Using IGV to identify true somatic variants from the false variants

http://www.broadinstitute.org/igv

A FAQ, sample files and a user guide are available on IGV website

If you use IGV in your publication:
James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer (2011) Nature Biotechnology 29: 24–26

Part 1: How to use IGV to visualize variants

Emilie Chautard

OICR
Ontario Institute
for Cancer Research

What can you do with IGV?

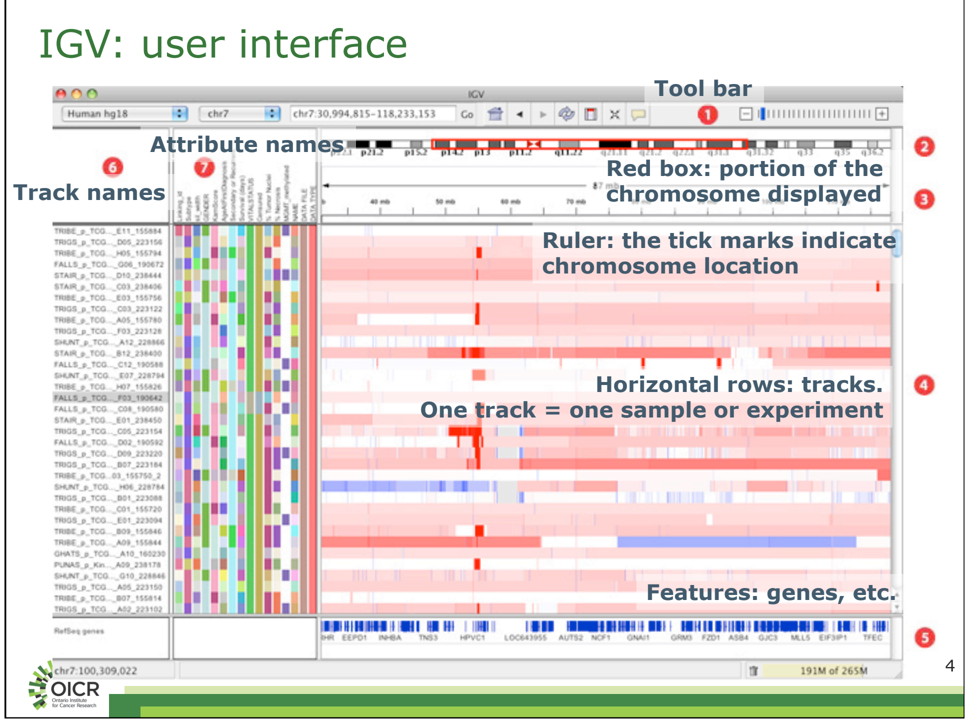
Visualization of different genomic data types:
- aligned sequence reads
- mutations
- copy number
- RNA interference screens
- gene expression
- methylation and genomic annotations

List of supported data formats:
http://www.broadinstitute.org/software/igv/FileFormats

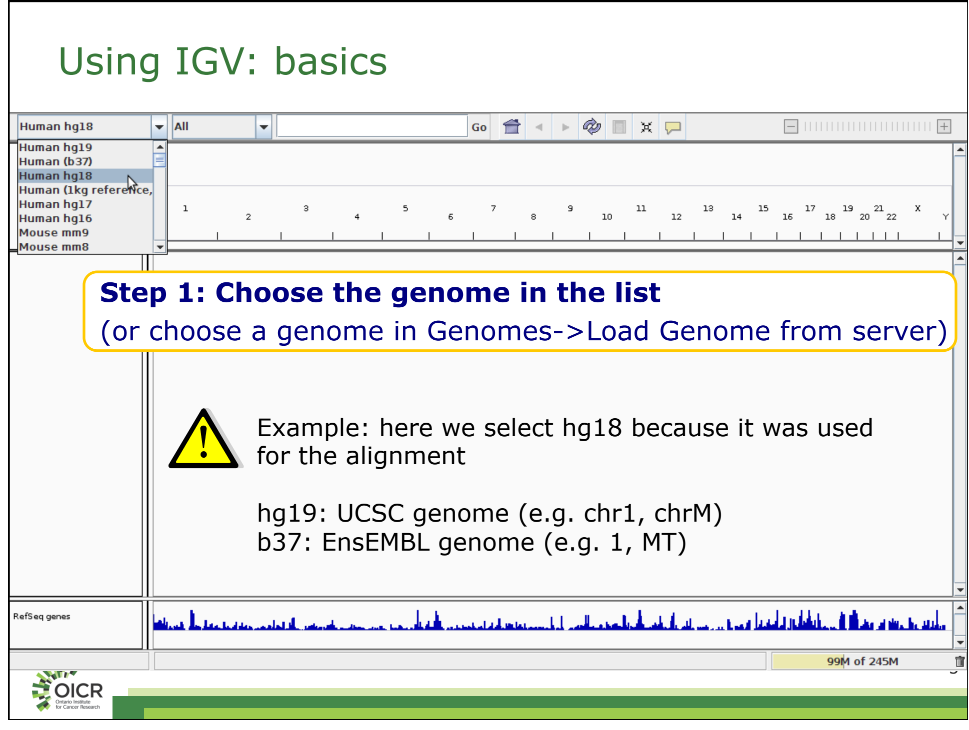
For this example:
- *.bam for the alignment file
- *.gtf for the genome annotation data

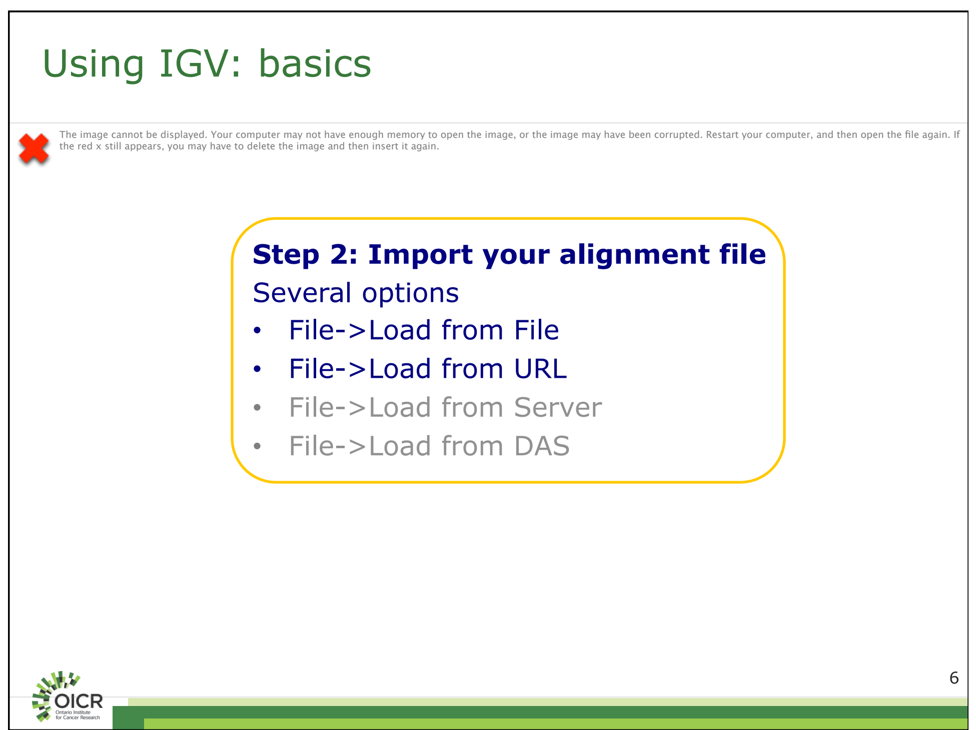
3

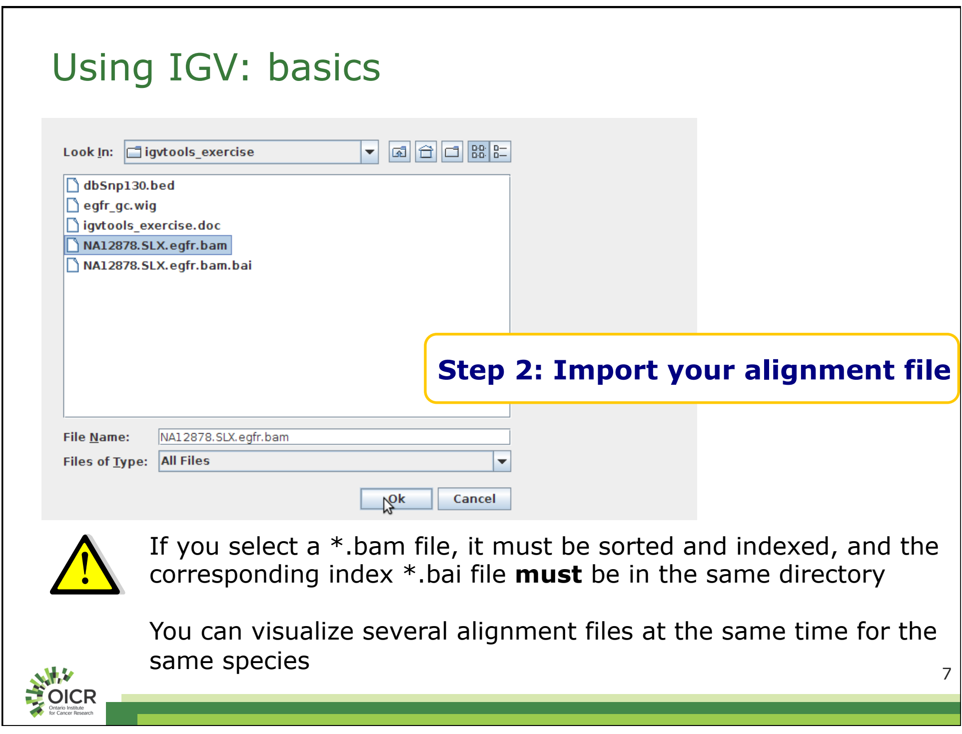
IGV: user interface

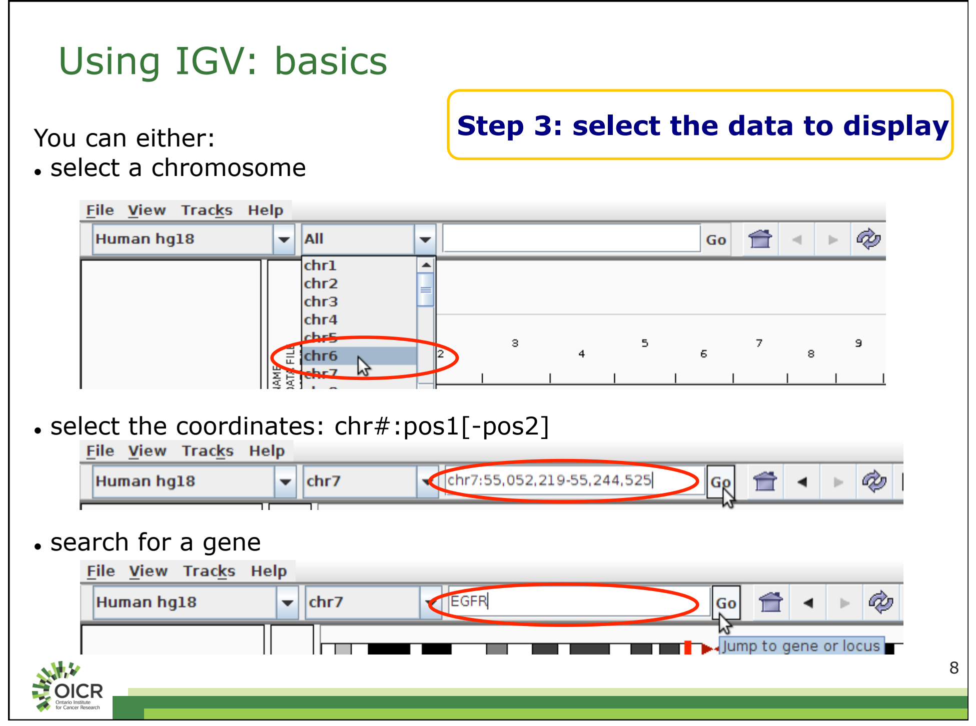


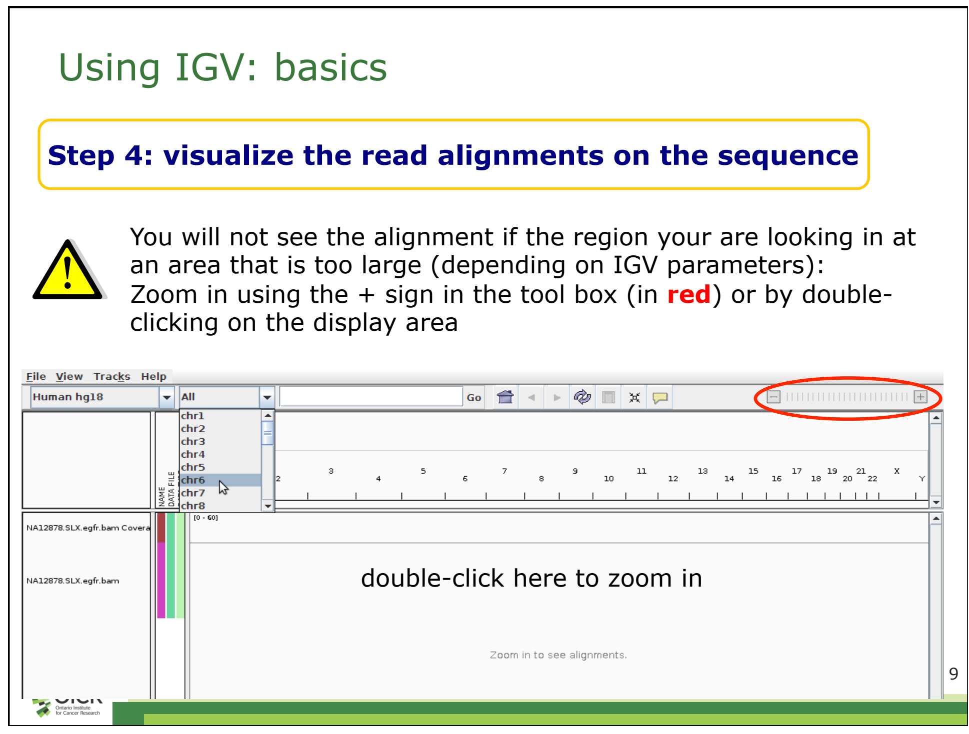
4


Using IGV: basics
Human hg18
Human hg19
Human (b37)
Human hg18
Human (1kg reference,
Human hg17
Human hg16
Mouse mm9
Mouse mm8
All
Go
Step 1: Choose the genome in the list
(or choose a genome in Genomes->Load Genome from server)
Example: here we select hg18 because it was used for the alignment
hg19: UCSC genome (e.g. chr1, chrM)
b37: EnsEMBL genome (e.g. 1, MT)
RefSeq genes
99M of 245M
OICR

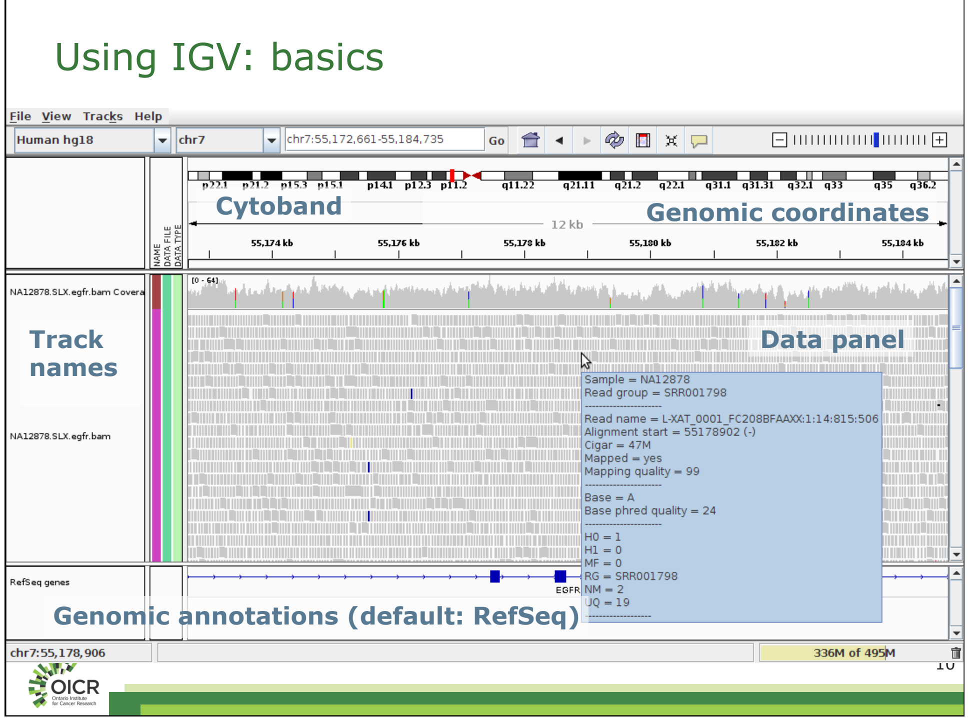

Using IGV: basics
The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.
Step 2: Import your alignment file
Several options
• File->Load from File
• File->Load from URL
• File->Load from Server
• File->Load from DAS
6
OICR


Using IGV: basics
Look In: igvtools_exercise
dbSnp130.bed
egfr_gc.wig
igvtools_exercise.doc
NA12878.SLX.egfr.bam
NA12878.SLX.egfr.bam.bai
Step 2: Import your alignment file
File Name: NA12878.SLX.egfr.bam
Files of Type: All Files
Ok
Cancel
If you select a *.bam file, it must be sorted and indexed, and the corresponding index *.bai file must be in the same directory
You can visualize several alignment files at the same time for the same species
7
OICR

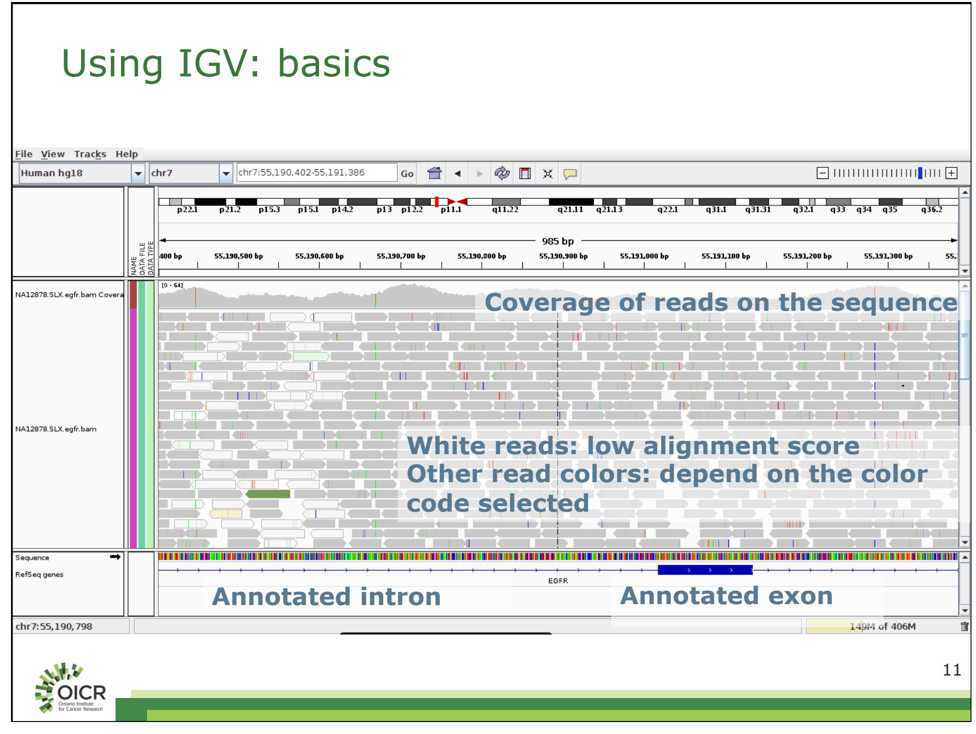

Using IGV: basics
Step 3: select the data to display
You can either:
• select a chromosome
File View Tracks Help
Human hg18
All
Go
chr1
chr2
chr3
chr4
chr5
chr6
chr7
• select the coordinates: chr#:pos1[-pos2]
Human hg18
chr7
chr7:55,052,219-55,244,525
Go
• search for a gene
Human hg18
chr7
EGFR
Go
Jump to gene or locus
8
OICR

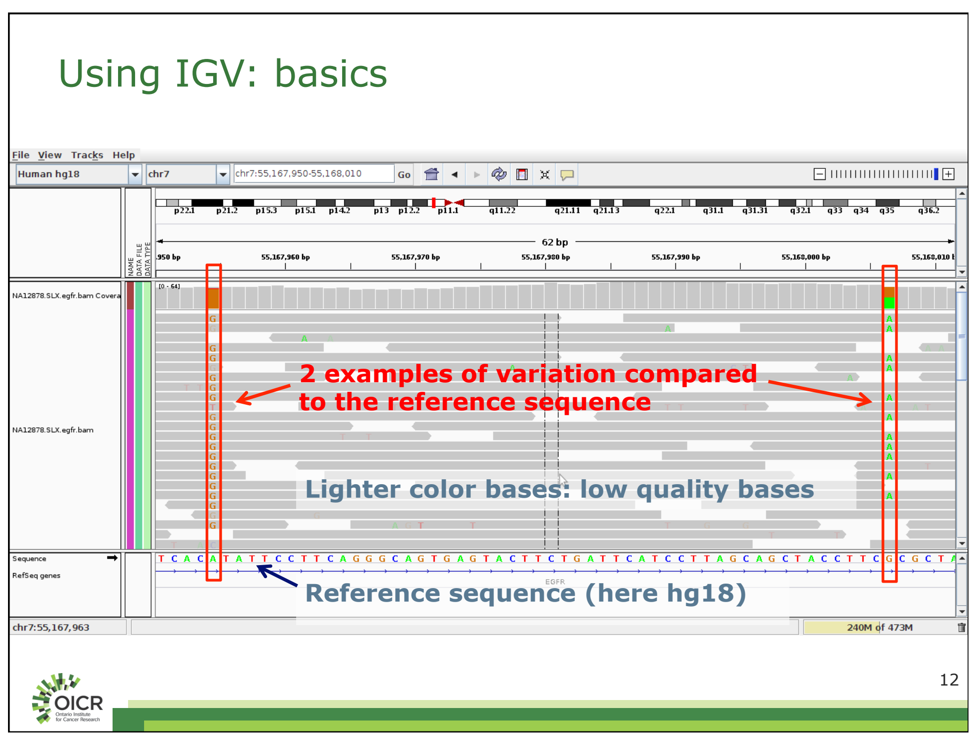


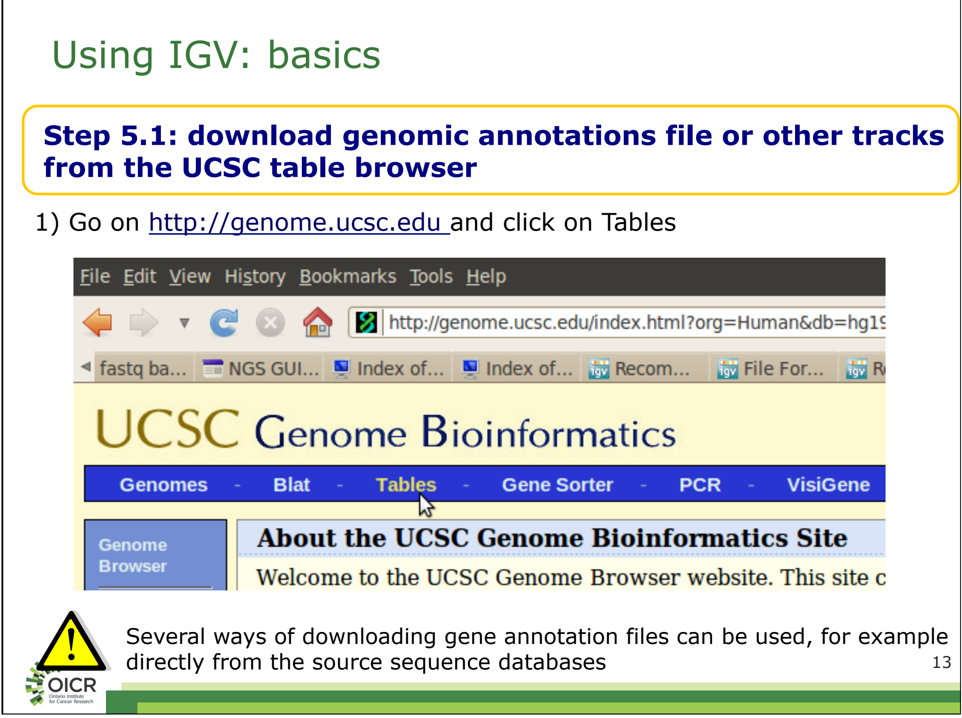
Using IGV: basics
Step 4: visualize the read alignments on the sequence
You will not see the alignment if the region your are looking in at an area that is too large (depending on IGV parameters):
Zoom in using the + sign in the tool box (in red) or by double-clicking on the display area
File View Tracks Help
Human hg18
All
Go
NA12878.SLX.egfr.bam Covera
NA12878.SLX.egfr.bam
double-click here to zoom in
Zoom in to see alignments.
9

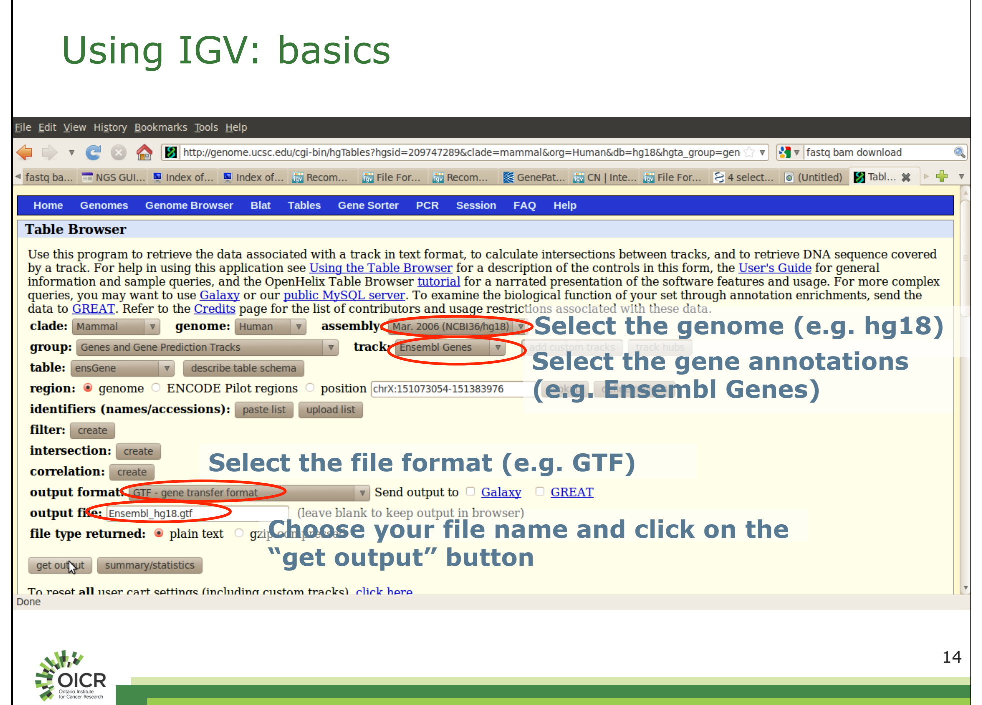


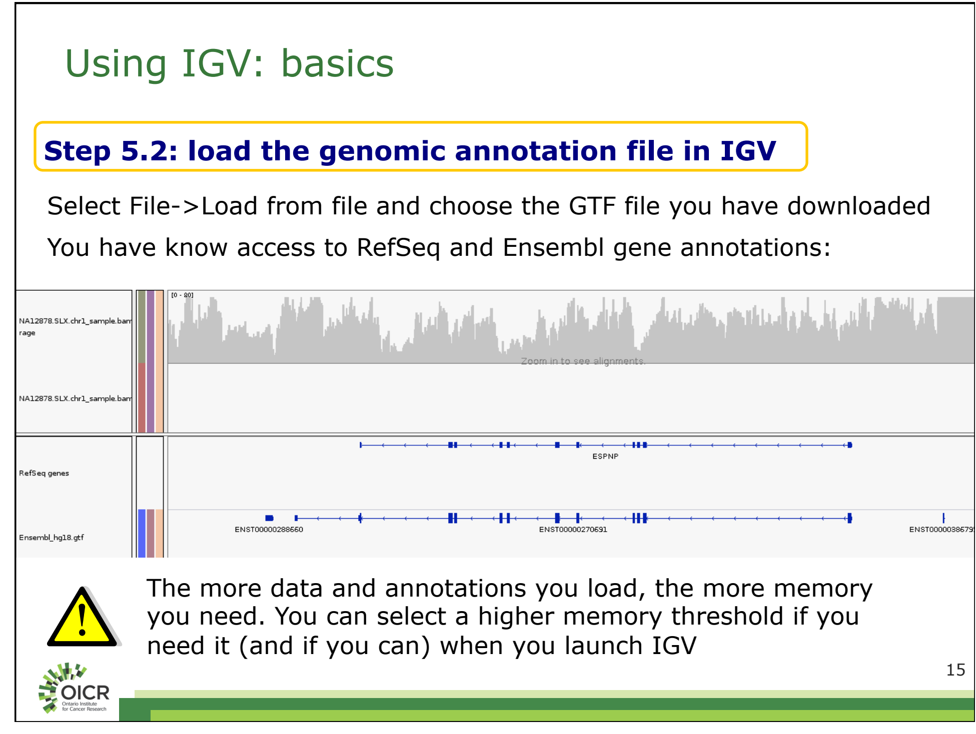
Using IGV: basics
File View Tracks Help
Human hg18
chr7
chr7:55,172,661-55,184,735
Go
Cytoband
Genomic coordinates
12 kb
55,174 kb
55,176 kb
55,178 kb
55,180 kb
55,182 kb
55,184 kb
NA12878.SLX.egfr.bam Covera
Track names
NA12878.SLX.egfr.bam
Data panel
Sample = NA12878
Read group = SRR001798
Read name = L-XAT_0001_FC208BFAAXX:1:14:815:506
Alignment start = 55178902 (-)
Cigar = 47M
Mapped = yes
Mapping quality = 99
Base = A
Base phred quality = 24
H0 = 1
H1 = 0
MF = 0
RG = SRR001798
NM = 2
UQ = 19
RefSeq genes
EGFR
Genomic annotations (default: RefSeq)
chr7:55,178,906
336M of 495M
10


Using IGV: basics
Coverage of reads on the sequence
White reads: low alignment score
Other read colors: depend on the color code selected
Annotated intron
Annotated exon
11

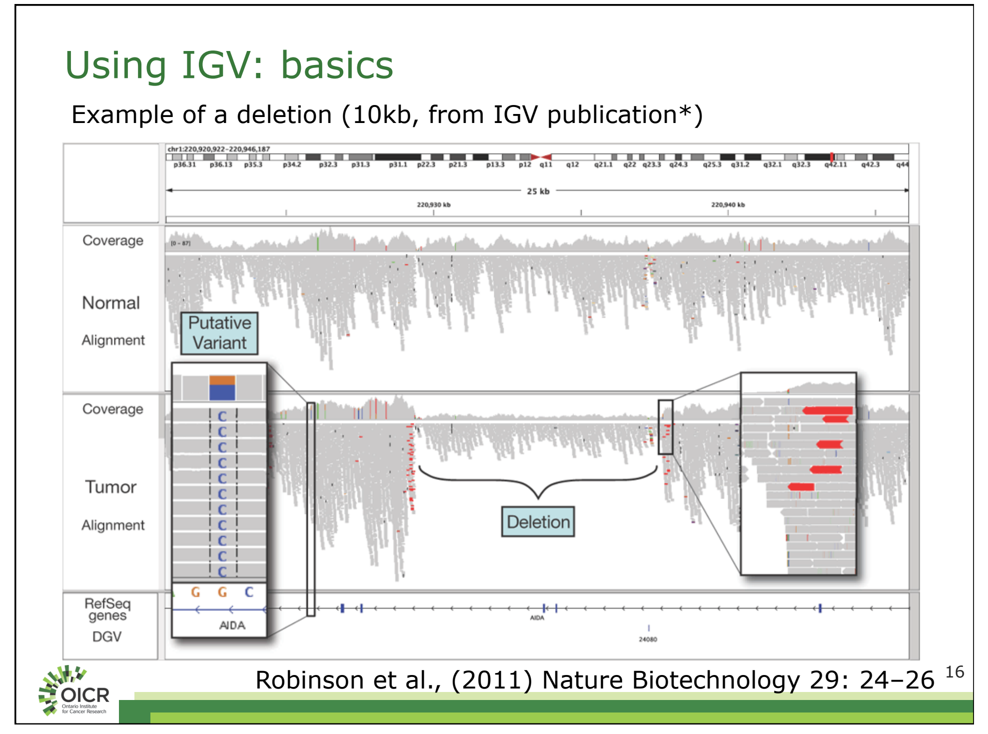

Using IGV: basics
2 examples of variation compared to the reference sequence
Lighter color bases: low quality bases
Reference sequence (here hg18)
12


Using IGV: basics
Step 5.1: download genomic annotations file or other tracks from the UCSC table browser
1) Go on http://genome.ucsc.edu and click on Tables
File Edit View History Bookmarks Tools Help
http://genome.ucsc.edu/index.html?org=Human&db=hg19
UCSC Genome Bioinformatics
Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene
Genome Browser
About the UCSC Genome Bioinformatics Site
Welcome to the UCSC Genome Browser website. This site c
Several ways of downloading gene annotation files can be used, for example directly from the source sequence databases
OICR
13

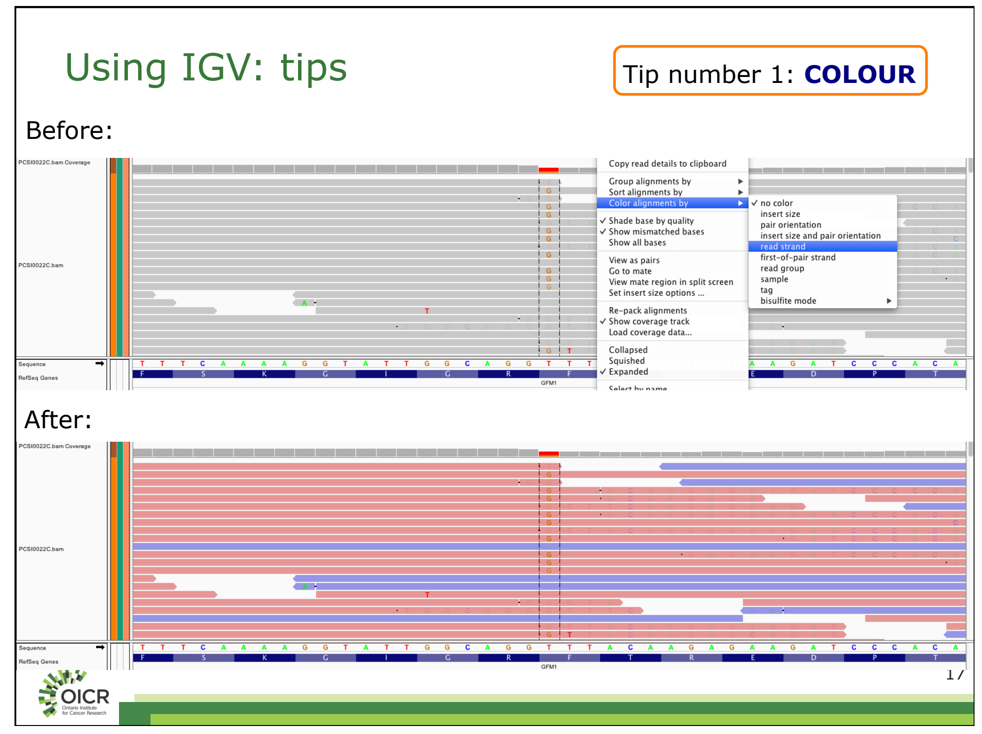

Using IGV: basics
Table Browser
Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
clade: Mammal genome: Human assembly: Mar. 2006 (NCBI36/hg18)
Select the genome (e.g. hg18)
group: Genes and Gene Prediction Tracks track: Ensembl Genes
Select the gene annotations (e.g. Ensembl Genes)
table: ensGene
region: genome ENCODE Pilot regions position chrX:151073054-151383976
identifiers (names/accessions): paste list upload list
filter: create
intersection: create
correlation: create
Select the file format (e.g. GTF)
output format: GTF - gene transfer format Send output to Galaxy GREAT
output file: Ensembl_hg18.gtf (leave blank to keep output in browser)
file type returned: plain text
Choose your file name and click on the "get output" button
get output summary/statistics
OICR
14

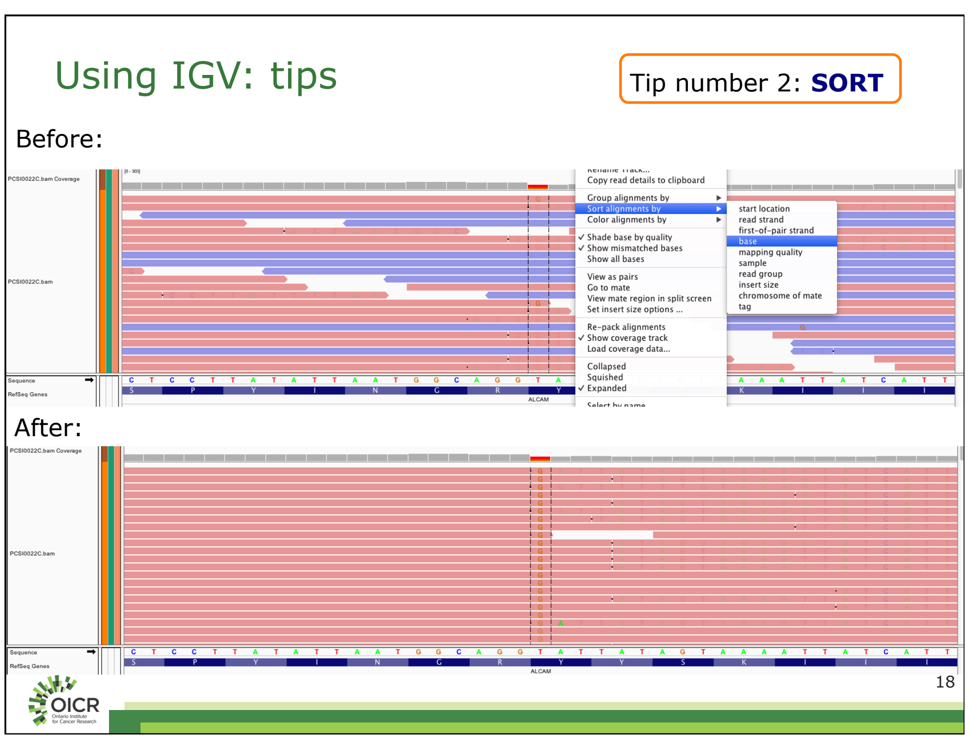


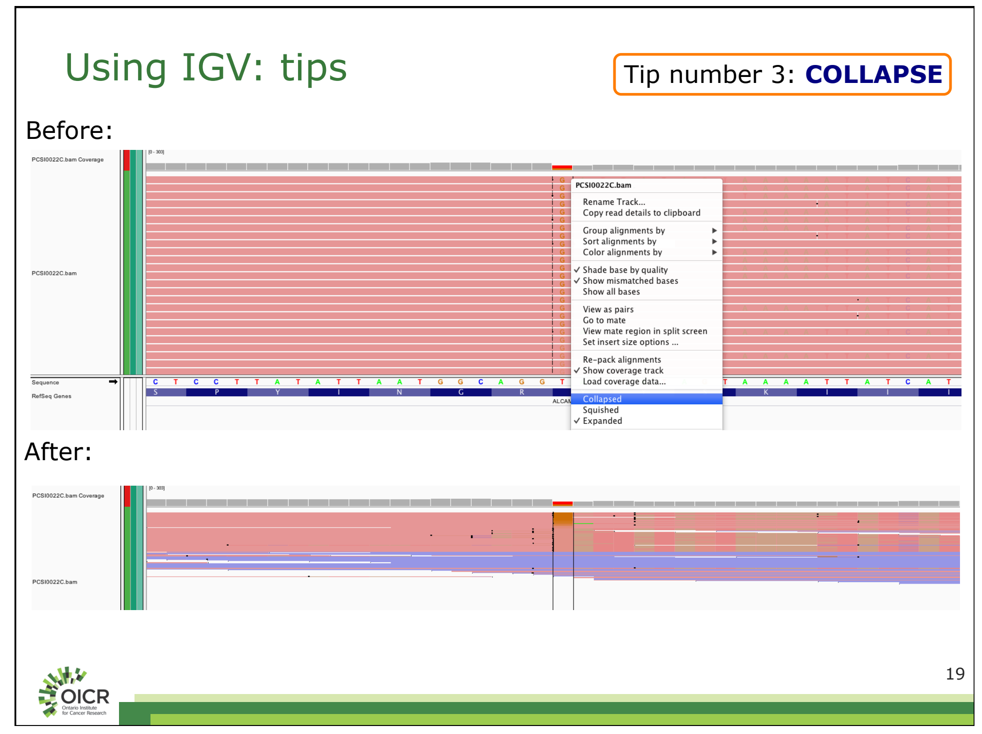
Using IGV: basics
Step 5.2: load the genomic annotation file in IGV
Select File->Load from file and choose the GTF file you have downloaded
You have know access to RefSeq and Ensembl gene annotations:
RefSeq genes
Ensembl_hg18.gtf
ESPNP
The more data and annotations you load, the more memory you need. You can select a higher memory threshold if you need it (and if you can) when you launch IGV
OICR
15

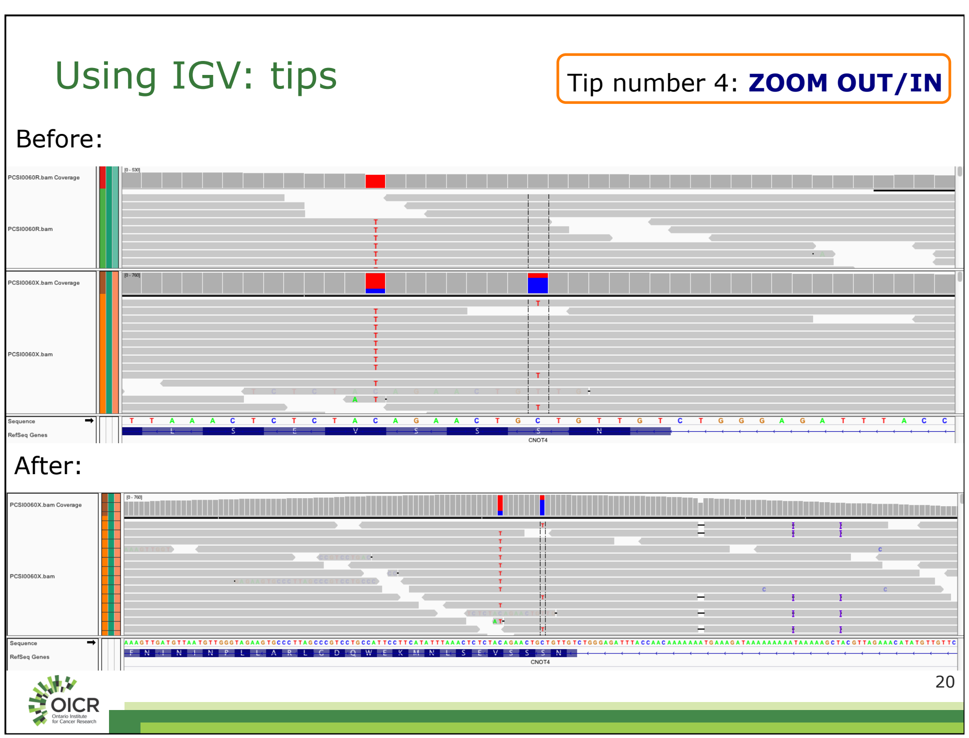


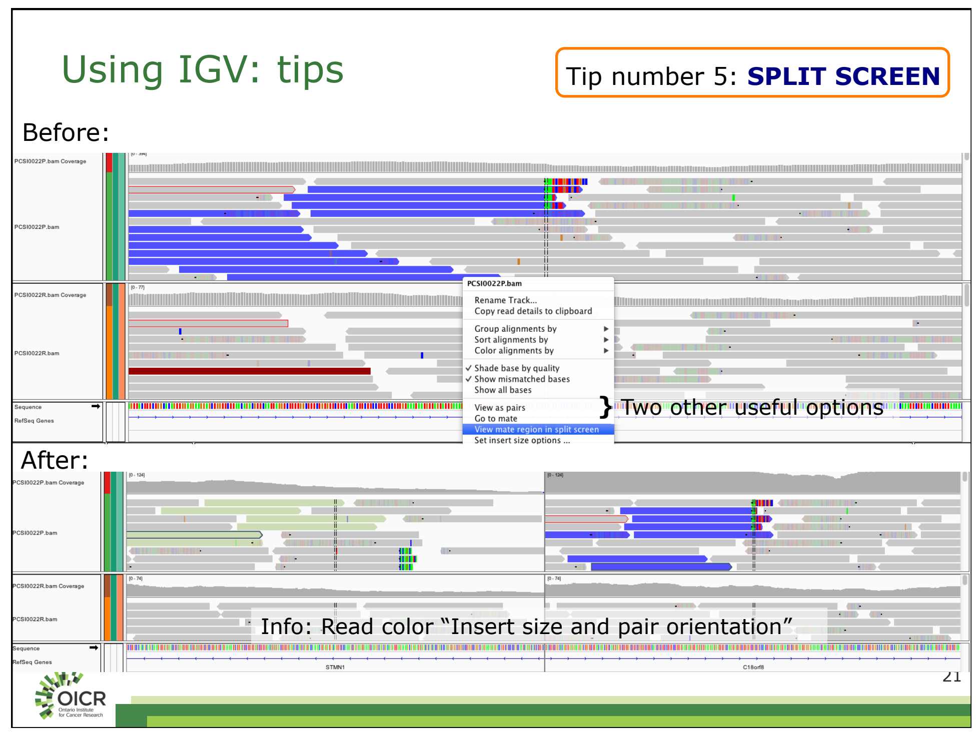
Using IGV: basics
Example of a deletion (10kb, from IGV publication*)
25 kb
Coverage
Normal
Alignment
Putative Variant
Coverage
Tumor
Alignment
Deletion
RefSeq genes
DGV
G G C
AIDA
Robinson et al., (2011) Nature Biotechnology 29: 24–26
OICR
16

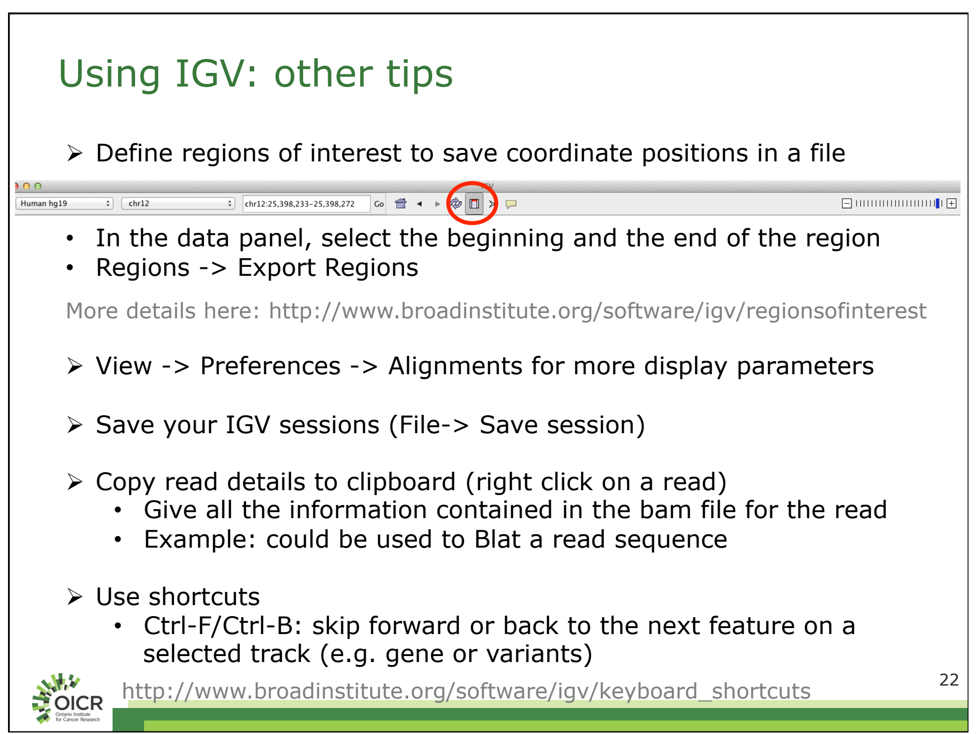

Using IGV: tips
Tip number 1: COLOUR
Before:
Copy read details to clipboard
Group alignments by
Sort alignments by
Color alignments by
Shade base by quality
Show mismatched bases
Show all bases
View as pairs
Go to mate
View mate region in split screen
Set insert size options ...
Re-pack alignments
Show coverage track
Load coverage data...
Collapsed
Squished
Expanded
no color
insert size
pair orientation
insert size and pair orientation
read strand
first-of-pair strand
read group
sample
tag
bisulfite mode
After:
OICR
17


Using IGV: tips
Tip number 2: SORT
Before:
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
Squished
Expanded
start location
read strand
first-of-pair strand
base
mapping quality
sample
read group
insert size
chromosome of mate
tag
After:
OICR
18


Using IGV: tips
Tip number 3: COLLAPSE
Before:
PCSI0022C.bam
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
Squished
Expanded
After:
19
OICR


Using IGV: tips
Tip number 4: ZOOM OUT/IN
Before:
After:
20
OICR


Using IGV: tips
Tip number 5: SPLIT SCREEN
Before:
Two other useful options
After:
Info: Read color "Insert size and pair orientation"
OICR
21


Using IGV: other tips
Define regions of interest to save coordinate positions in a file
In the data panel, select the beginning and the end of the region
Regions -> Export Regions
More details here: http://www.broadinstitute.org/software/igv/regionsofinterest
View -> Preferences -> Alignments for more display parameters
Save your IGV sessions (File-> Save session)
Copy read details to clipboard (right click on a read)
Give all the information contained in the bam file for the read
Example: could be used to Blat a read sequence
Use shortcuts
Ctrl-F/Ctrl-B: skip forward or back to the next feature on a selected track (e.g. gene or variants)
http://www.broadinstitute.org/software/igv/keyboard_shortcuts
OICR
22