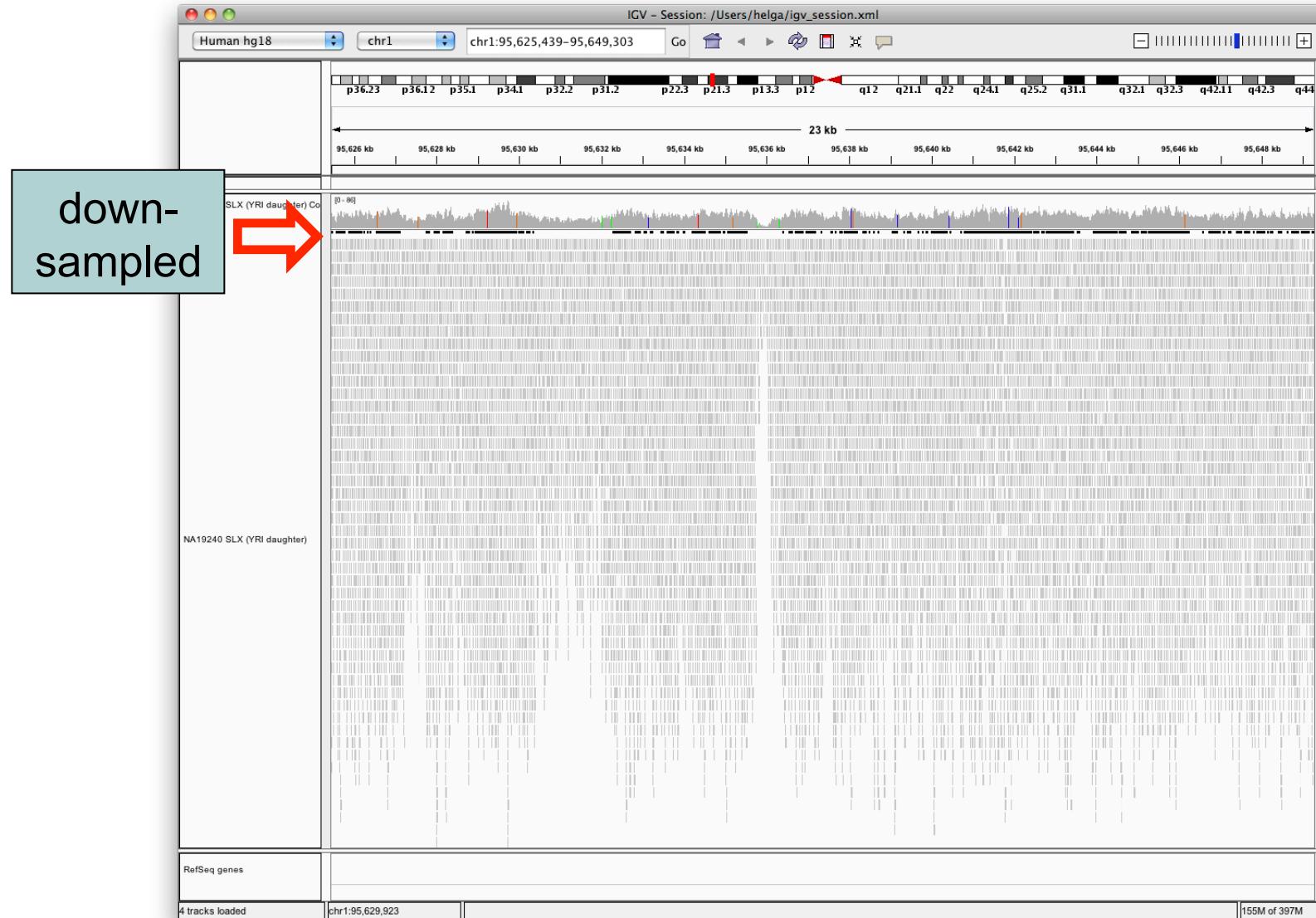
For very deep coverage files: use low visibility threshold value



Viewing alignments



Integrative
Genomics
Viewer
ALIGMENT



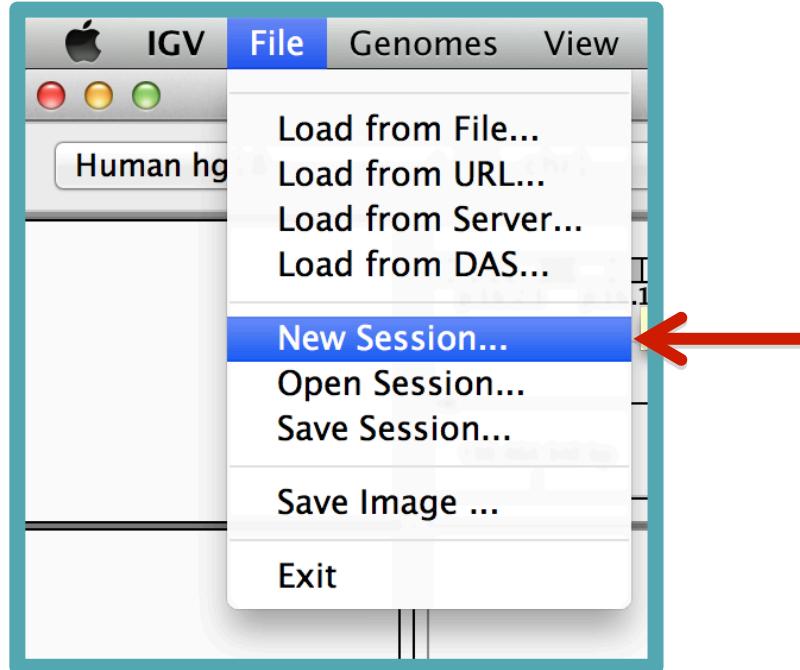
Viewing SNPs



Hands-on exercise

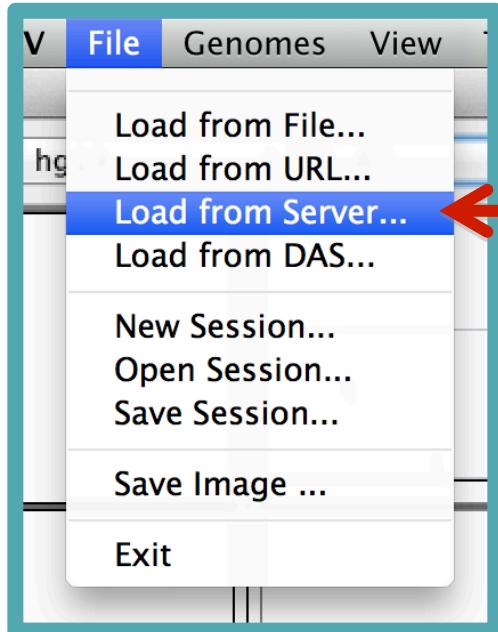
- Load alignments from whole genome sequencing
- View sites where SNPs were called
- Sort and color to highlight patterns

Viewing SNPs



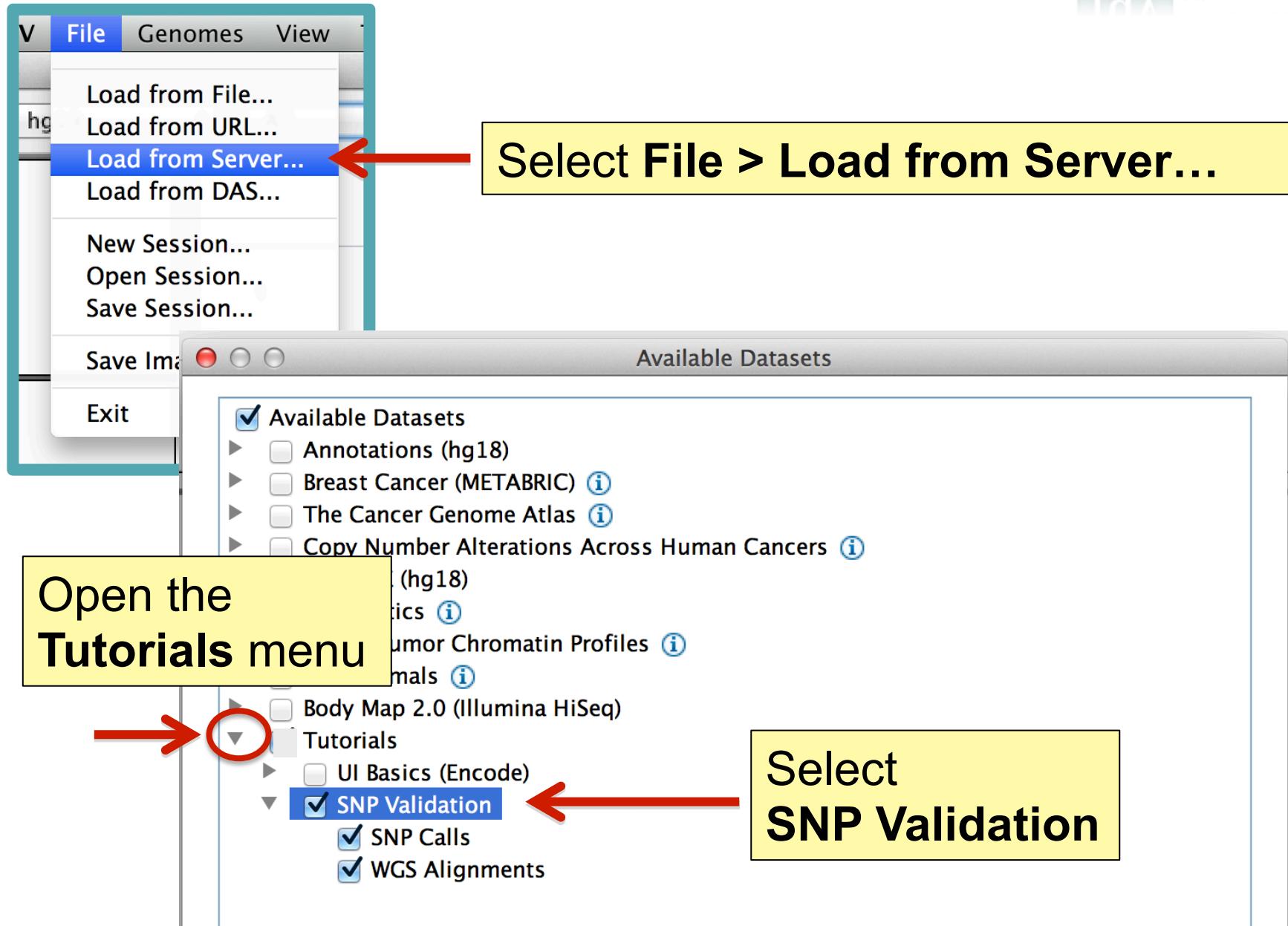
Before we start:
Select File > New Session
to clear IGV window

Viewing SNPs

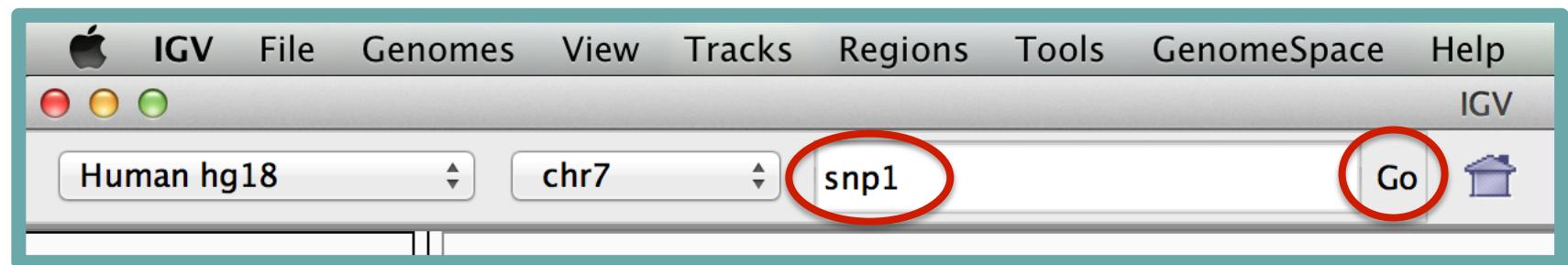


Select **File > Load from Server...**

Viewing SNPs



Viewing SNPs

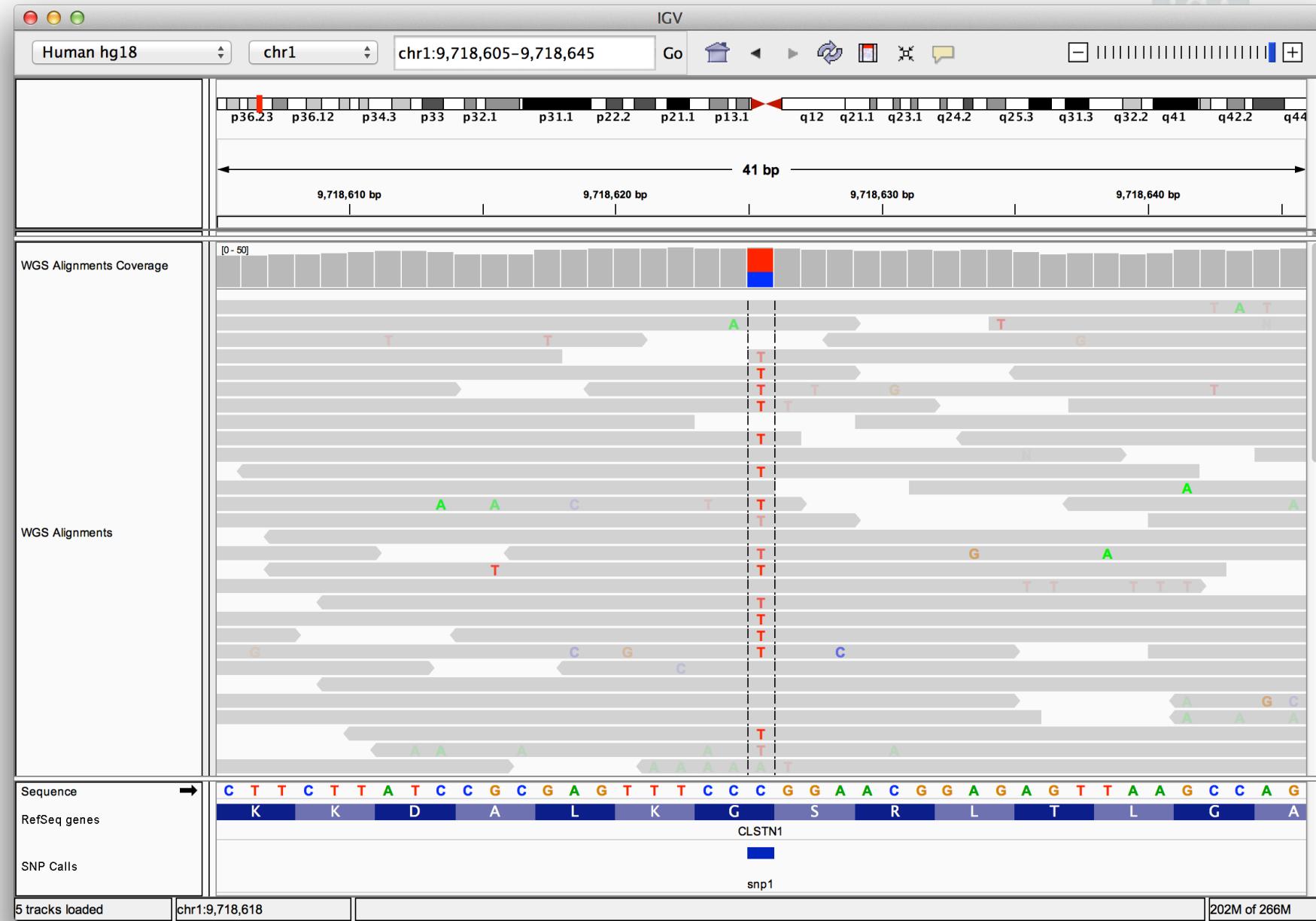


Type “snp1” in the **Search Box**
and click **Go**



Integrative
Genomics
Viewer

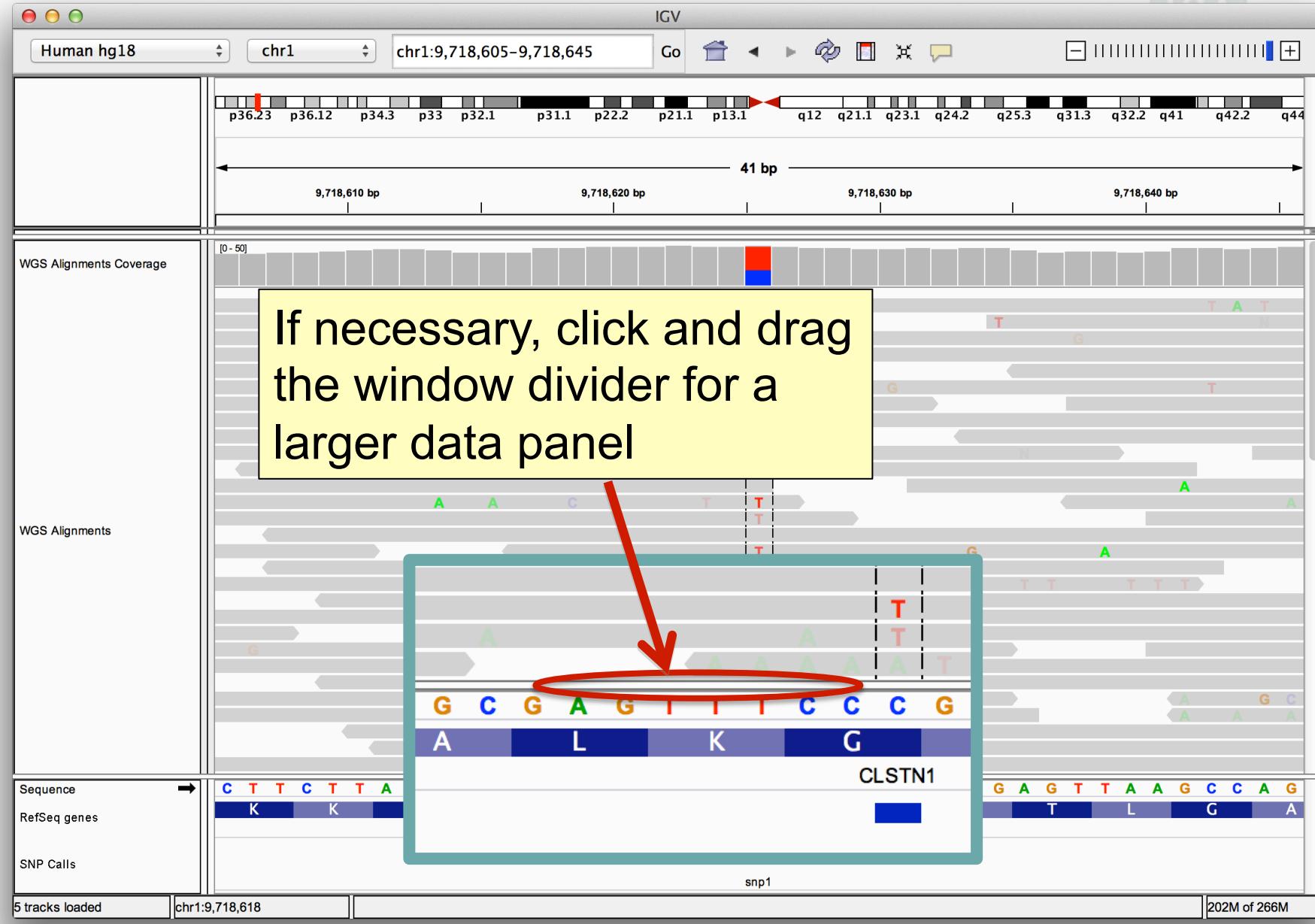
Viewing SNPs





Integrative
Genomics
Viewer

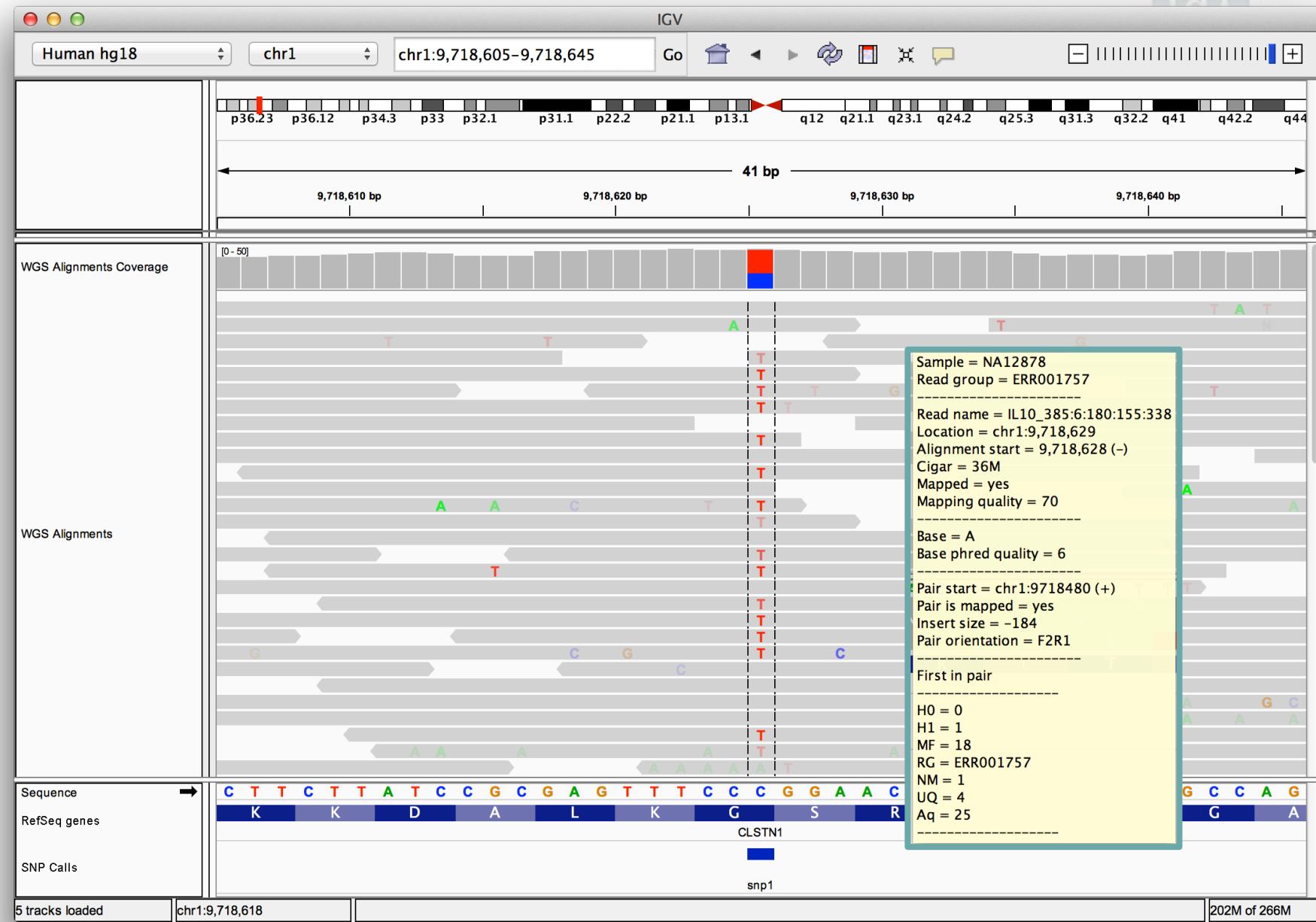
Viewing SNPs





Integrative
Genomics
Viewer

Viewing SNPs



Viewing SNPs



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

WGS Alignments

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

Sequence →

RefSeq genes

SNP Calls

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

IGV

WGS Alignments

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

CLSTN1

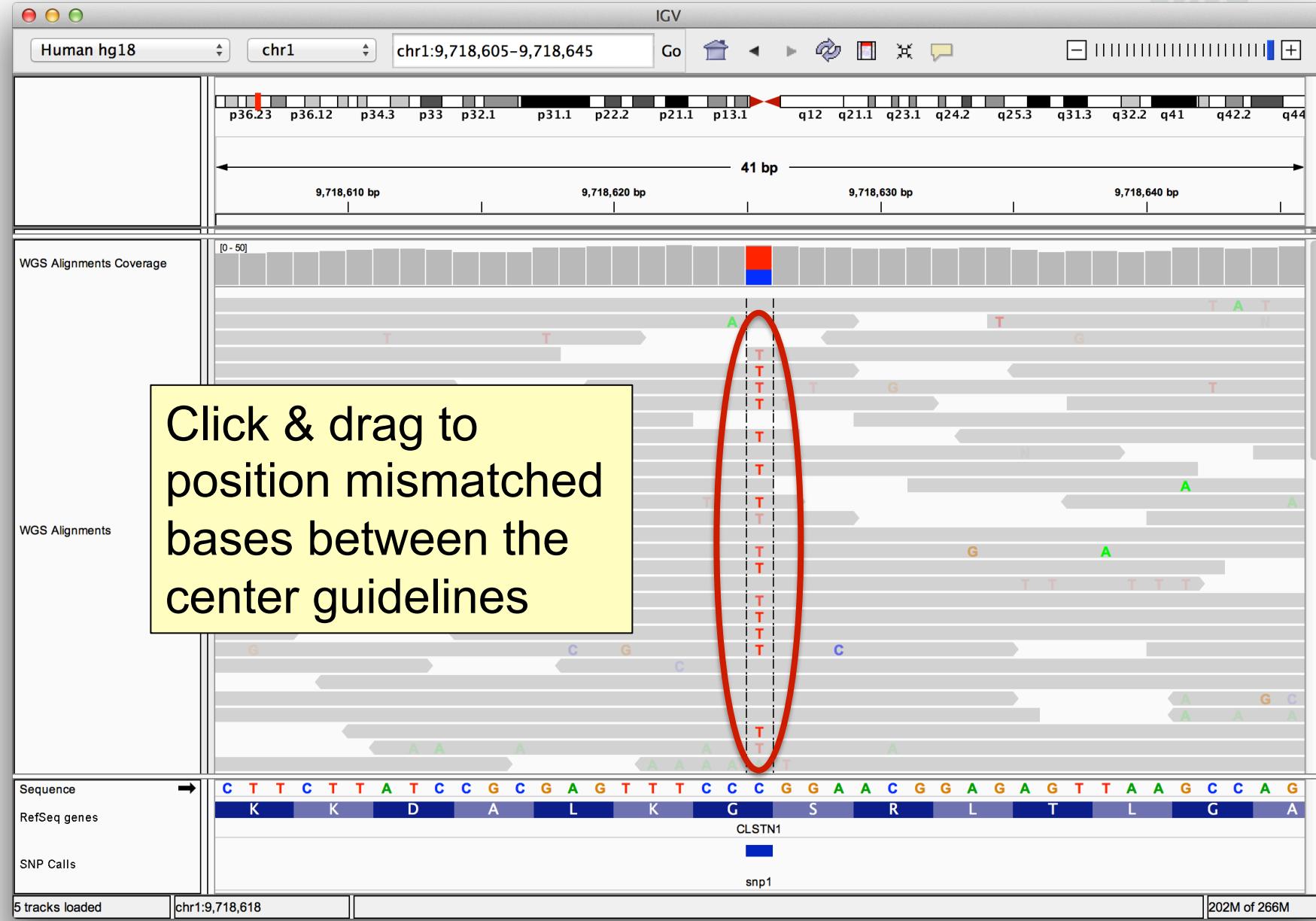
snp1

D E



Integrative
Genomics
Viewer

Viewing SNPs





Integrative
Genomics
Viewer

Viewing SNPs

IGV

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

WGS Alignments Coverage

WGS

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

On Mac:
Right-click = Control-click

Sequence RefSeq genes SNP Calls

CLSTN1

snp1

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

WGS Alignments

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

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

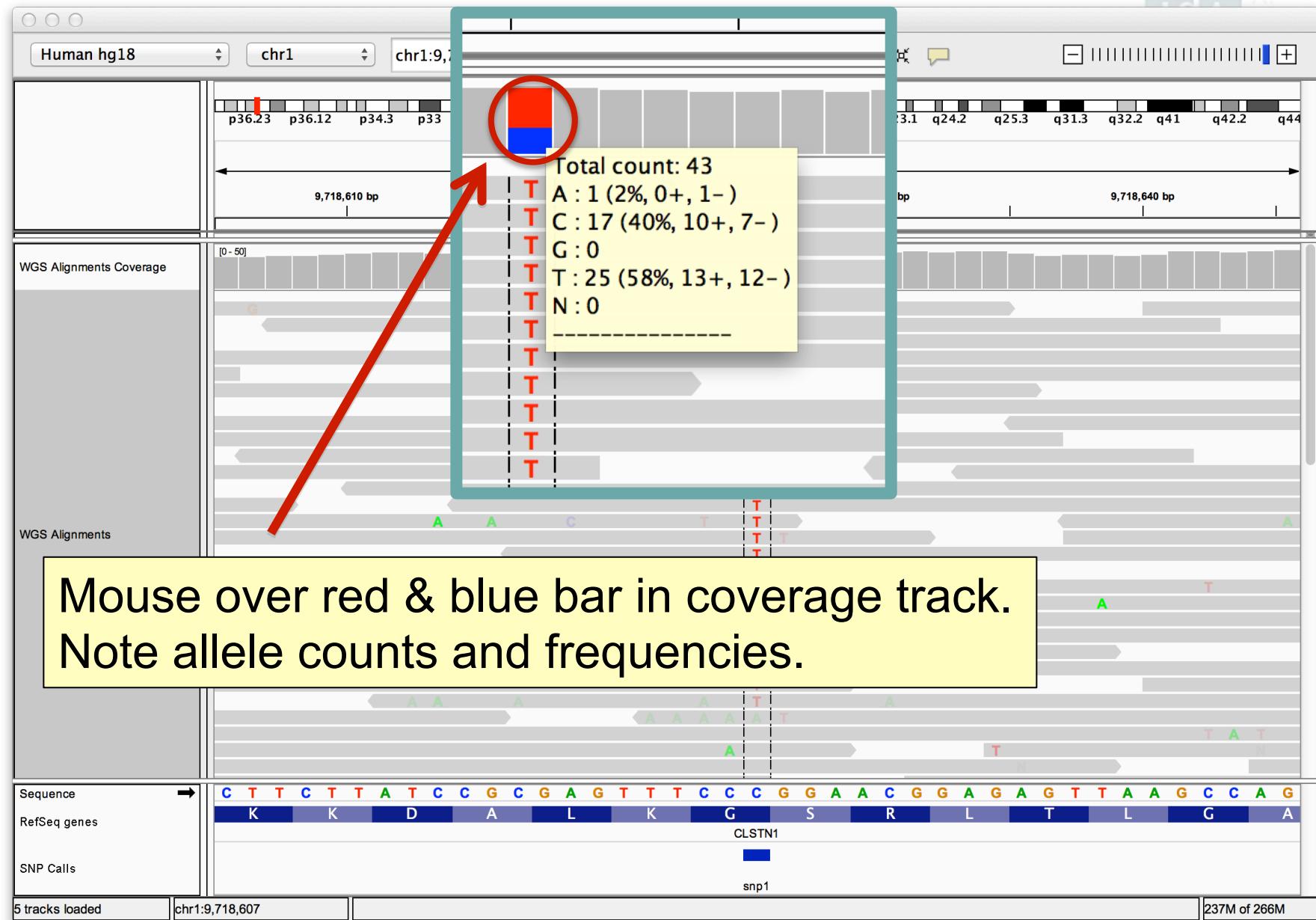


Integrative
Genomics
Viewer

Viewing SNPs



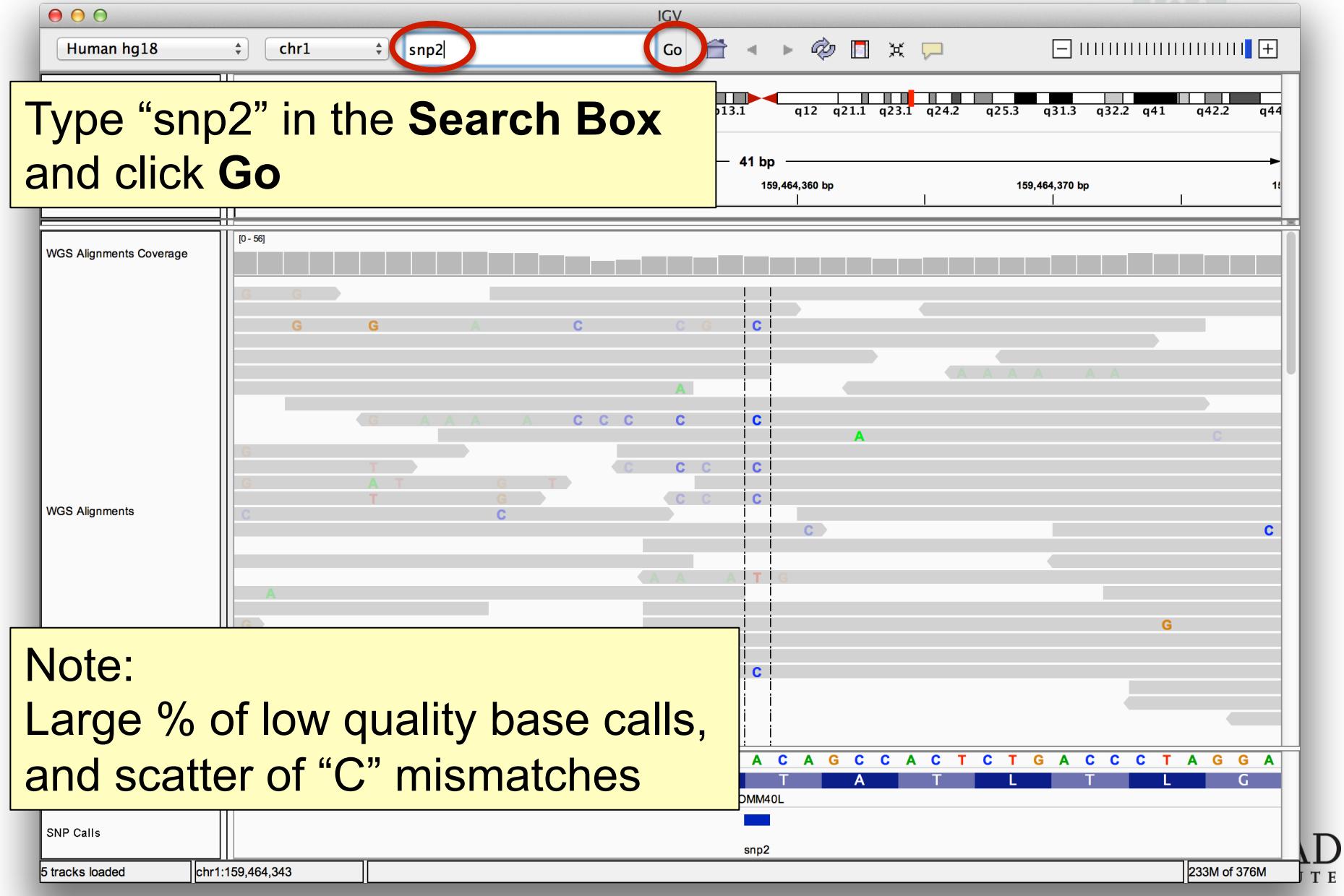
Viewing SNPs





Integrative
Genomics
Viewer

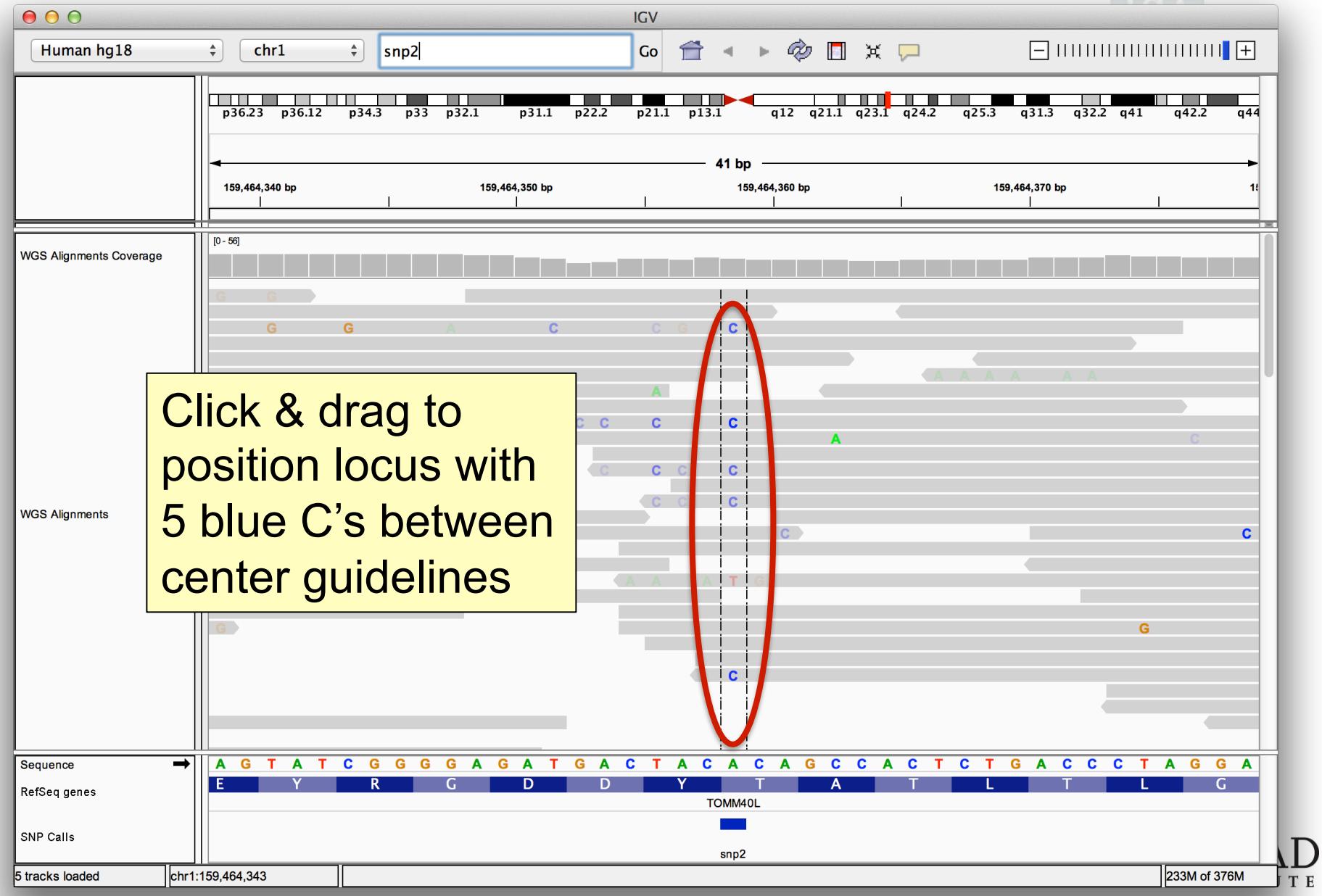
Viewing SNPs





Integrative
Genomics
Viewer

Viewing SNPs



Viewing SNPs



IGV

Human hg18 chr1 snp2 Go

WGS Alignments Coverage

WGS Alignments

Sequence RefSeq genes SNP Calls

5 tracks loaded chr1:159,464,343 snp2 233M of 376M

Right-click on alignments and select Shade base by quality

WGS Alignments

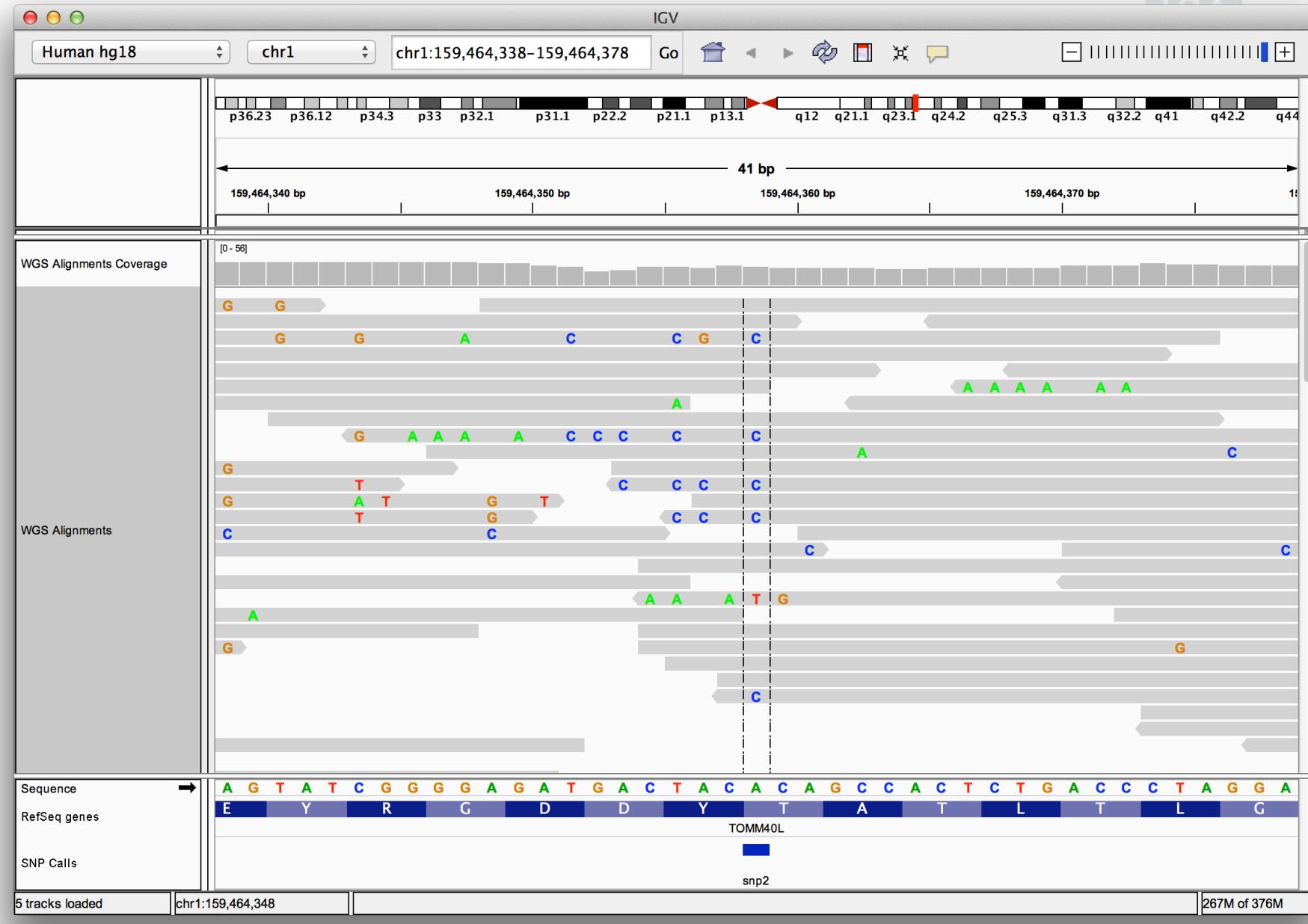
- Rename Track...
- Copy read details to clipboard
- Group alignments by ►
- Sort alignments by ►
- Color alignments by ►
- Shade base by quality
- Show mismatched bases
- Show all bases
- View as pairs
- Go to mate
- View mate region in split screen
- Set insert size options ...
- Re-pack alignments

41 bp
159,464,340 bp 159,464,350 bp 159,464,360 bp

TOMM40L

D E

Viewing SNPs



Viewing SNPs



Human hg18 chr1.snp2

WGS Alignments Coverage

WGS Alignments

Sequence RefSeq genes SNP Calls

5 tracks loaded chr1:159,464,343 D E

233M of 376M

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

WGS Alignments

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

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



Integrative
Genomics
Viewer

Viewing SNPs

Human hg18 chr1.snp2

WGS Alignments Coverage

WGS Alignments

Sequence RefSeq genes SNP Calls

5 tracks loaded chr1:159,464,343

159,464,340 bp

p36.23 p36.12 p34.3 p33 p32.1

159,464

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 ...

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

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

TOMM40L

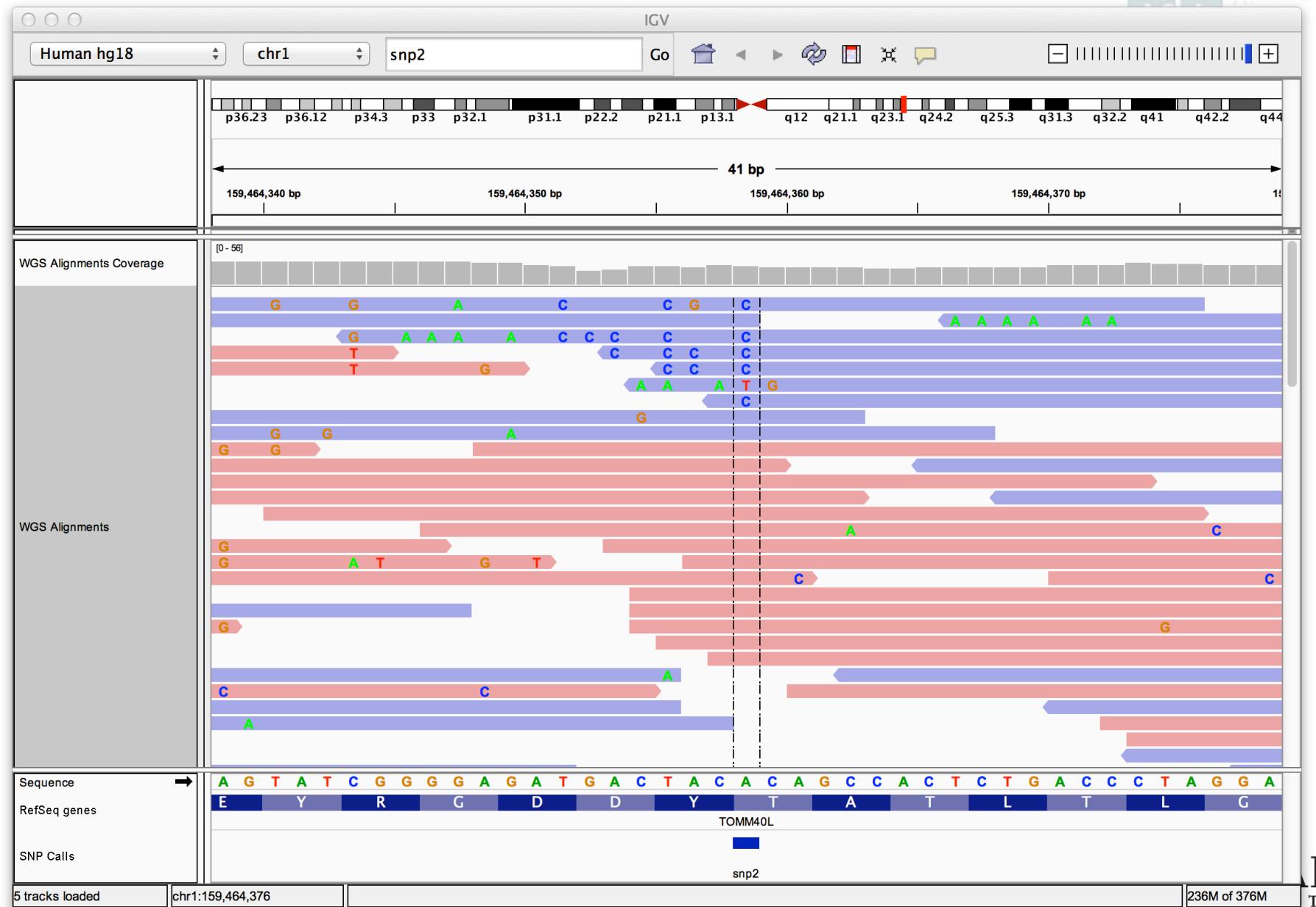
snp2

233M of 376M



Integrative
Genomics
Viewer

Viewing SNPs



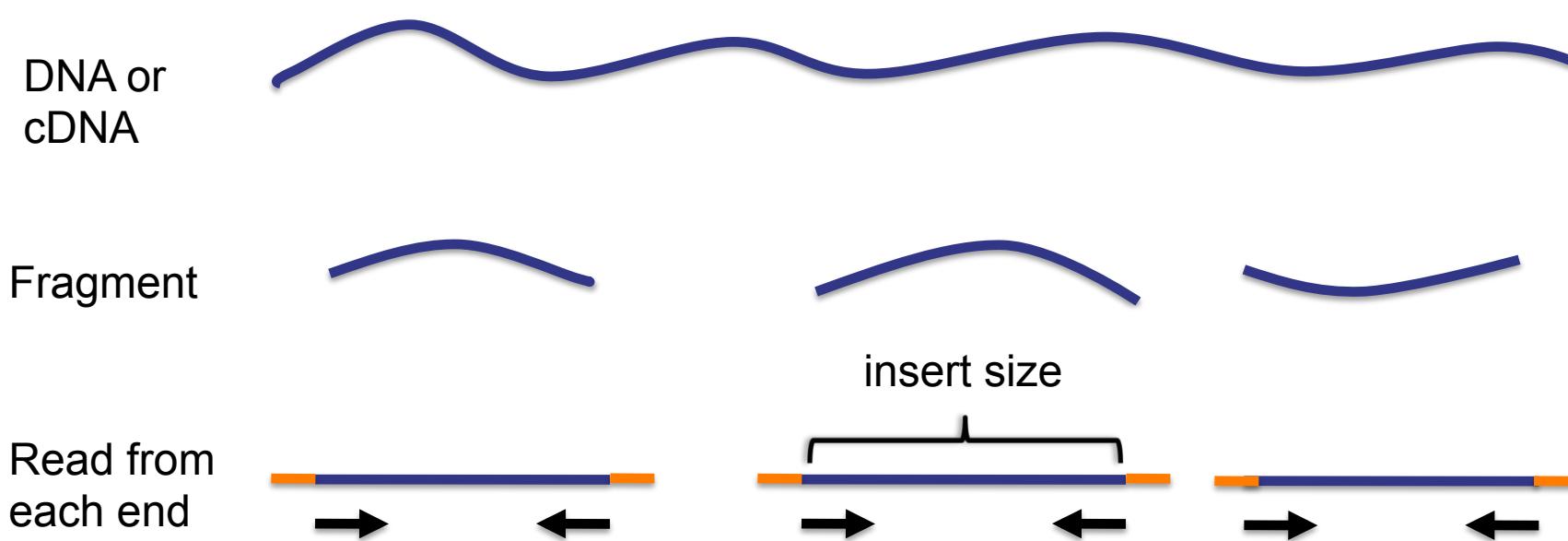
Viewing Structural Events

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)

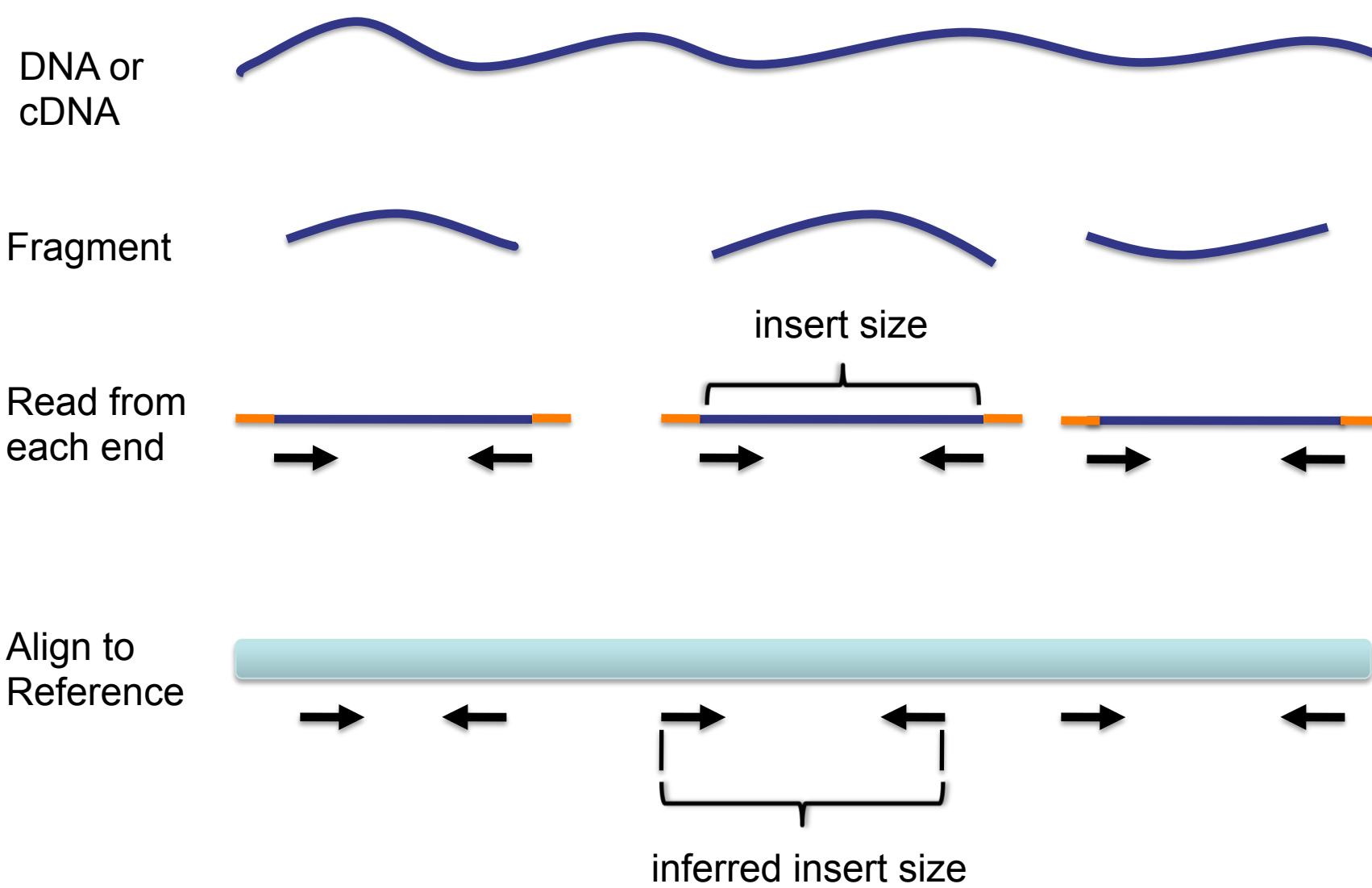
Paired-end sequencing



Paired-end sequencing



Integrative
Genomics
Viewer
IGV



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



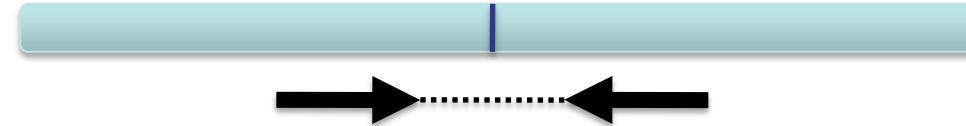
Deletion



Reference
Genome



Subject



Deletion



Reference
Genome



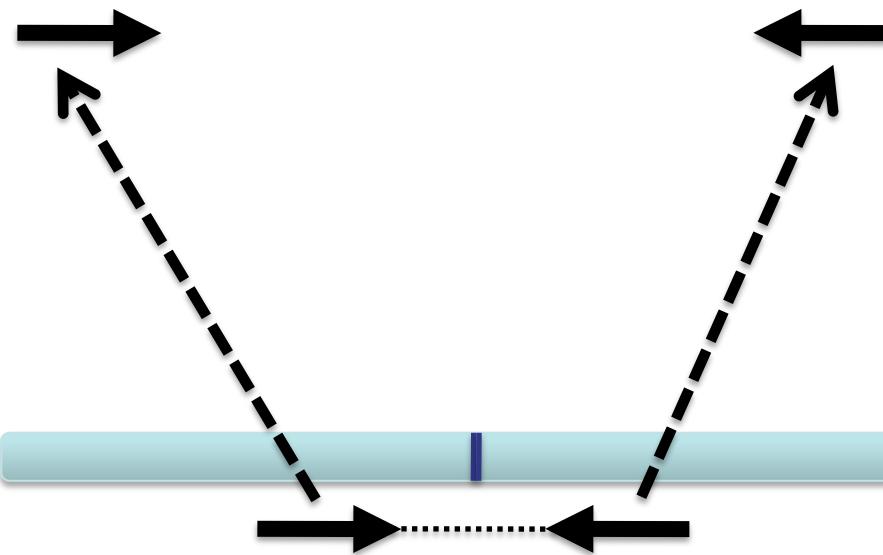
Subject



Deletion



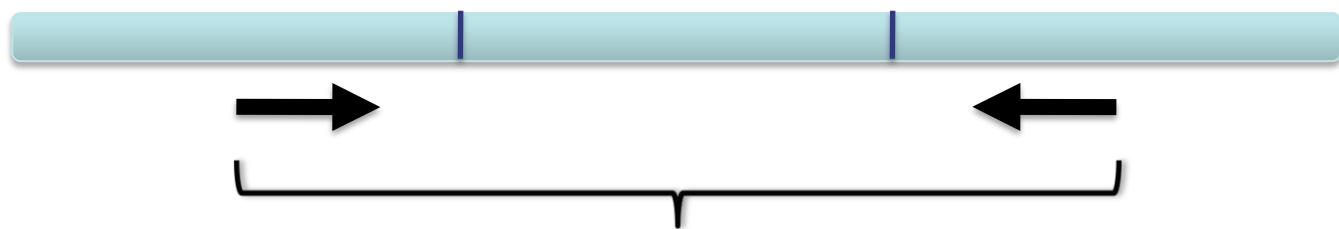
Reference
Genome



Deletion

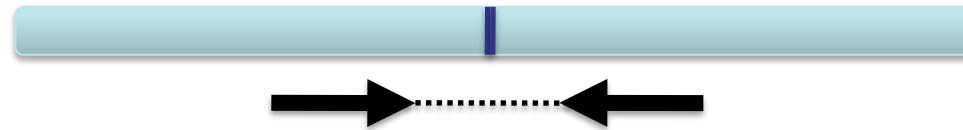


Reference
Genome



inferred insert size

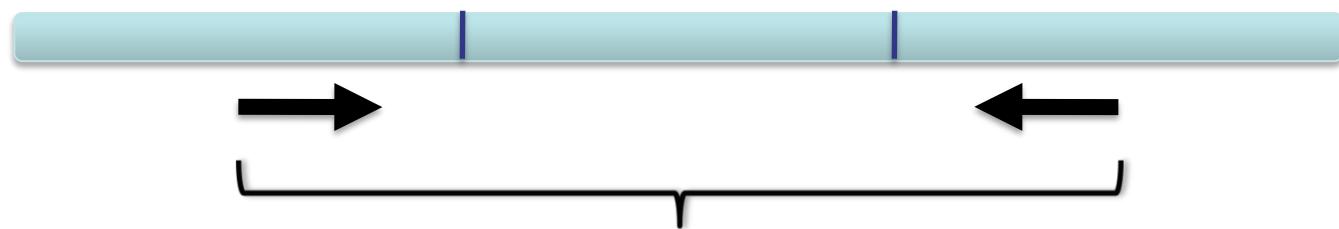
Subject



Deletion

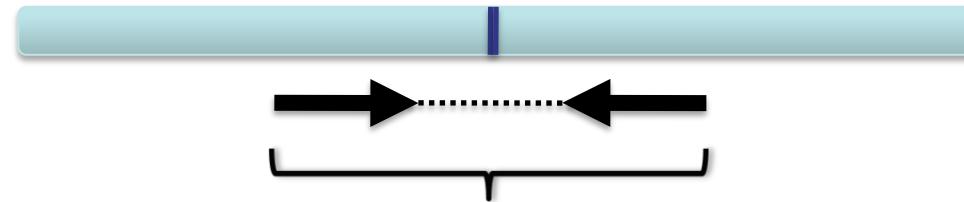


Reference
Genome



inferred insert size

Subject



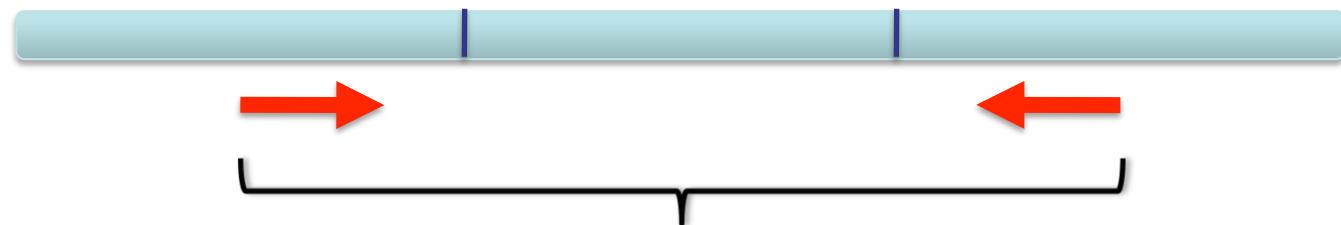
expected insert size

Deletion



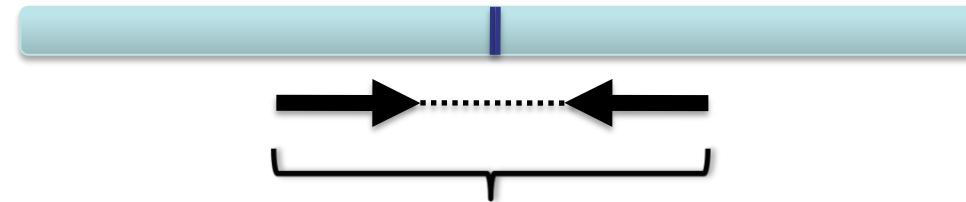
Inferred insert size is > expected value

Reference
Genome



inferred insert size

Subject



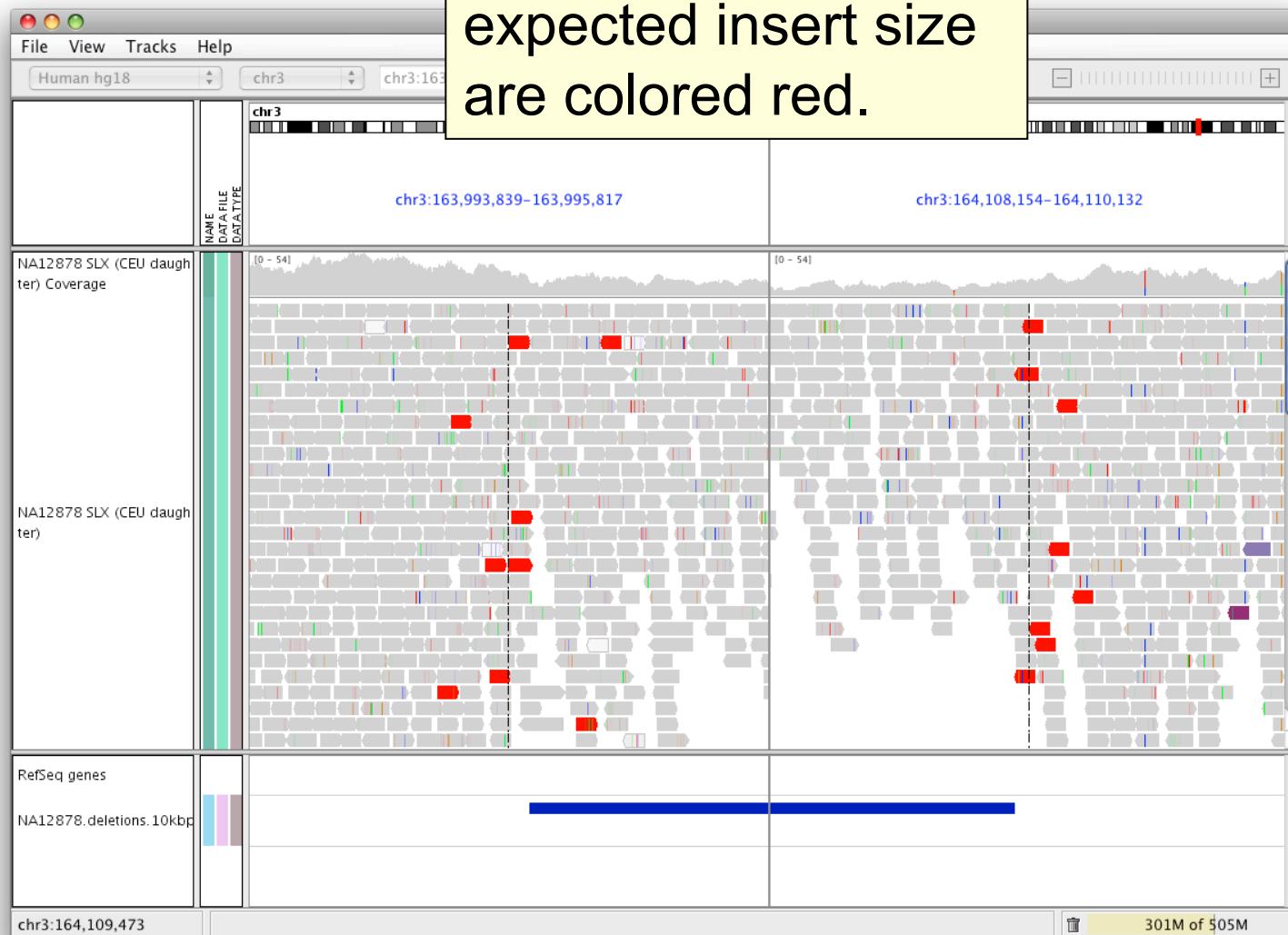
expected insert size

Deletion



Integrative
Genomics
Viewer
IGV

Pairs with larger than
expected insert size
are colored red.

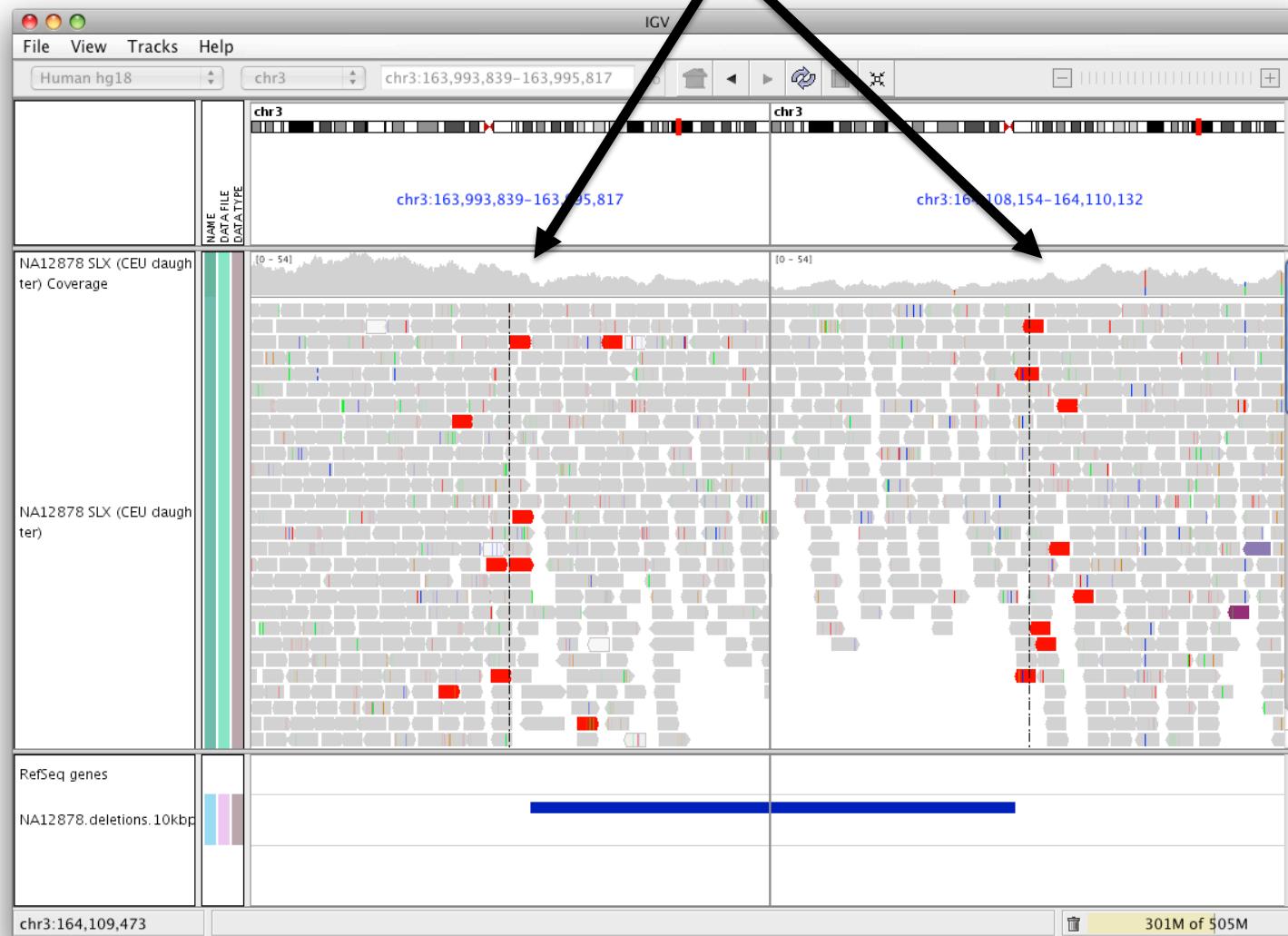


Deletion



Integrative
Genomics
Viewer
IGV

Note drop in coverage

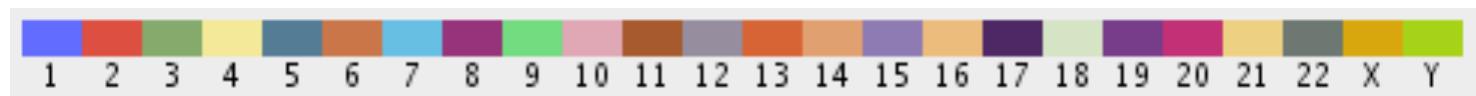


Insert size color scheme



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

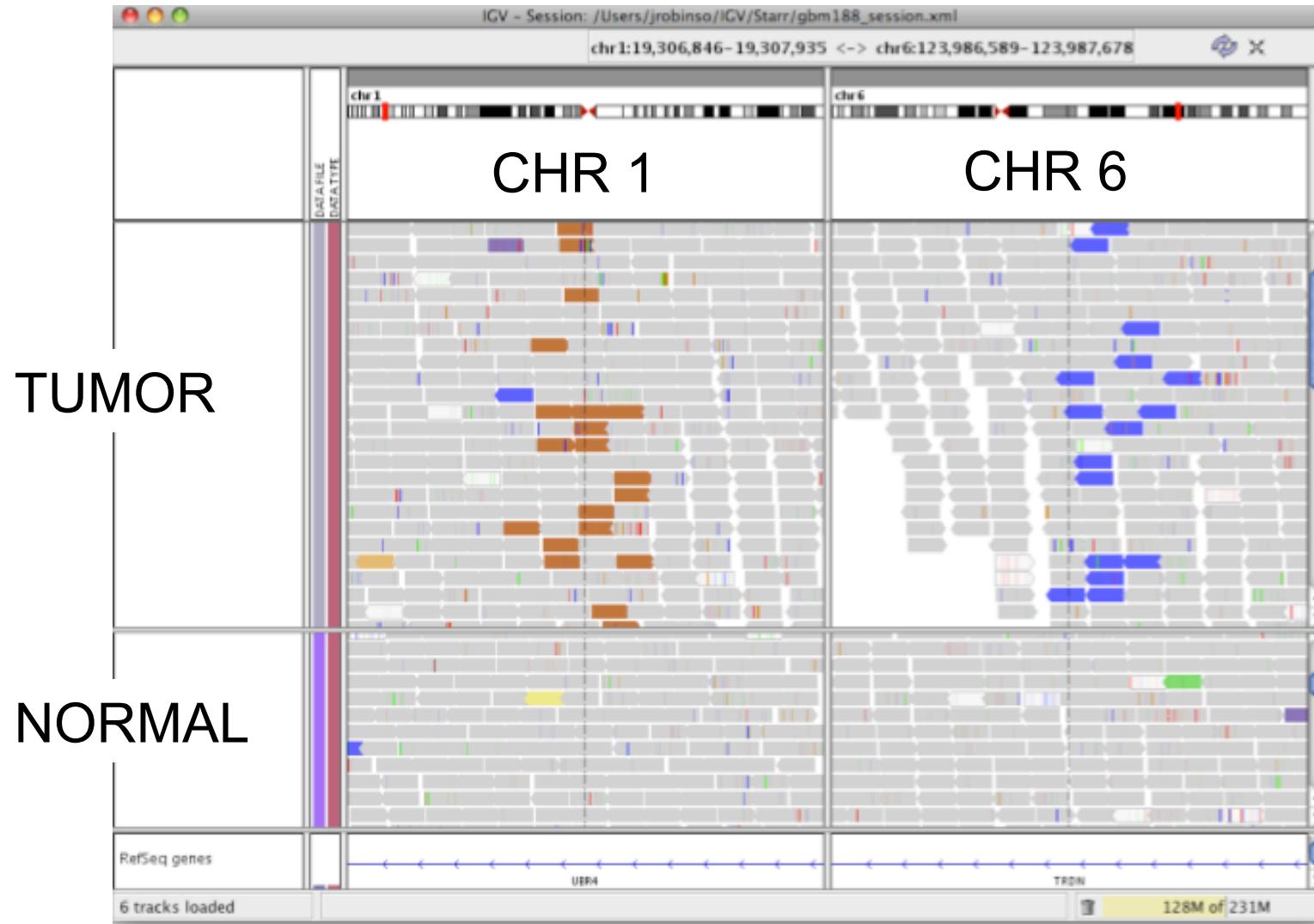
Each end colored by chromosome of its mate



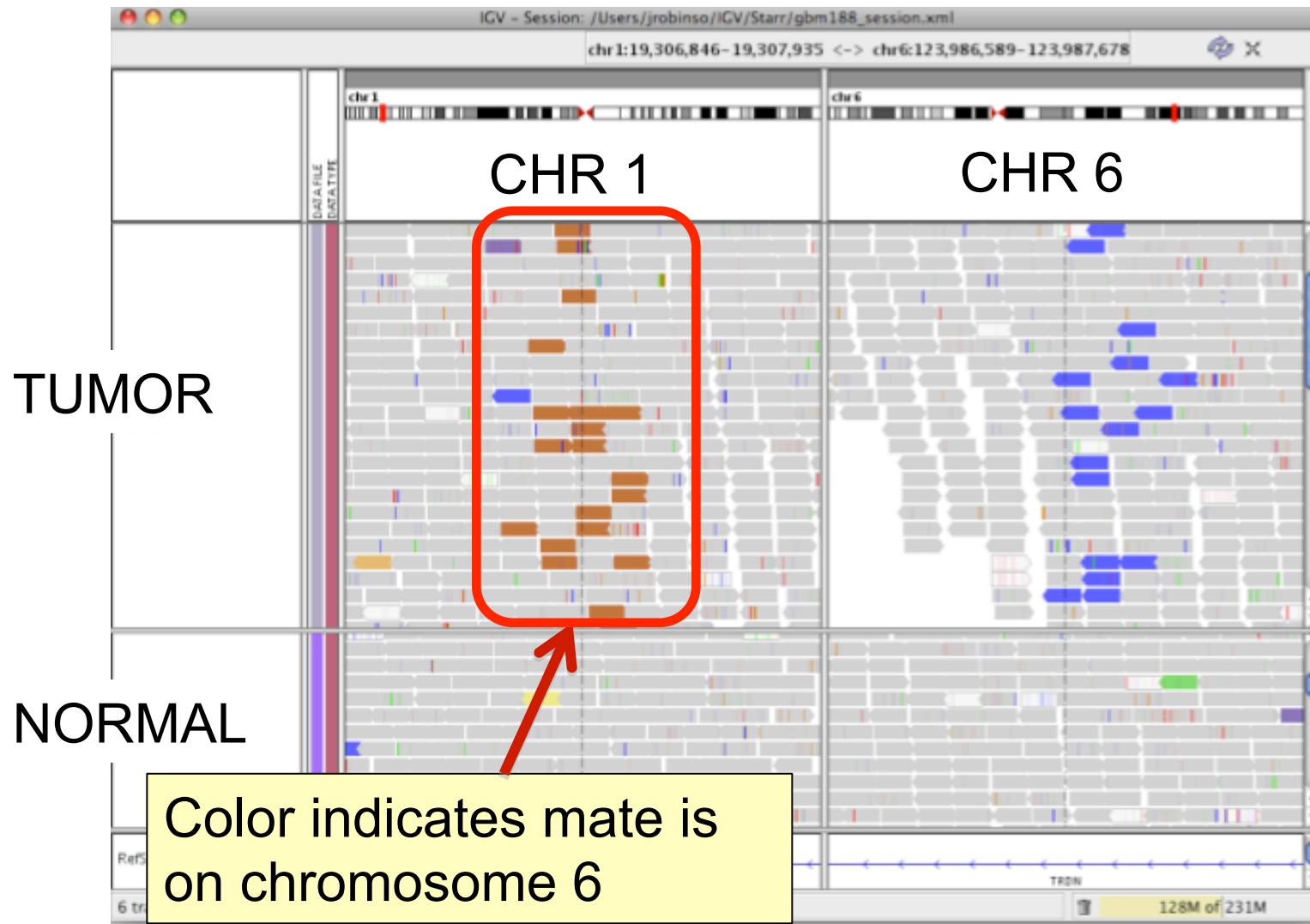
Rearrangement



Integrative
Genomics
Viewer
ALIVE



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



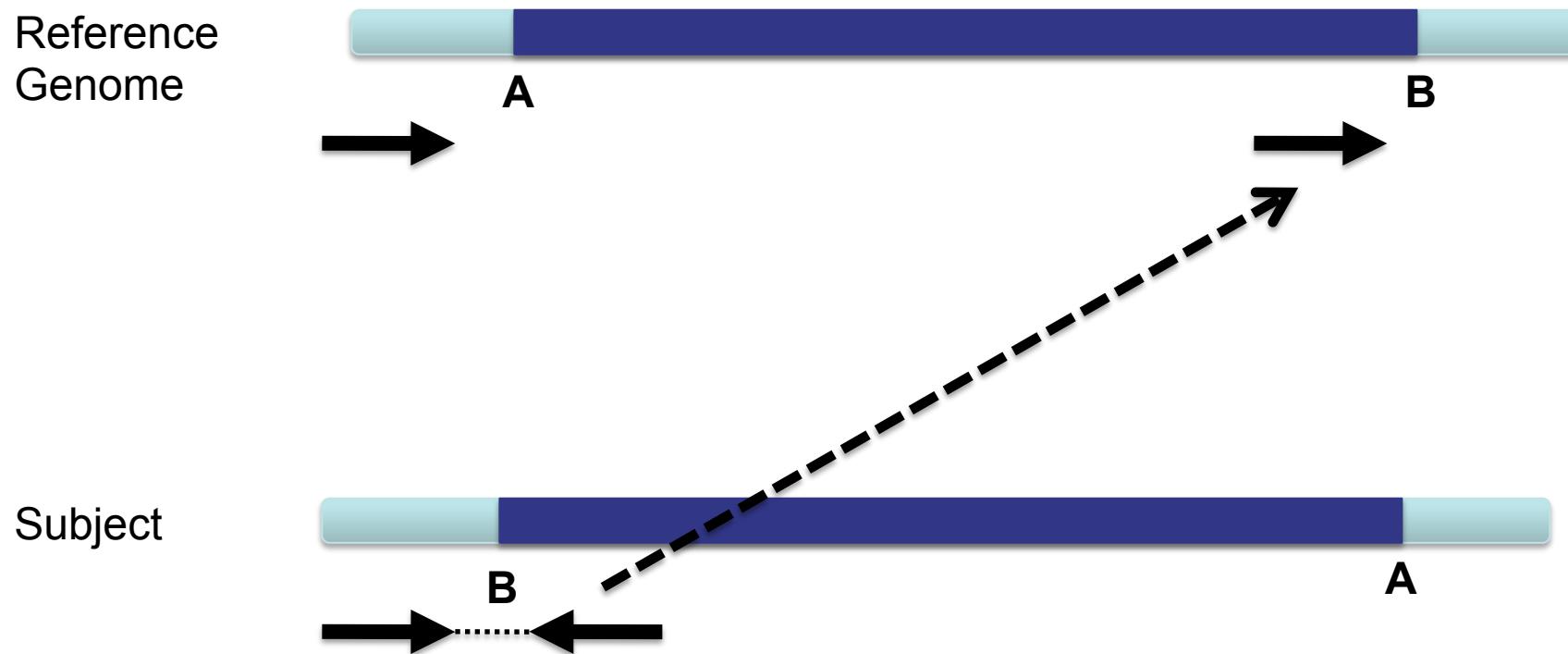
Reference
Genome



Subject



Inversion



Inversion



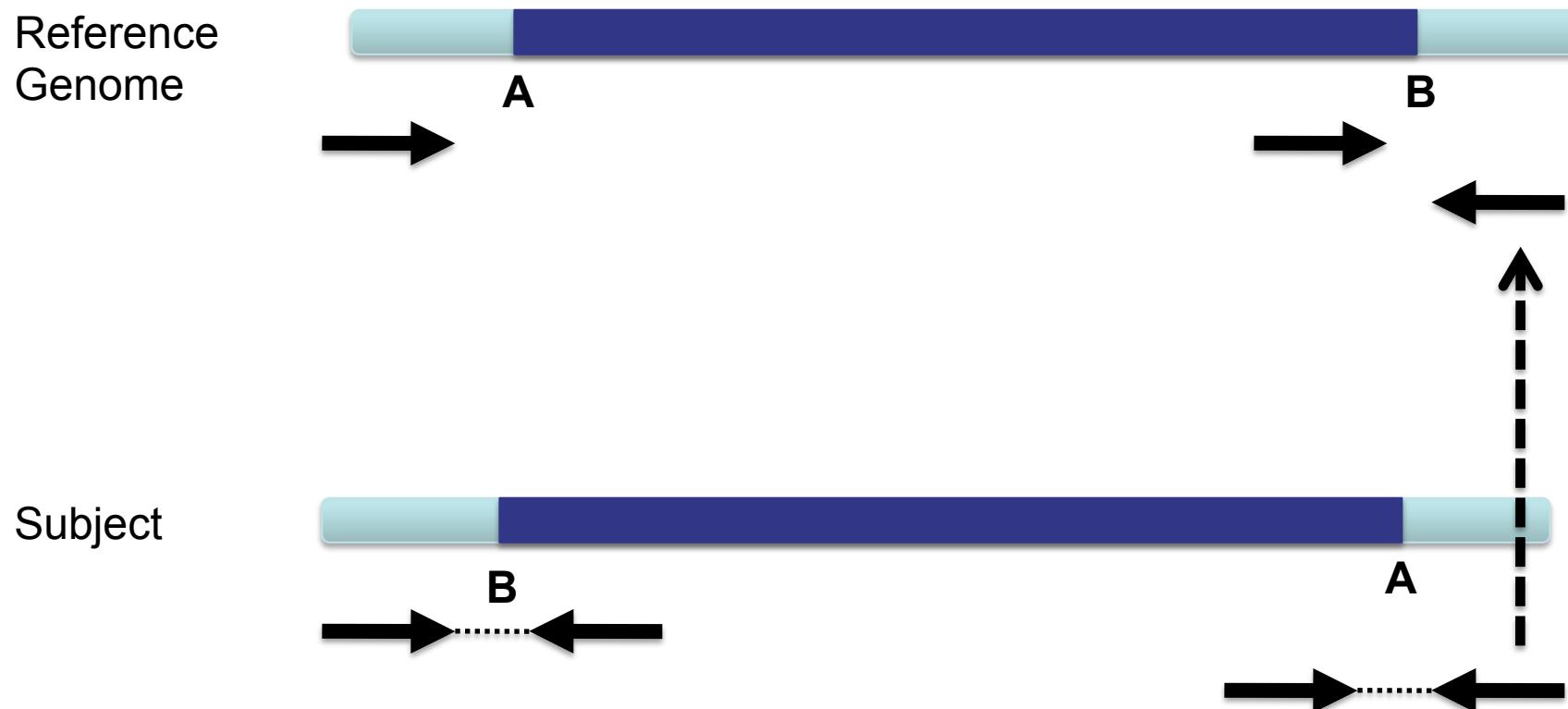
Reference
Genome



Subject



Inversion



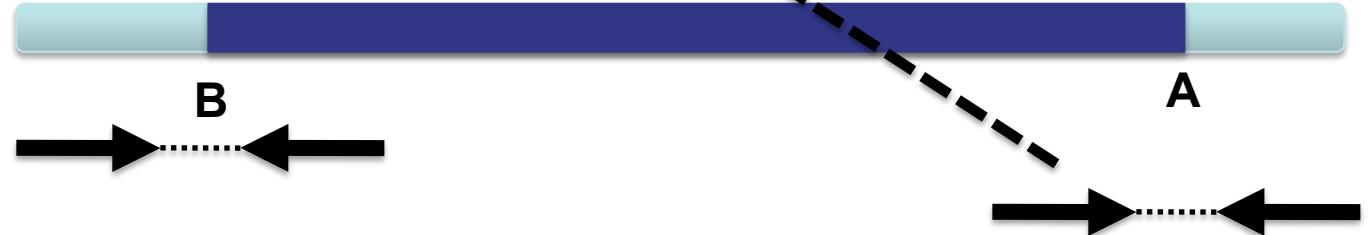
Inversion



Reference
Genome



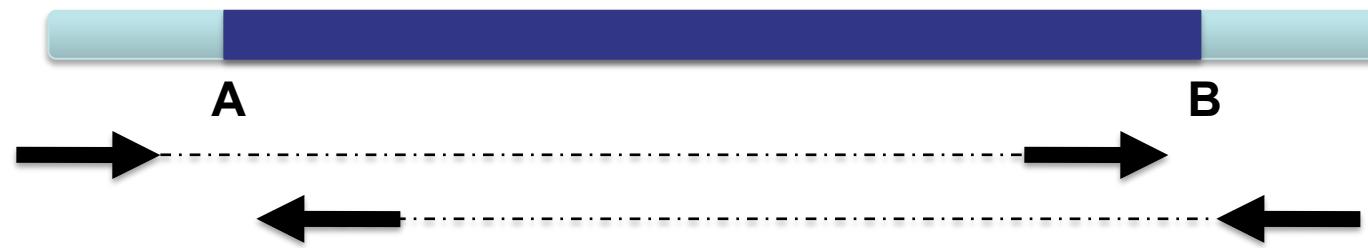
Subject



Inversion



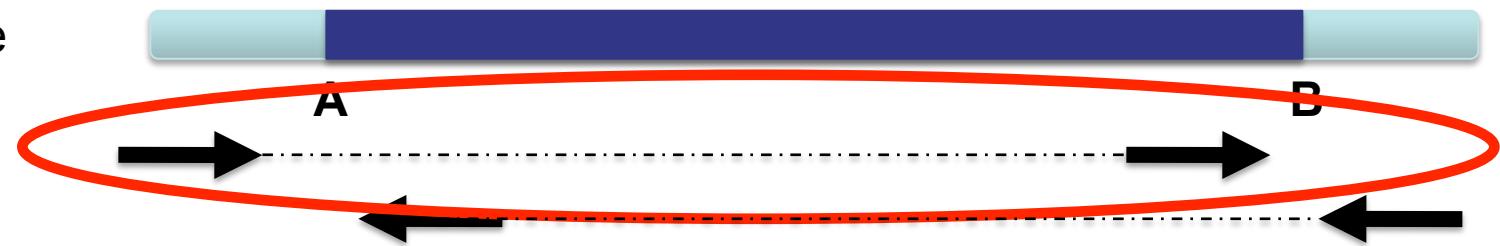
Reference
Genome



Inversion



Reference
Genome

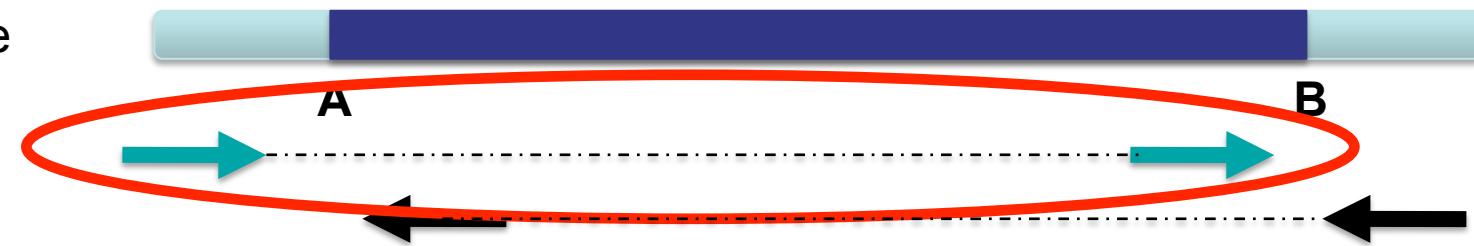


Anomaly: expected orientation of pair is
inward facing (→ ←)

Inversion



Reference
Genome

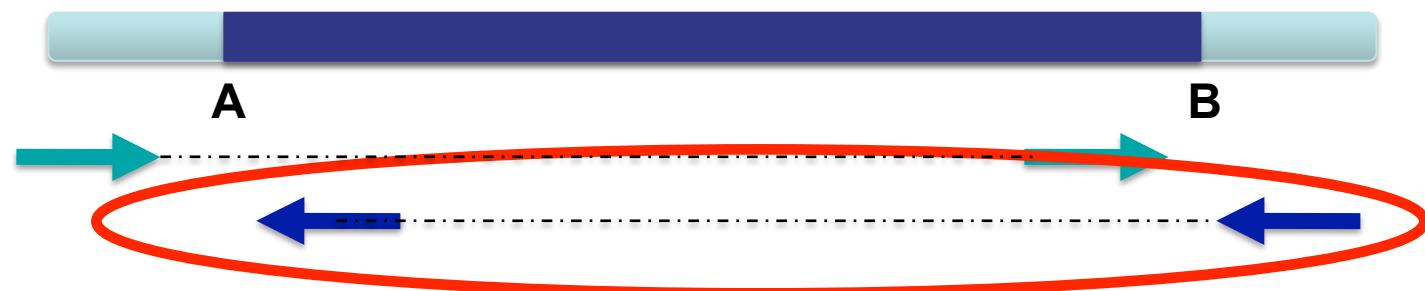


“Left” side pair

Inversion



Reference
Genome



“Right” side pair

Color by pair orientation



Integrative
Genomics
Viewer
ALGVIEWER

NA12878 WGS

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

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

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

- Re-pack alignments

no color

insert size

pair orientation

insert size and pair orientation

read strand

first-of-pair strand

read group

sample

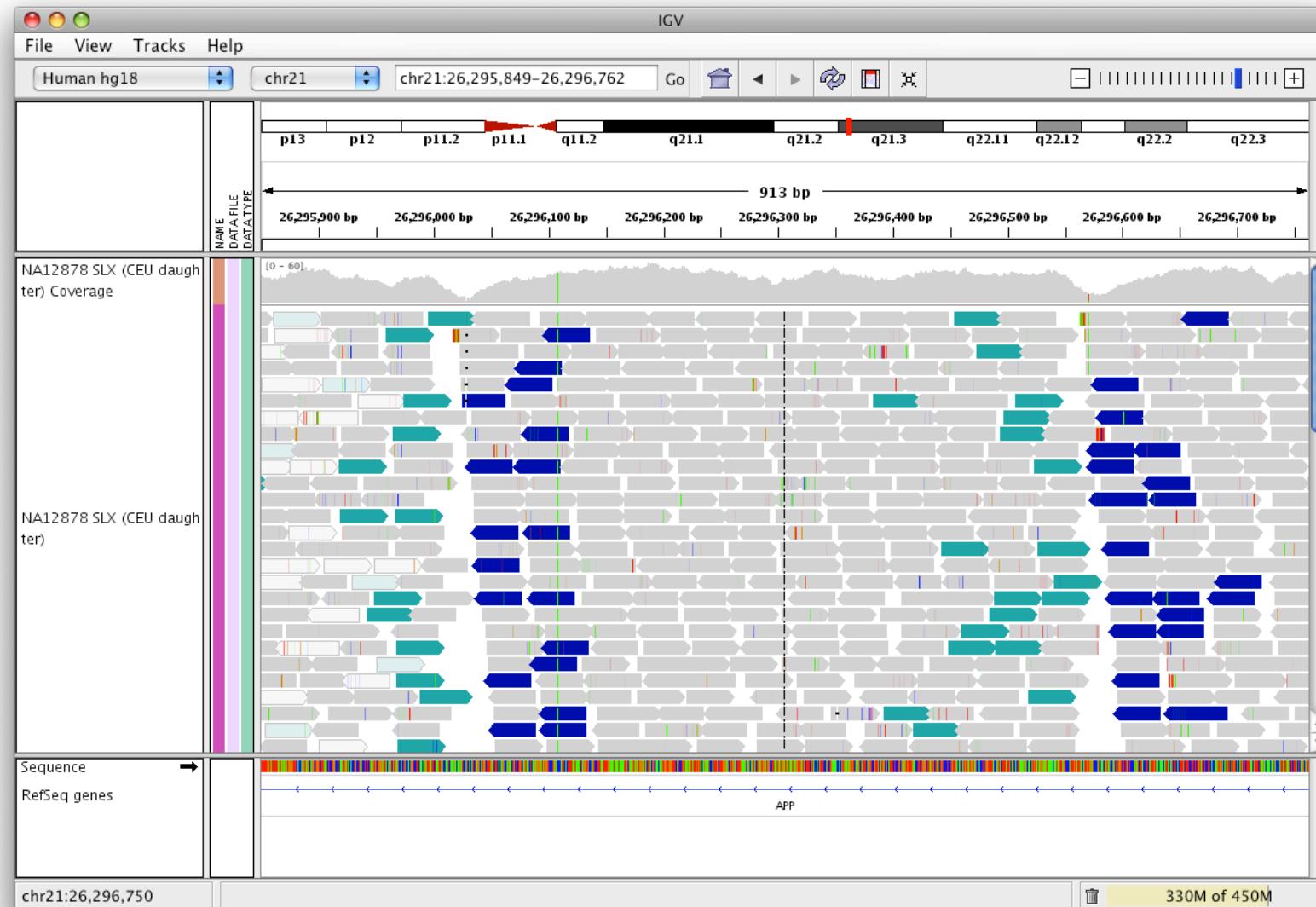
tag

bisulfite mode

Inversion



Integrative
Genomics
Viewer
ALGEMEL

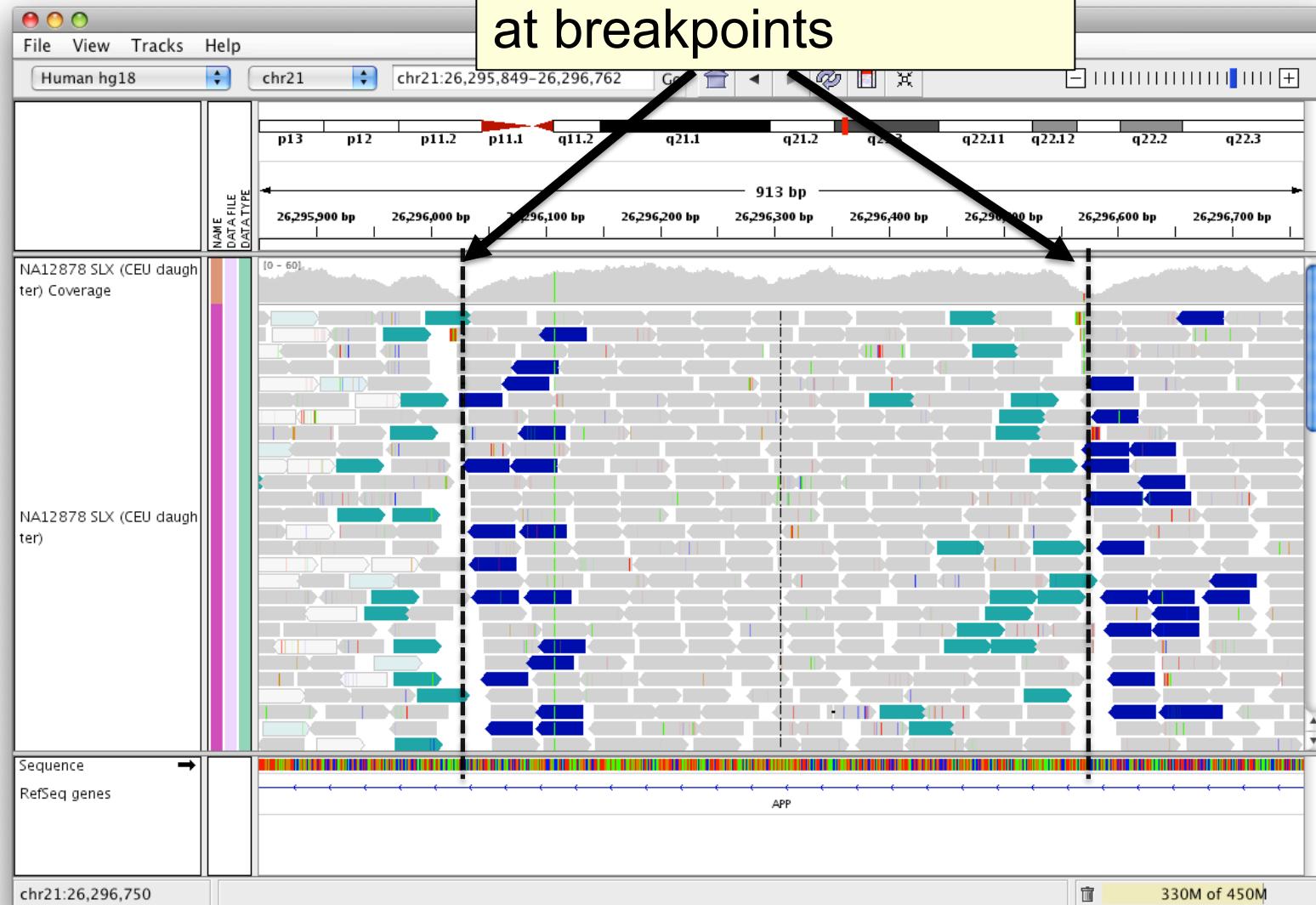


Inversion

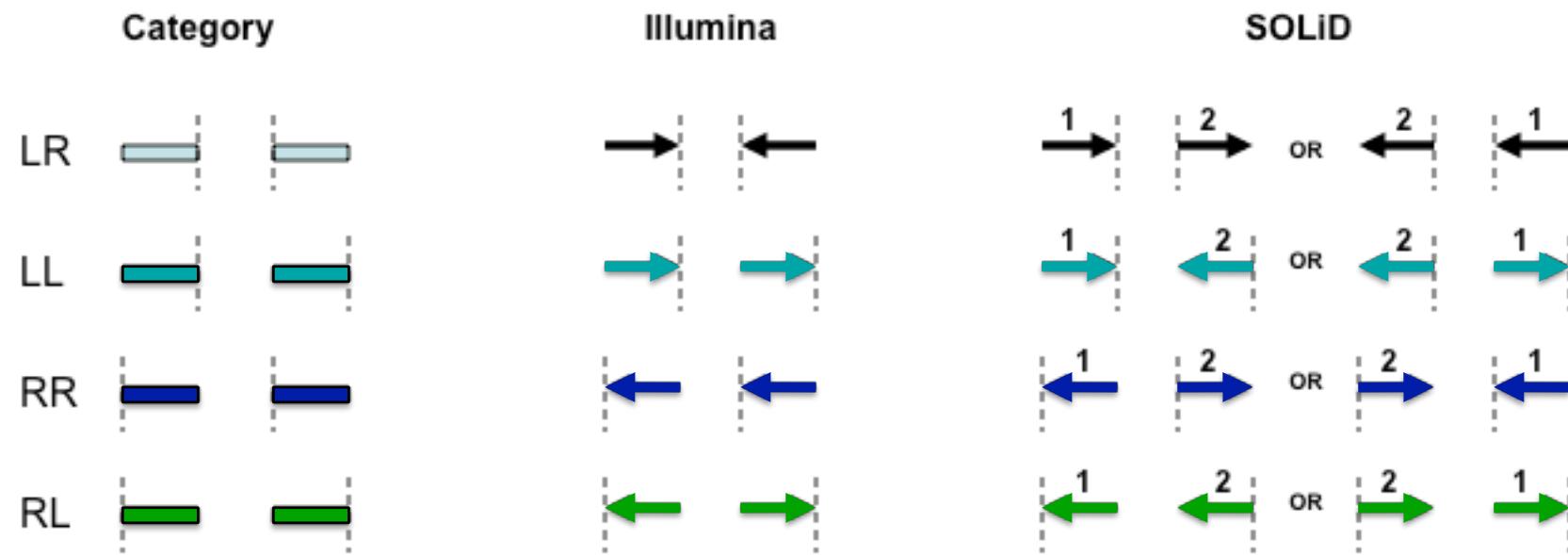


Integrative
Genomics
Viewer
IGV

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

Viewing RNA-Seq Data



- Tune preferences
- Examine tissue-specific alternative splicing
- Supplementary views

igvtools - coverage

Tophat - insertions, deletions, junctions

Cufflinks - transcripts

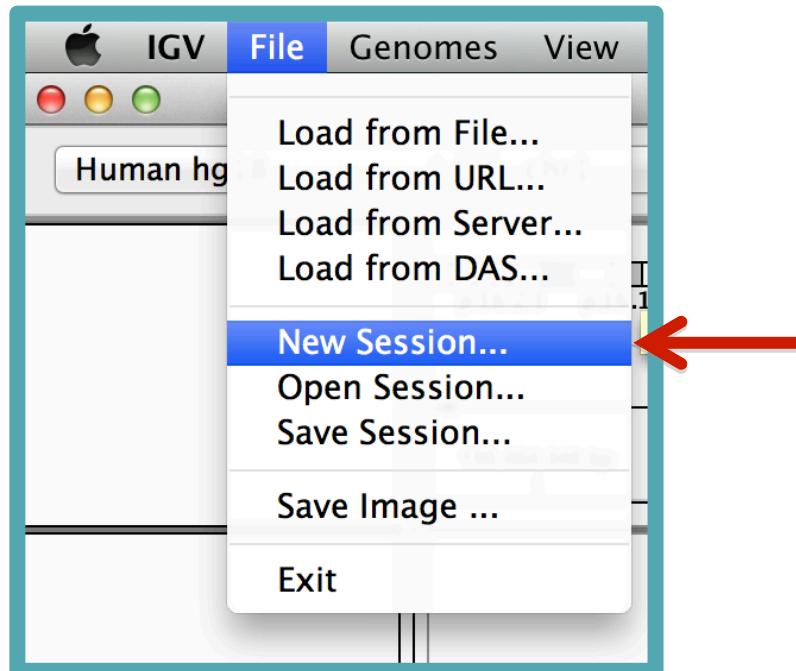
Data



- Download data from

[http://www.broadinstitute.org/igvdata/tutorials/
epi2014/data.zip](http://www.broadinstitute.org/igvdata/tutorials/epi2014/data.zip)

- Unzip



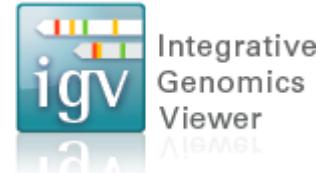
Before we start:
Select File > New Session
to clear IGV window

RNA-Seq Setup

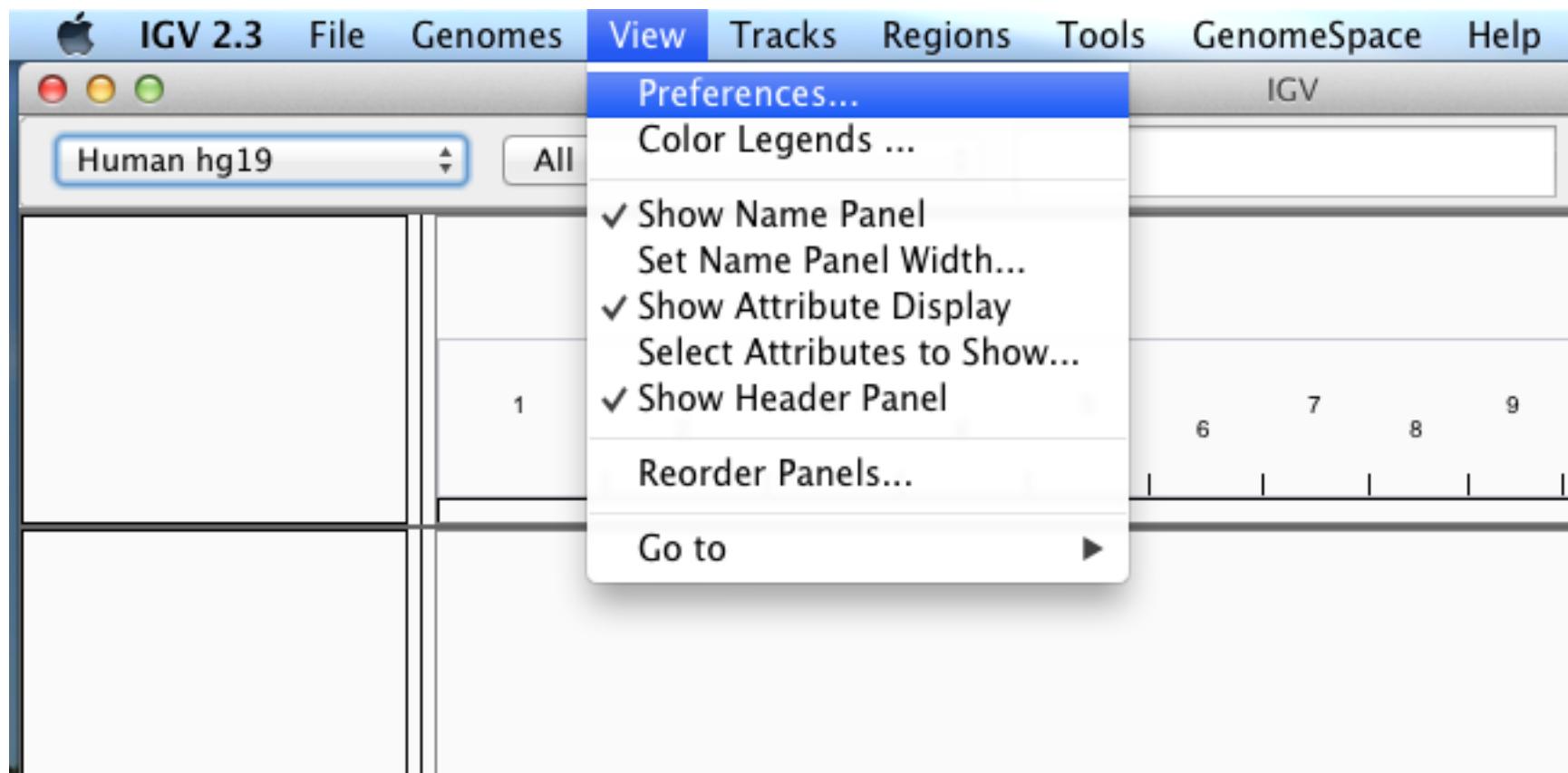
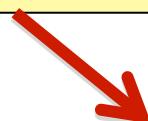


- Tune settings for RNA.

RNA-Seq Setup



Select View > Preferences...



RNA-Seq Setup



Integrative
Genomics
Viewer
ALGEME

Click Alignments tab

Human hg18

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

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 Show center line

Filter vendor failed reads Show coverage track

Filter secondary alignments Show soft-clipped bases

Flag unmapped pairs Flag zero-quality alignments

Shade mismatched bases by quality: 5 to 20

Flag insertions larger than: bases

Filter alignments by read group URL or path to filter file

Splice Junction Track Options

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

Show flanking regions

Insert Size Options

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

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

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

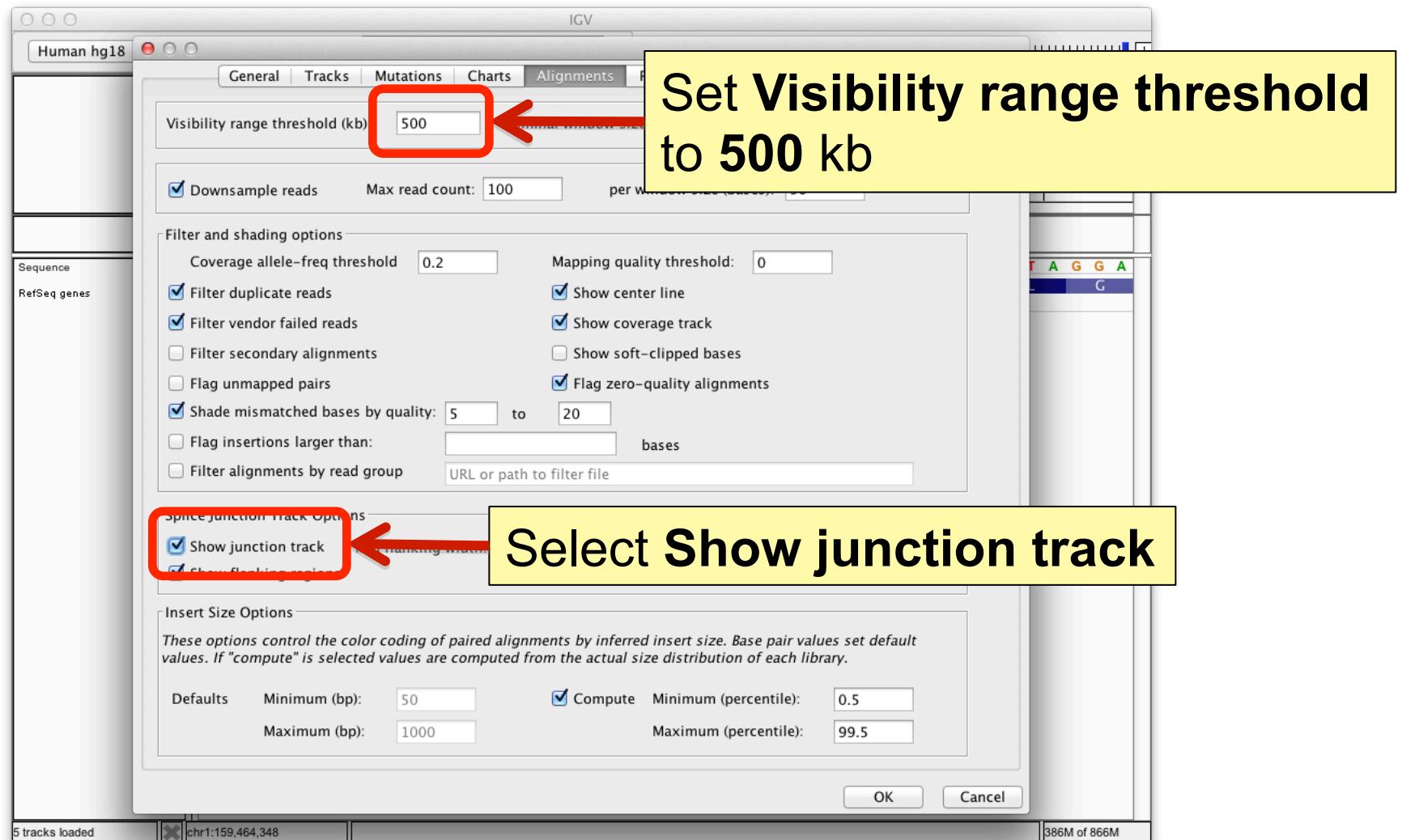
OK Cancel

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

RNA-Seq Setup



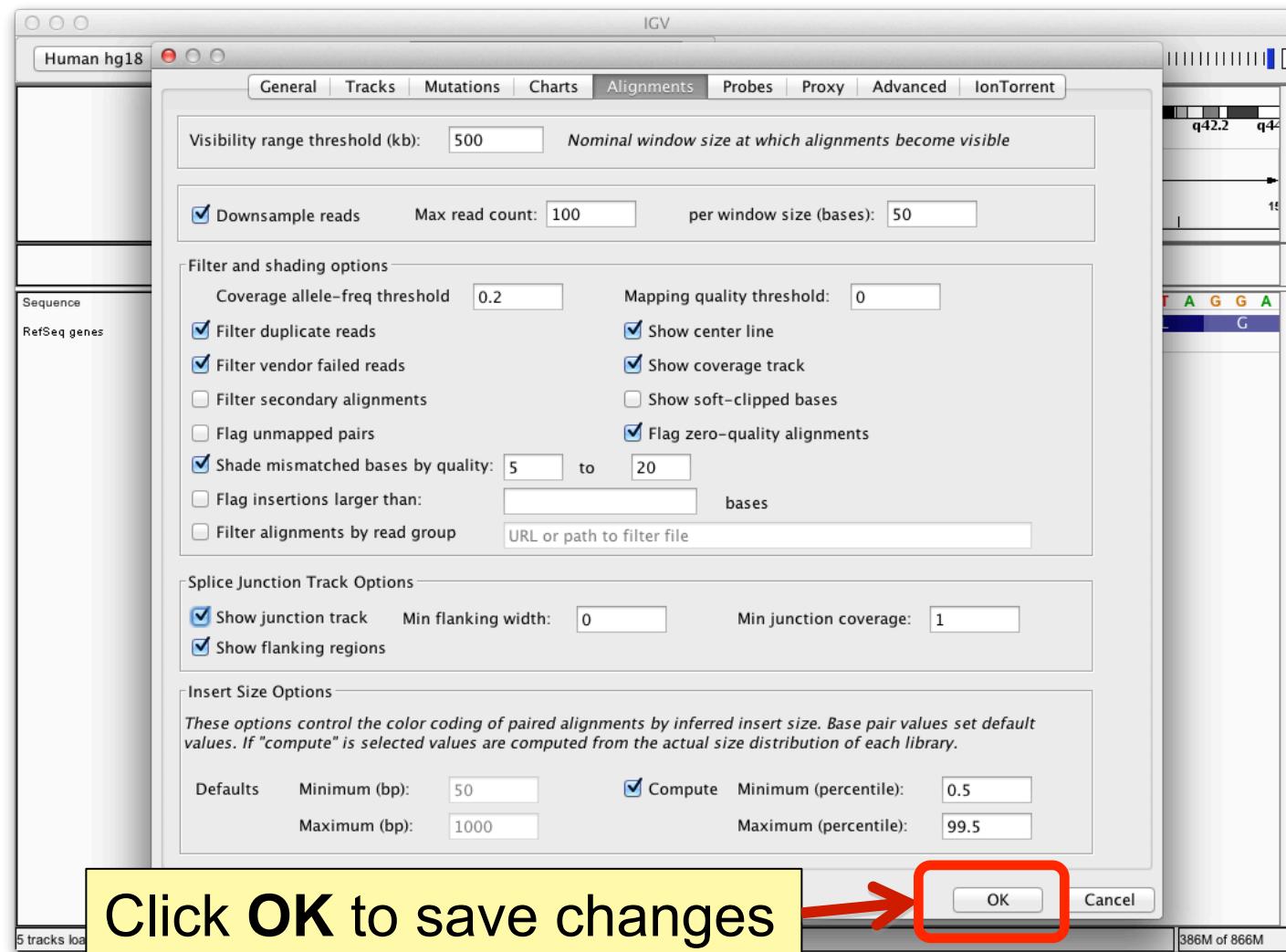
Integrative
Genomics
Viewer
ALMEL



RNA-Seq Setup



Integrative
Genomics
Viewer
ALGEMEL



RNA-Seq



Hands-on exercise

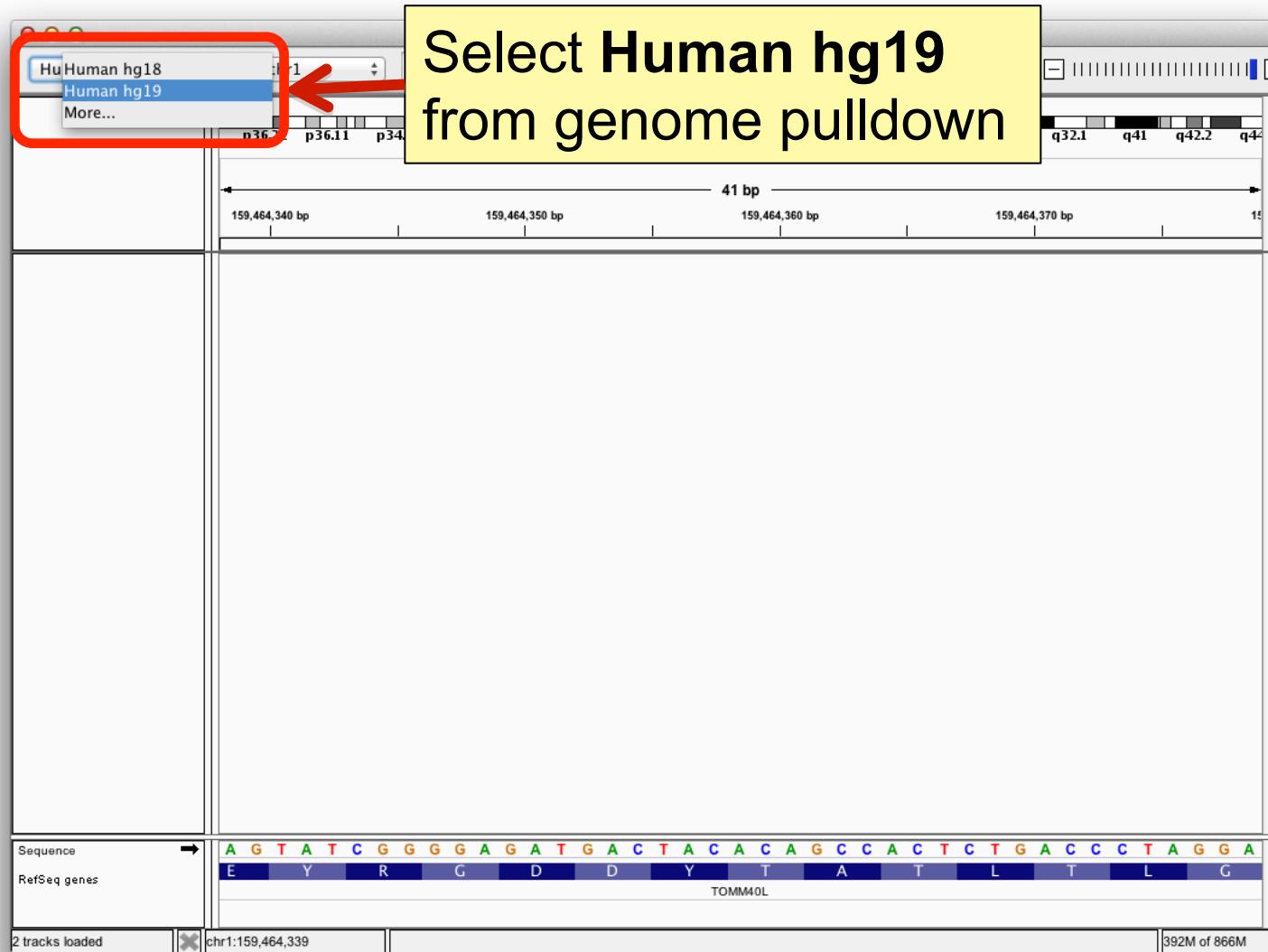
- Examine tissue-specific alternative splicing.
- Data: Illumina BodyMap 2.0

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

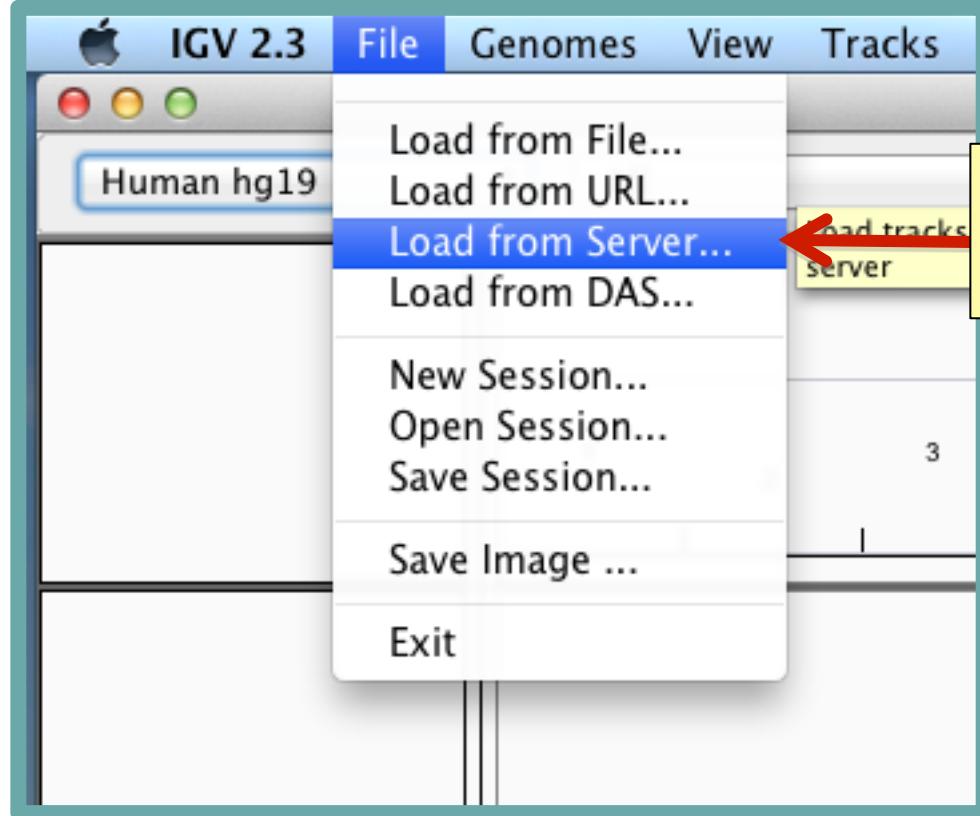
RNA-seq alignments



Integrative
Genomics
Viewer
ALIGMENT

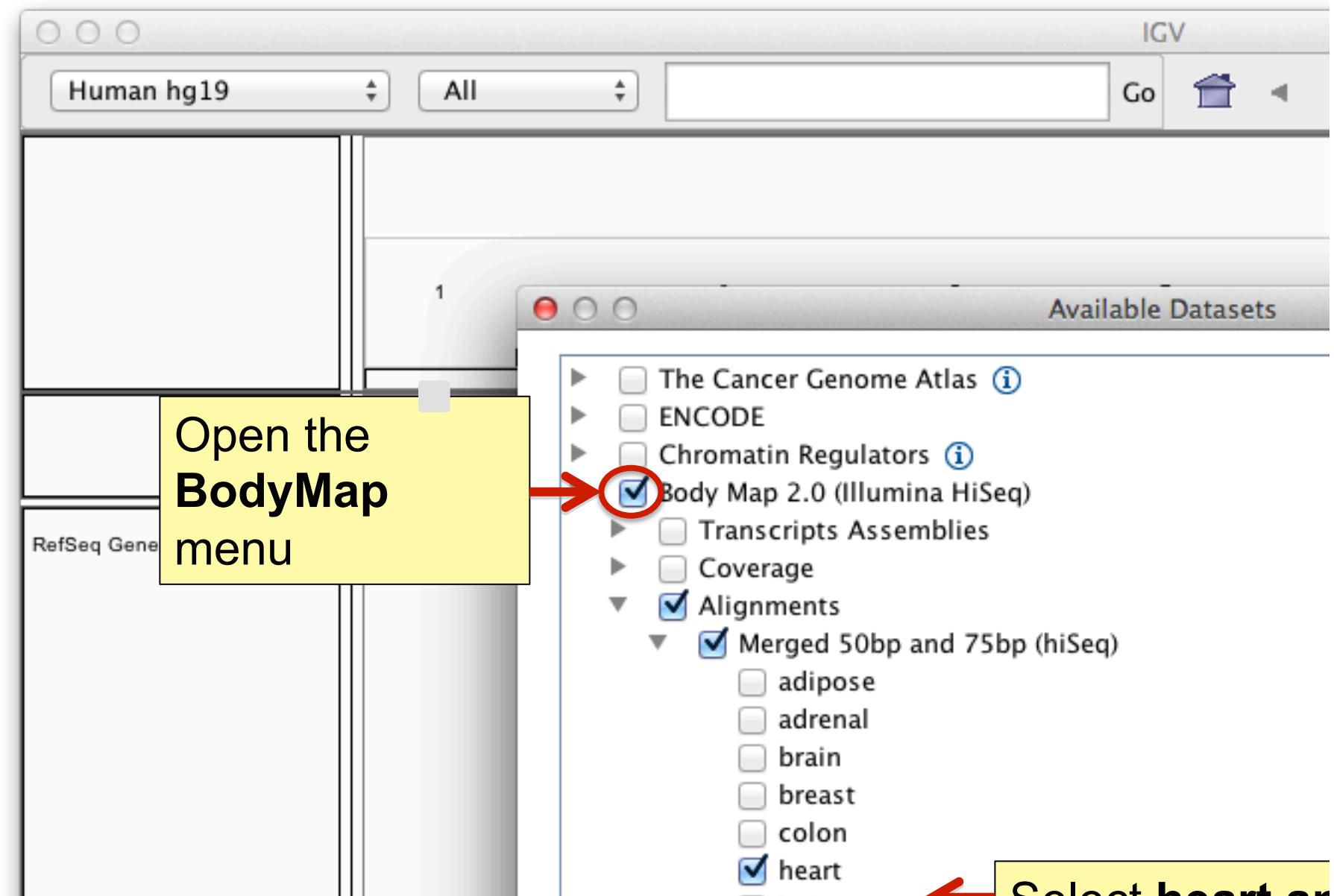


RNA-seq alignments



Select:
File > Load from Server...

RNA-seq alignments



The screenshot shows the IGV software interface. At the top, there are dropdown menus for 'Human hg19' and 'All', a search bar, and a 'Go' button. Below the search bar is a small icon of a house. A yellow callout box with black text says 'Open the BodyMap menu'. To the right, a floating window titled 'Available Datasets' lists various genomic datasets. A red arrow points to the 'Body Map 2.0 (Illumina HiSeq)' entry, which has a checked checkbox. Another red arrow points to the 'heart' entry under the 'Alignments/Merged 50bp and 75bp (hiSeq)' section, also with a checked checkbox. A yellow callout box at the bottom right says 'Select heart or' with a red arrow pointing to the 'heart' checkbox.

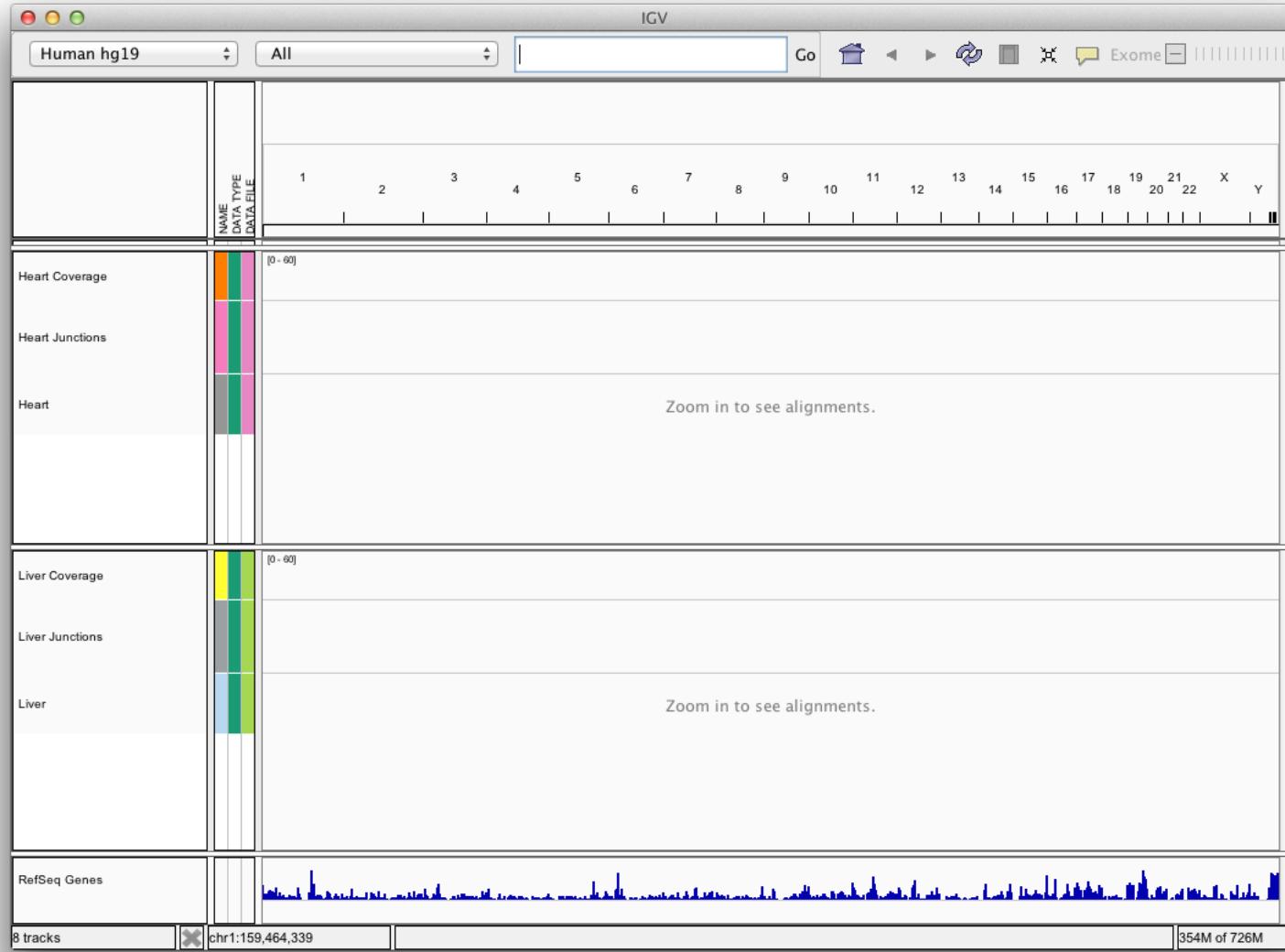
Available Datasets

- The Cancer Genome Atlas [\(i\)](#)
- ENCODE
- Chromatin Regulators [\(i\)](#)
- Body Map 2.0 (Illumina HiSeq)
- Transcripts Assemblies
- Coverage
- ▼ Alignments
 - Merged 50bp and 75bp (hiSeq)
 - adipose
 - adrenal
 - brain
 - breast
 - colon
 - heart

RNA-seq alignments



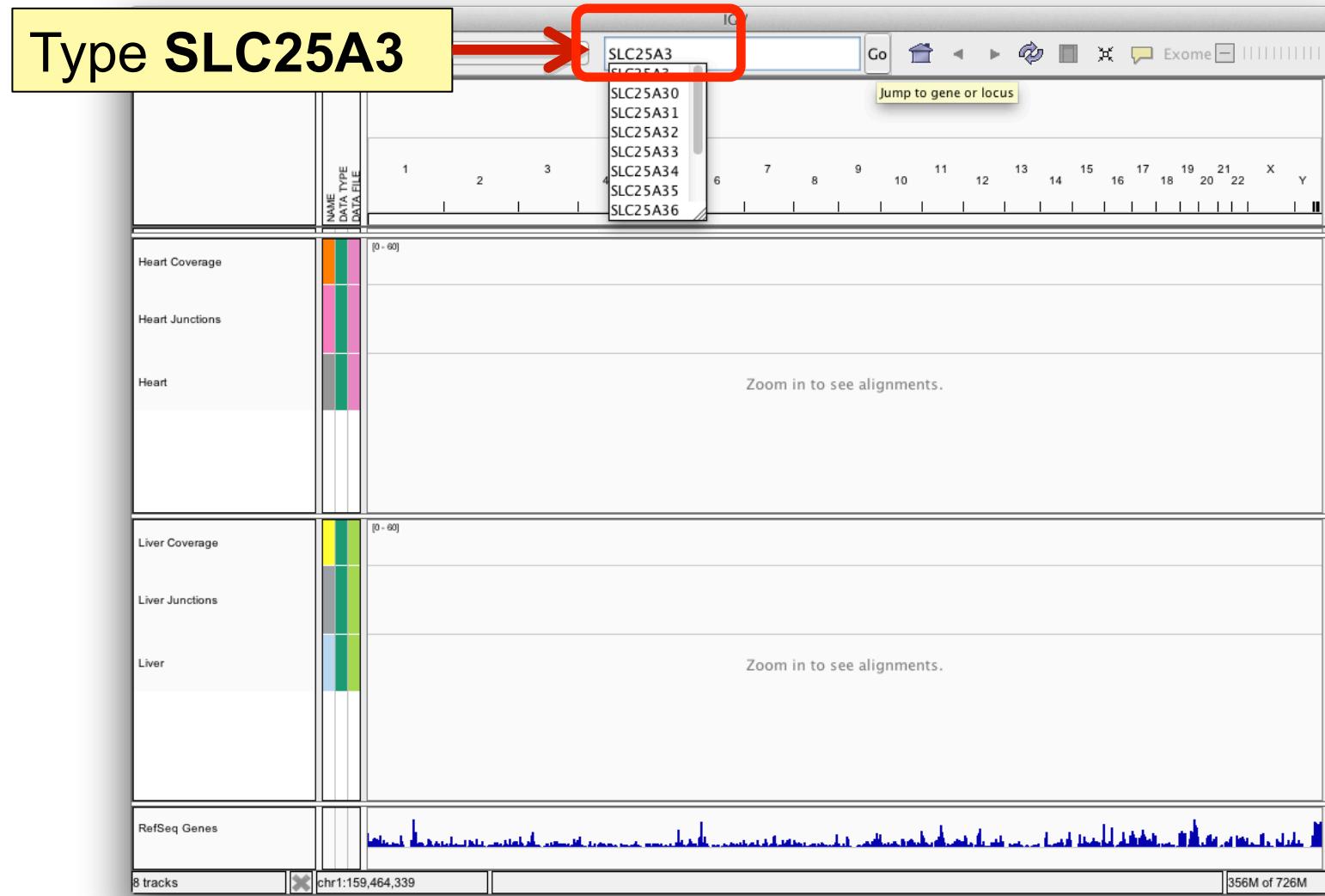
Integrative
Genomics
Viewer
ALIVE



RNA-seq alignments



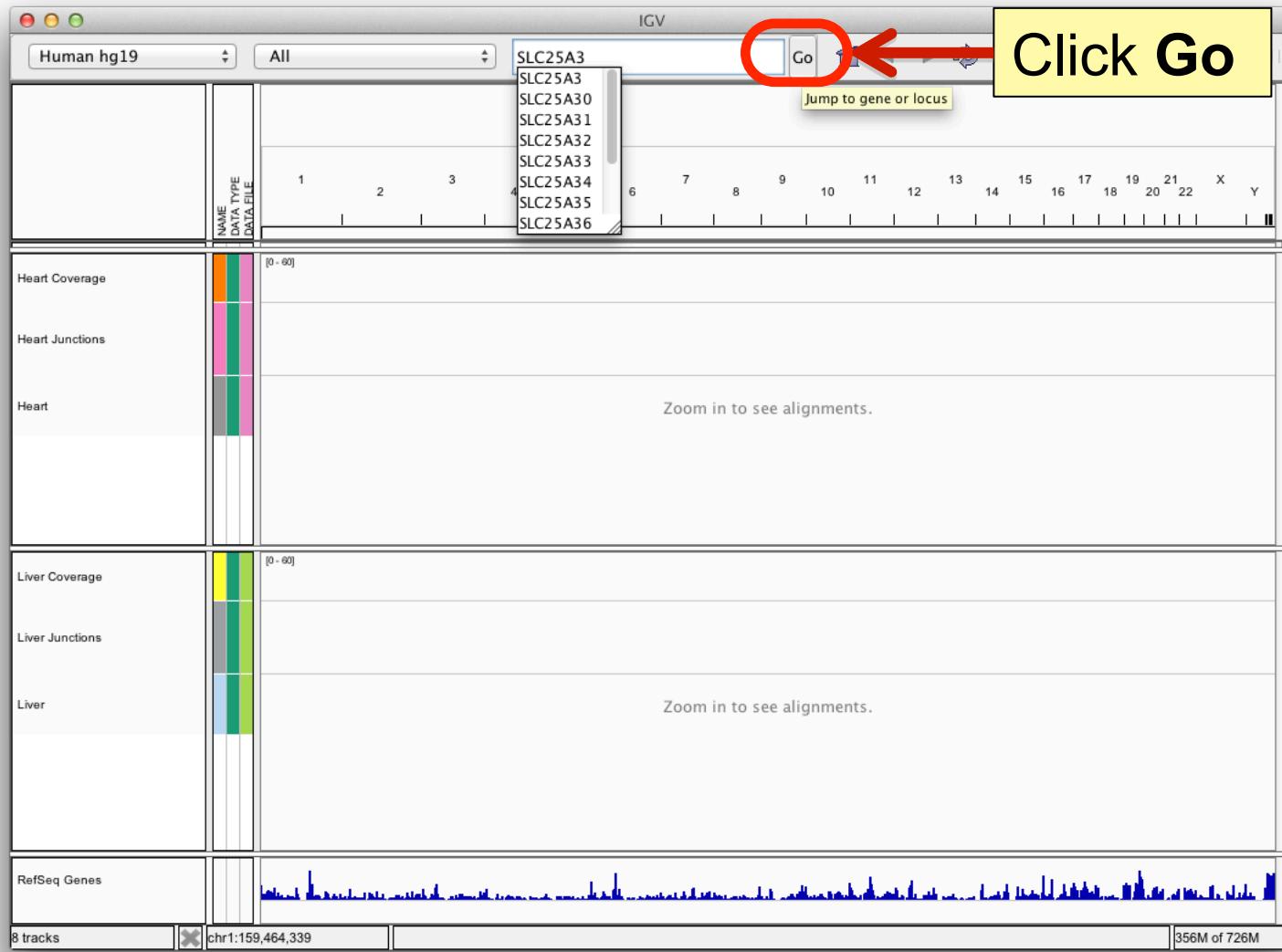
Integrative
Genomics
Viewer
ALIGMENT



RNA-seq alignments



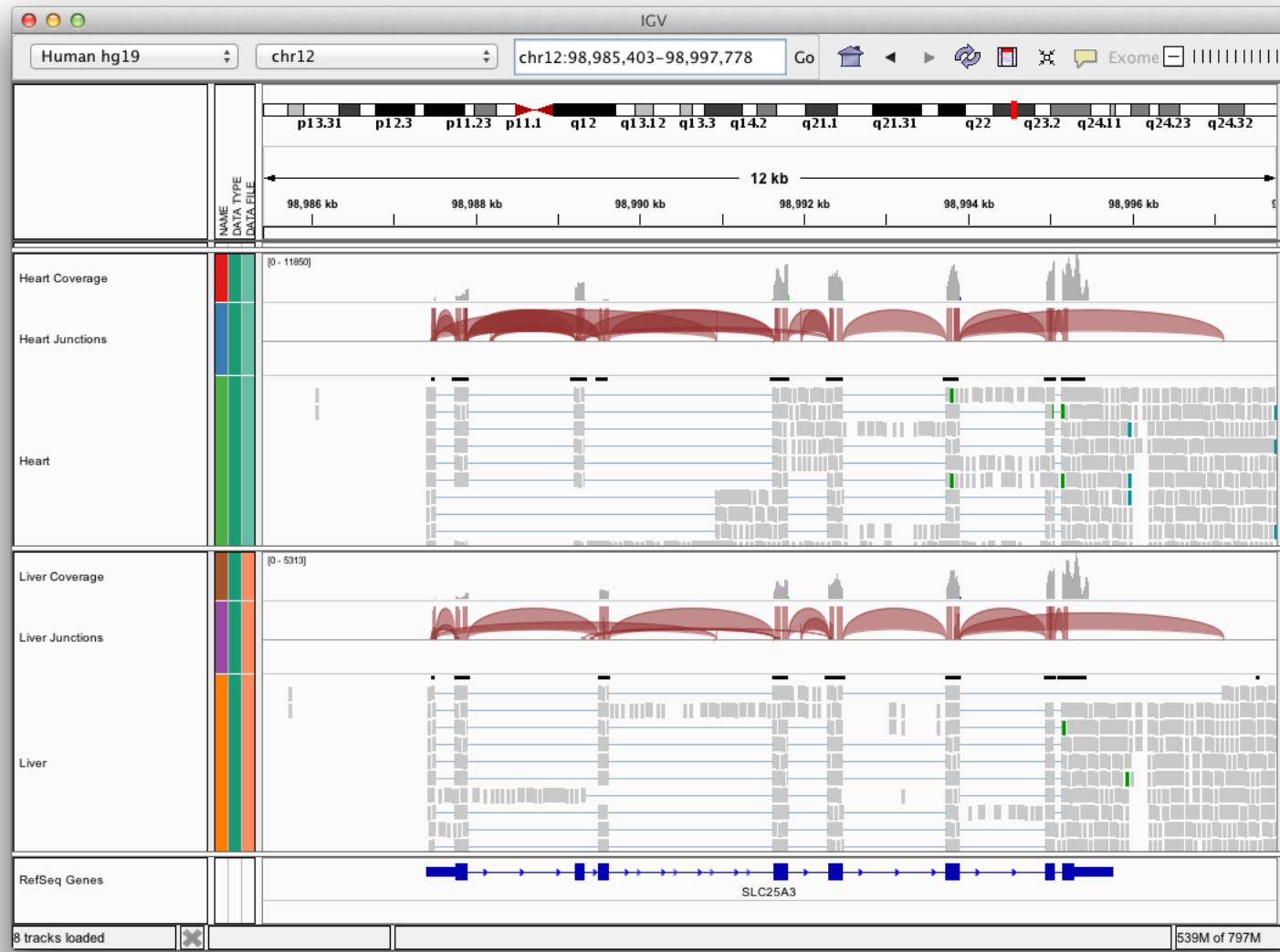
Integrative
Genomics
Viewer
ALIGMENT



RNA-seq alignments



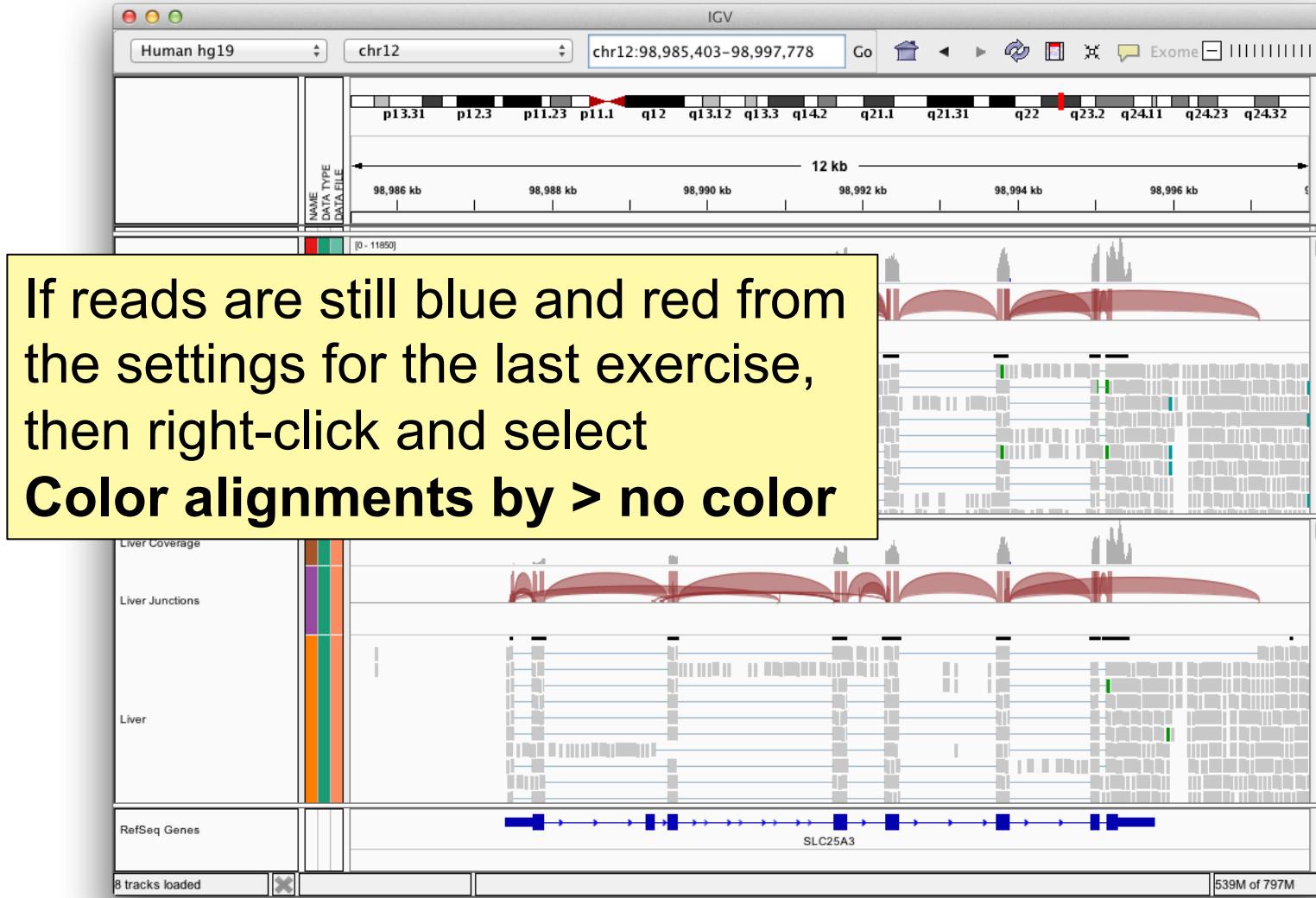
Integrative
Genomics
Viewer
ALIVE



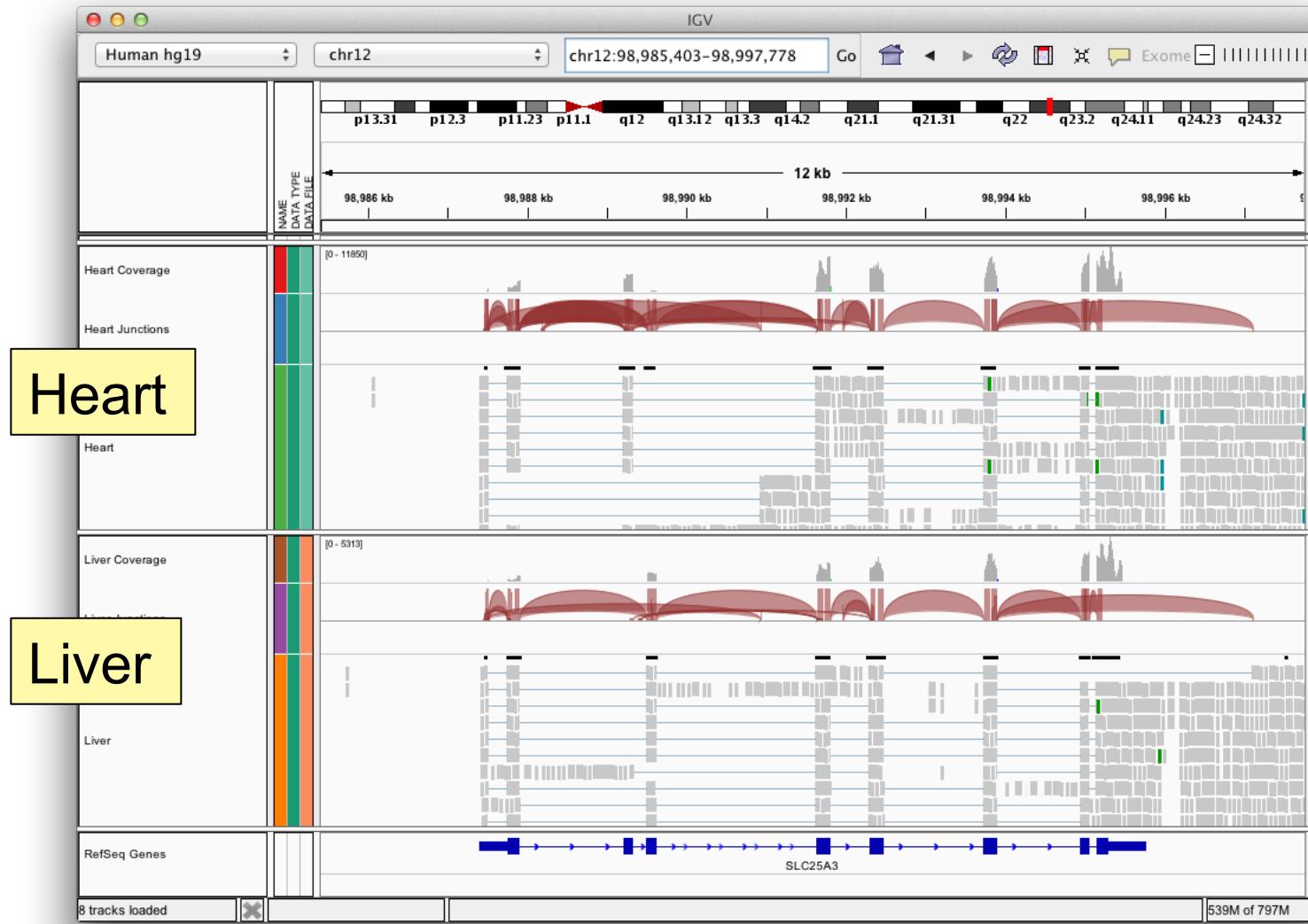
RNA-seq alignments



Integrative
Genomics
Viewer
ALIGMENT



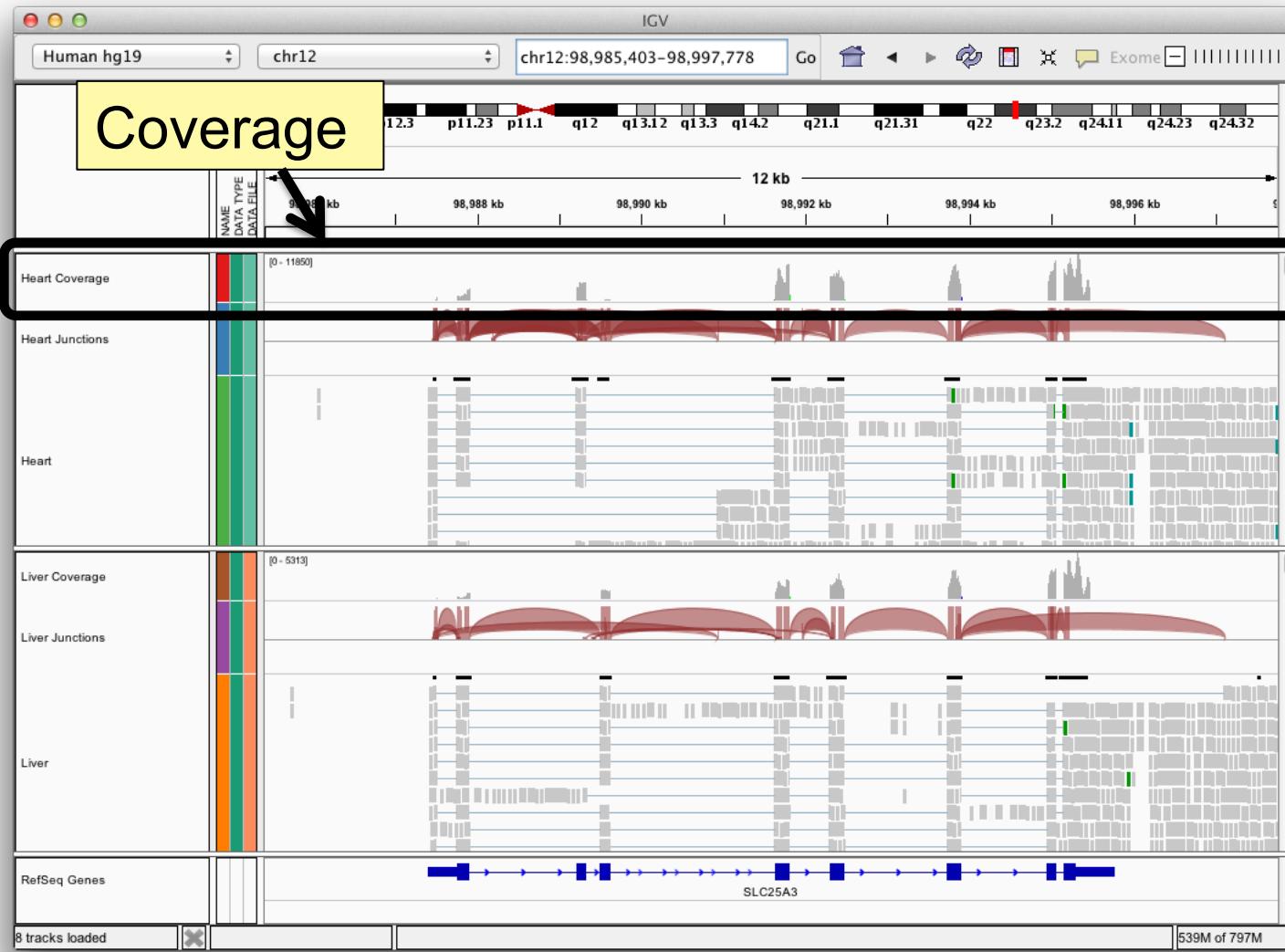
RNA-seq alignments



RNA-seq alignments



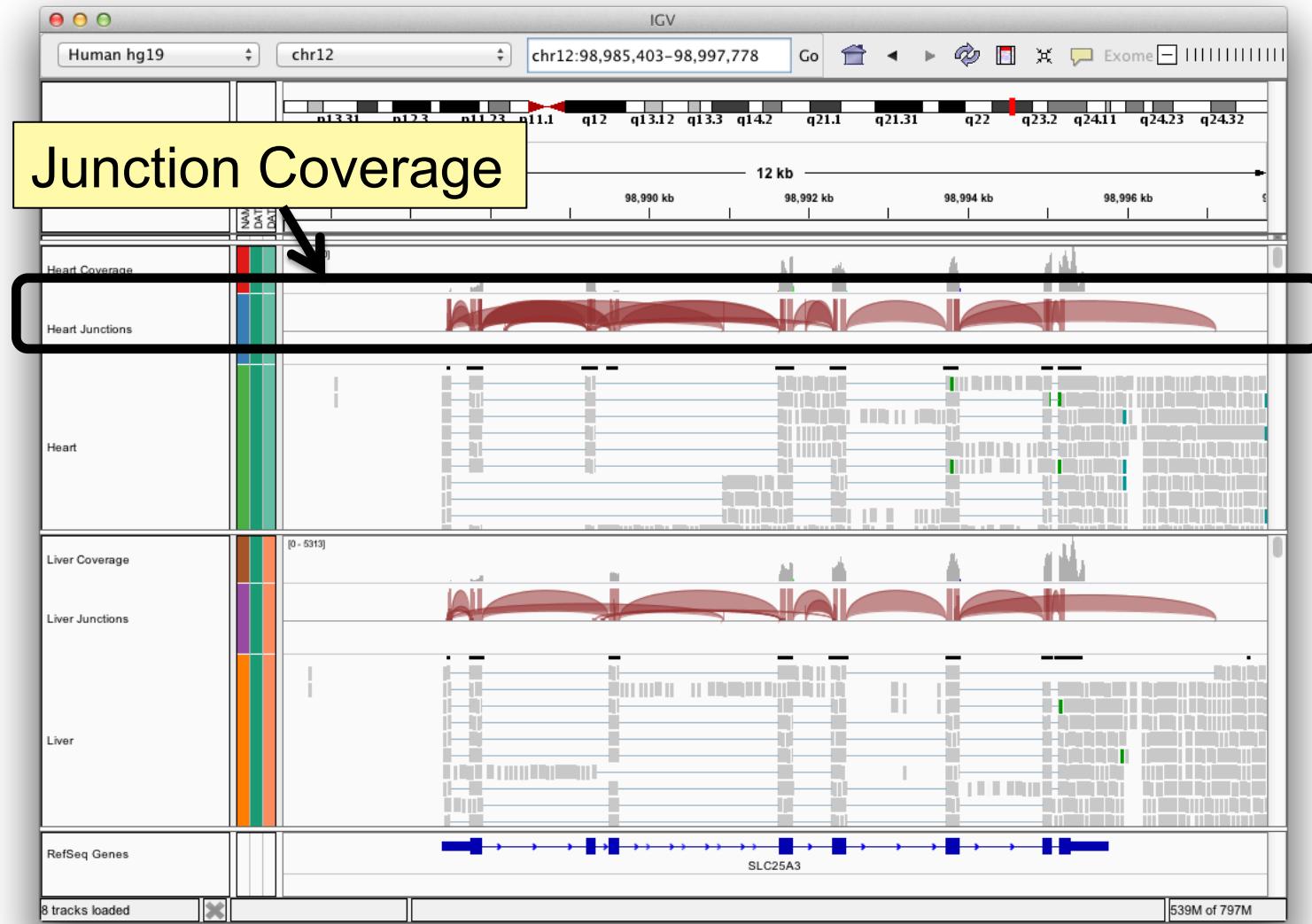
Integrative
Genomics
Viewer



RNA-seq alignments



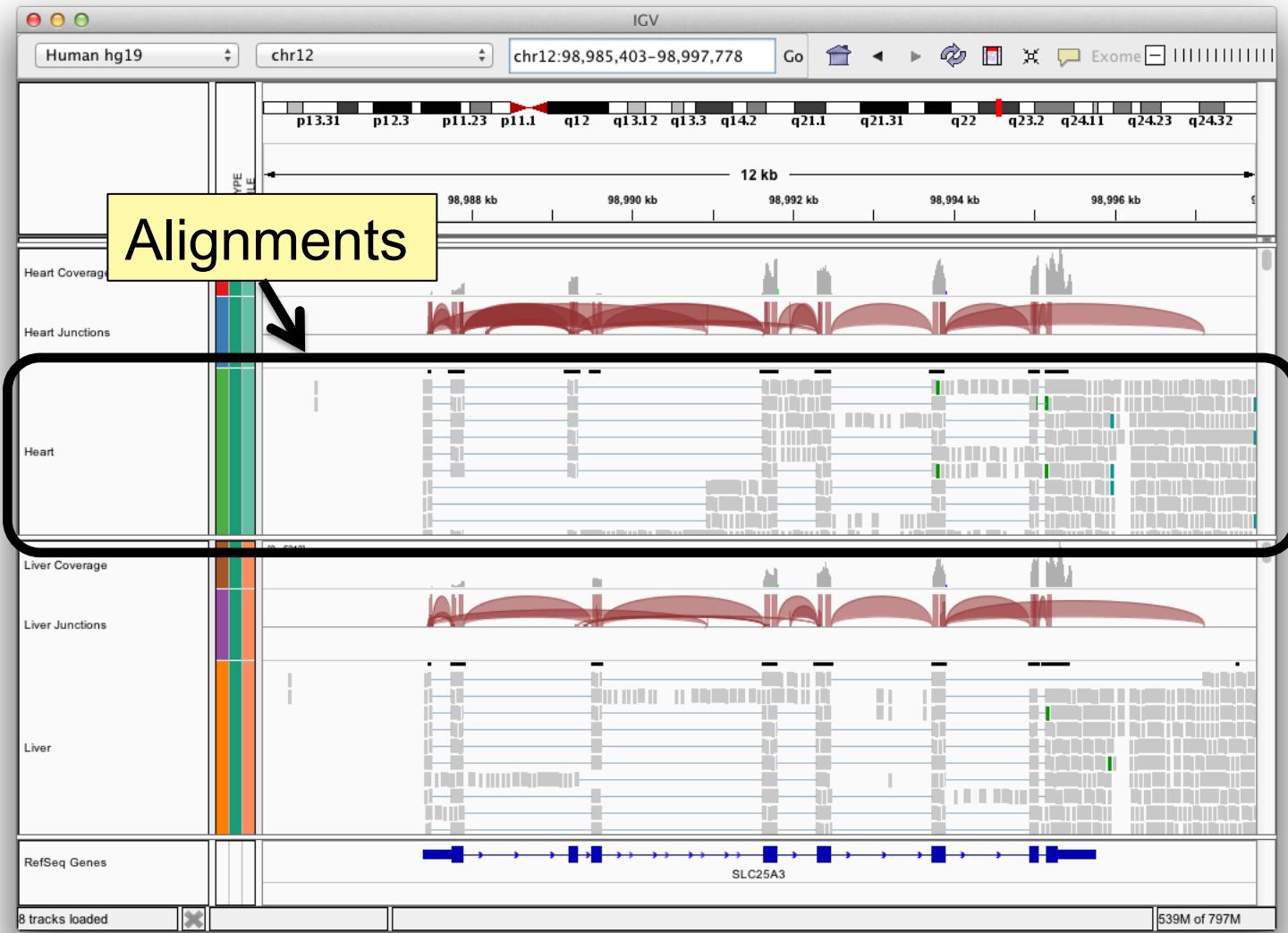
Integrative
Genomics
Viewer
ALIGMENT



RNA-seq alignments



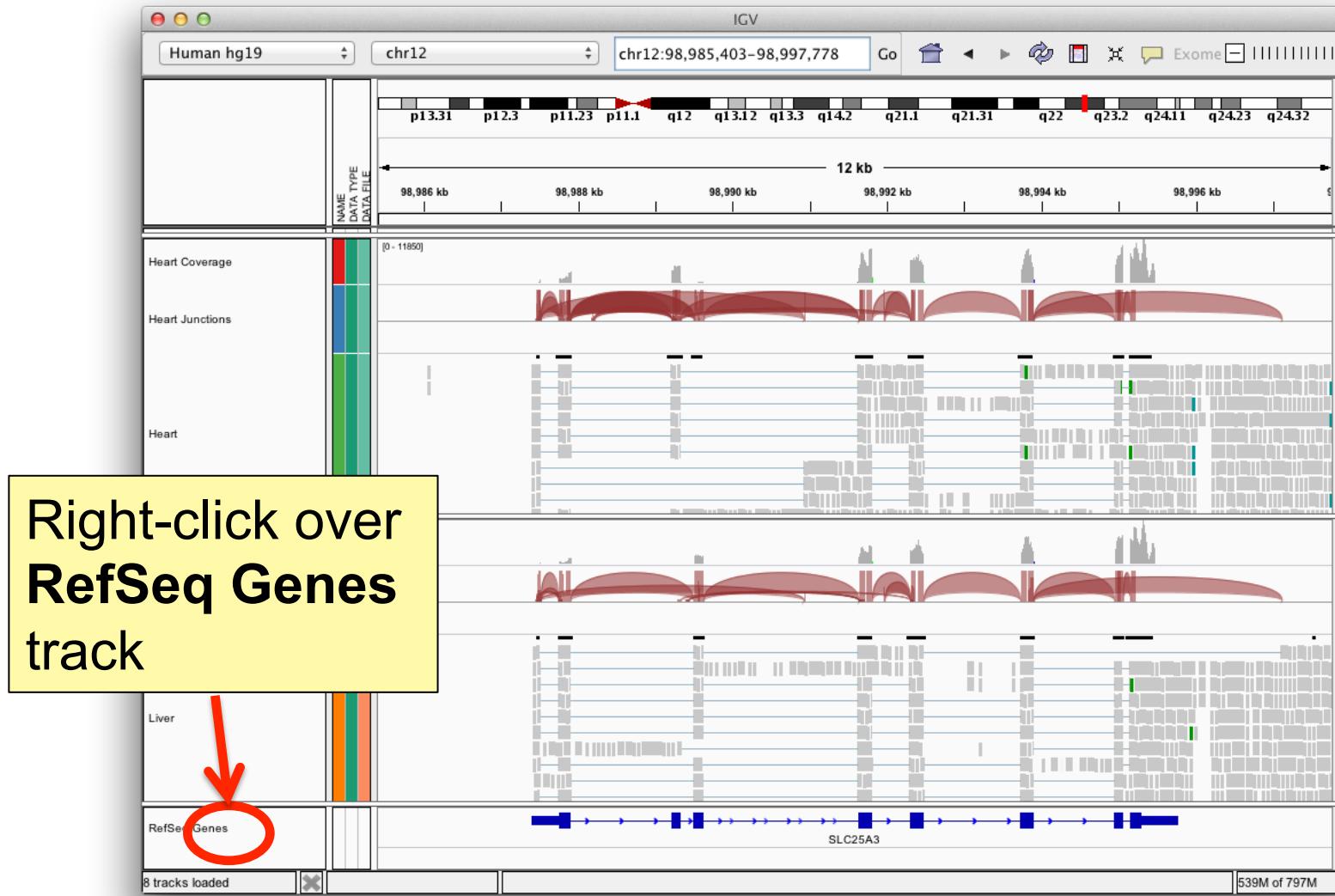
Integrative
Genomics
Viewer



RNA-seq alignments



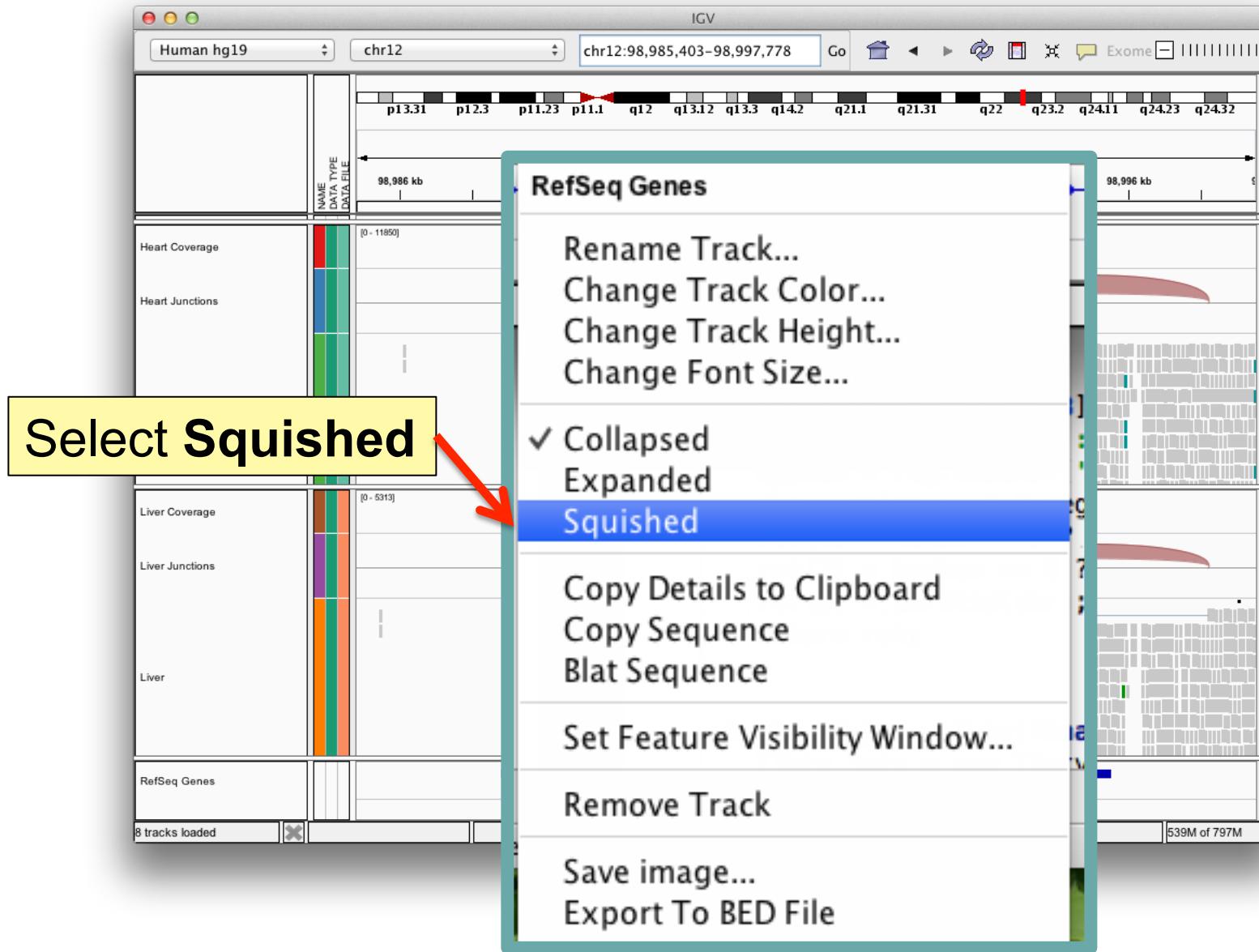
Integrative
Genomics
Viewer
ALGVIEWER



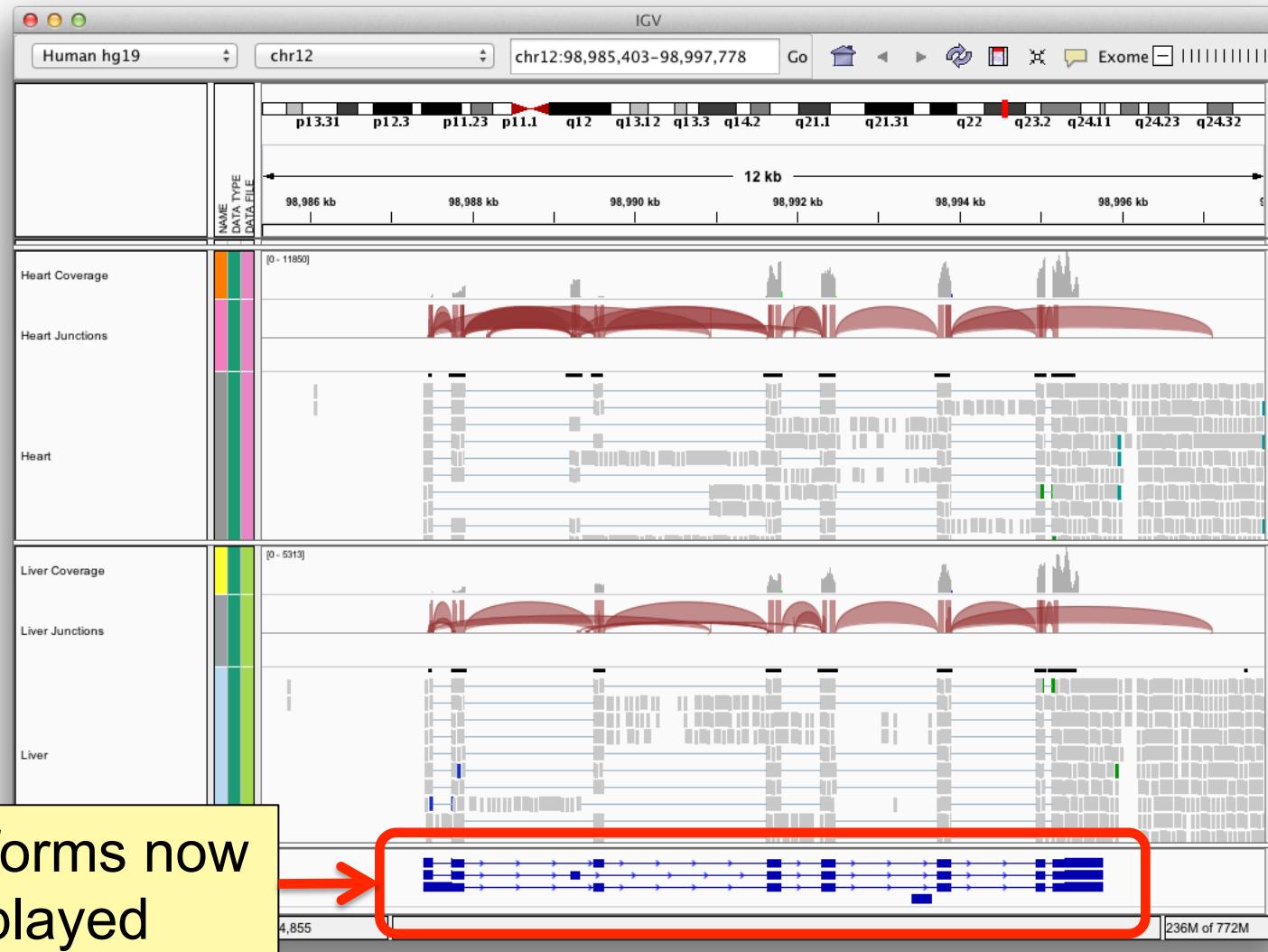
RNA-seq alignments



Integrative
Genomics
Viewer
ALGEMEL



RNA-seq alignments

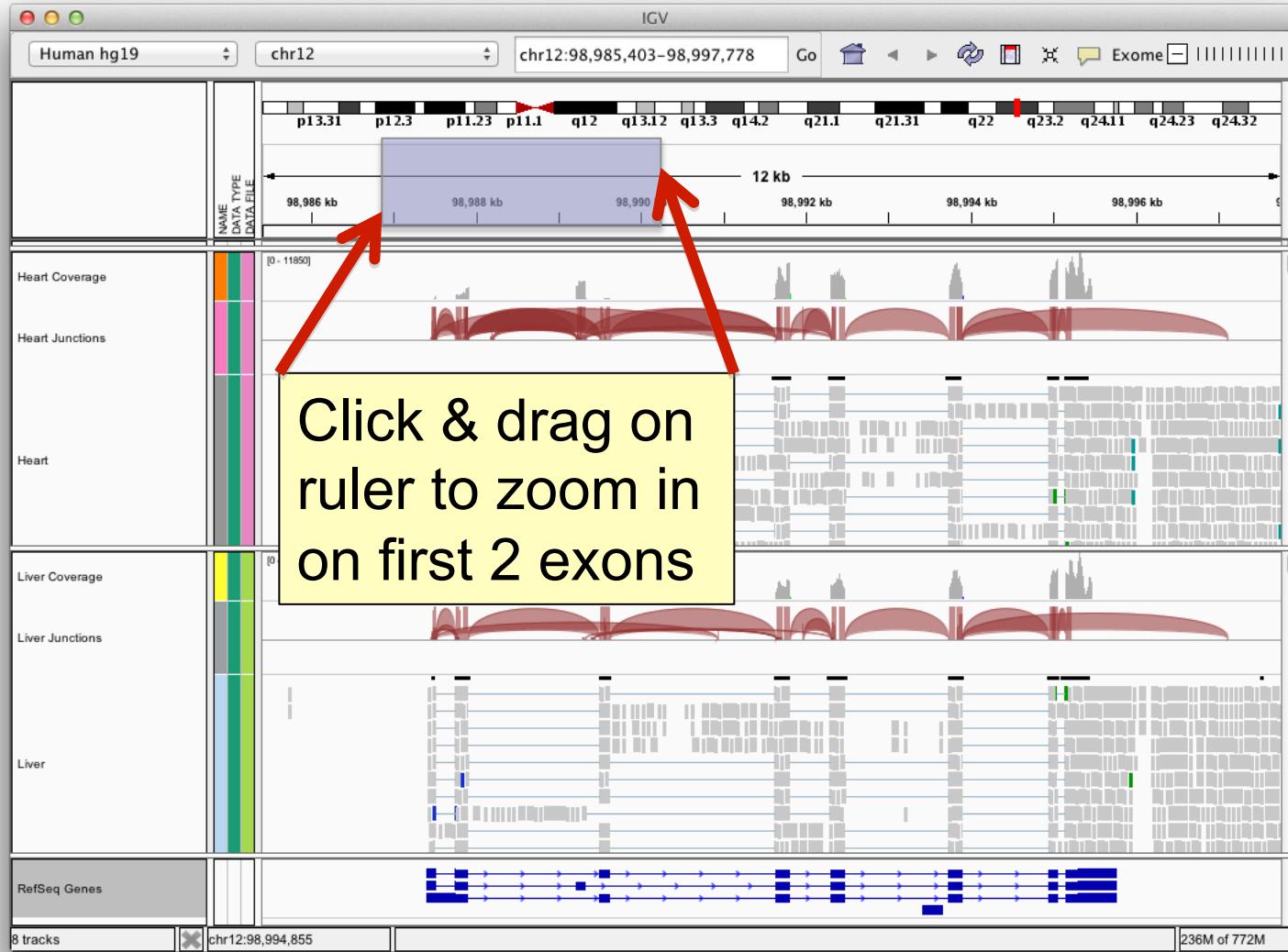


Isoforms now
displayed

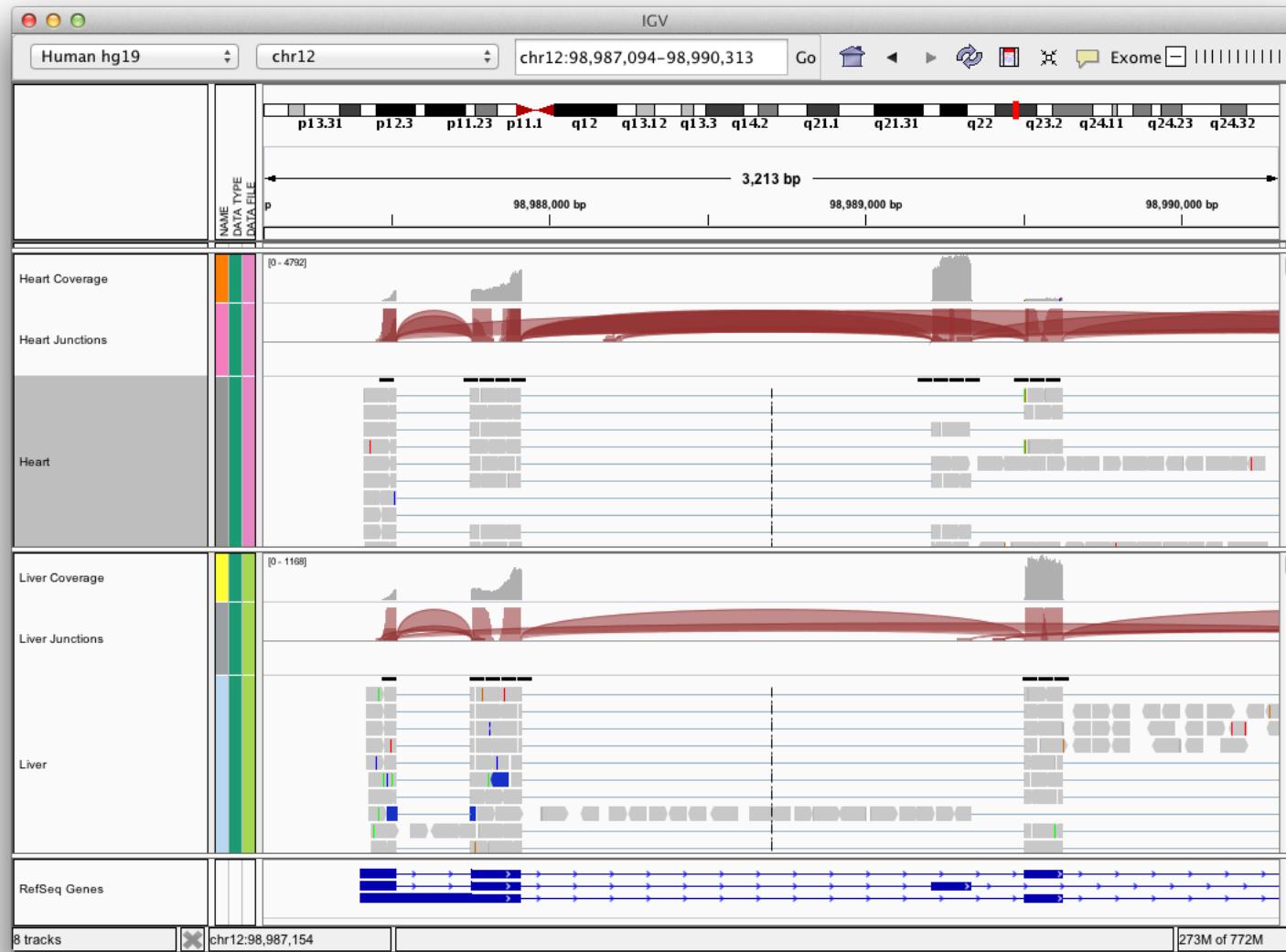
RNA-seq alignments



Integrative
Genomics
Viewer
ALIGMENT



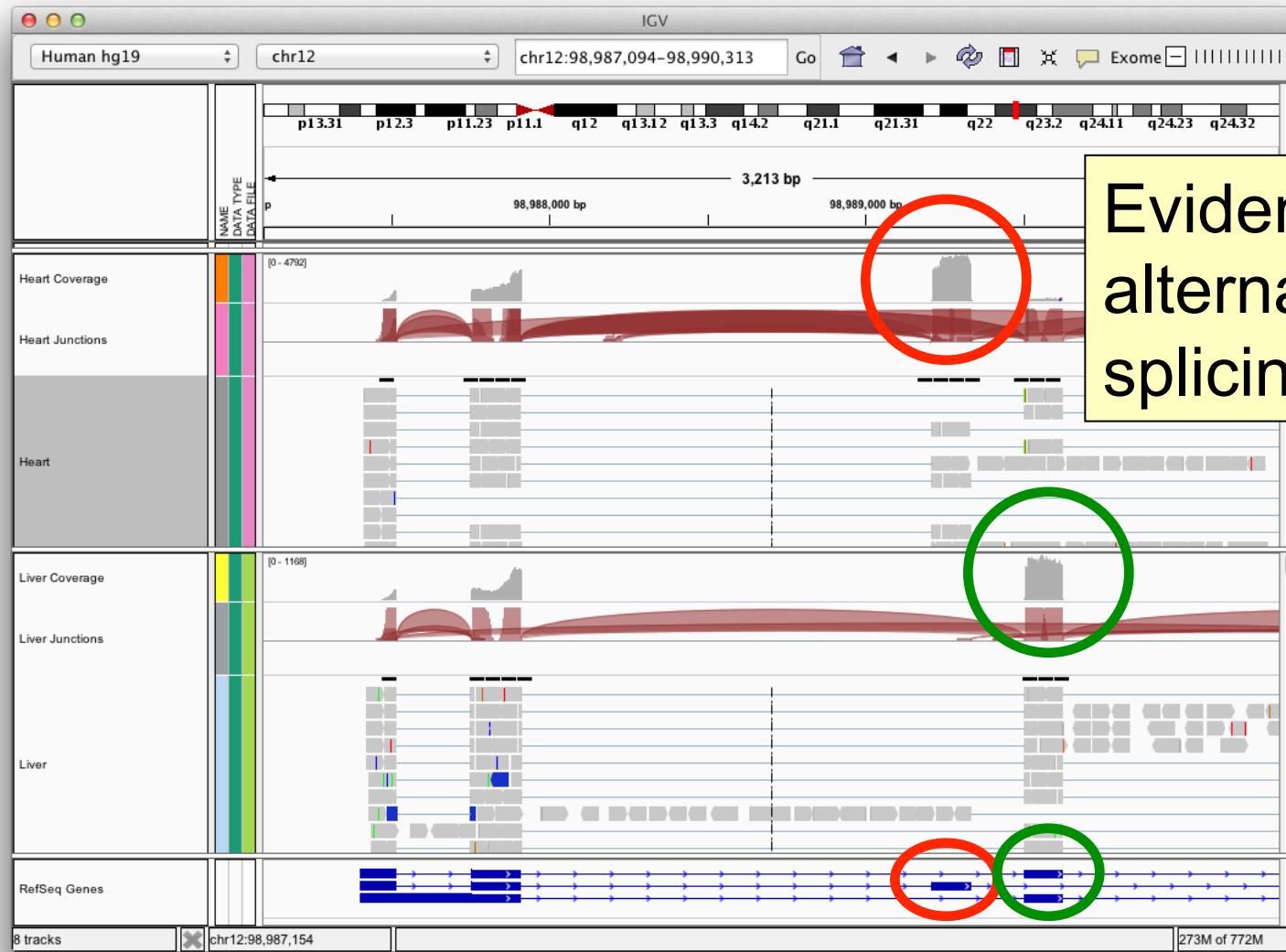
RNA-seq alignments



RNA-seq alignments



Integrative
Genomics
Viewer
ALIVEL



Evidence of
alternative
splicing

Sashimi plot



Viewing RNA splicing with Sashimi Plots

Reference:

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

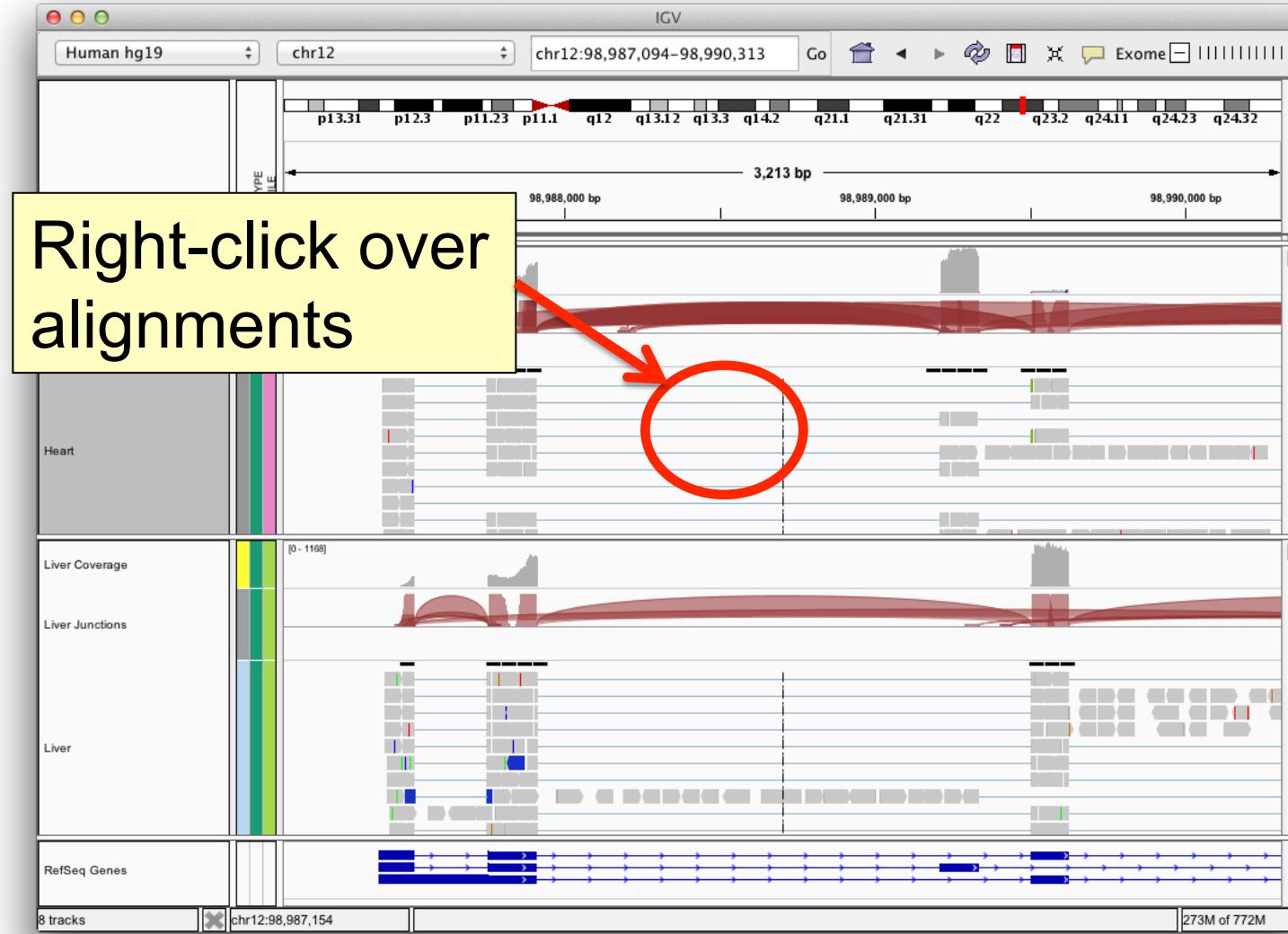
***Sashimi plots: Quantitative visualization of
RNA sequencing read alignments.***

arXiv:1306.3466 [q-bio.GN], 2013

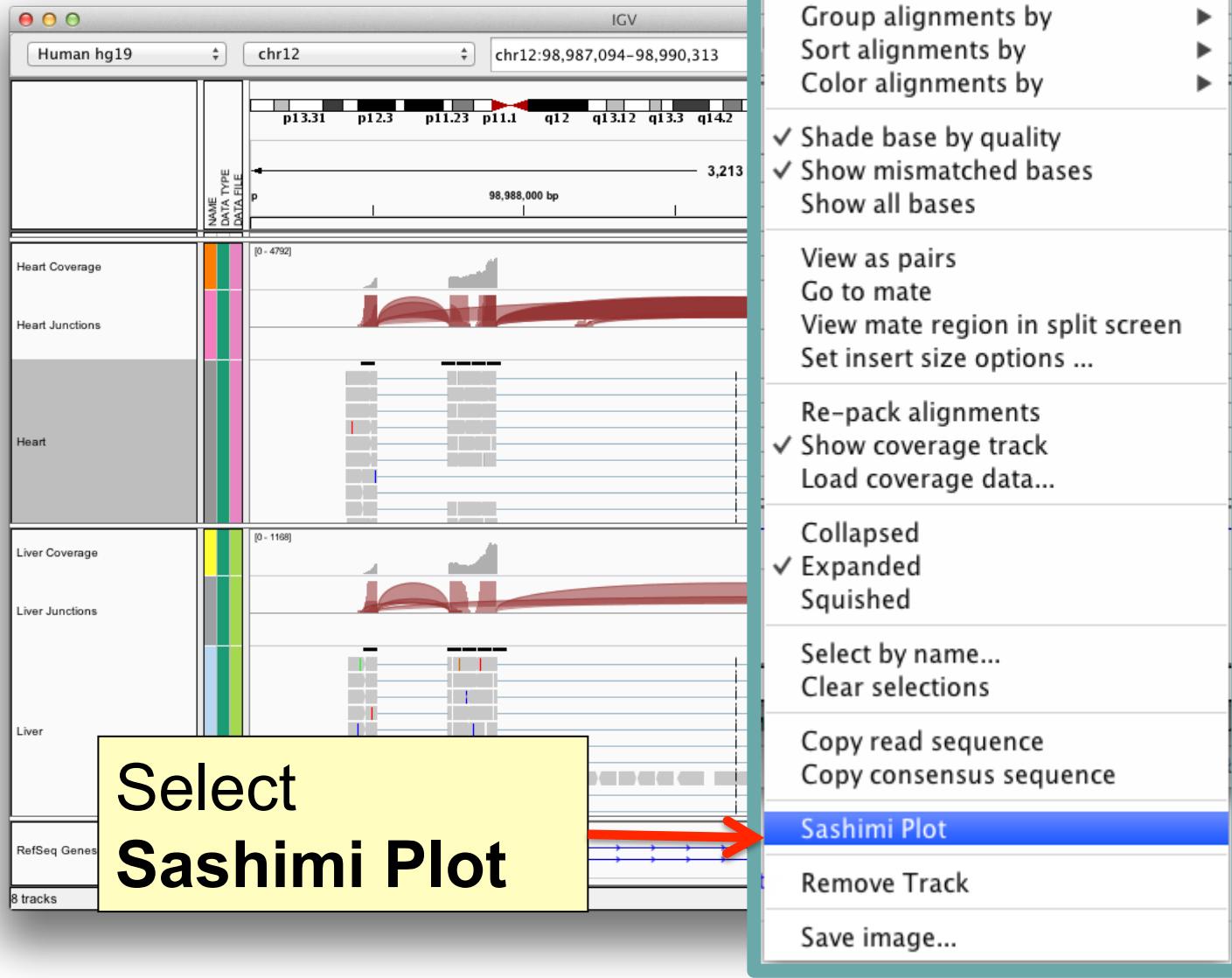
RNA-seq alignments



Integrative
Genomics
Viewer
ALMEL



RNA-seq alignments

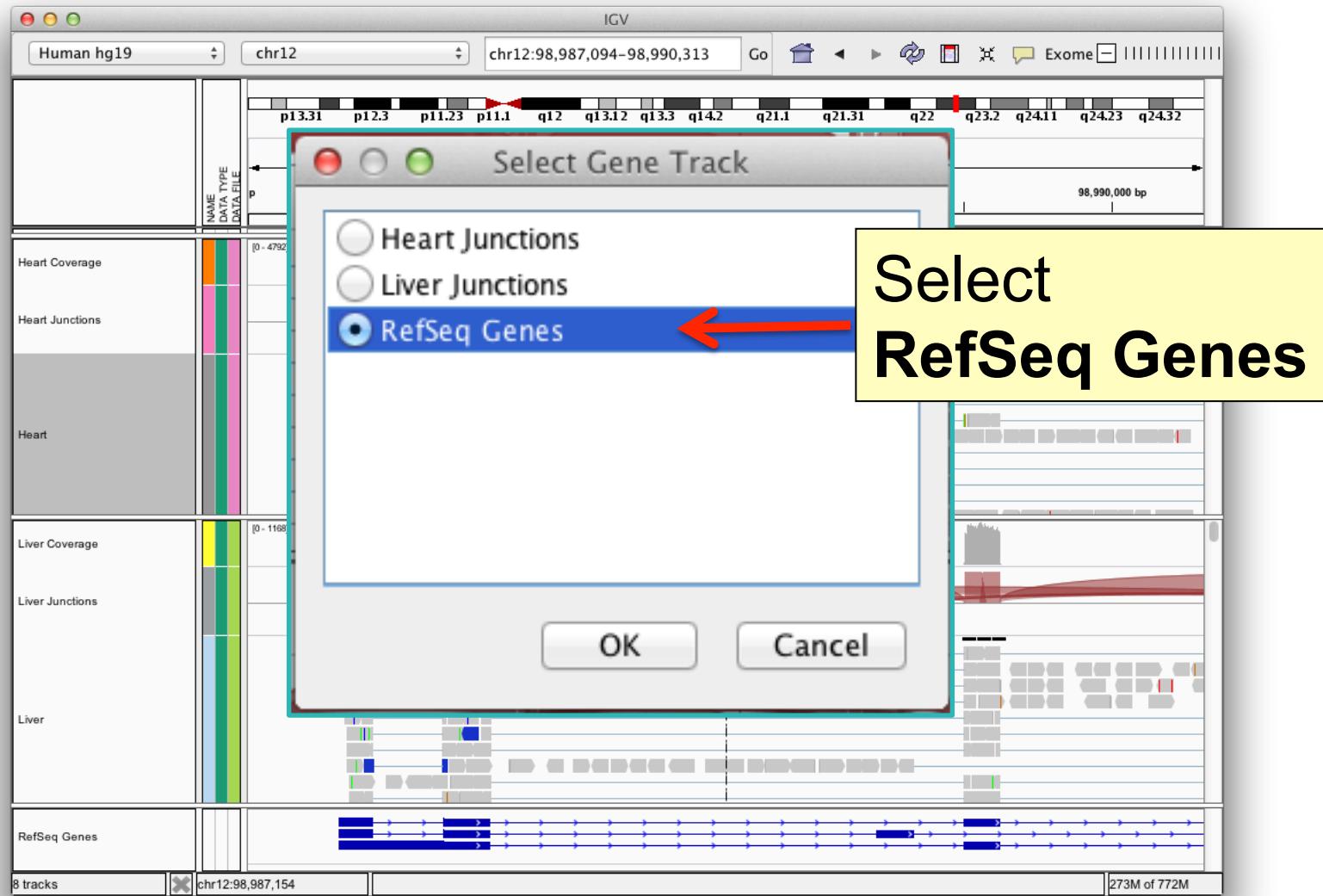


Integrative
Genomics
Viewer
ALMEL

RNA-seq alignments



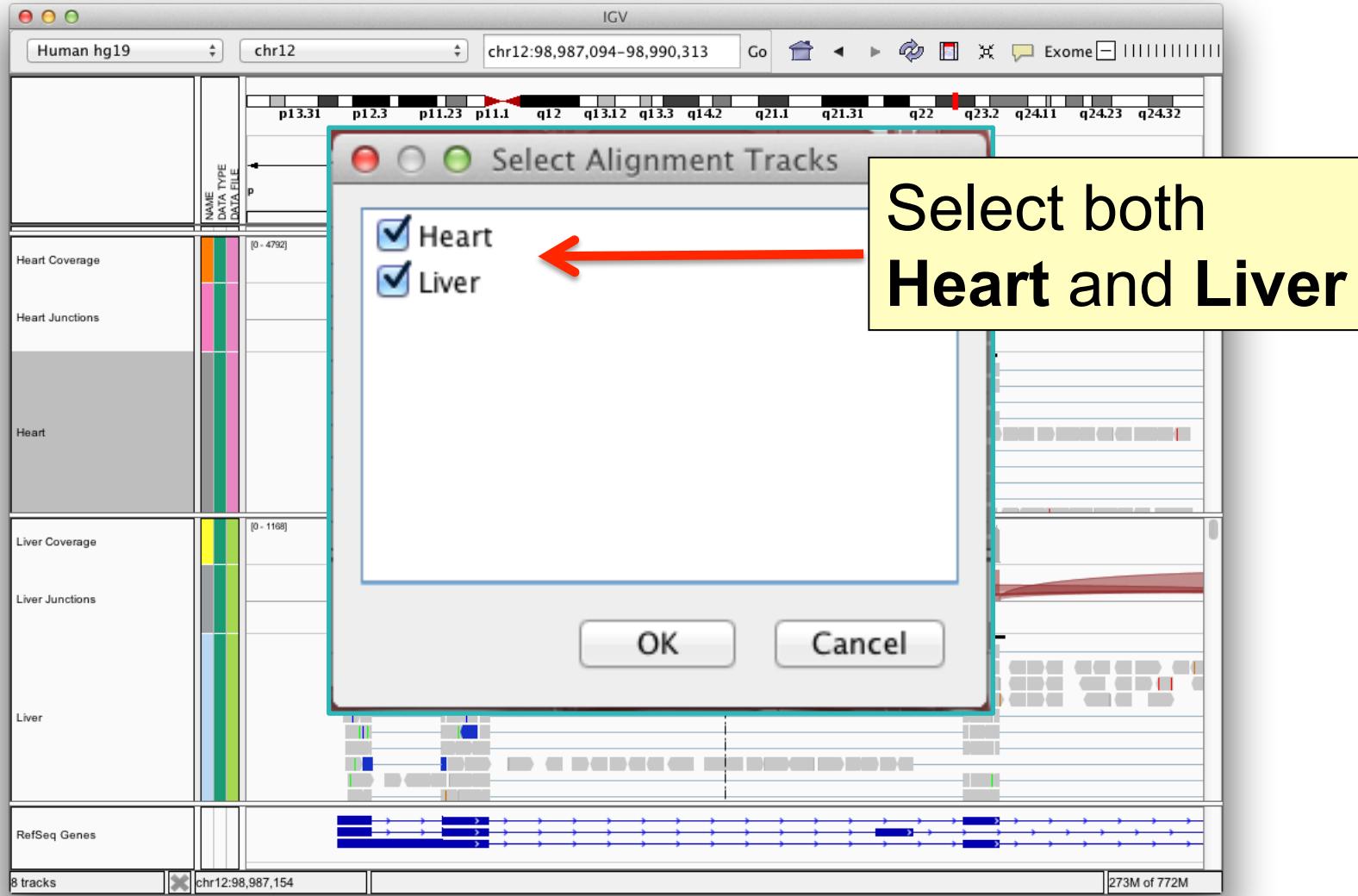
Integrative
Genomics
Viewer
ALMEL



RNA-seq alignments



Integrative
Genomics
Viewer
ALIVE



RNA-seq alignments



Integrative
Genomics
Viewer
ALGEME



RNA-seq alignments



Integrative
Genomics
Viewer
ALMEL



A little busy.
Let's filter out
low-count events

RNA-seq alignments



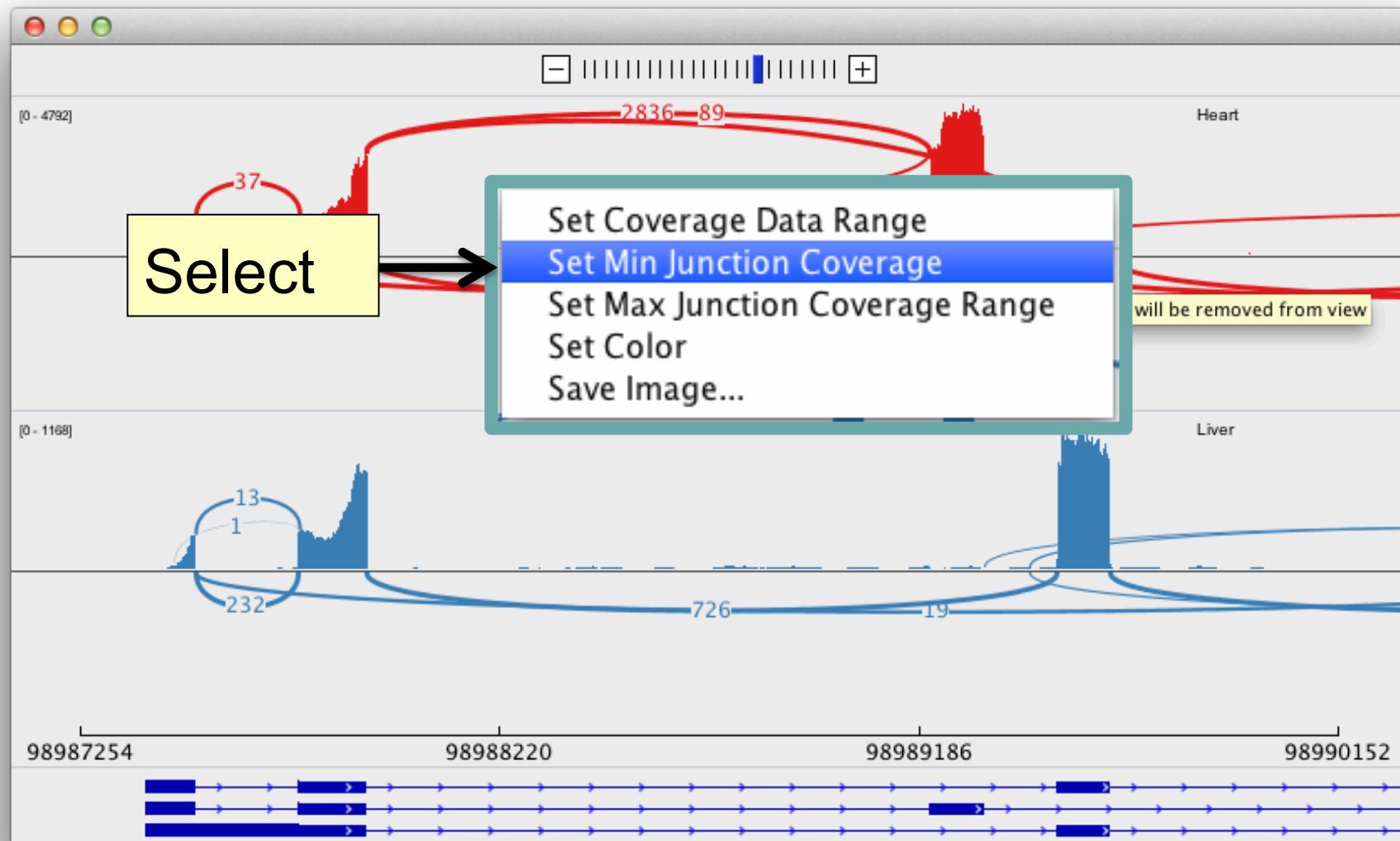
Integrative
Genomics
Viewer
IGV



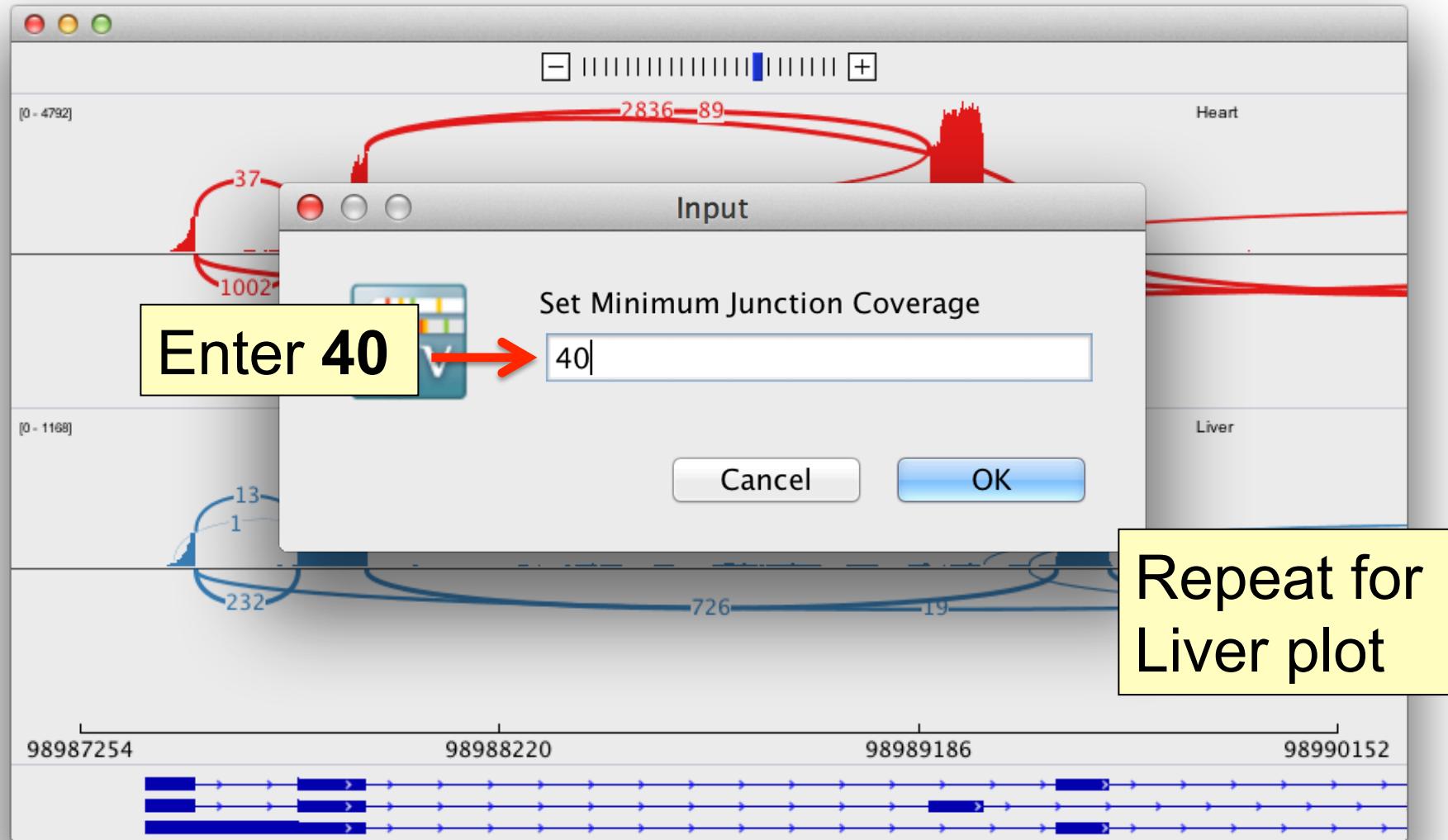
RNA-seq alignments



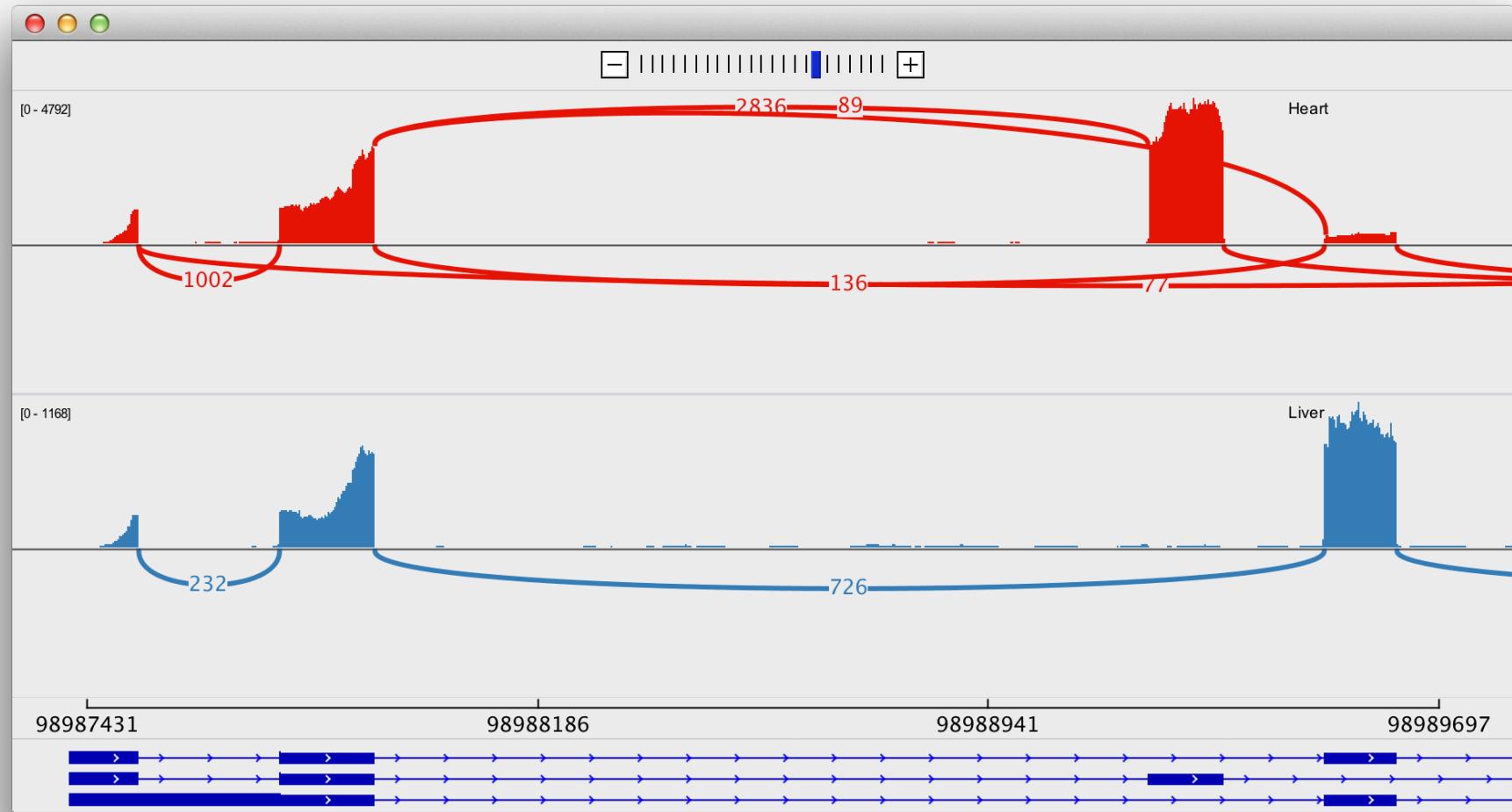
Integrative
Genomics
Viewer
ALMEL



RNA-seq alignments



RNA-seq alignments



Viewing RNA-Seq Data



- Tune preferences
- Examine tissue-specific alternative splicing
- **Supplementary views**

igvtools - coverage

Tophat - insertions, deletions, junctions

Cufflinks - transcripts

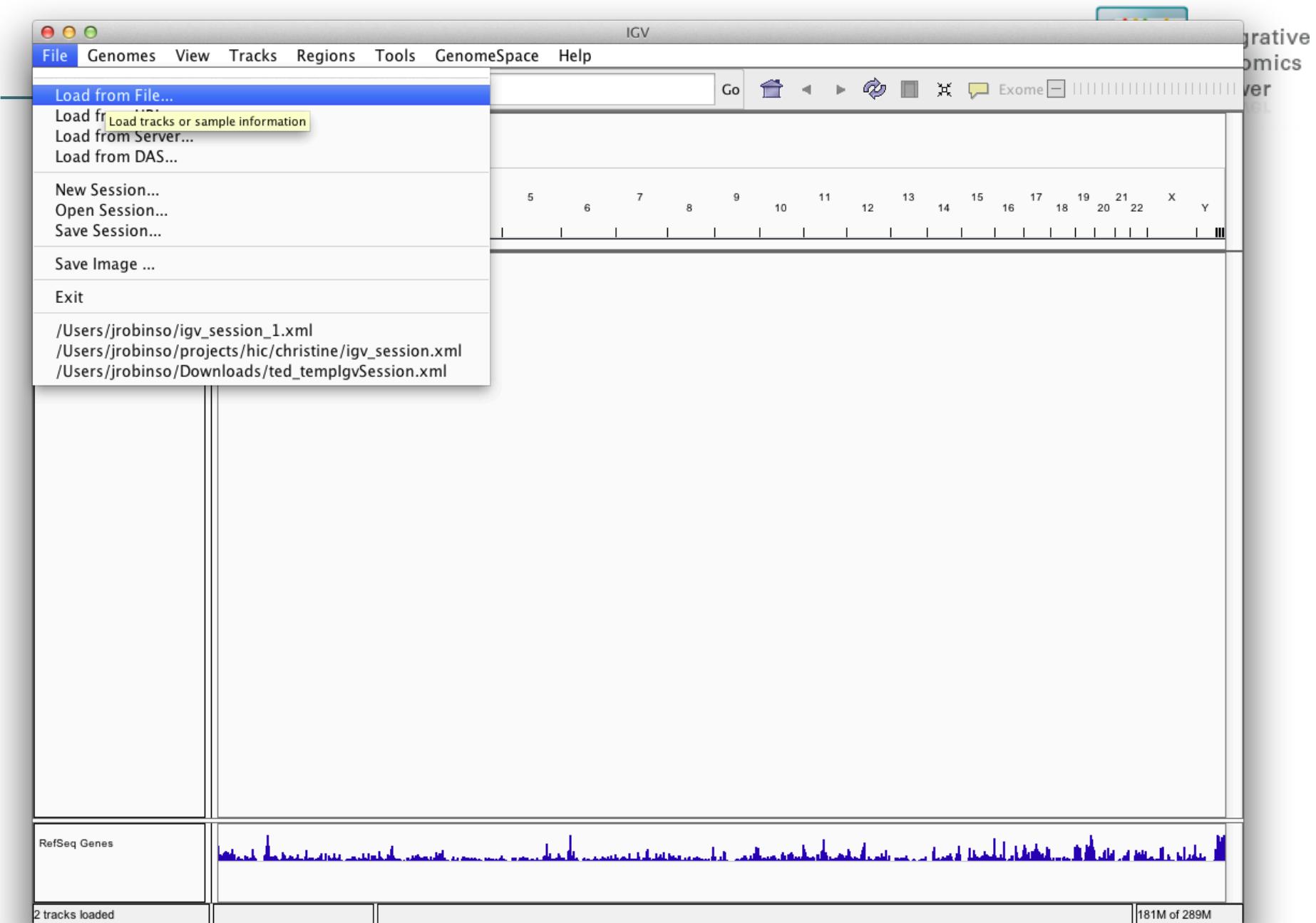
SETUP

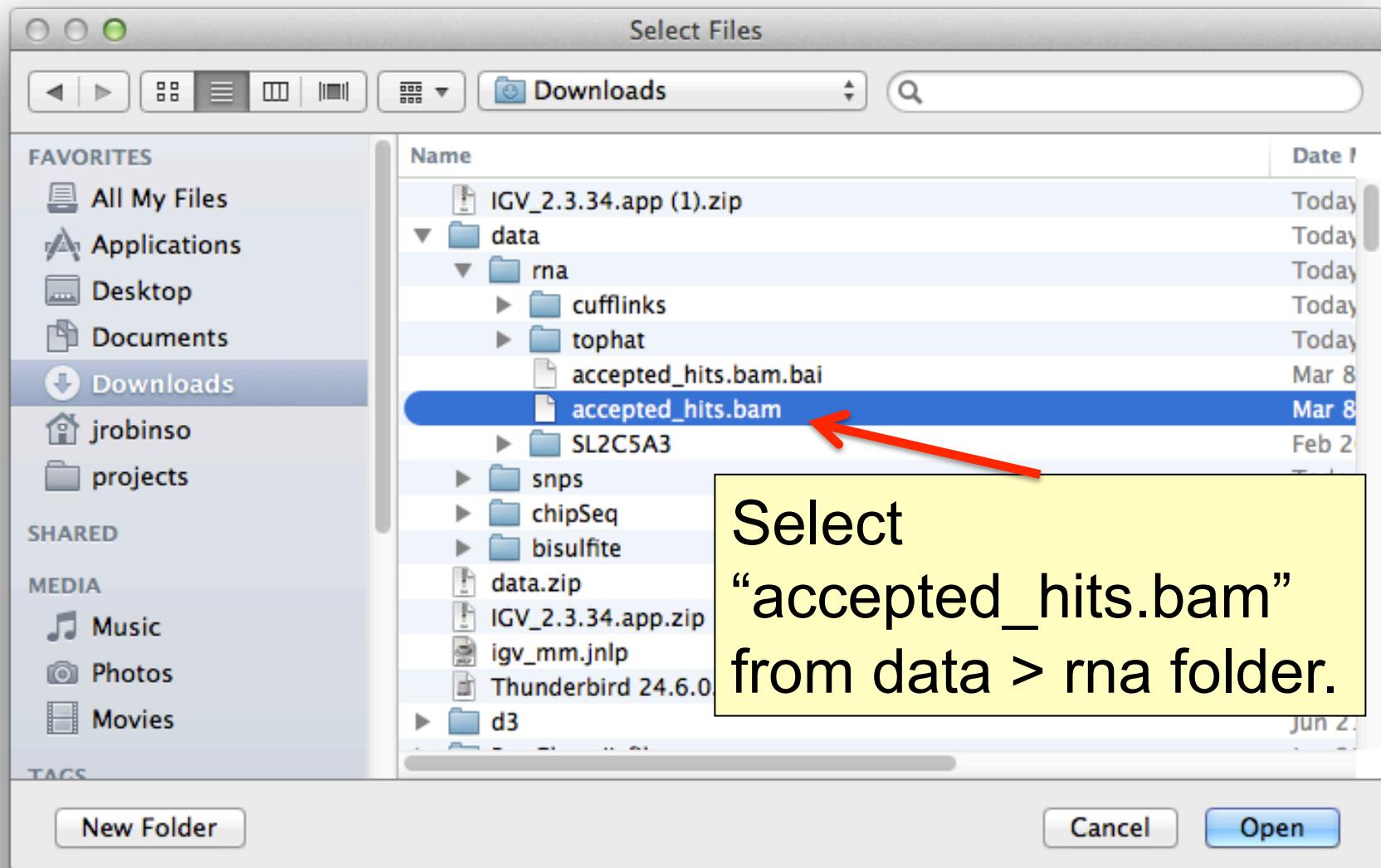


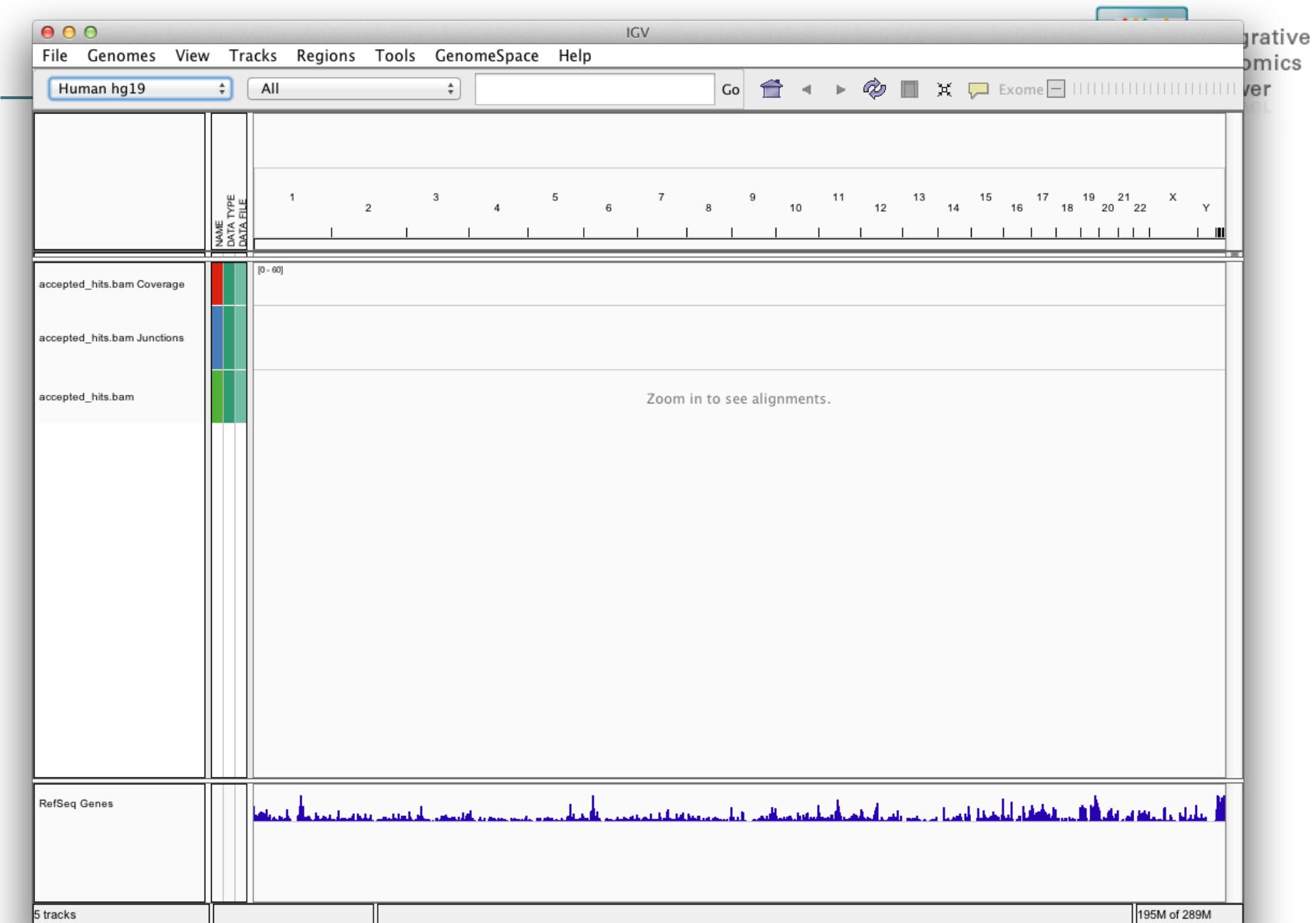
Select “File > New session”

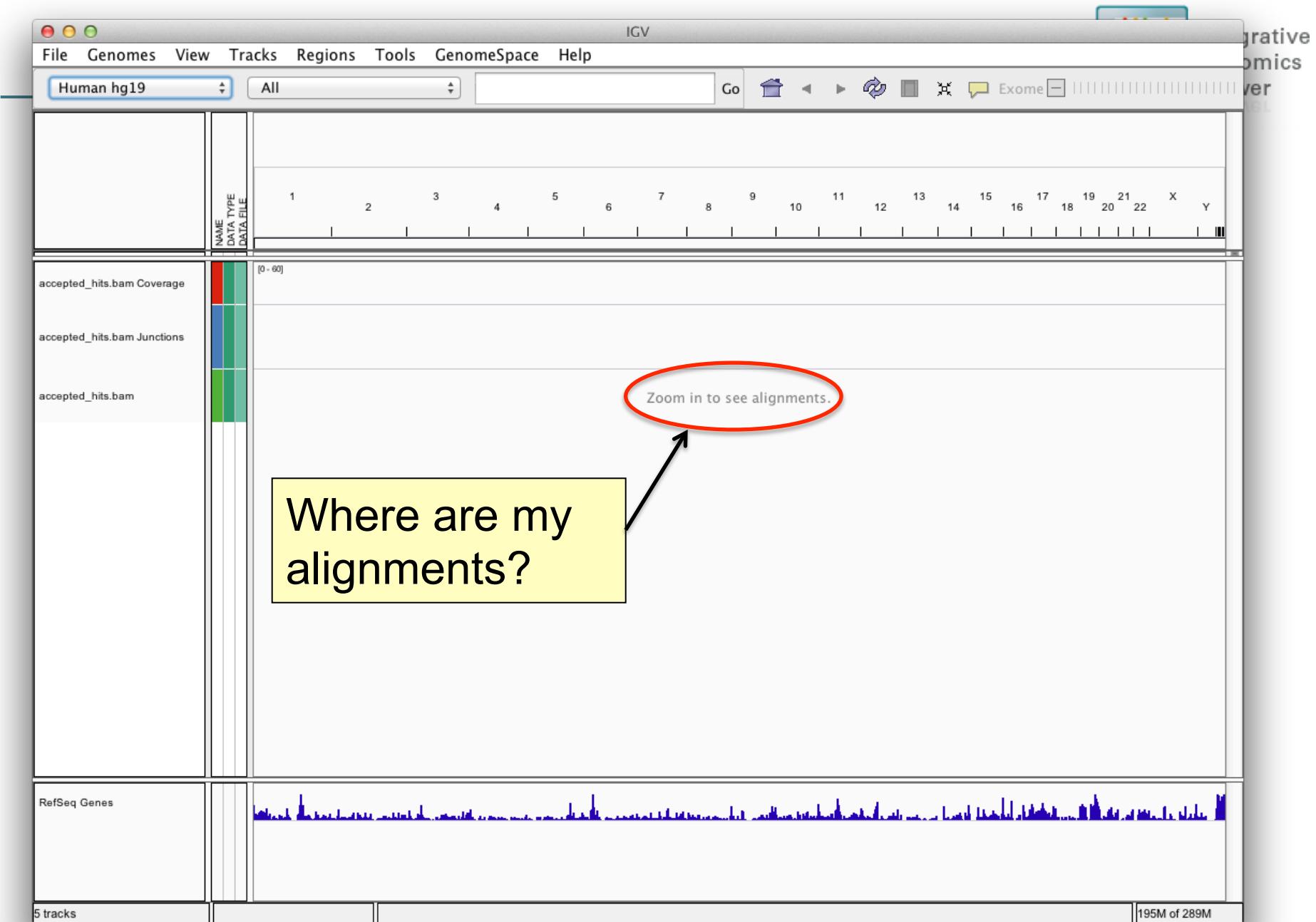
Click the “Home” icon to return to whole genome view.

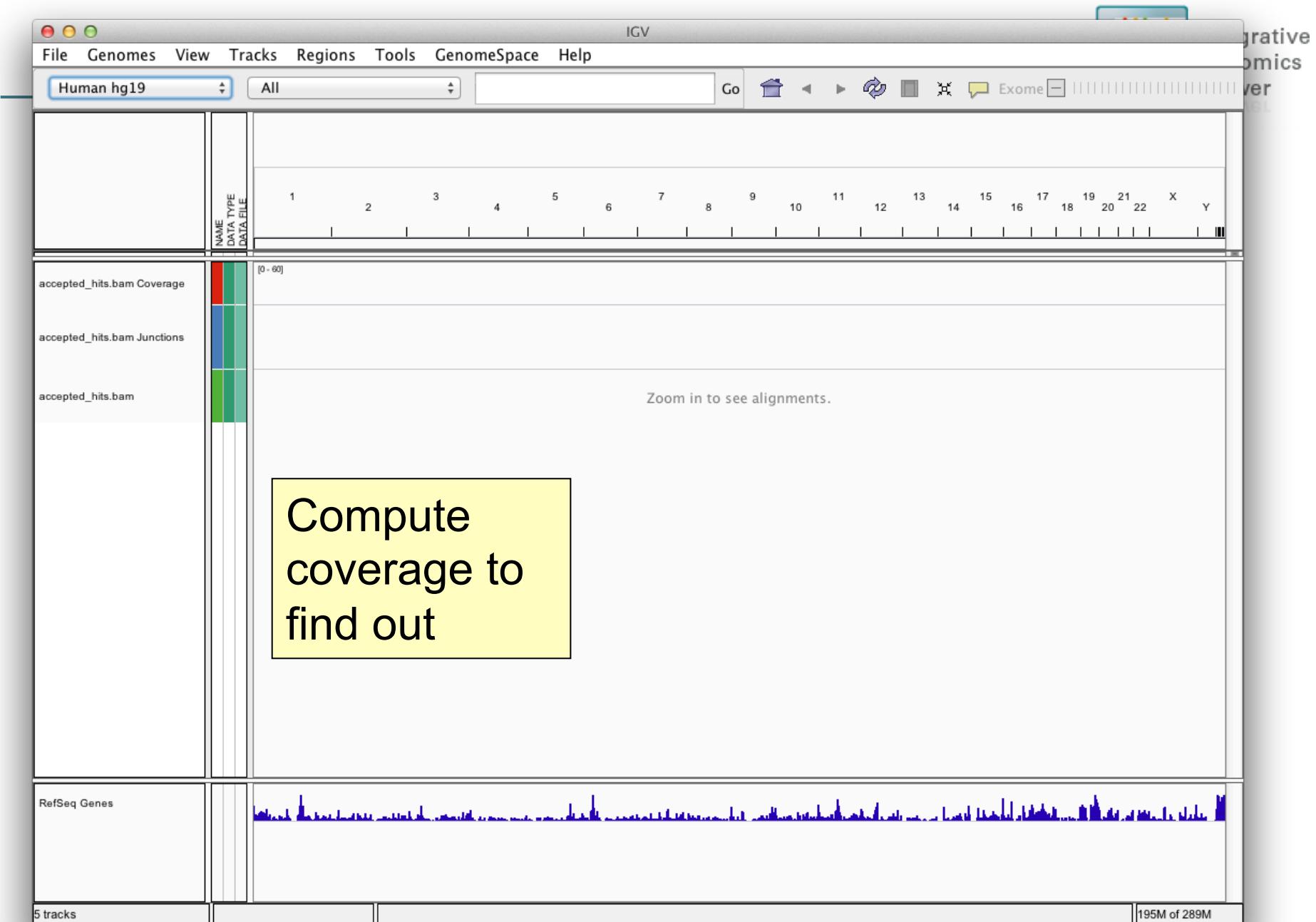


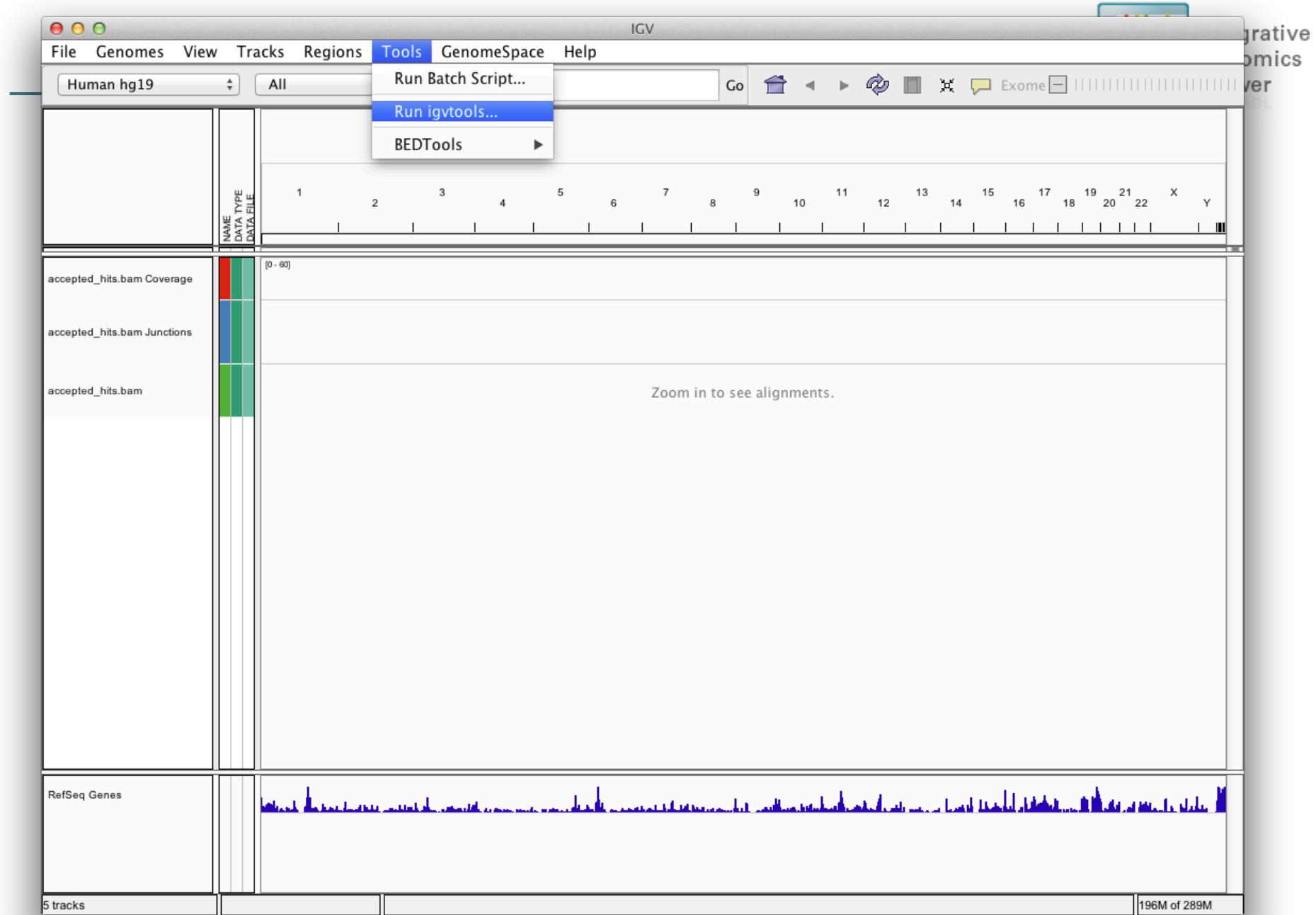












Command ✓ Tile
Input File Count
Output File
Genome hg19

Browse
Browse
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

Command Count

Input File

Output File

Genome hg19

Tile and Count Options

Zoom Levels

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

Probe to Loci Mapping

Window Size 25

Sort Options

Temp Directory

Max Records 500000

Messages

Command Count

Input File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam

Output File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam.tdf

Genome hg19

Tile and Count Options

Zoom Levels 7

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

Probe to Loci Mapping

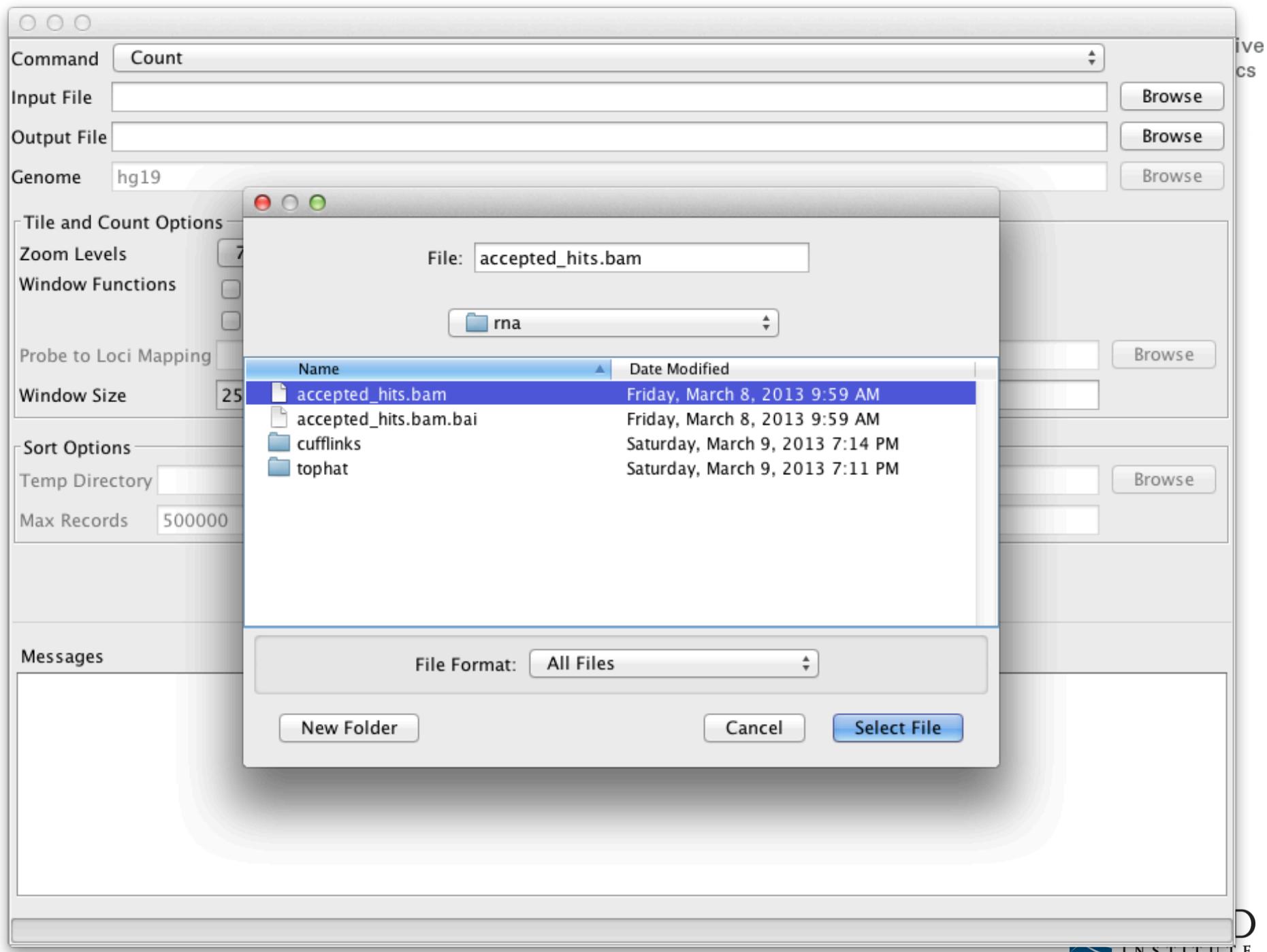
Window Size 25

Sort Options

Temp Directory

Max Records 500000

Messages



Command Count

Input File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam

Output File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam.tdf

Genome hg19

Tile and Count Options

Zoom Levels 7

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

Probe to Loci Mapping

Window Size 25

Sort Options

Temp Directory

Max Records 500000

Messages

Command Count

Input File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam

Output File /Users/jrobinso/projects/tutorial/rna/accepted_hits.bam.tdf

Genome hg19

Tile and Count Options

Zoom Levels 7

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

Probe to Loci Mapping

Window Size 25

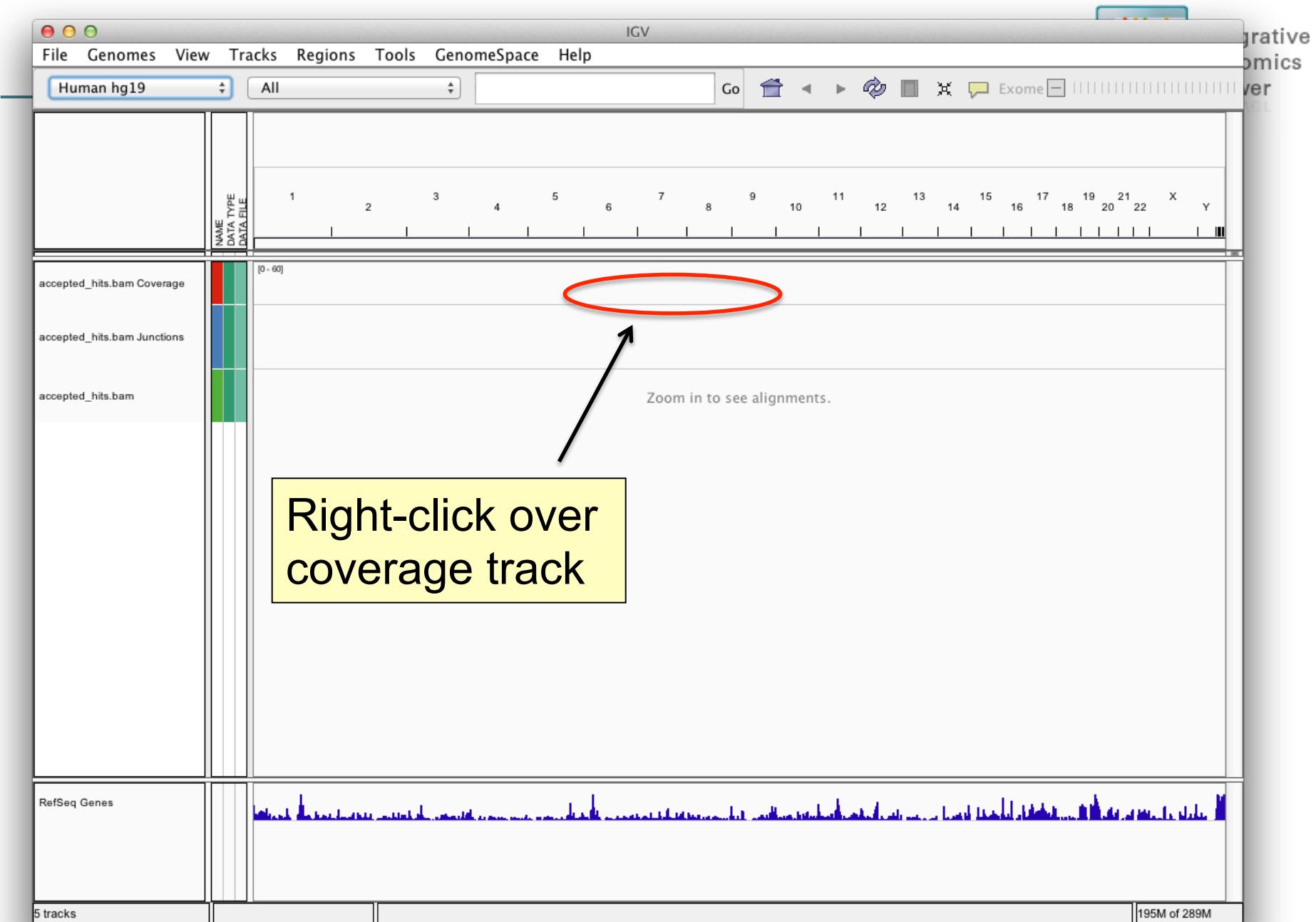
Sort Options

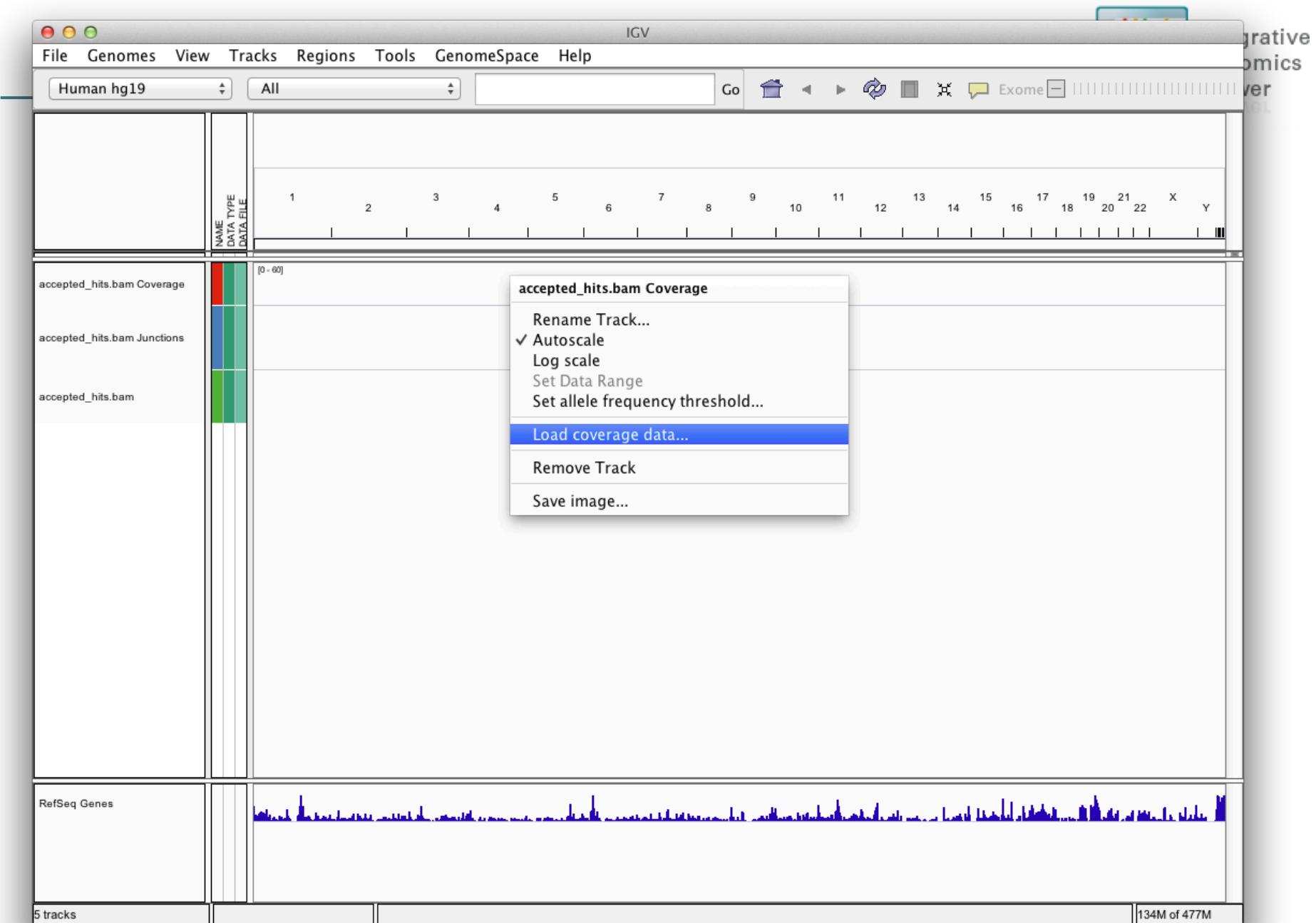
Temp Directory

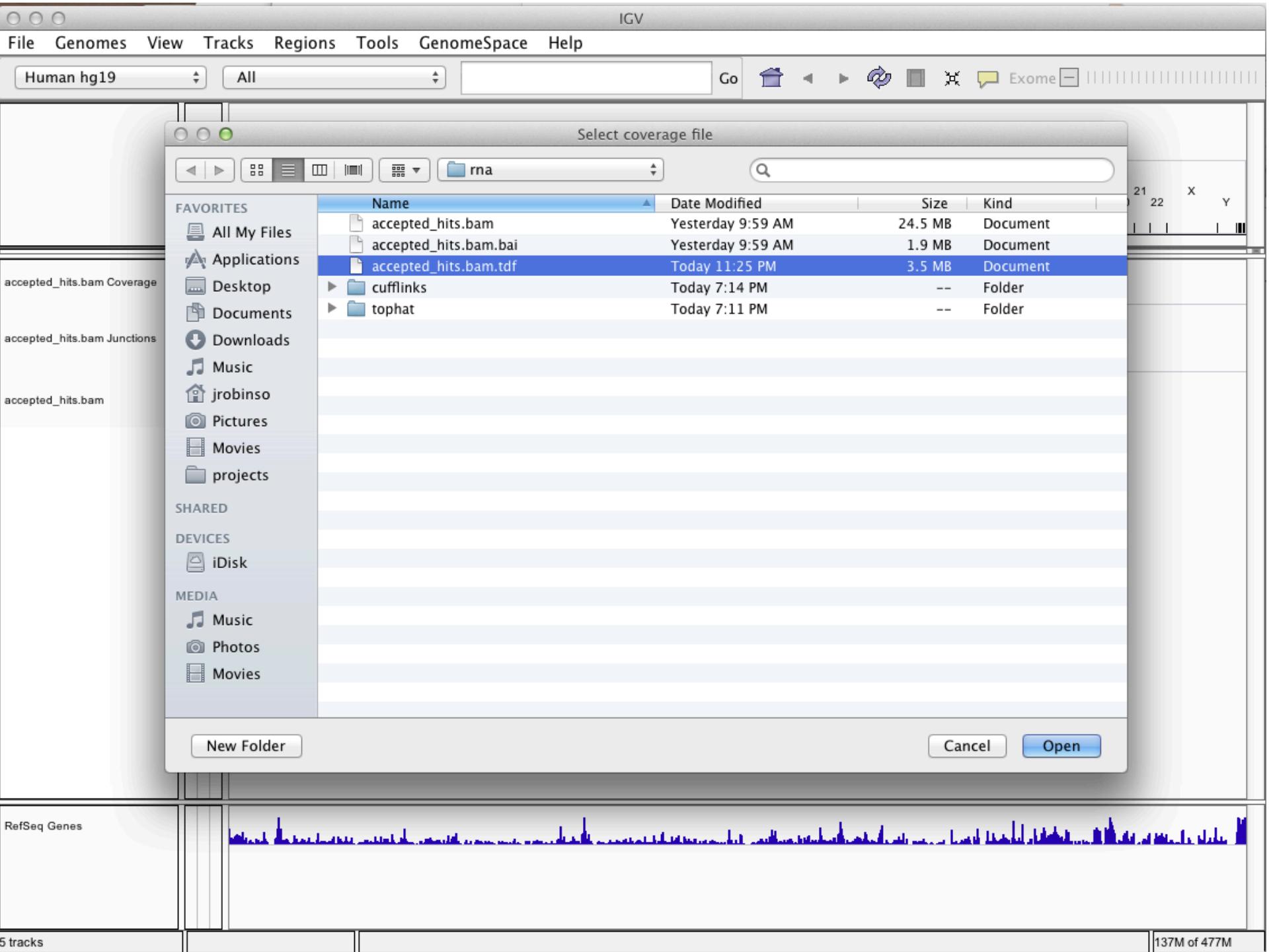
Max Records 500000

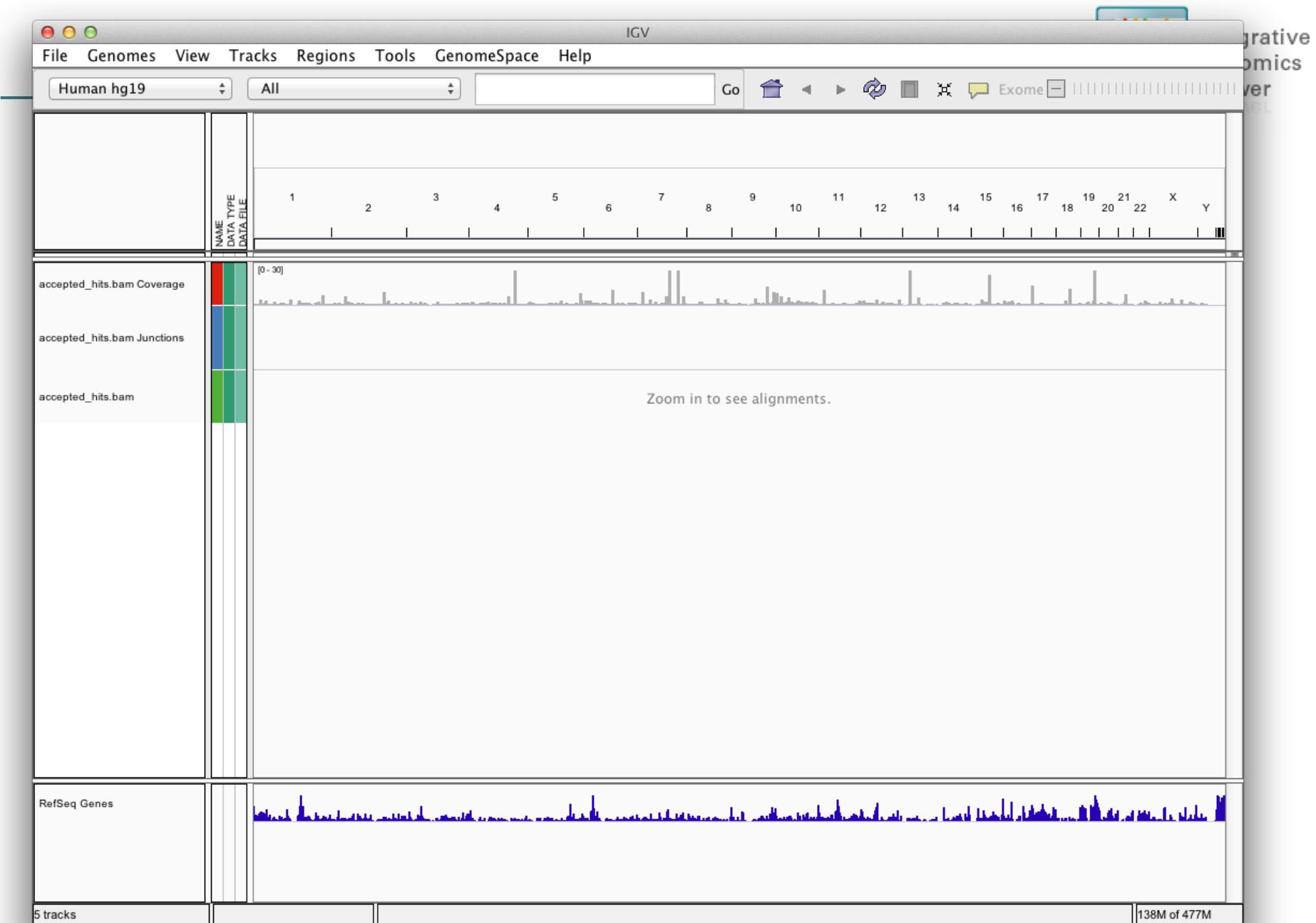
Messages

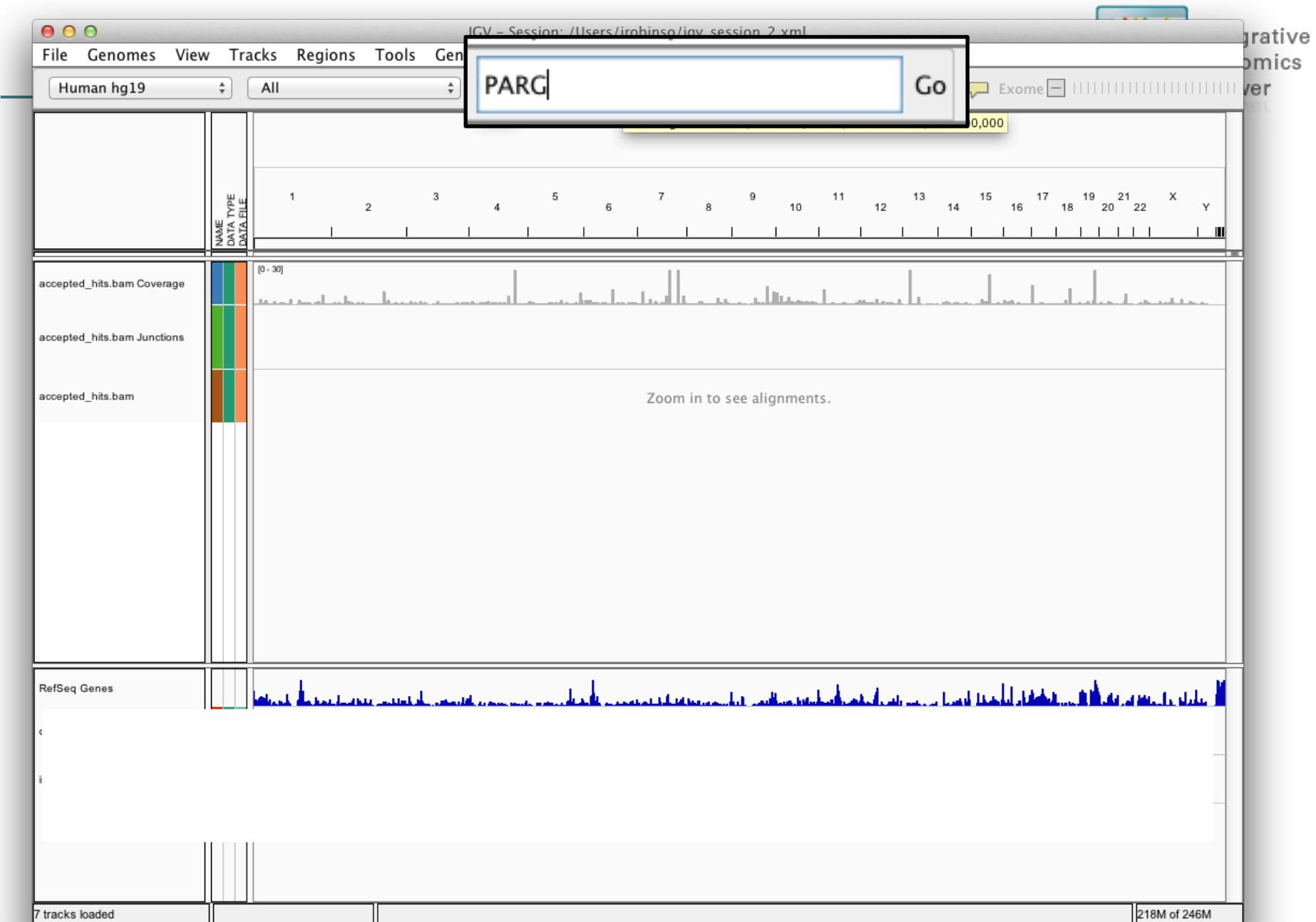
```
Processing chromosome chr9
Processing chromosome chrM
Processing chromosome chrX
Processing chromosome chrY
Group idx: 1
100.0%
```

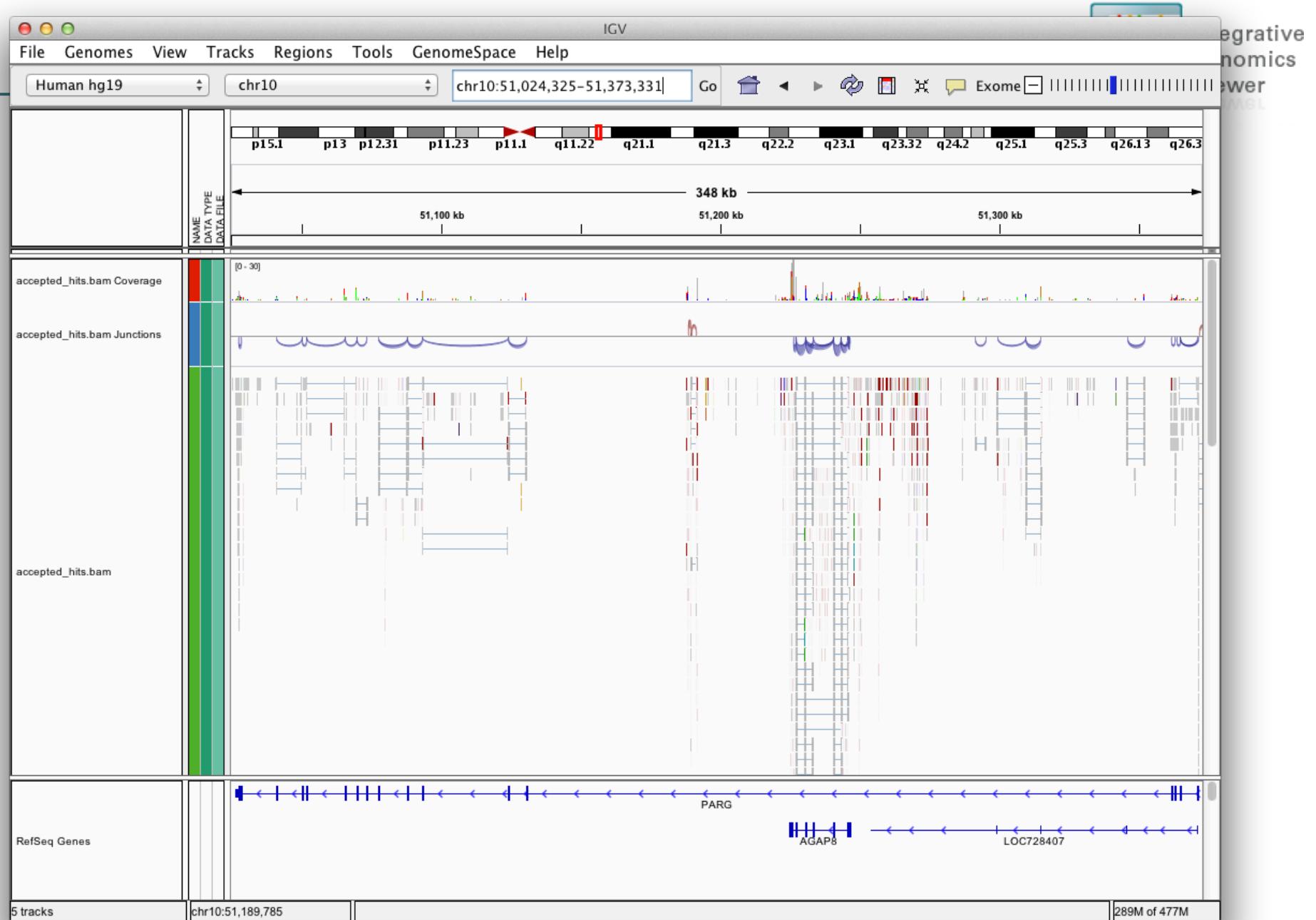












Viewing RNA-Seq Data



- Tune preferences
- Examine tissue-specific alternative splicing
- **Supplementary views**

igvtools - coverage

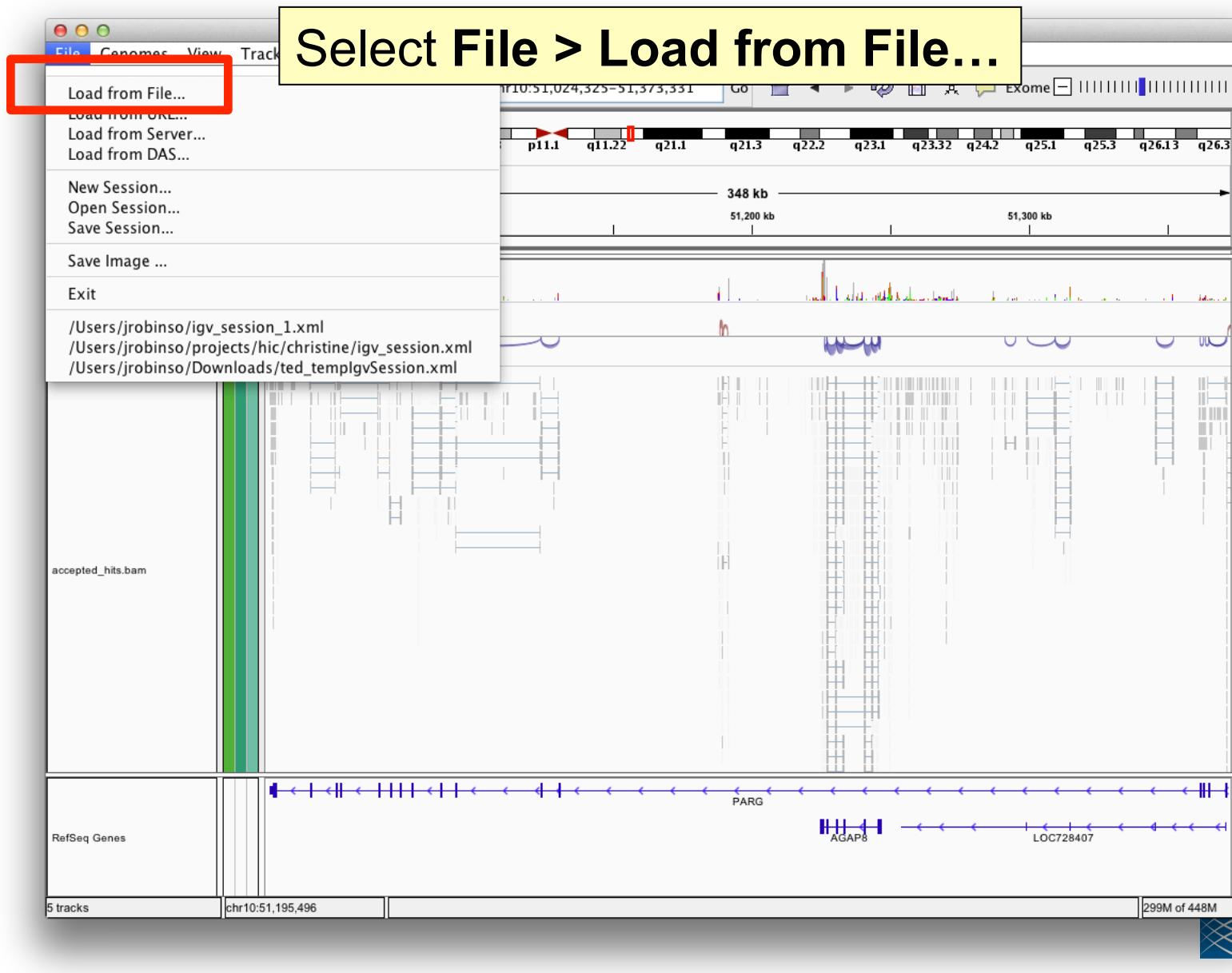
TopHat - insertions, deletions, junctions

Cufflinks - transcripts

TopHat output



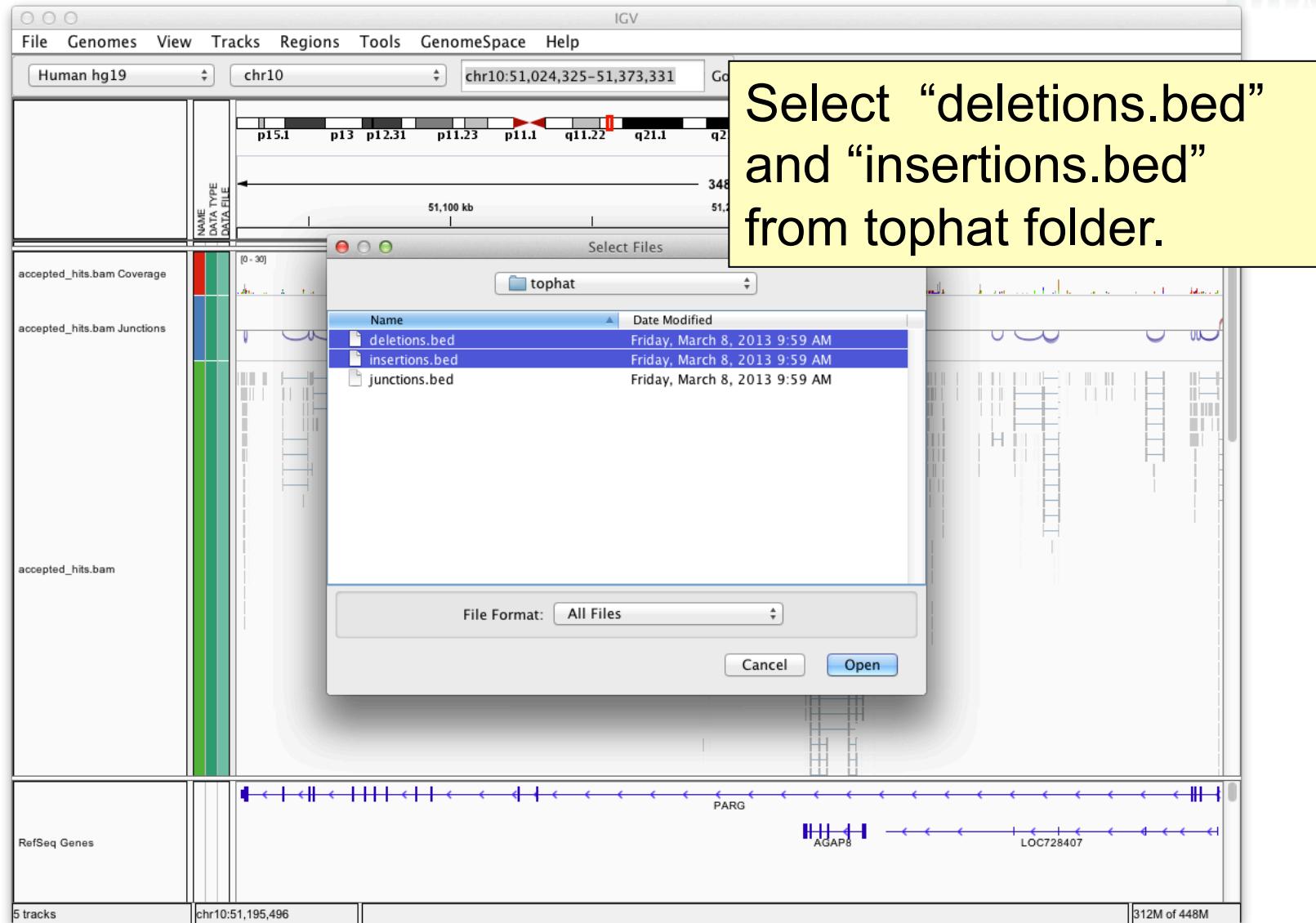
Integrative
Genomics
Viewer
ALIGMENT



TopHat output



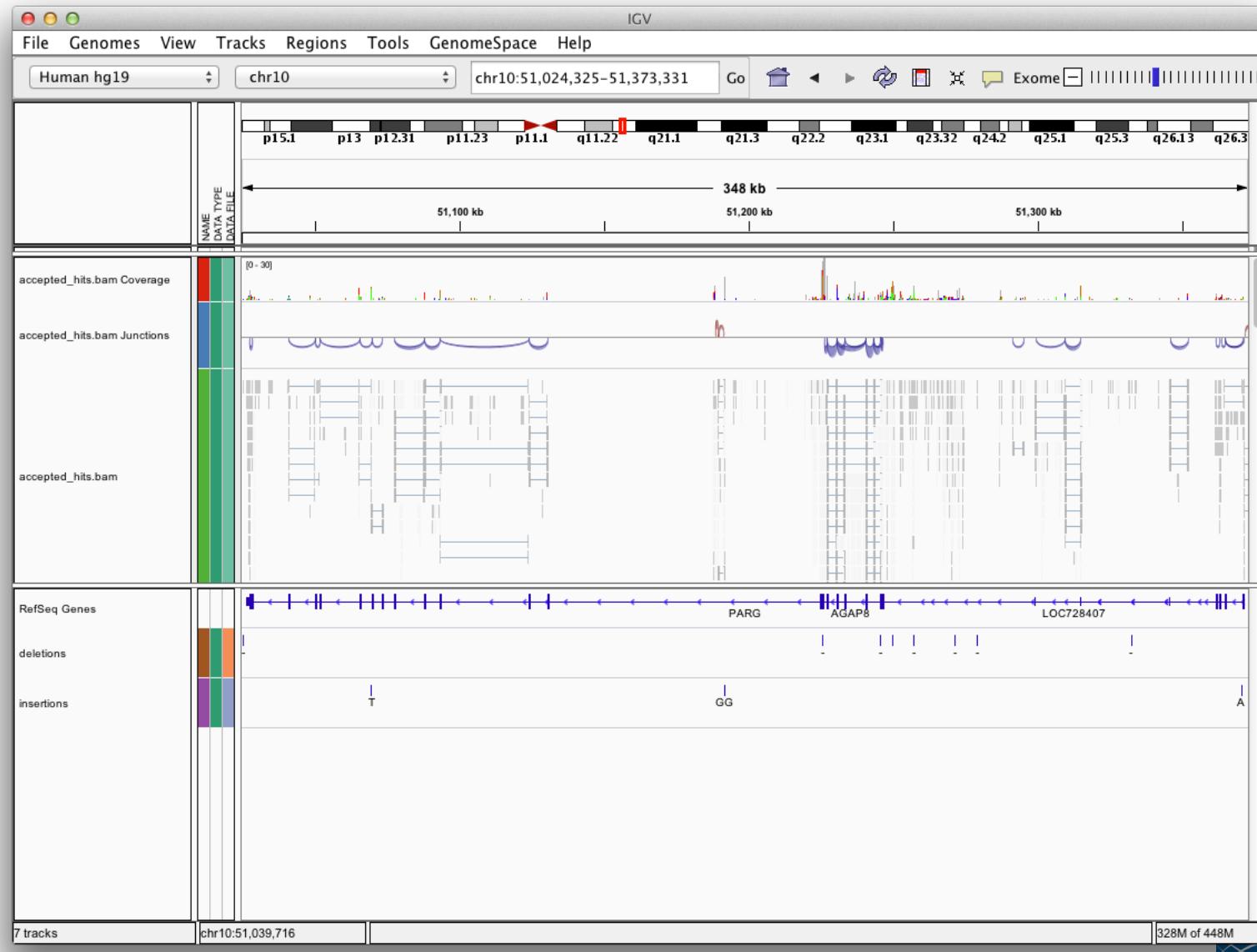
Integrative
Genomics
Viewer
IGV



TopHat output



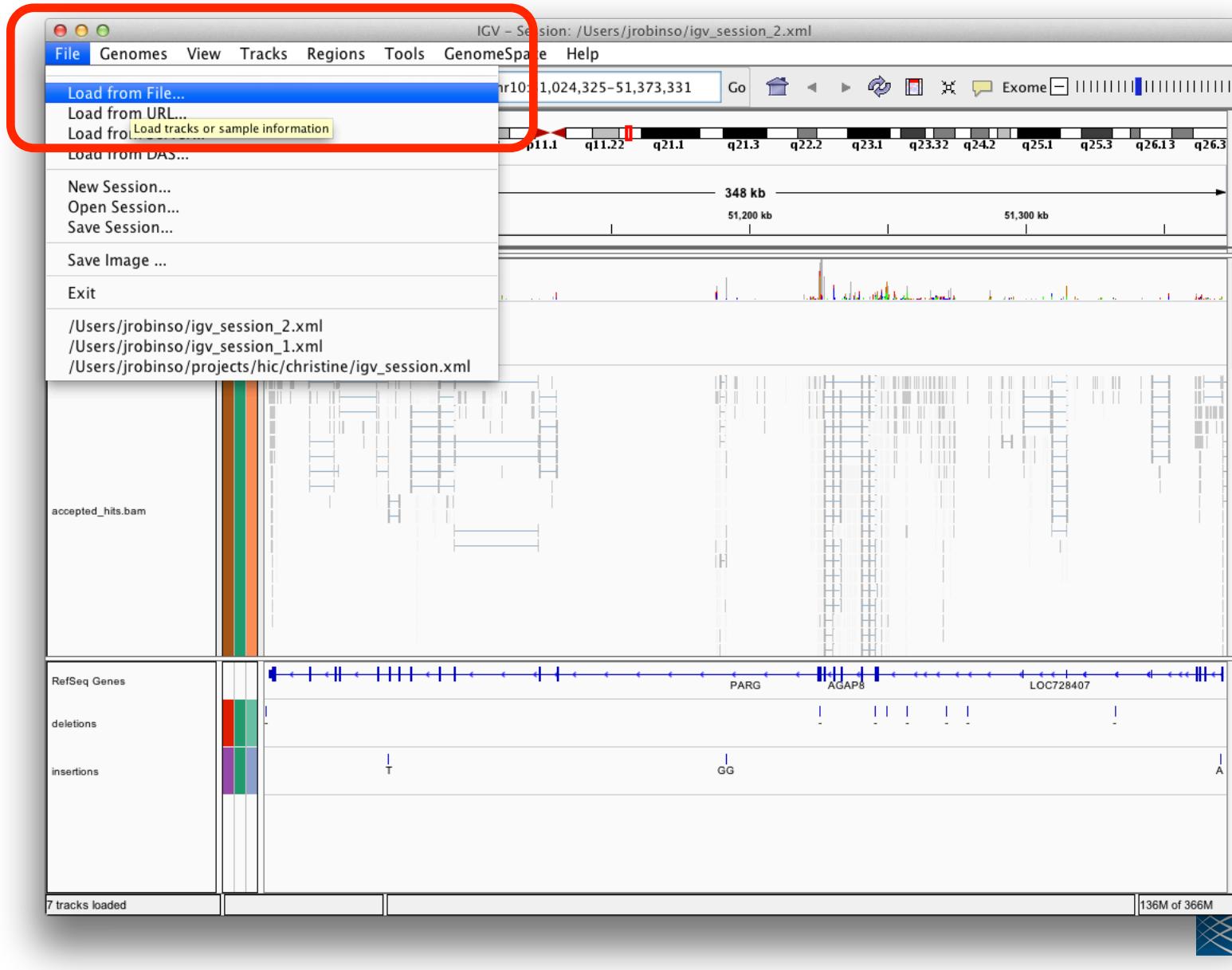
Integrative
Genomics
Viewer
IGV



TopHat output



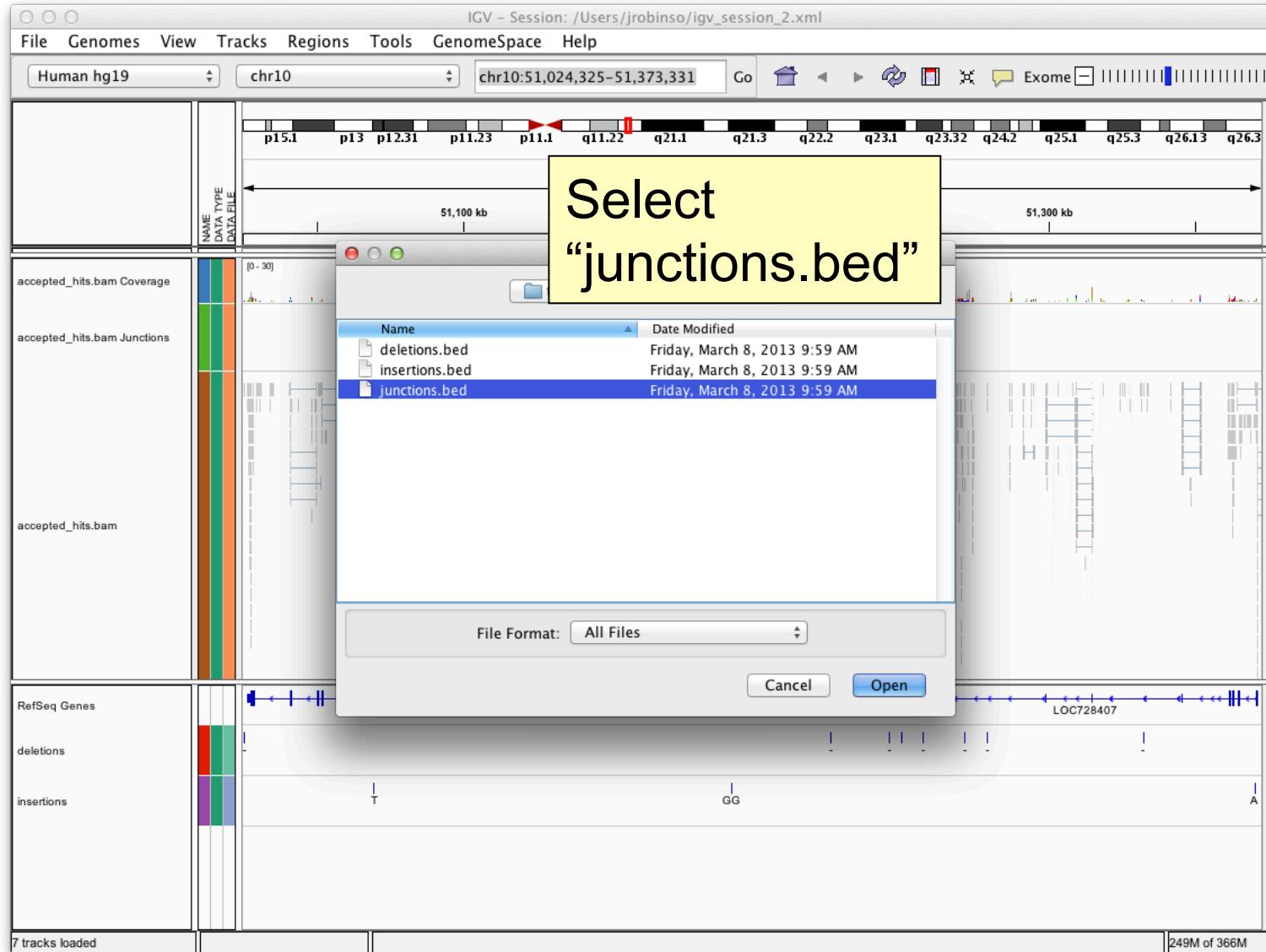
Integrative
Genomics
Viewer
ALGEMEL



TopHat output



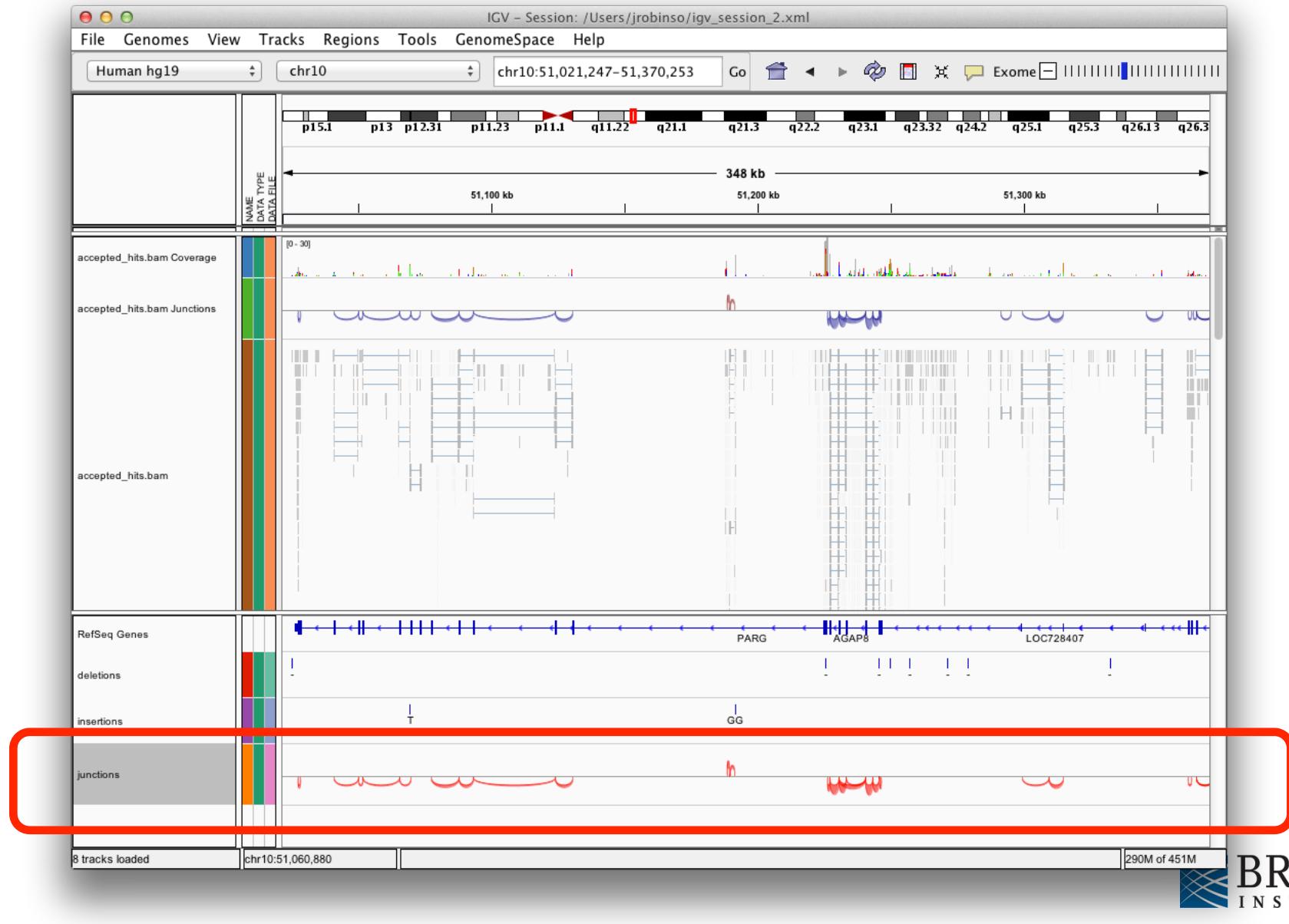
Integrative
Genomics
Viewer
ALMEL



TopHat output



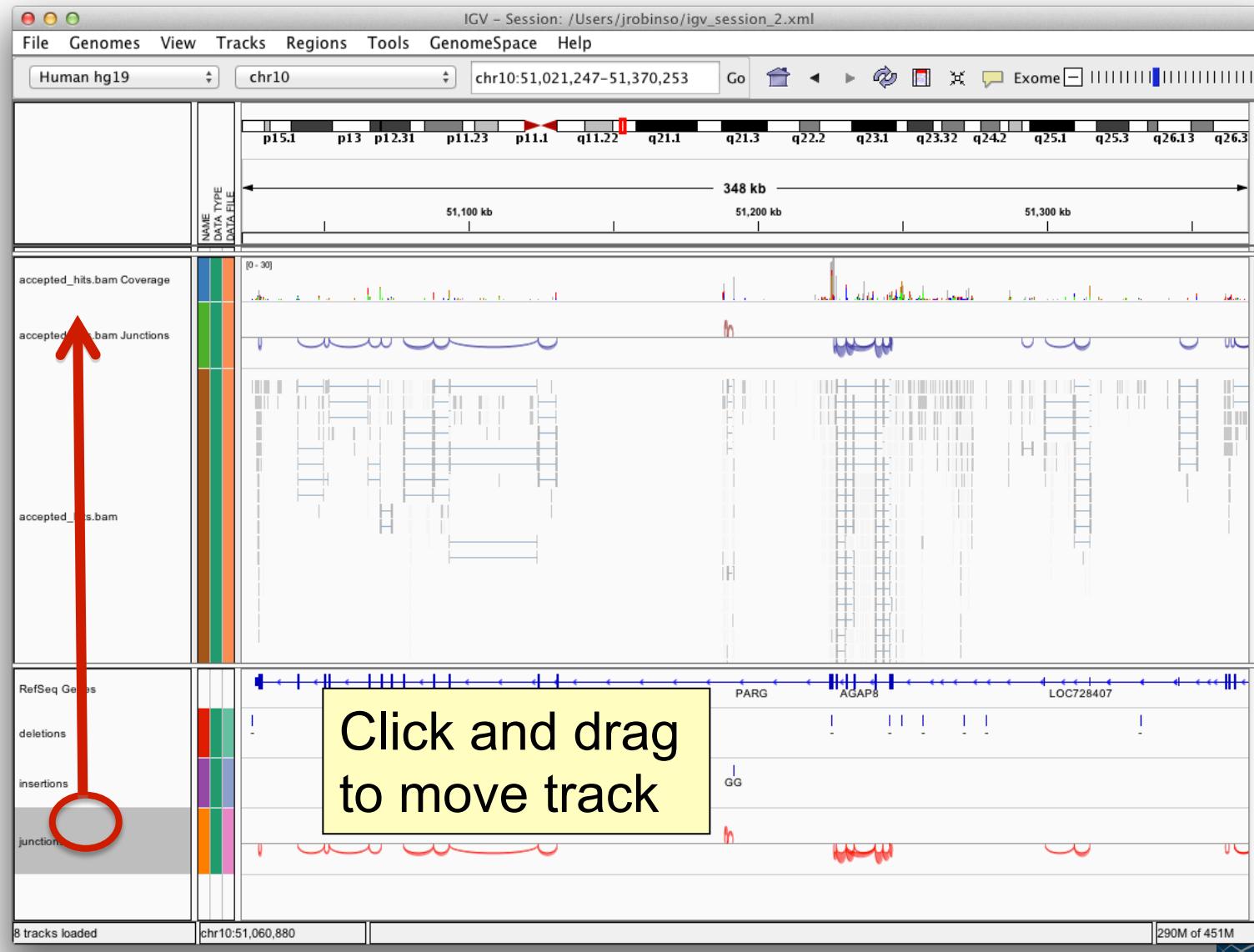
Integrative
Genomics
Viewer
ALGEME



TopHat output



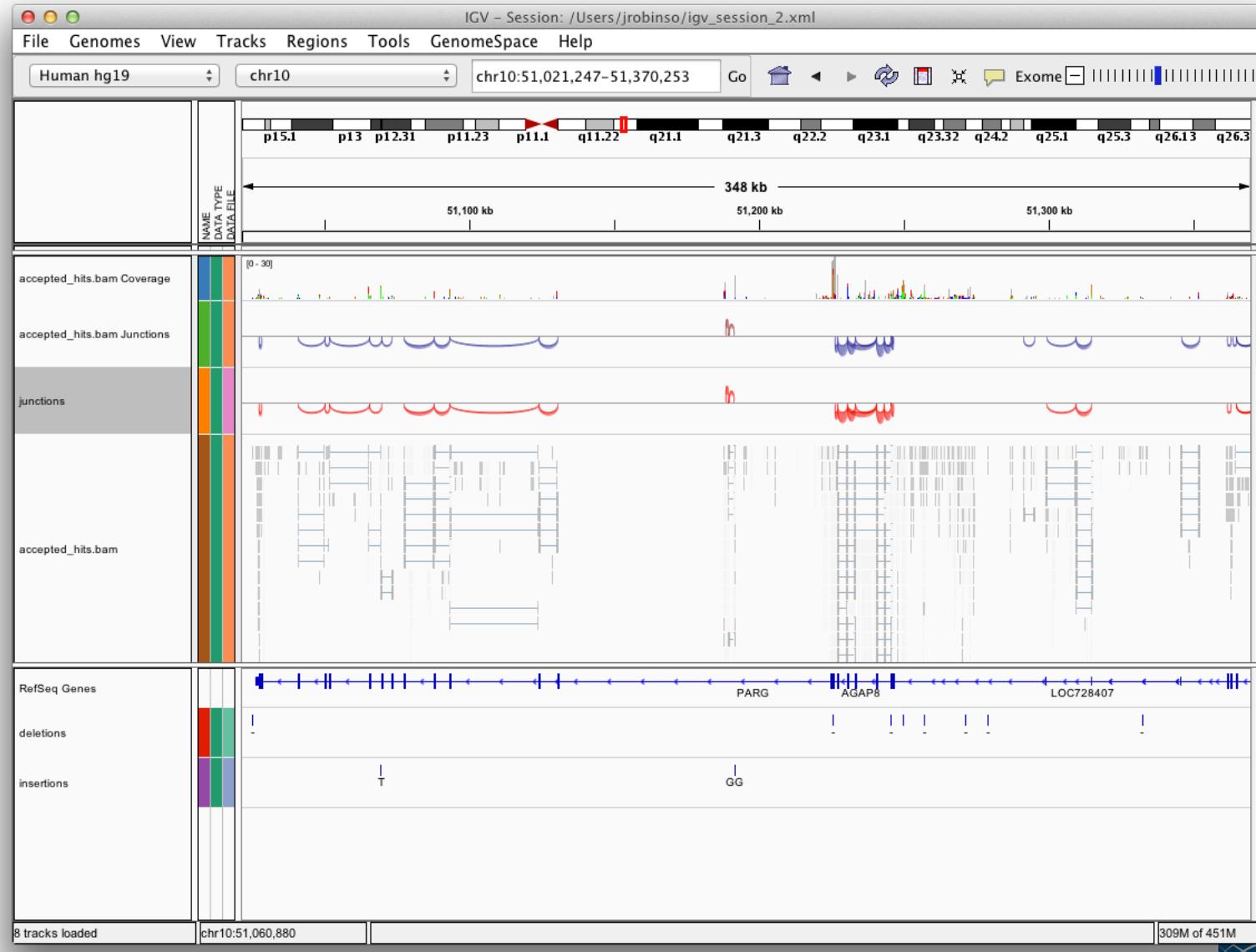
Integrative
Genomics
Viewer
ALGEME



TopHat output



Integrative
Genomics
Viewer
ALGEMEEL



Viewing RNA-Seq Data



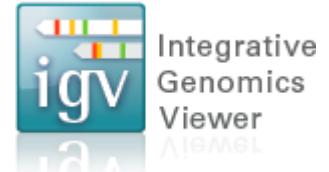
- Tune preferences
- Examine tissue-specific alternative splicing
- **Supplementary views**

igvtools - coverage

TopHat - insertions, deletions, junctions

Cufflinks - transcripts

Viewing predicted transcripts



Files can be large

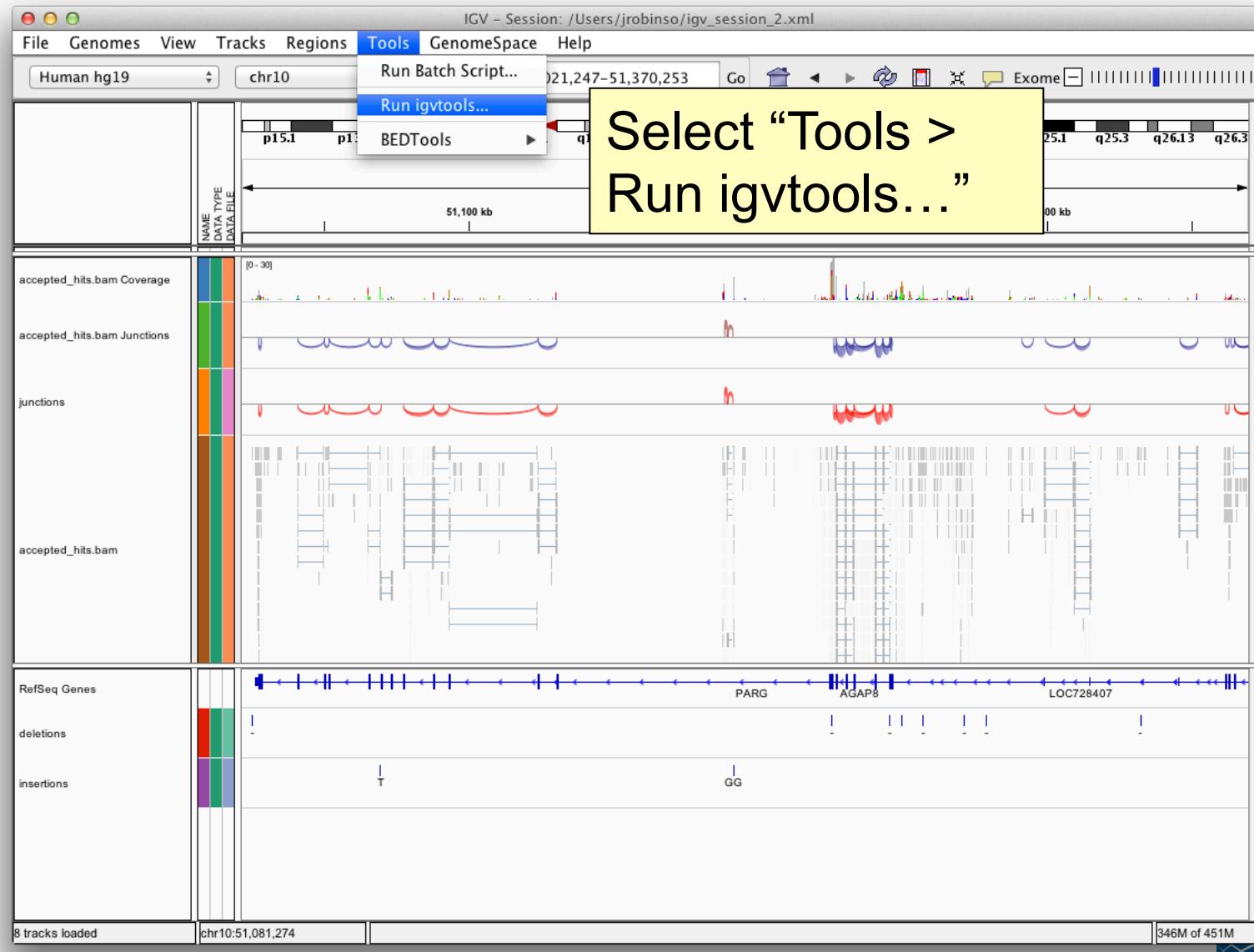
=> Index with igvtools

Note, this is useful for any large bed, gtf, or gff file, not just cufflinks output.

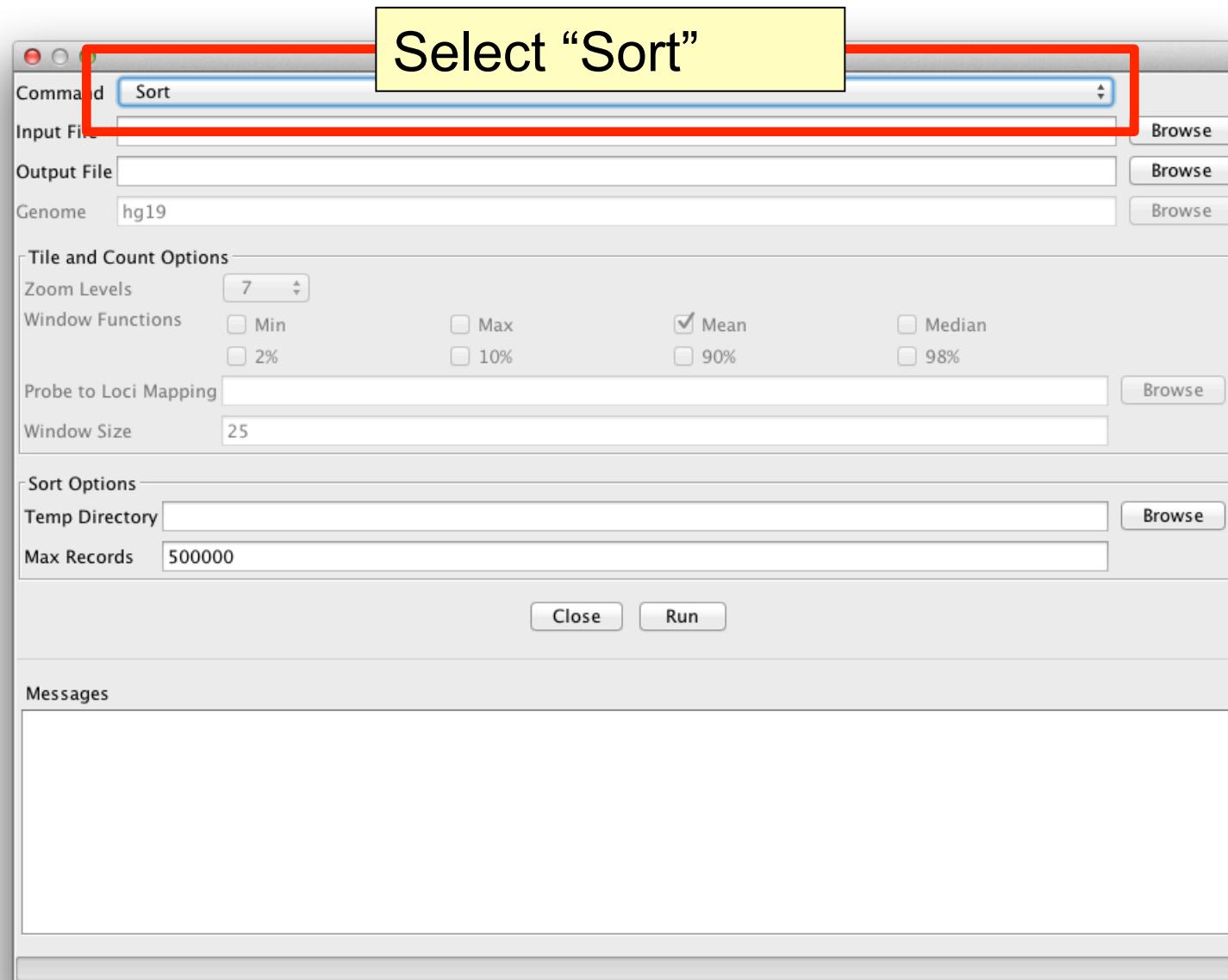
Indexing with igvtools



Integrative
Genomics
Viewer
ALGEME



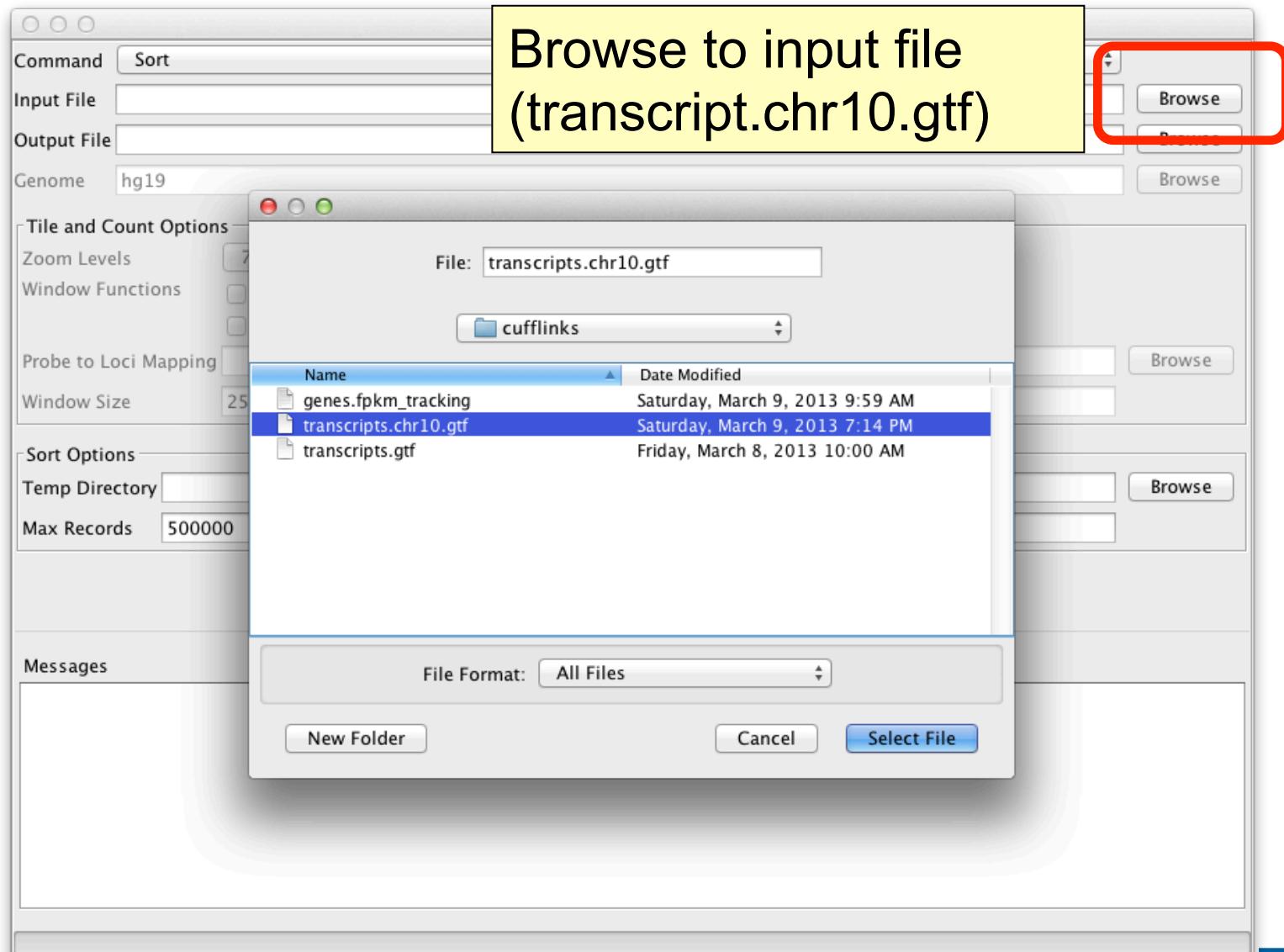
Indexing with igvtools



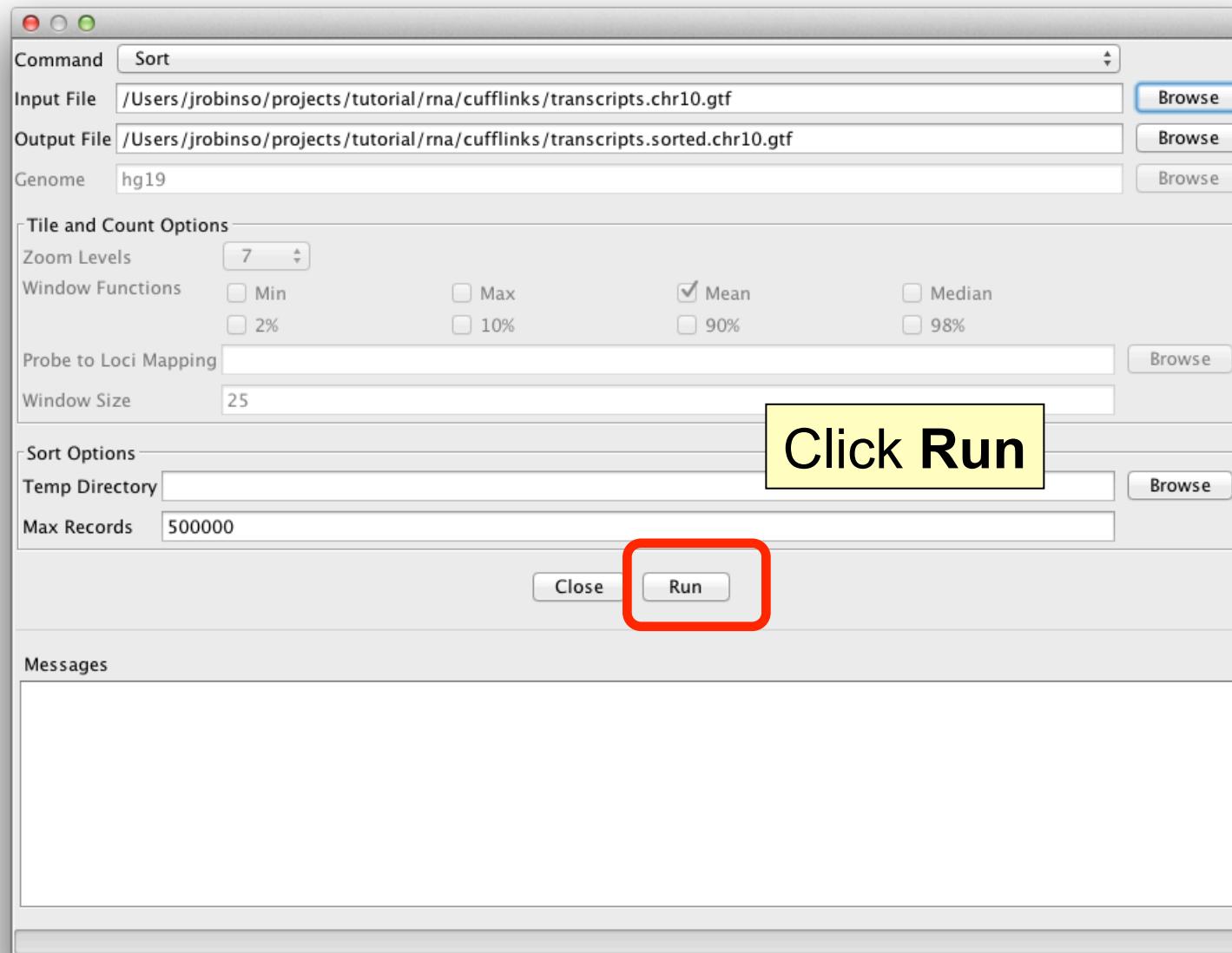
Indexing with igvtools



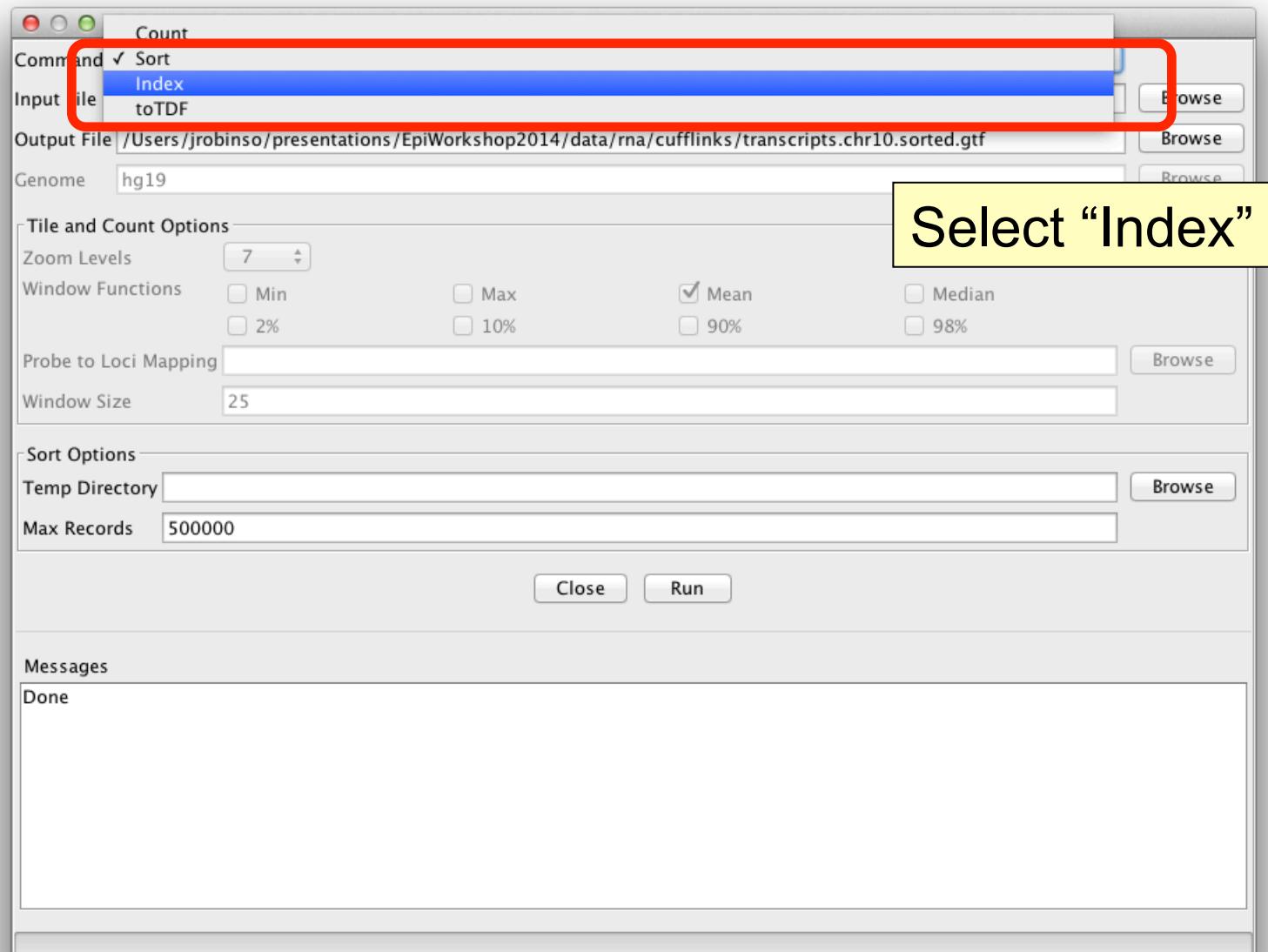
Integrative
Genomics
Viewer
ALGML



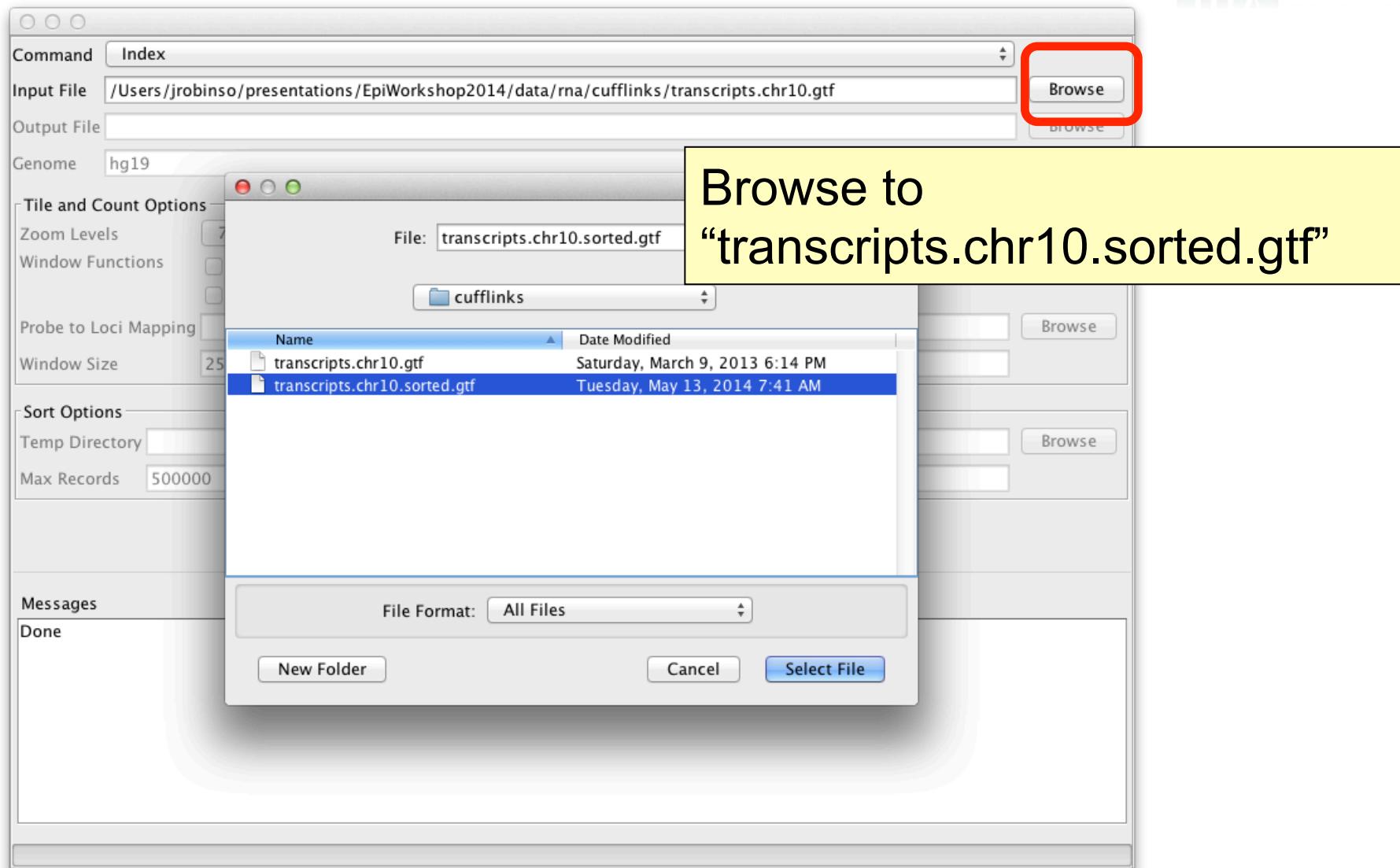
Indexing with igvtools



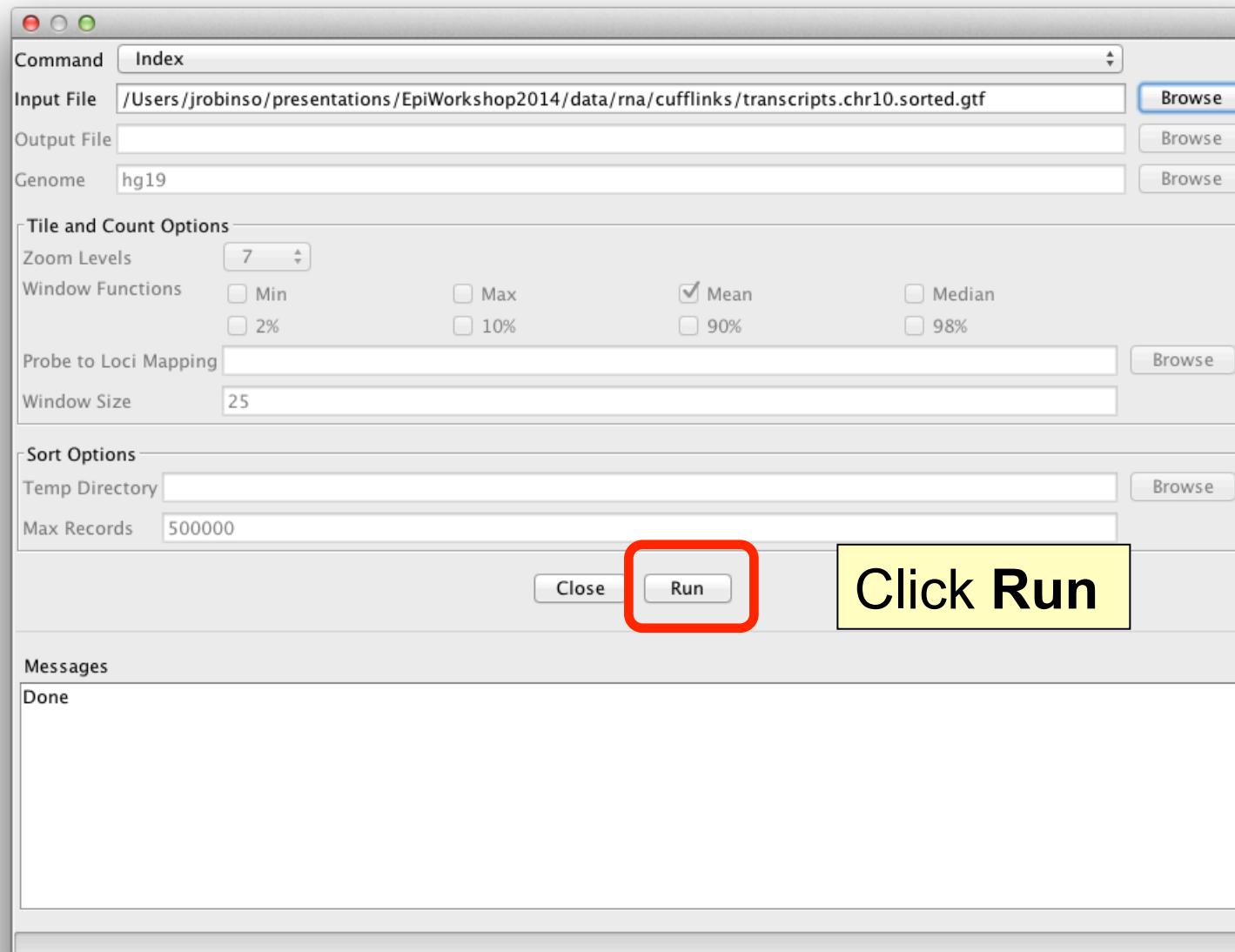
Indexing with igvtools



Indexing with igvtools



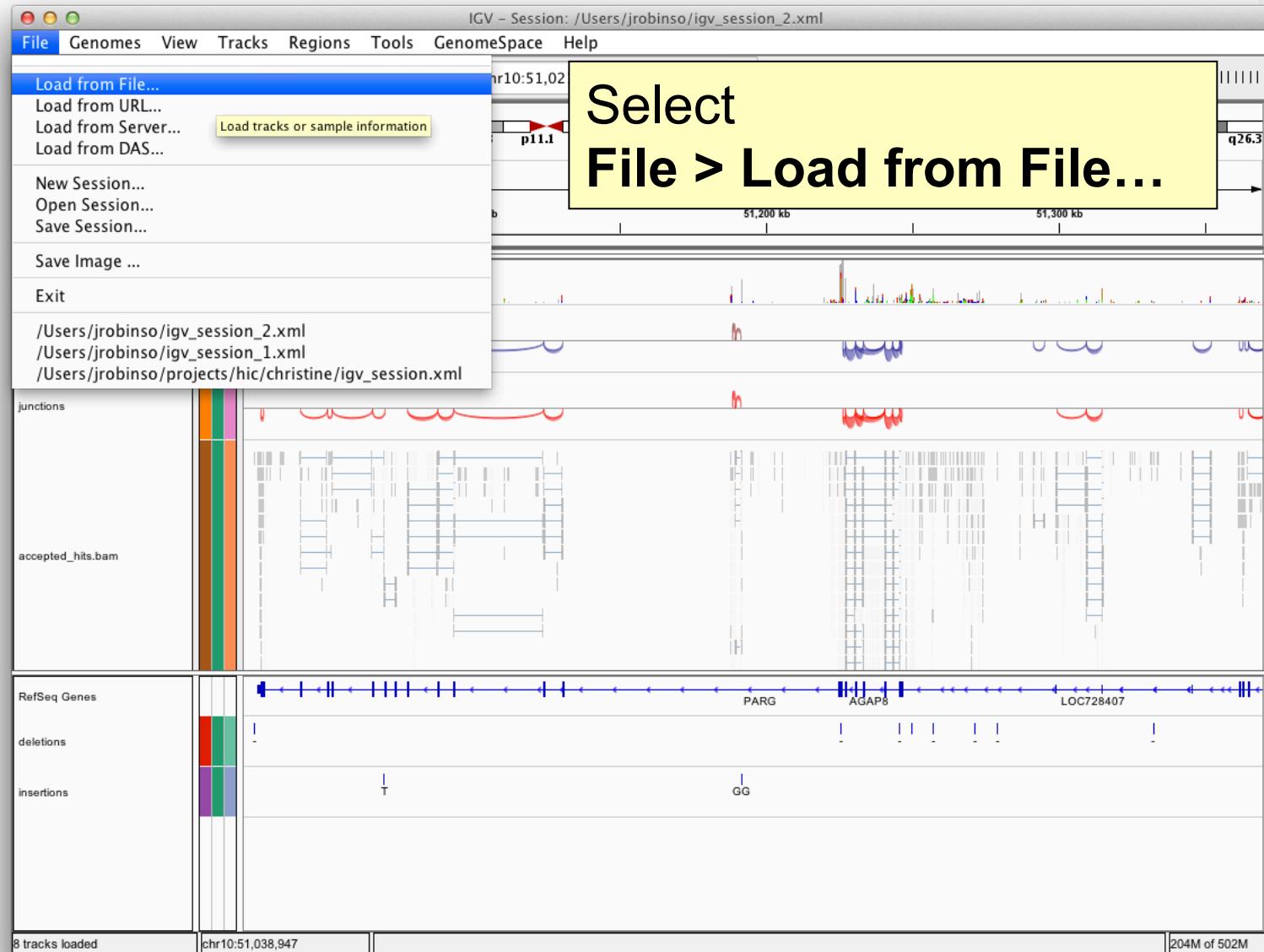
Indexing with igvtools



Indexing with igvtools



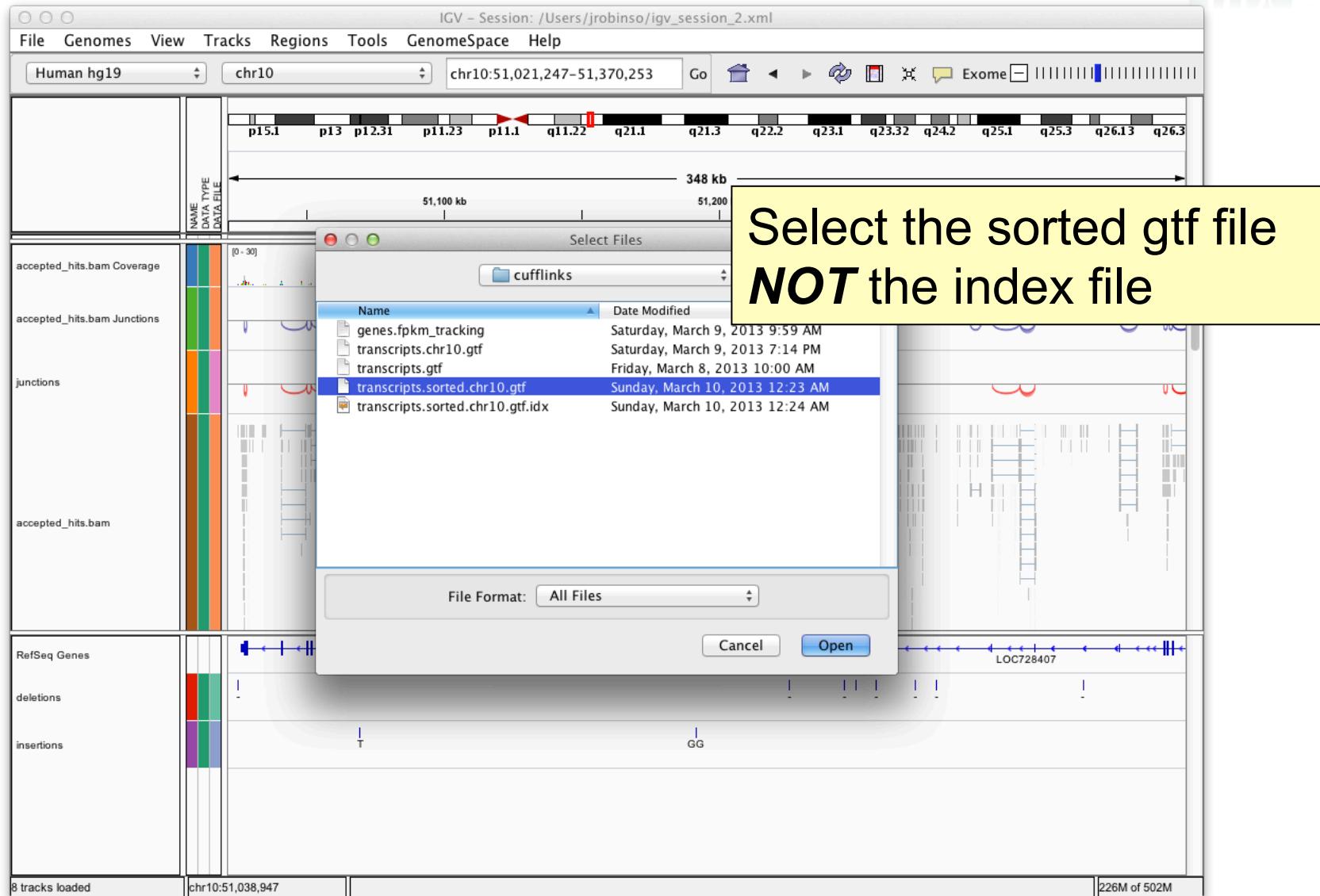
Integrative
Genomics
Viewer
ALGVIEWER



Indexing with igvtools



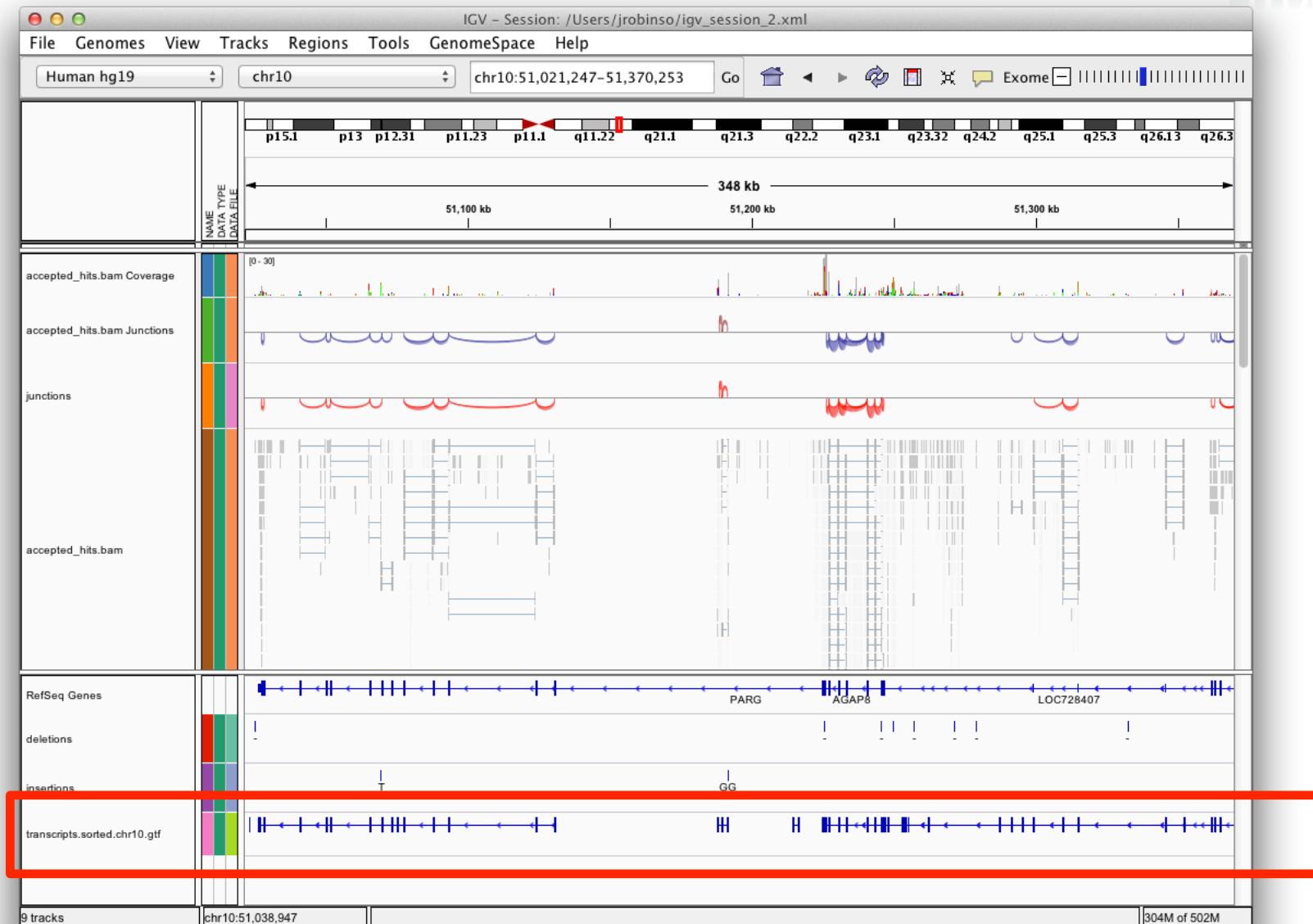
Integrative
Genomics
Viewer
ALGML



Indexing with igvtools



Integrative
Genomics
Viewer
ALGEME



Reset alignment preferences



Integrative
Genomics
Viewer
ALMEL

Set “Visibility window” to 30

Human hg18

File Genomes View Tracks Regions Tools GenomeSpace Help

IGV

General Tracks Mutations Charts Aliases

Visibility range threshold (kb): Nominal

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

Filter and shading options

Coverage allele-freq threshold: Mapping quality threshold:

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: to

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: Min junction coverage:

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): Compute Minimum (percentile):
Maximum (bp): Maximum (percentile):

OK Cancel

RefSeq genes

8 tracks loaded chr11:8,563,411 154M of 1,098M

ChIP-Seq

ChIP-Seq



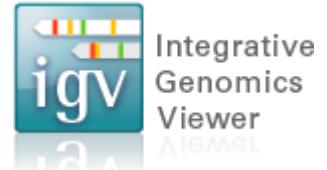
- File types
- Loading ENCODE data
- ChIP Seq options in igvtools
- Motif locator

ChIP-Seq file types



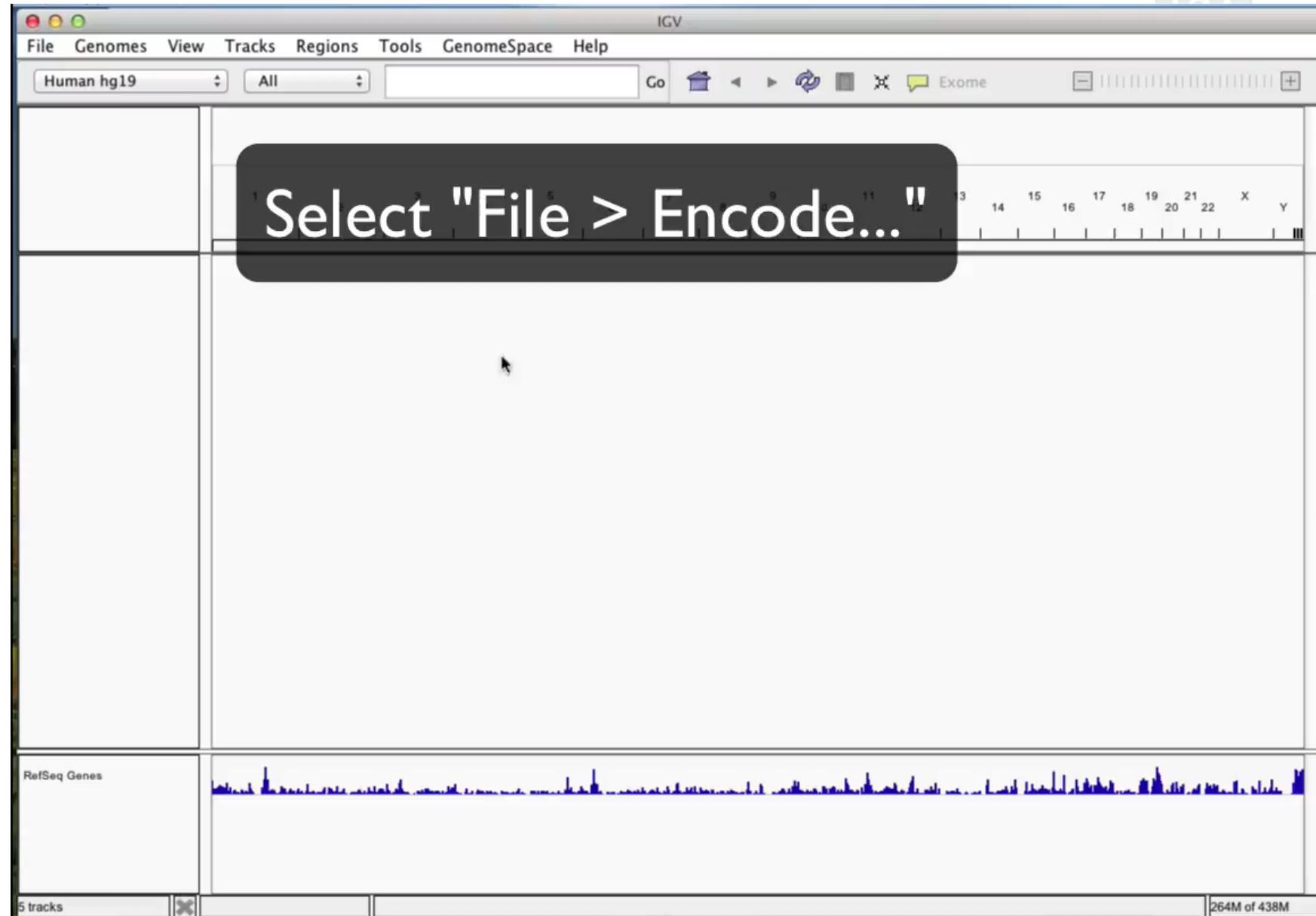
- Alignments
 - bam
- Signals
 - wig
 - bigWig
 - tdf
- Peaks
 - narrowPeak
 - broadPeak
 - bed

Loading ENCODE data



- Options for loading ENCODE data from UCSC:
 1. Download and use “Load from file...”
 2. Enter URL in “Load from url...”
 3. Drag and drop from web browser
 4. ENCODE file loader (*new*)

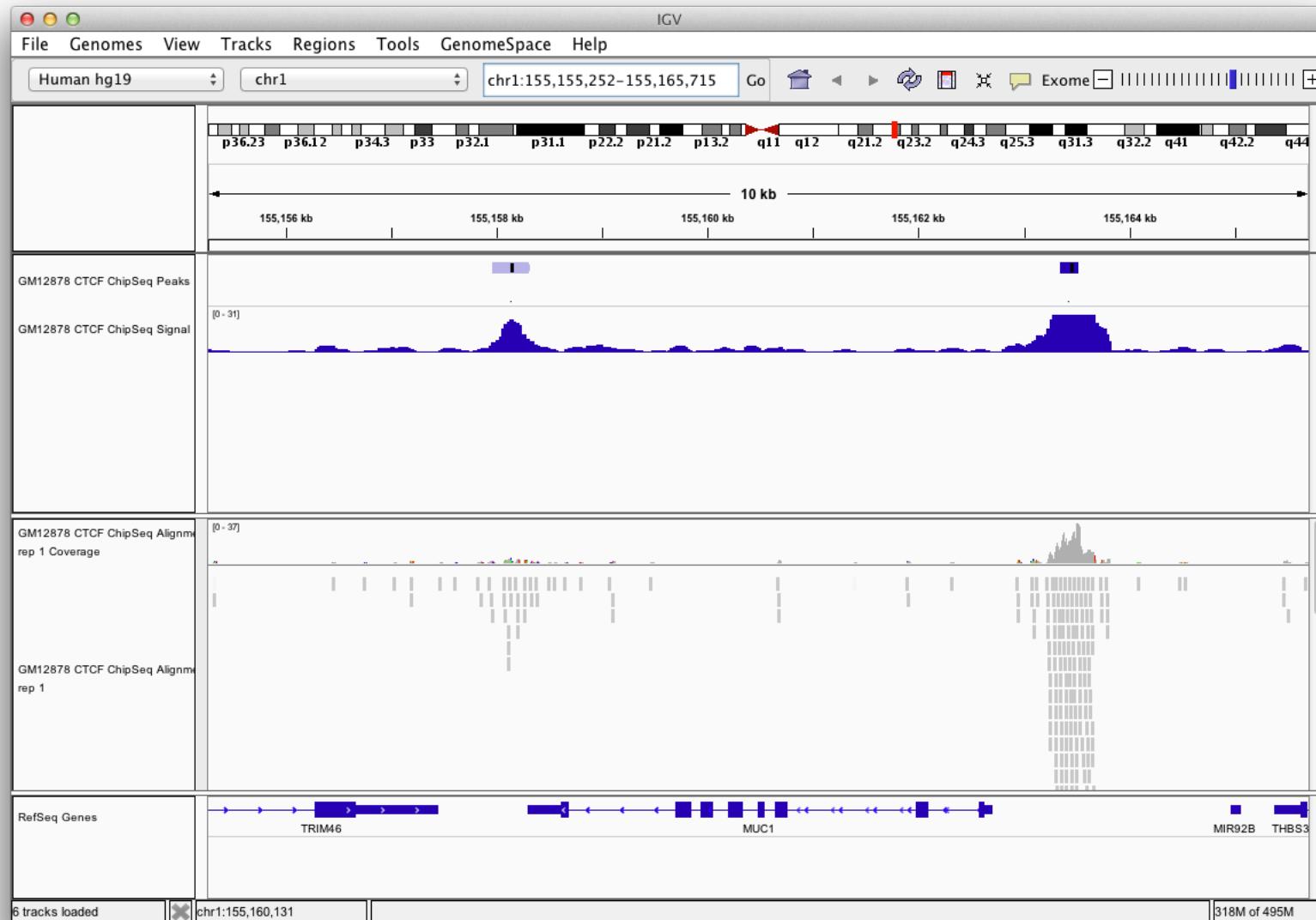
Encode Loader



Encode Loader



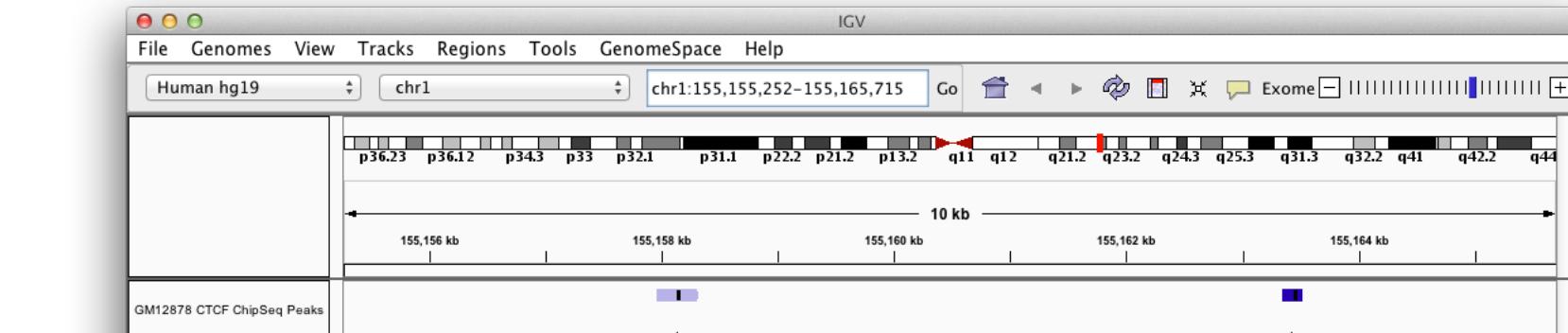
Integrative
Genomics
Viewer
IGV



Encode Loader

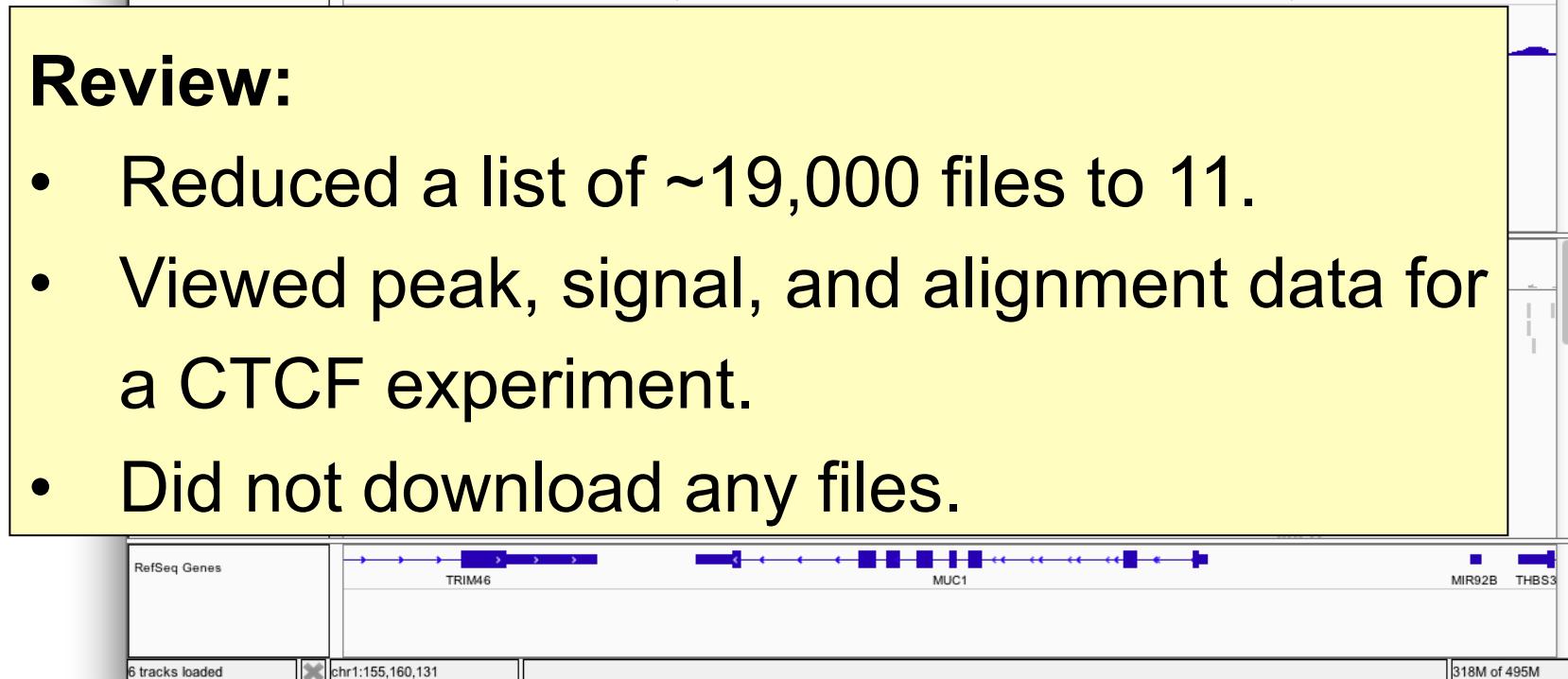


Integrative
Genomics
Viewer
IGV



Review:

- Reduced a list of ~19,000 files to 11.
- Viewed peak, signal, and alignment data for a CTCF experiment.
- Did not download any files.



Encode Loader



Practice

1. Open loader window by selecting “File > Load from ENCODE...”
2. Type terms in search box such as cell line, antibody, and signal type. For example

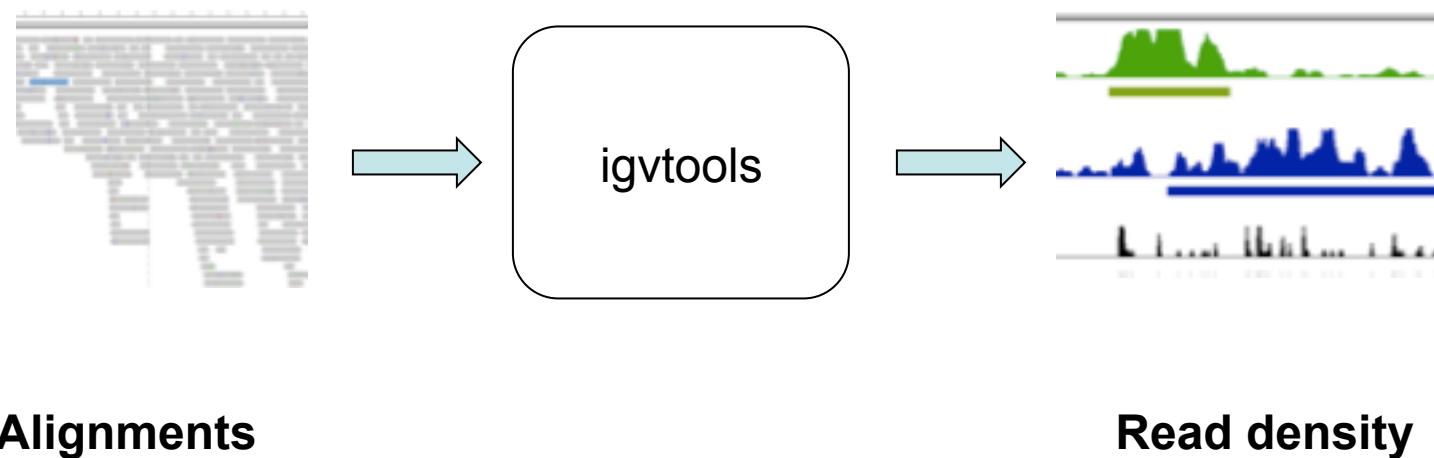
gm12878 h3k4me3 signal

3. Select a few results and load

igvtools count



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



igvtools ChIP-Seq options



“Count” options of particular interest for ChIP experiments

- **-e, --extFactor num**

The alignment s extended by the specified distance in the 3' direction prior to counting

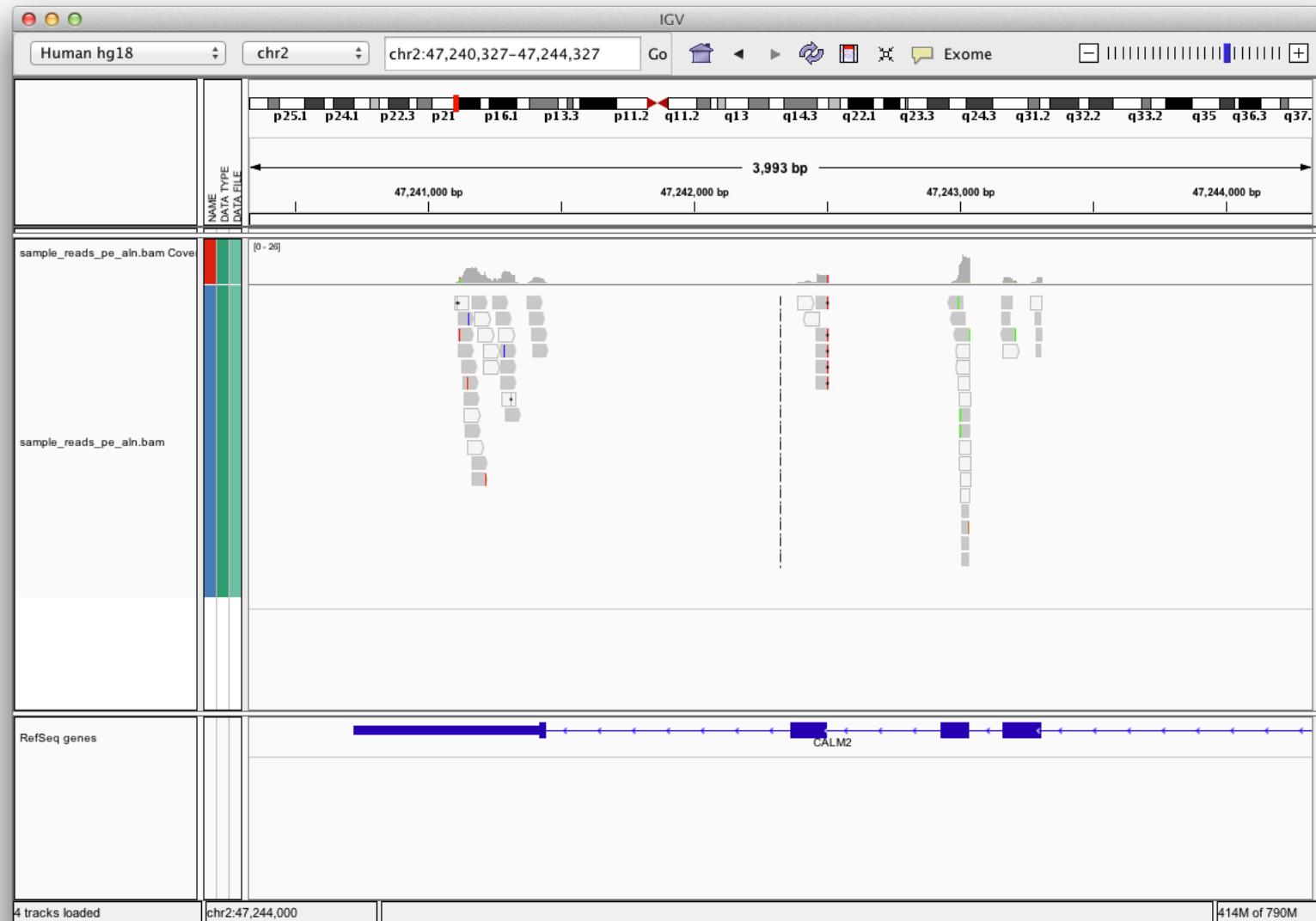
- **--pairs**

Compute coverage from paired alignments counting the entire insert as covered.

igvtools ChIP-Seq options



Integrative
Genomics
Viewer
ALGEME



igvtools ChIP-Seq options



```
~/bin/IGVTools/igvtools count --extFactor 1950  
sample_reads_pe_aln.bam extFactor.tdf hg18
```

```
~/bin/IGVTools/igvtools count --pairs  
sample_reads_pe_aln.bam pairs.tdf hg18
```

Acknowledgments

Development of IGV has been made possible by 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.