Canadian Bioinformatics Workshops

www.bioinformatics.ca

Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska 中文 華語 (台灣) isiZulu

creative commons

Attribution-Share Alike 2.5 Canada

You are free:

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

Free Cultural Works **APPROVED FOR**

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French

Learn how to distribute your work using this licence.

Introduction to IGV
The Integrative Genomics Viewer

Malachi Griffith, Obi Griffith, Jason Walker
Informatics for RNA-seq analysis
June 16 - 17, 2016
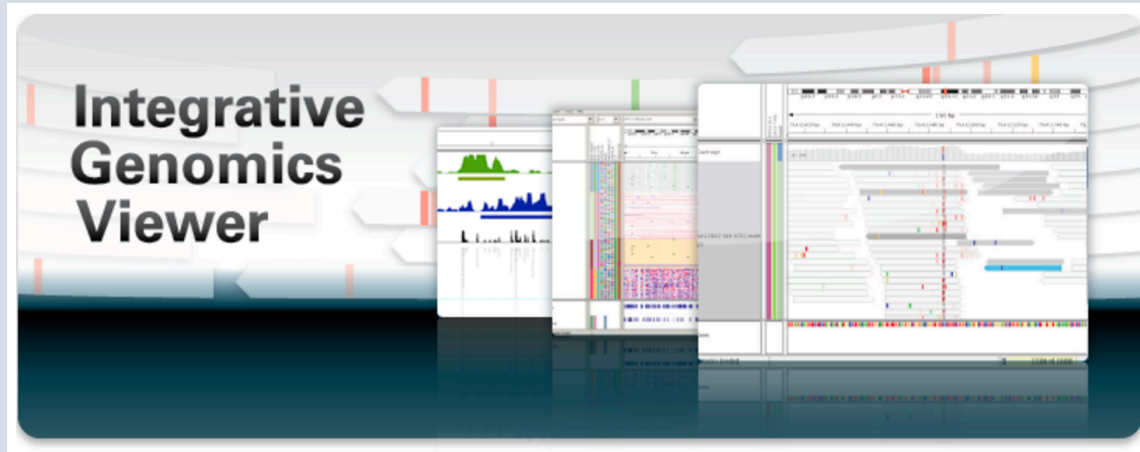
Visualization Tools in Genomics

- there are **over 40 different genome browsers**, which to use?

- depends on
  - task at hand
  - kind and size of data
  - data privacy

HT-seq Genome Browsers

Integrative
Genome
Viewer

UCSC
Genome Browser
Cancer Genome Browser

Trackster
(part of Galaxy)

Savant
Genome
Browser

- task at hand : visualizing HT-seq reads, especially good for inspecting variants
- kind and size of data : large BAM files, stored locally or remotely
- data privacy : run on the desktop, can keep all data private
- UCSC Genome Browser has been retro-fitted to display BAM files
- Trackster is a genome browser that can perform visual analytics on small windows of the genome, deploy full analysis with Galaxy

Integrative Genomics Viewer (IGV)

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

Epigenomics
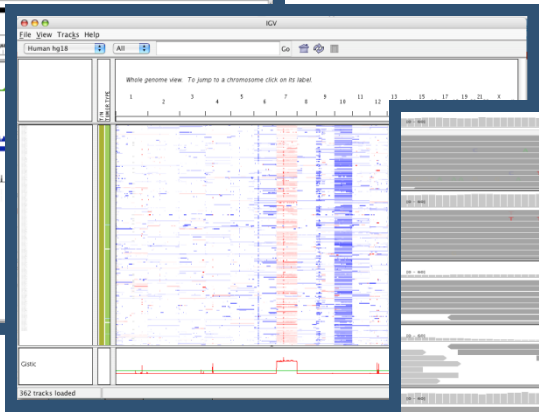
Microarrays
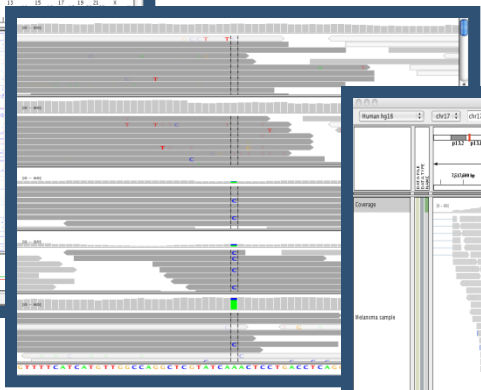
NGS alignments

RNA-Seq

mRNA, CNV, Seq

http://www.broadinstitute.org/igv
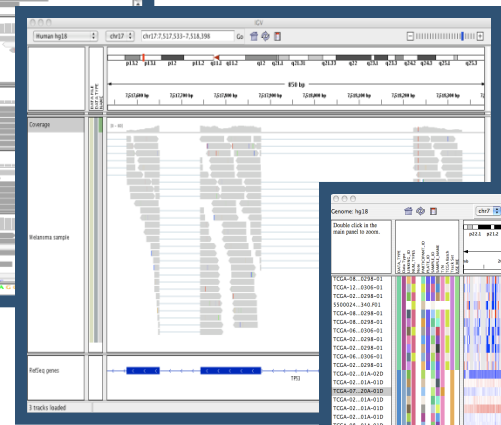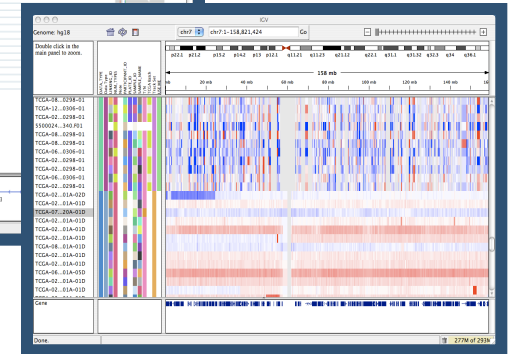
>85,000 registrations (2014)

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".

- Automation of specific tasks using command-line interface

IGV data sources
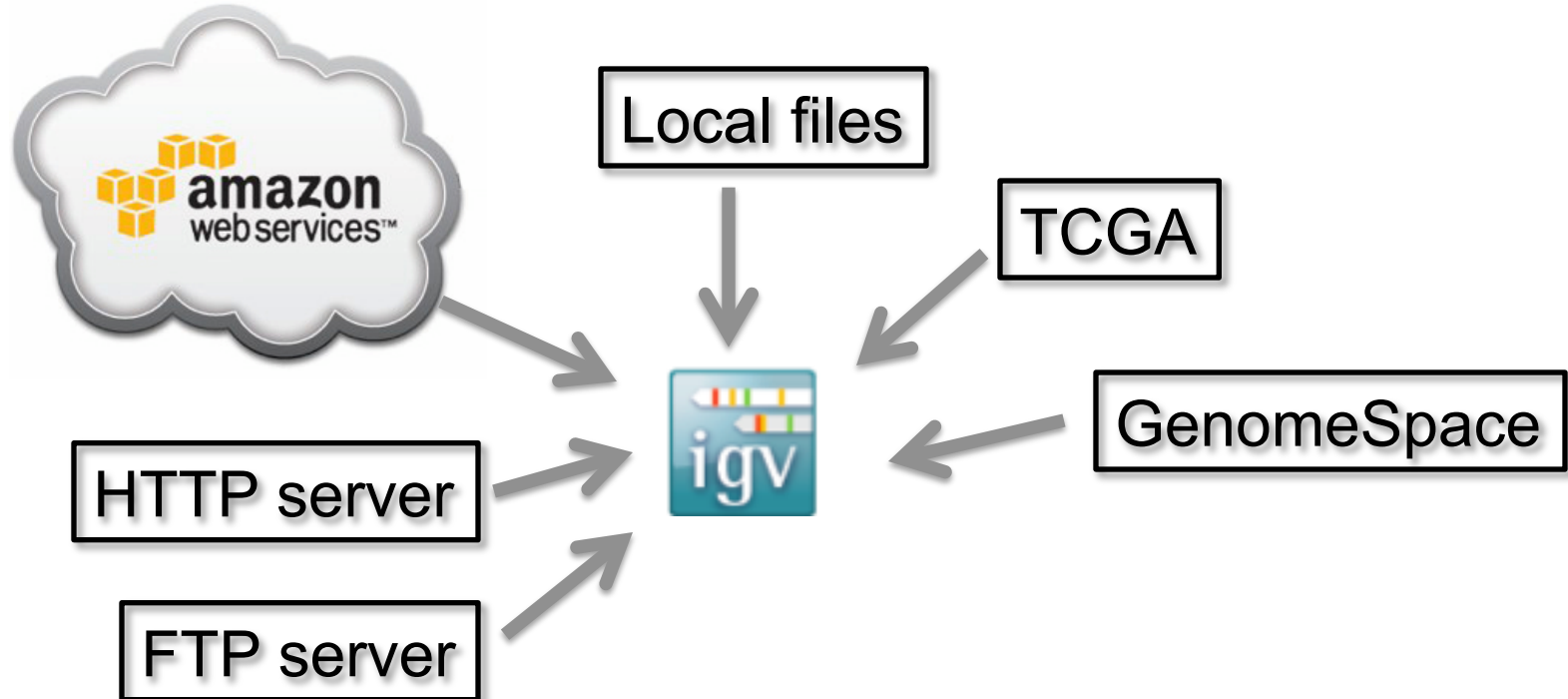
- View **local** files without uploading.

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

Using IGV: the basics

- Launch IGV

- Select a reference genome

- Load data

- Navigate through the data
  - WGS data
    - SNVs
    - structural variations

Launch IGV
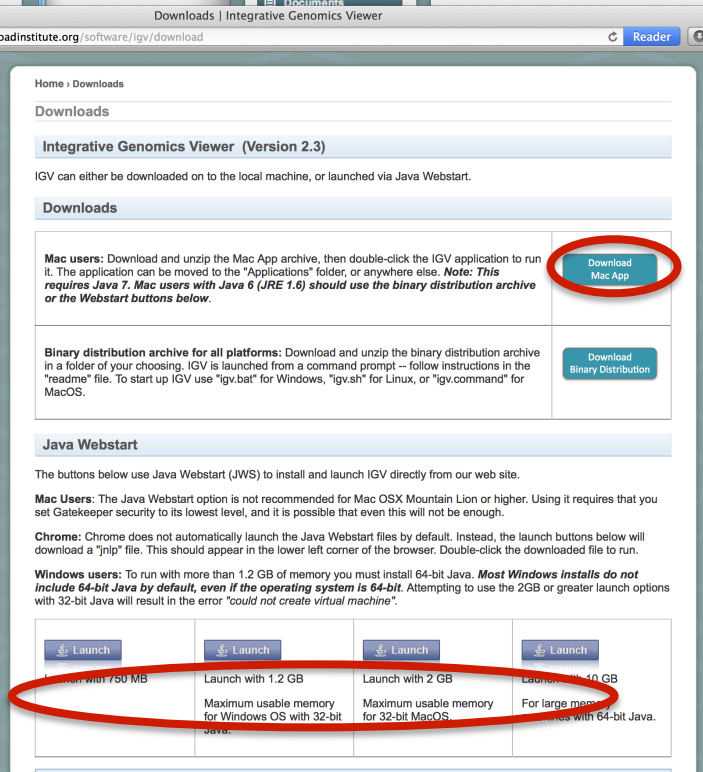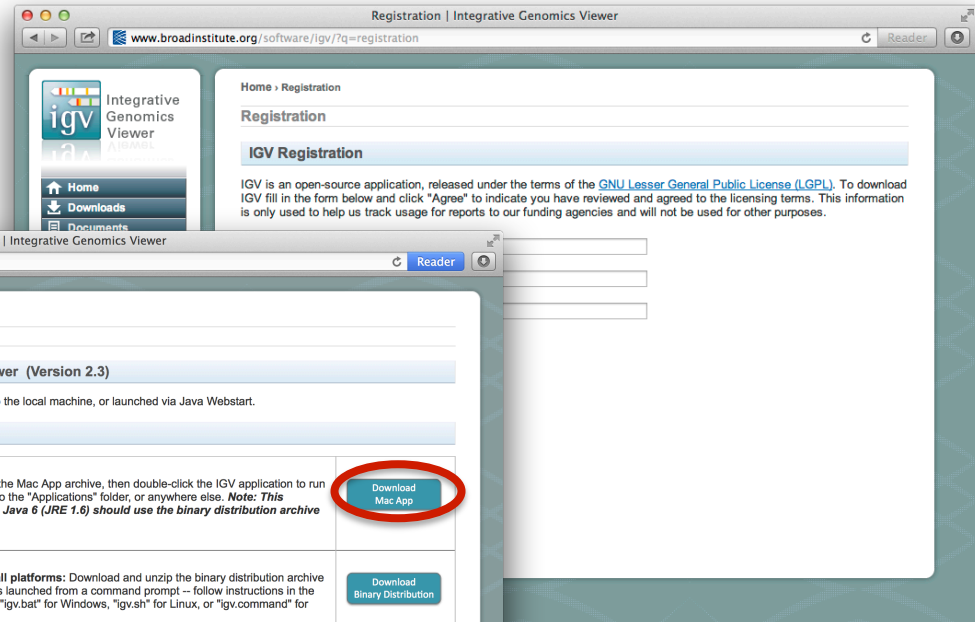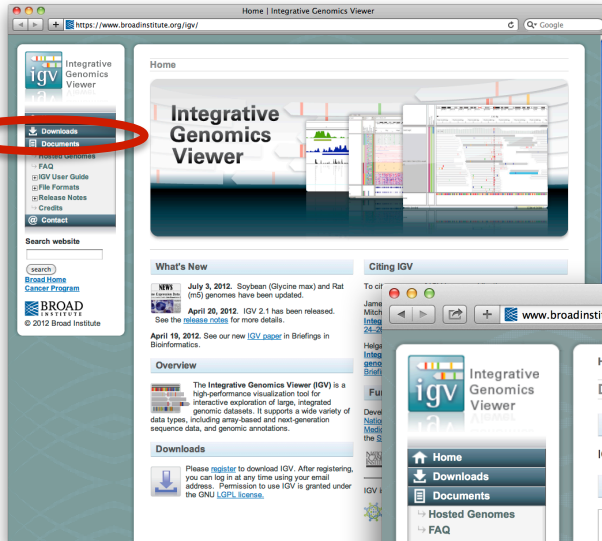
Launch IGV

1. Select genome from the drop-down menu

2. Load data

Load data

Select **File > Load from Server…**
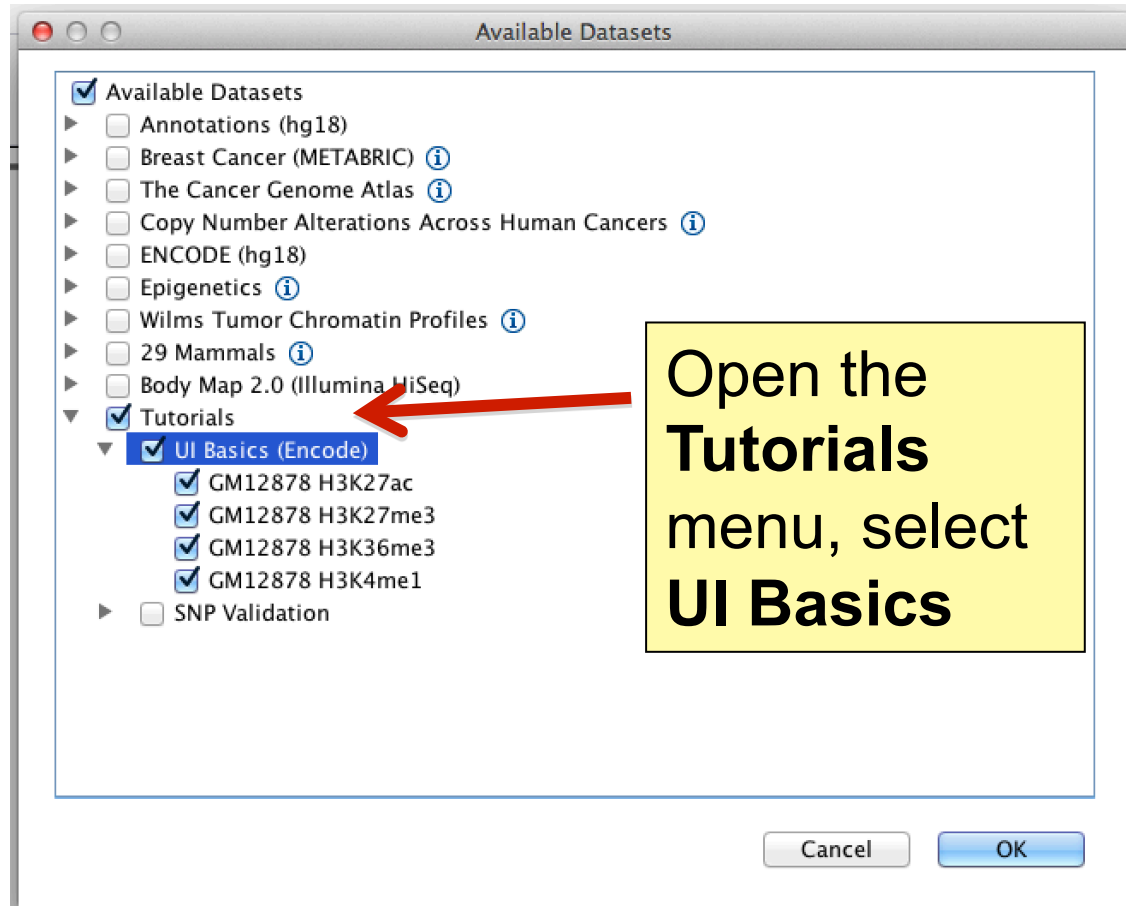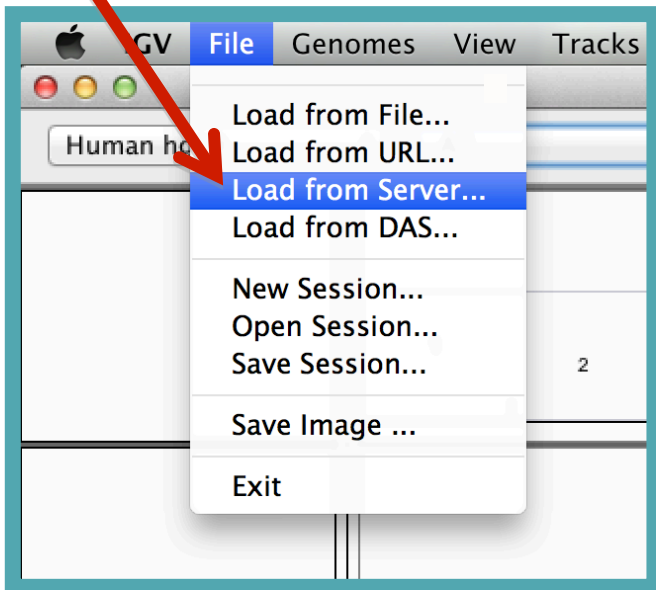
Open the **Tutorials** menu, select **UI Basics**

Screen layout
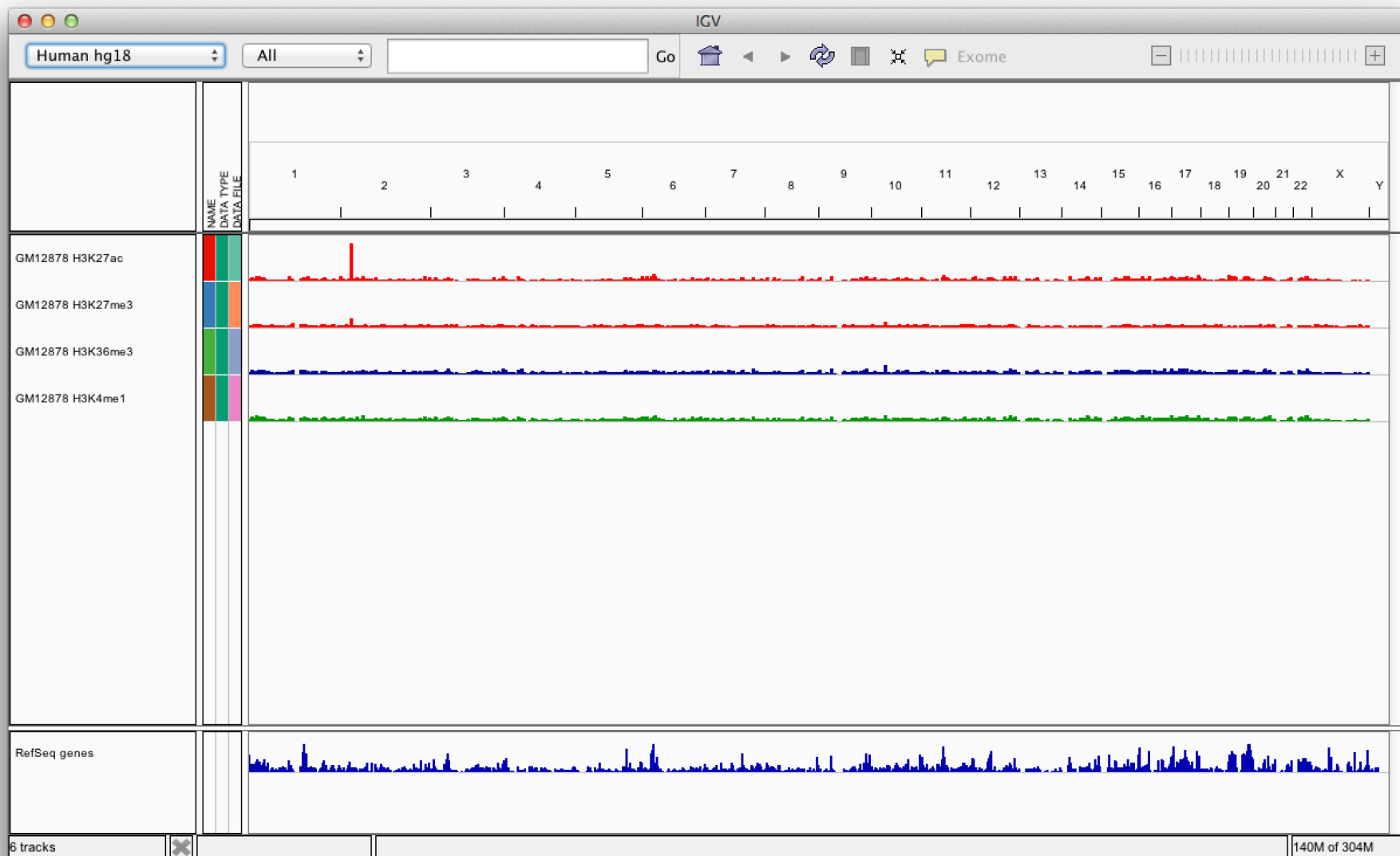
Screen layout
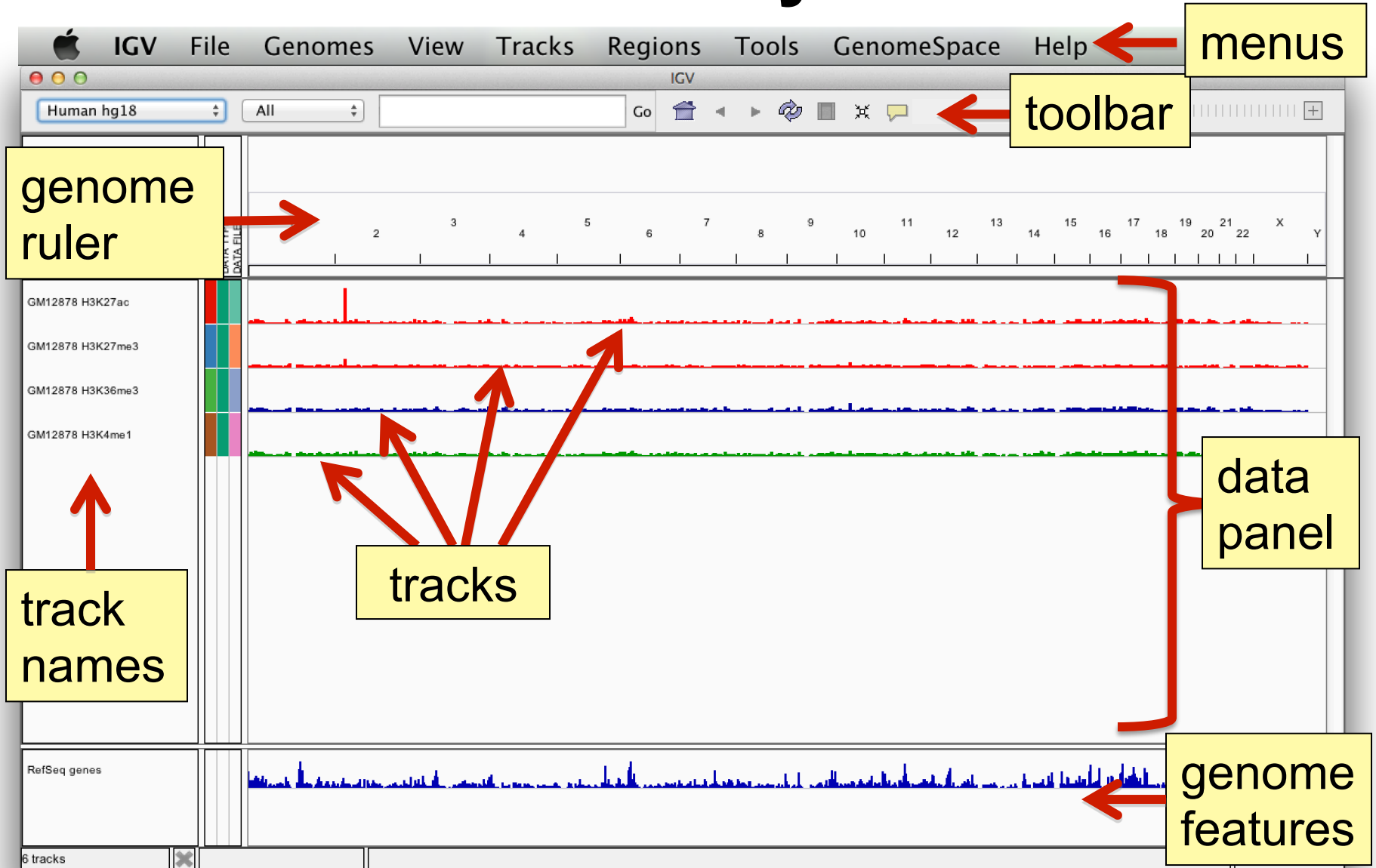
File formats and track types
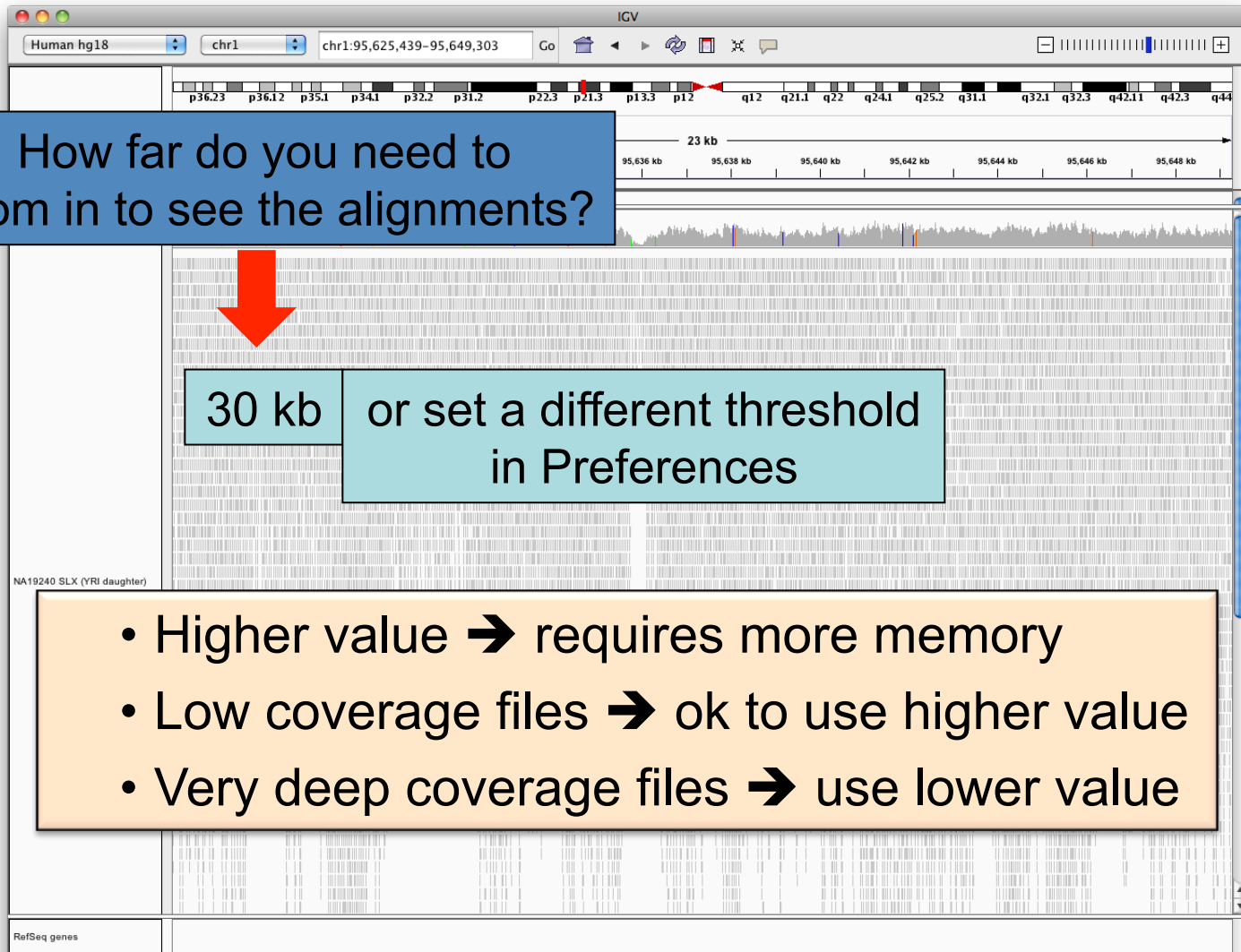
- The **file format** defines the track type.

- The **track type** determines the display options

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

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

Viewing alignments

Whole chromosome view
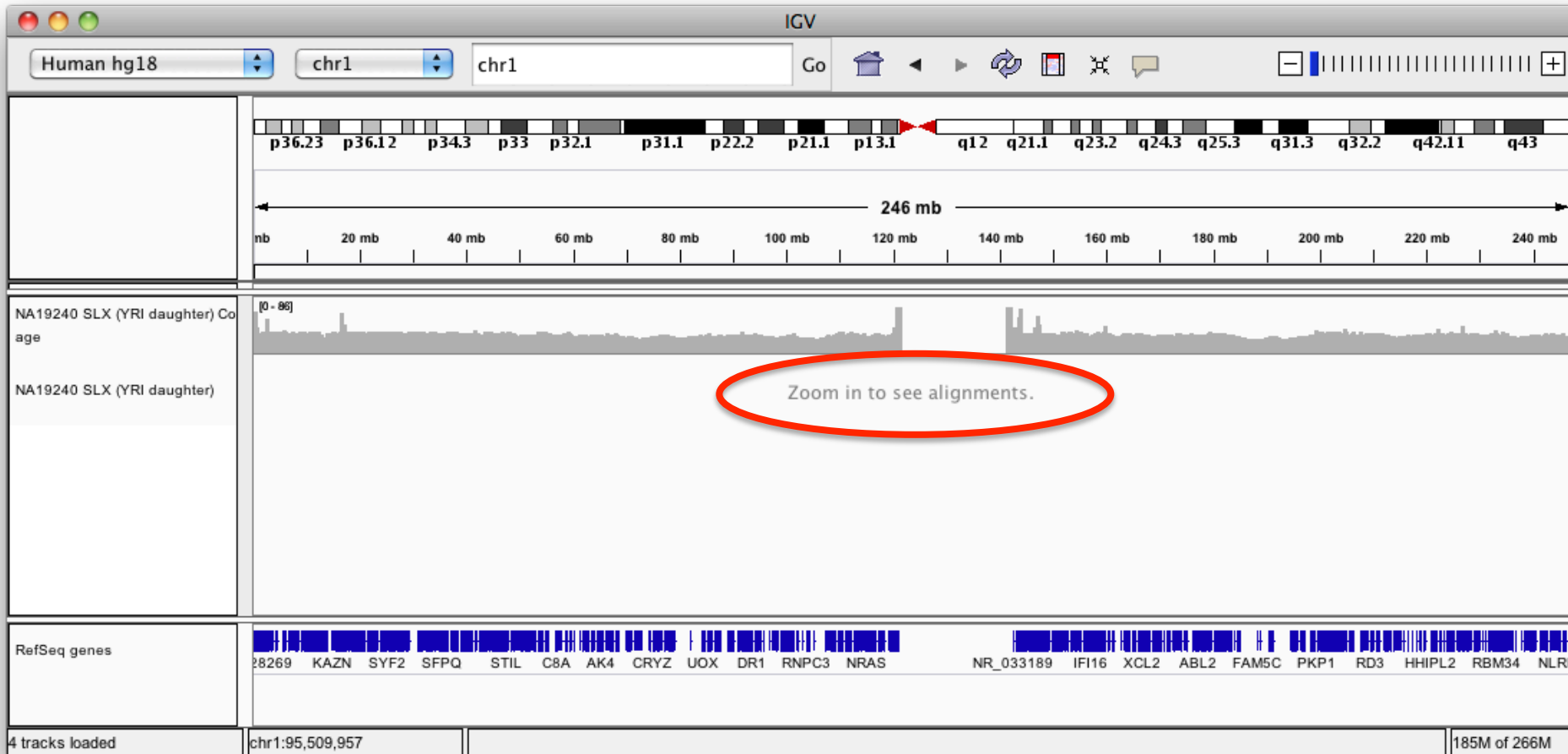
Viewing alignments – Zoom in

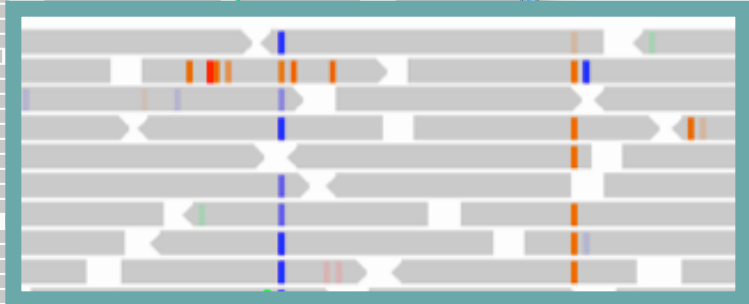How far do you need to zoom in to see the alignments?

30 kb | or set a different threshold in Preferences

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

Viewing alignments – Zoom in
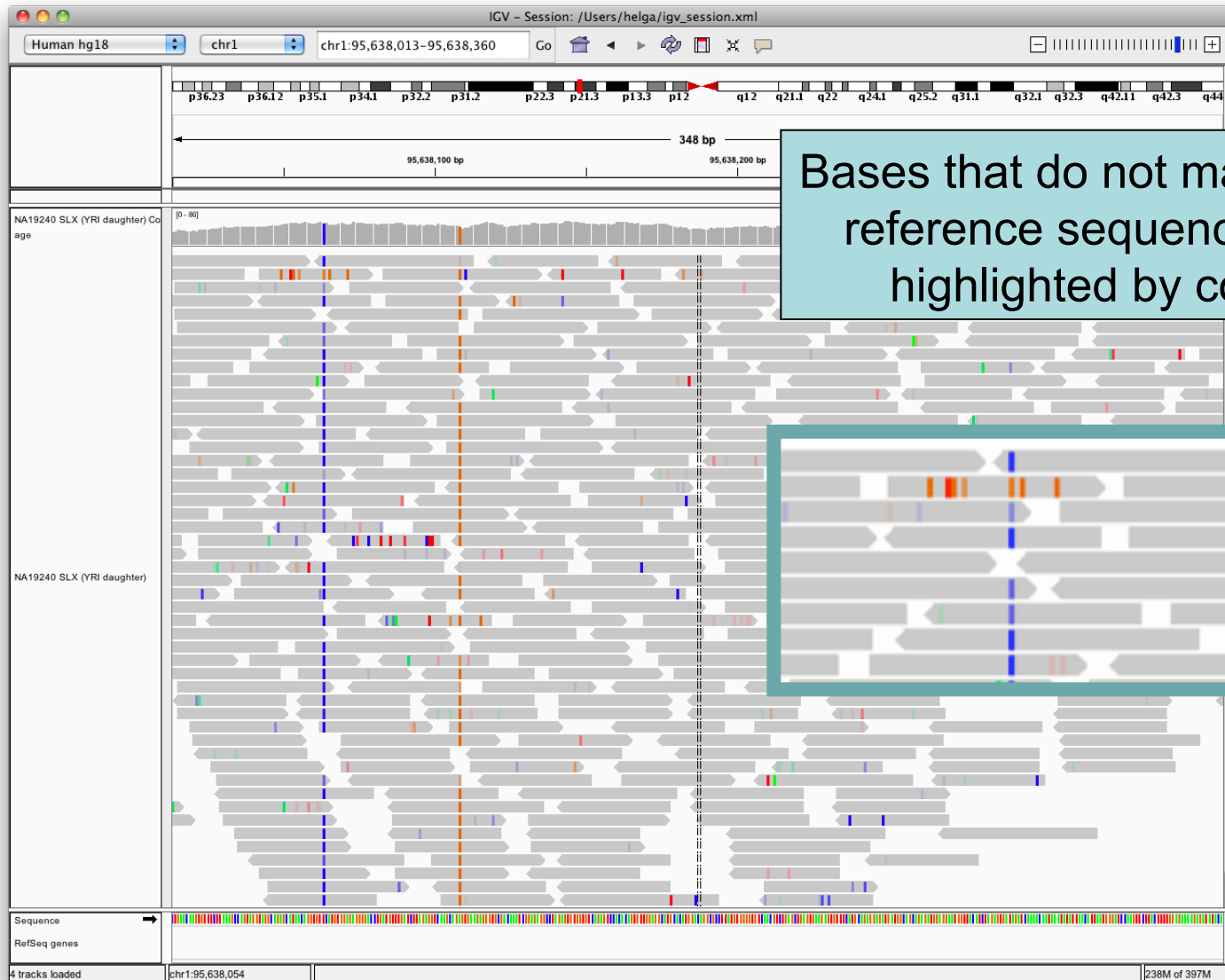
Bases that do not match the reference sequence are highlighted by color

SNVs and Structural variations

- Important metrics for evaluating the validity of SNVs:
  - Coverage
  - Amount of support
  - Strand bias / PCR artifacts
  - Mapping qualities
  - Base qualities

- Important metrics for evaluating SVs:
  - Coverage
  - Insert size
  - Read pair orientation

Viewing SNPs and SNVs
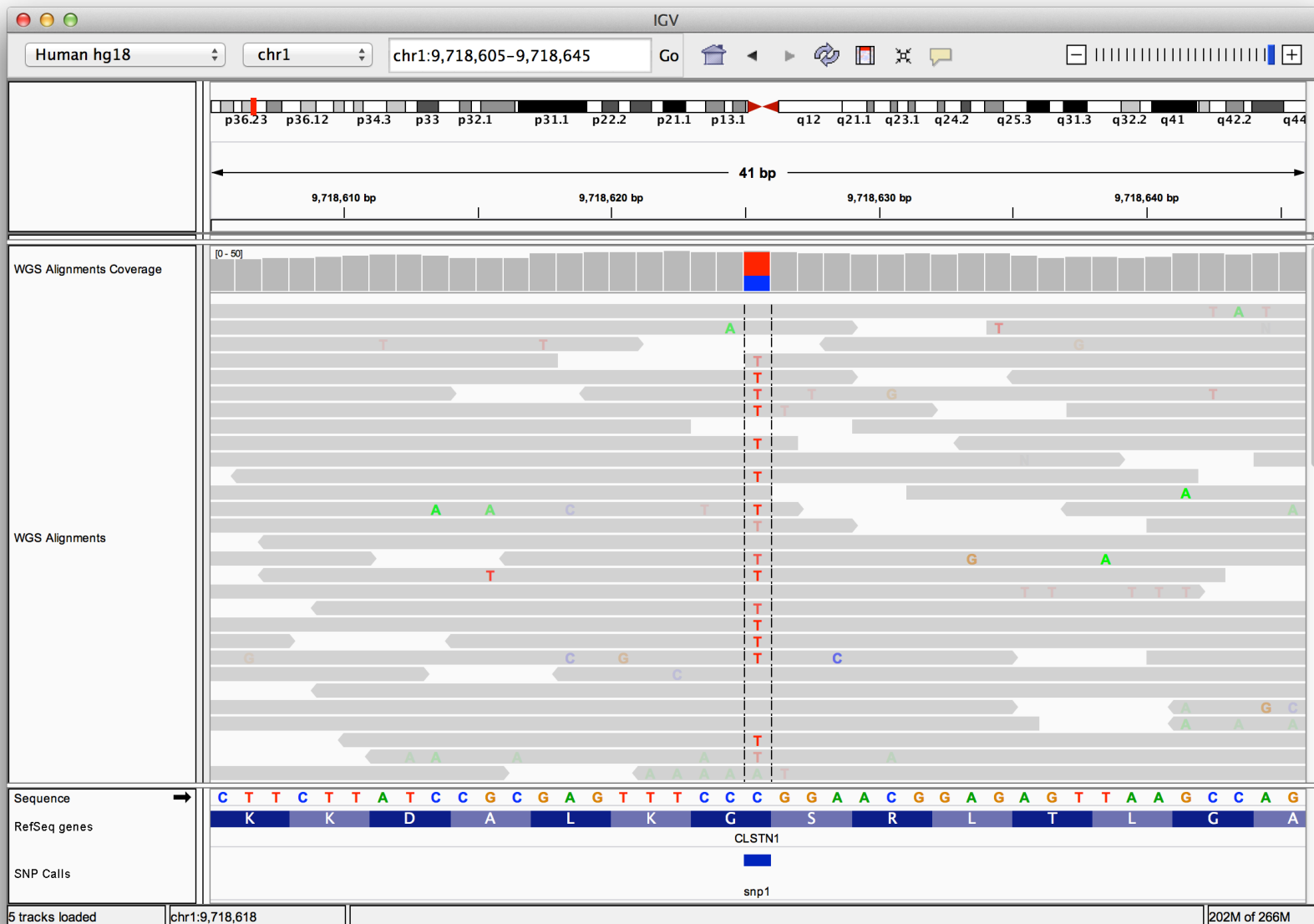
Viewing SNPs and SNVs

Viewing 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)

bioinformatics.ca

Paired-end sequencing

DNA or cDNA

Fragment and size select

Paired-end sequencing

DNA or
cDNA

Fragment and
size select

Read from
each end

insert size

Paired-end sequencing

DNA or cDNA

Fragment and size select

insert size

Read from each end

Align to Reference
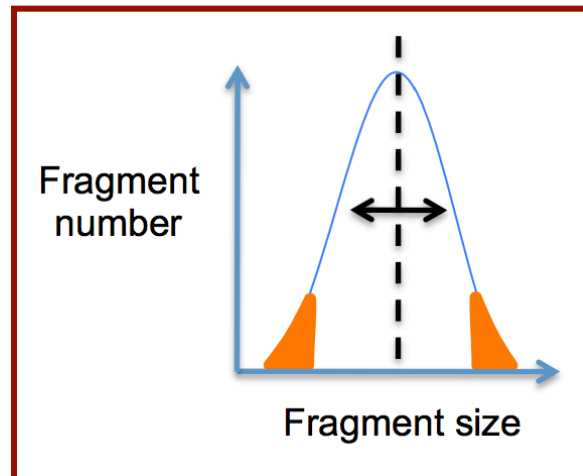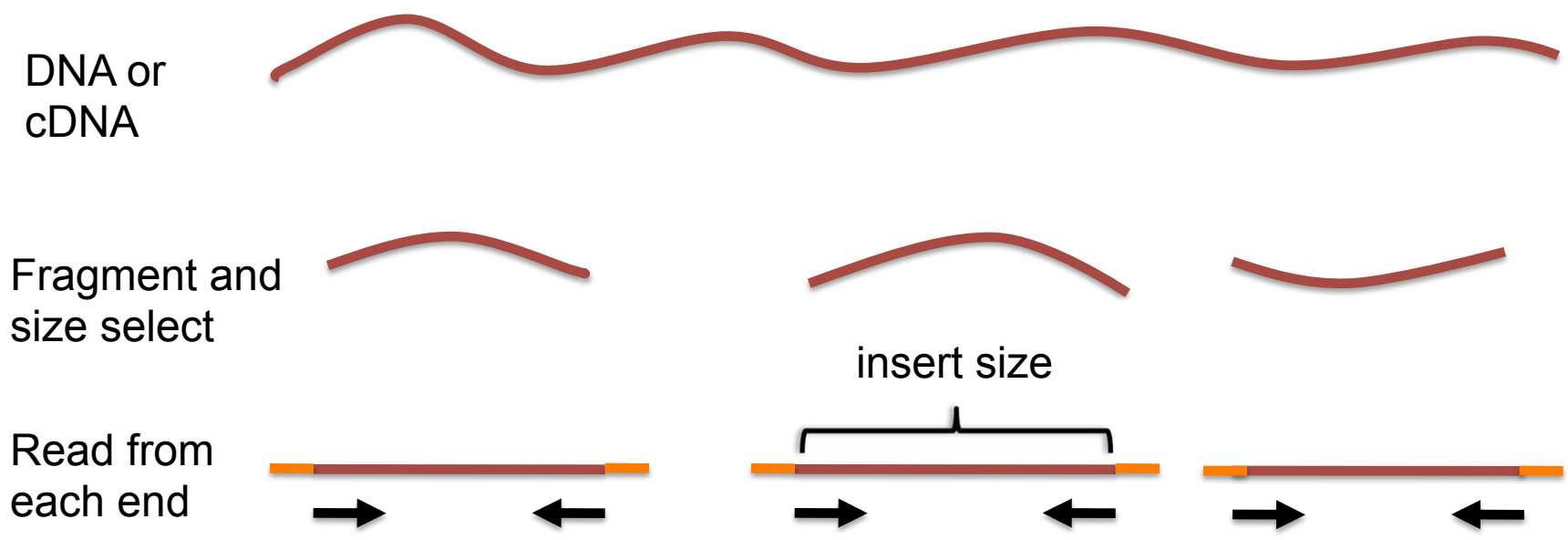
inferred insert size
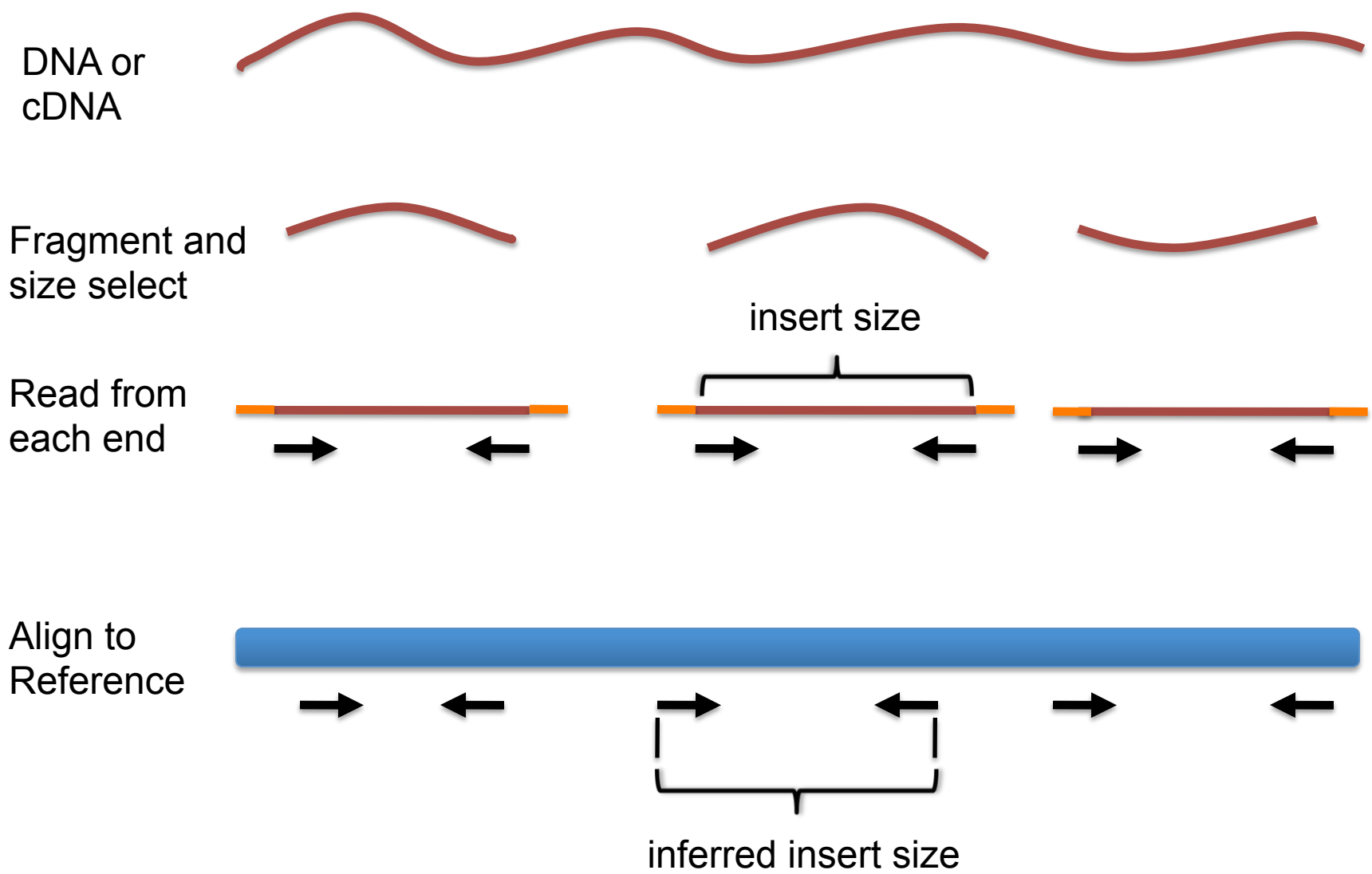
Interpreting inferred insert size

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
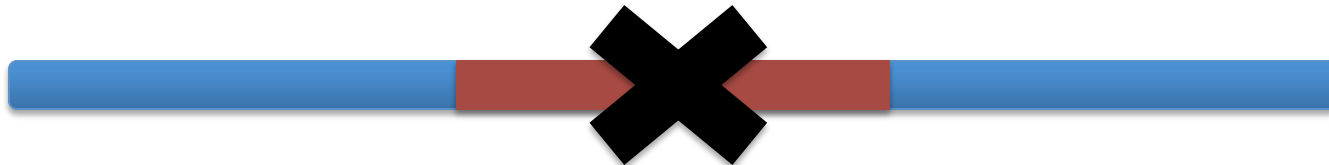
Reference Genome

Subject

Deletion

Reference
Genome

Subject

bioinformatics.ca

Deletion

Deletion

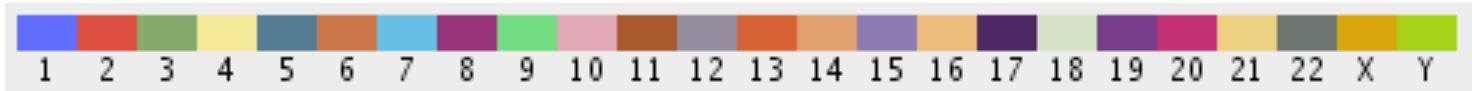# Color by insert size

Deletion

Note drop in coverage

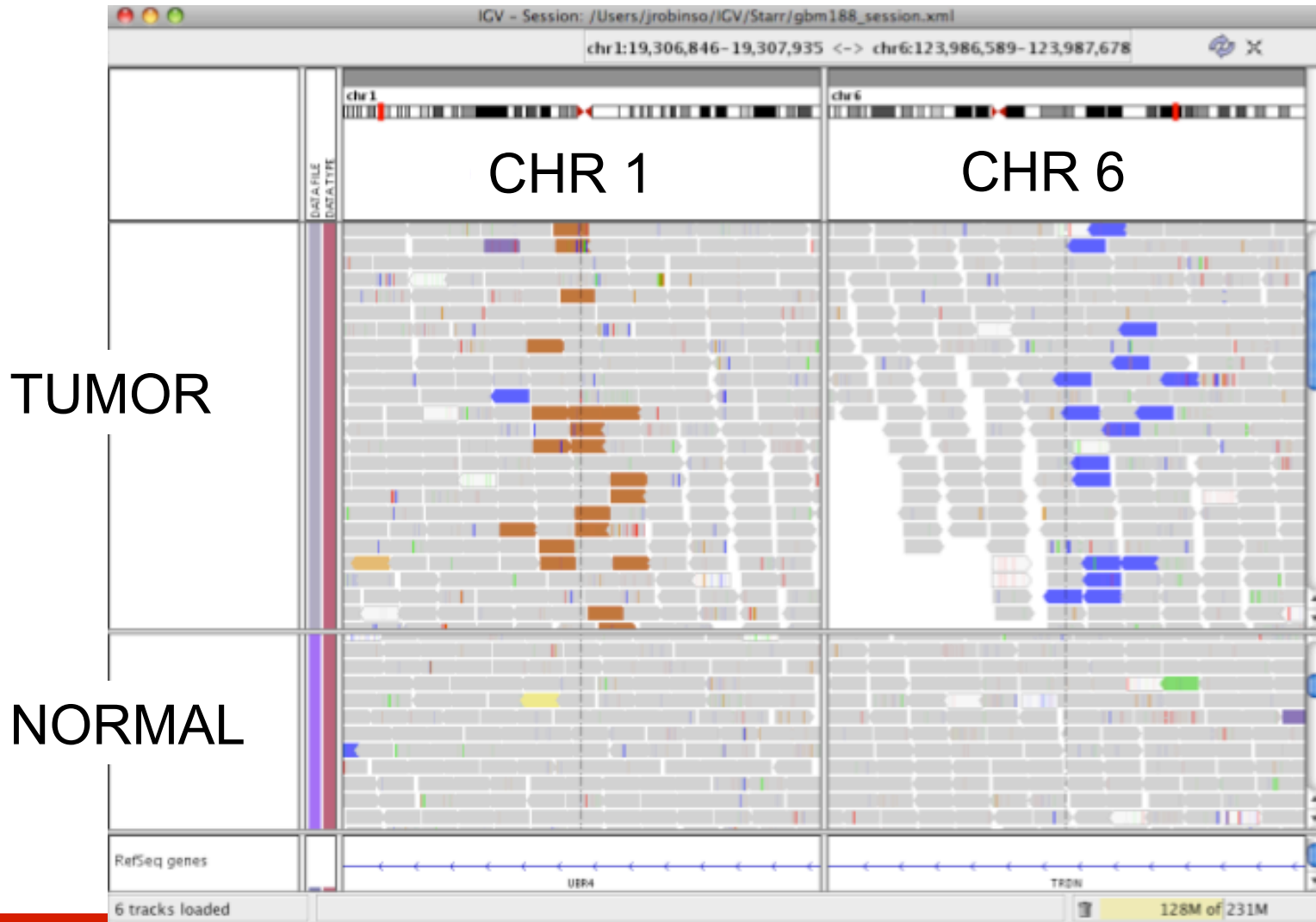Pairs with larger than expected insert size are colored red.

Insert size color scheme

- Smaller than expected insert size:

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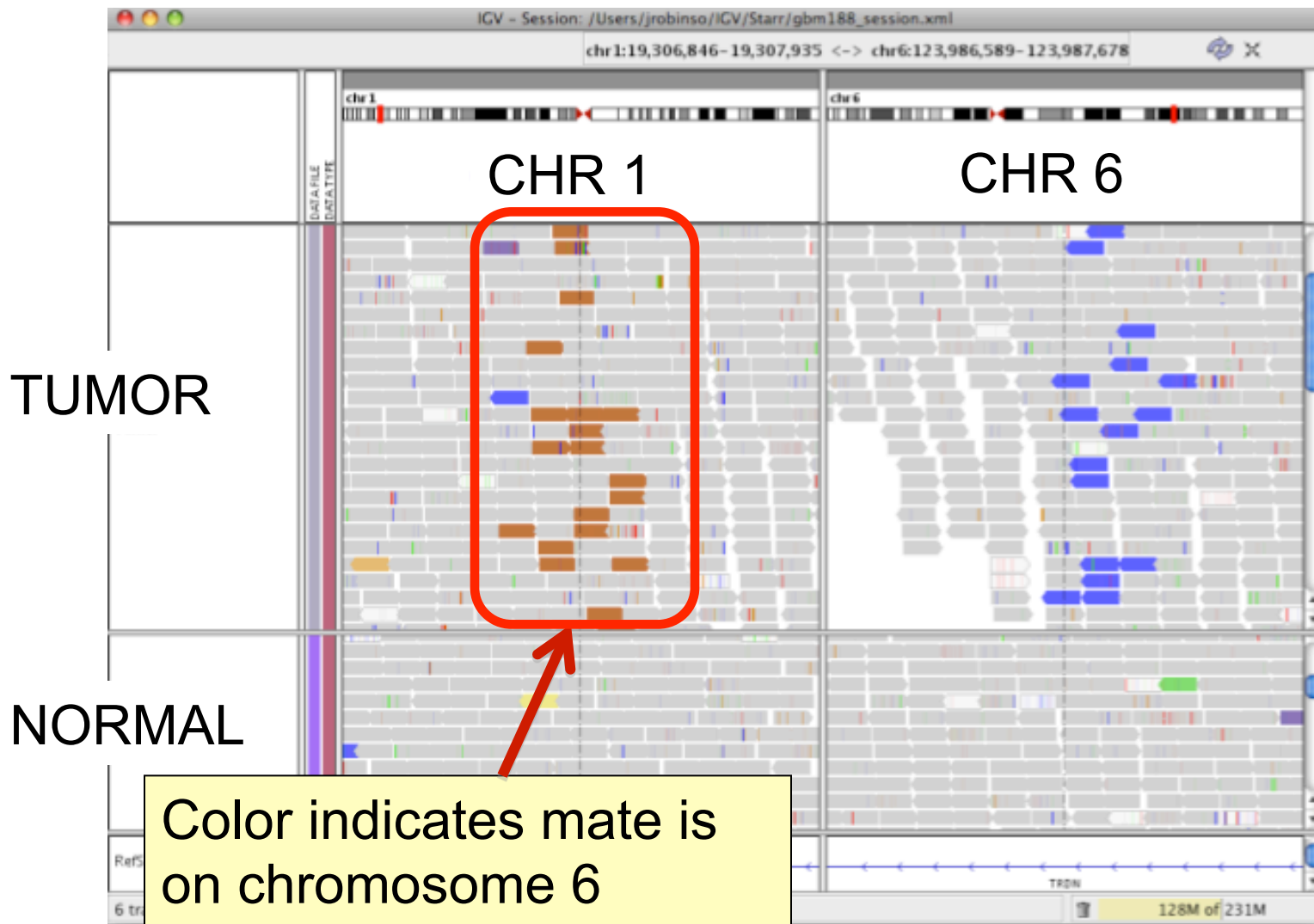
- Pairs on different chromosomes

  Each end colored by chromosome of its mate

Rearrangement

Rearrangement

Interpreting Read-Pair Orientations

Orientation of paired reads can reveal structural events:

- Inversions

- Duplications

- Translocations

- Complex rearrangements

Orientation is defined in terms of

- read strand, left *vs* right, *and*

- read order, first *vs* second

Inversion
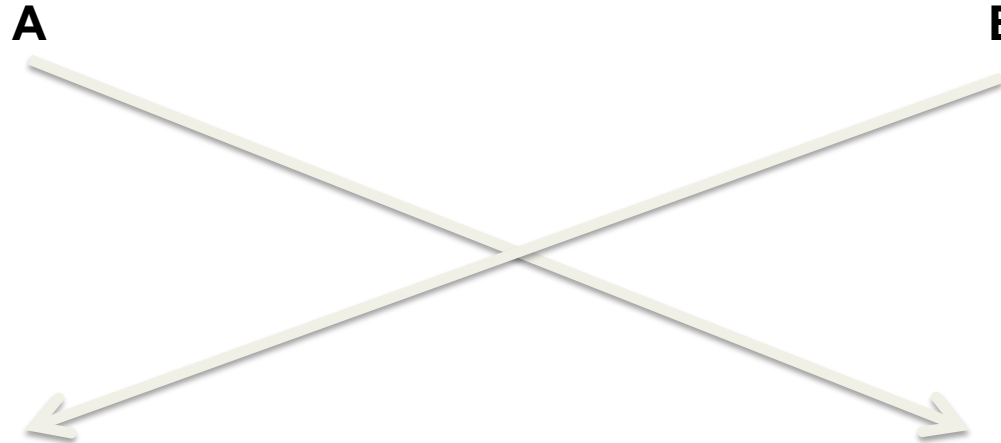
Reference
genome

Inversion

Reference
genome

A   B
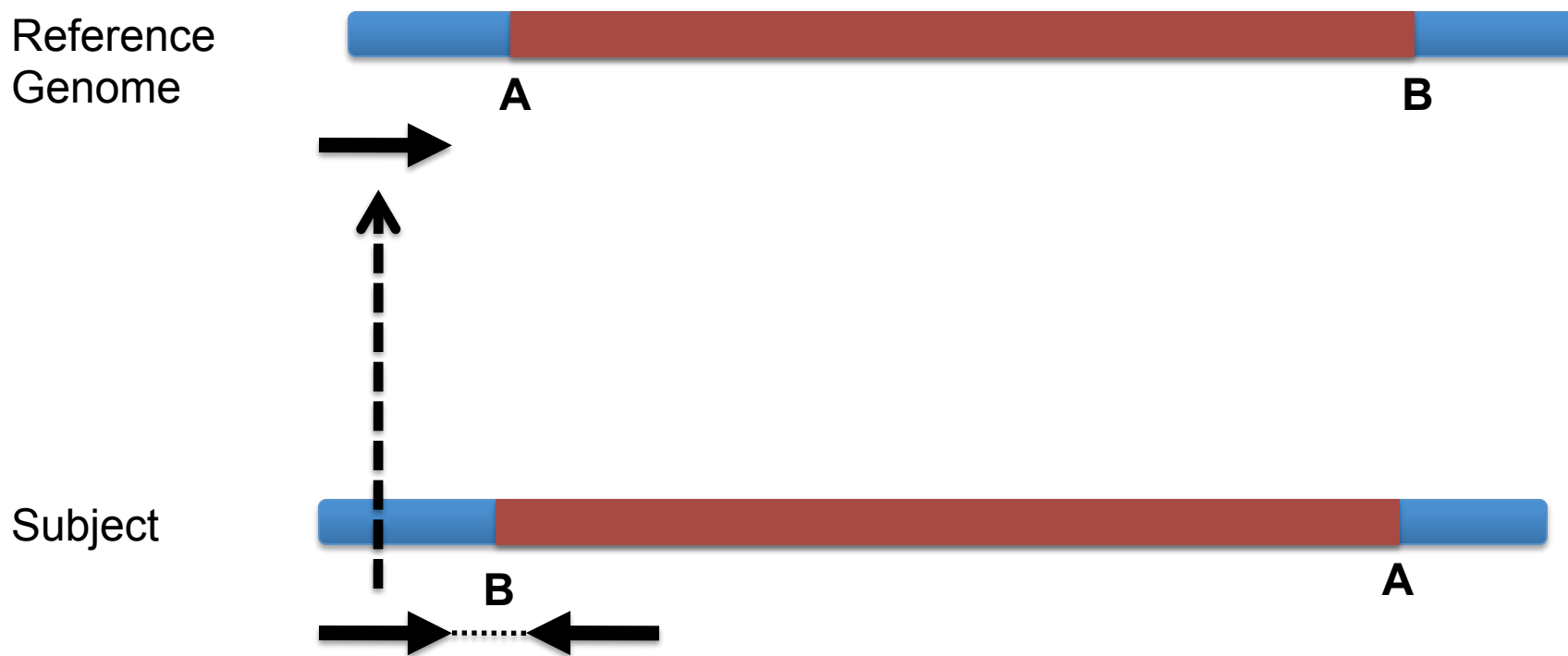
bioinformatics.ca

Inversion

Inversion

Reference Genome

Subject

Inversion

Inversion

Inversion

Reference Genome

A   B

Subject

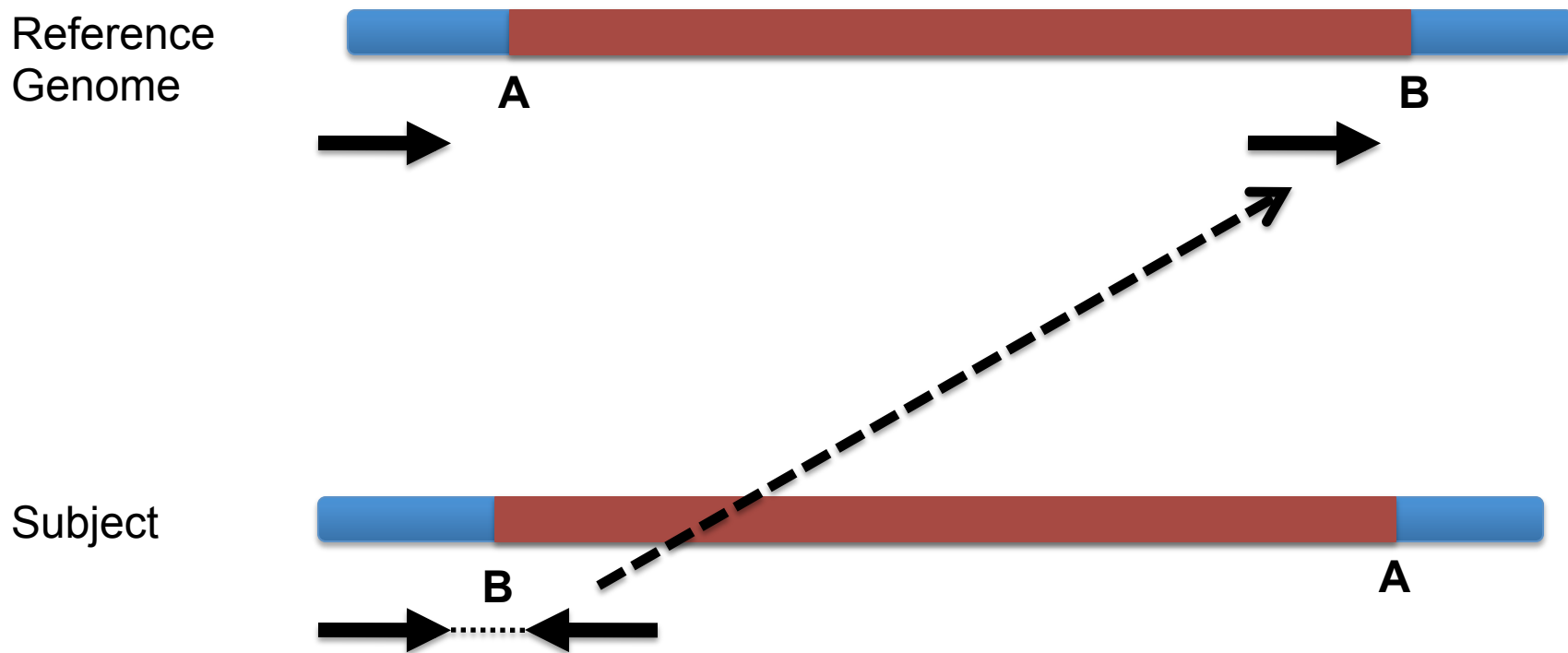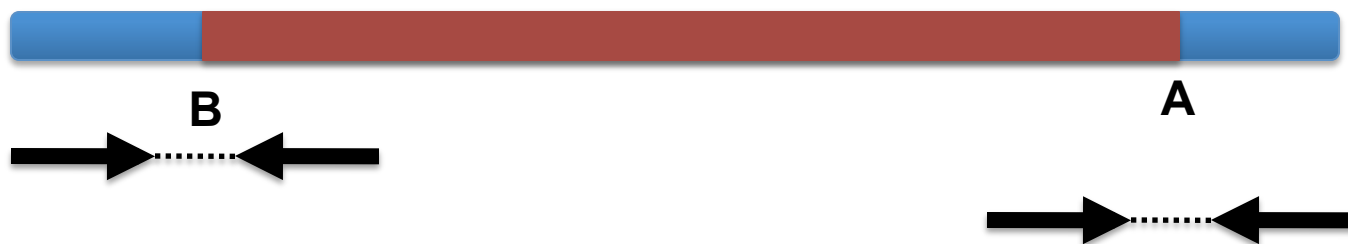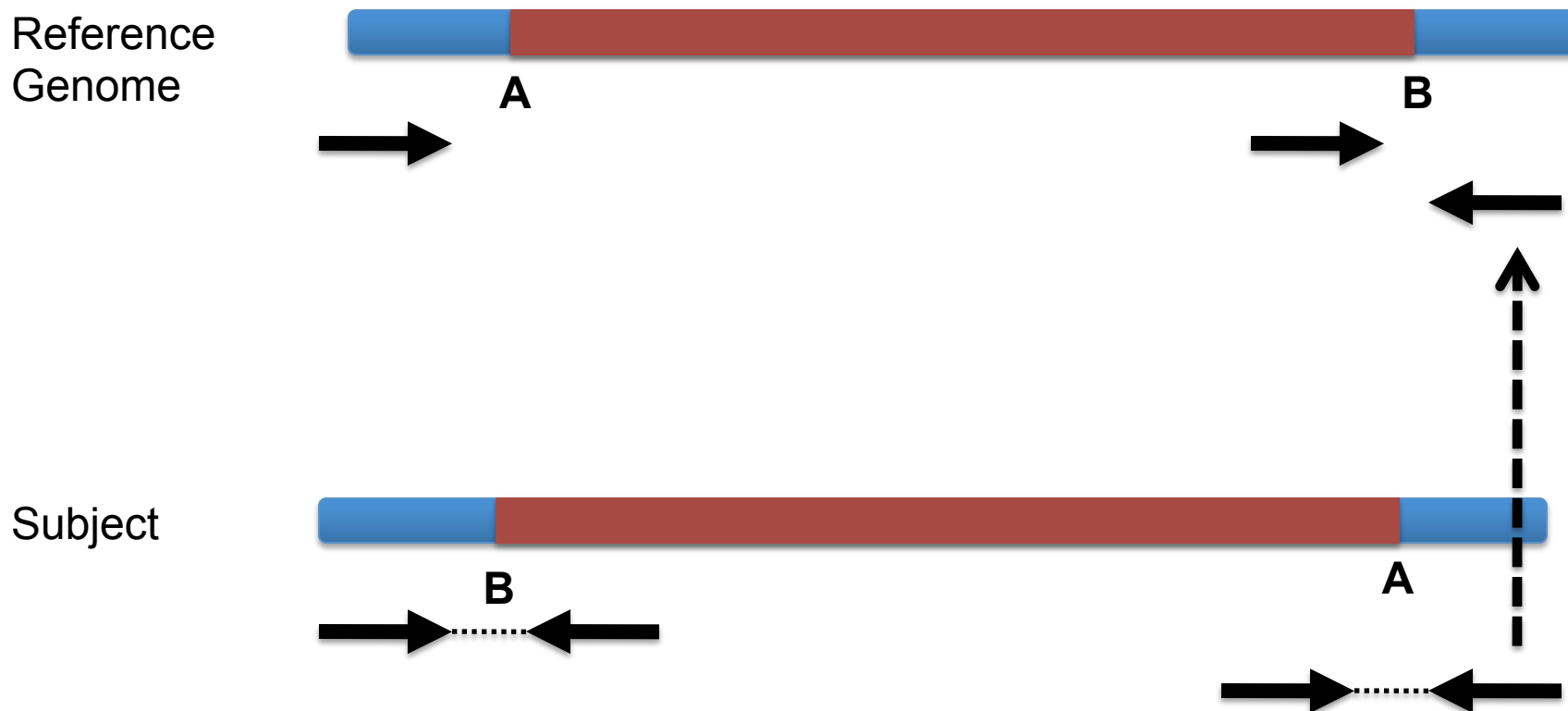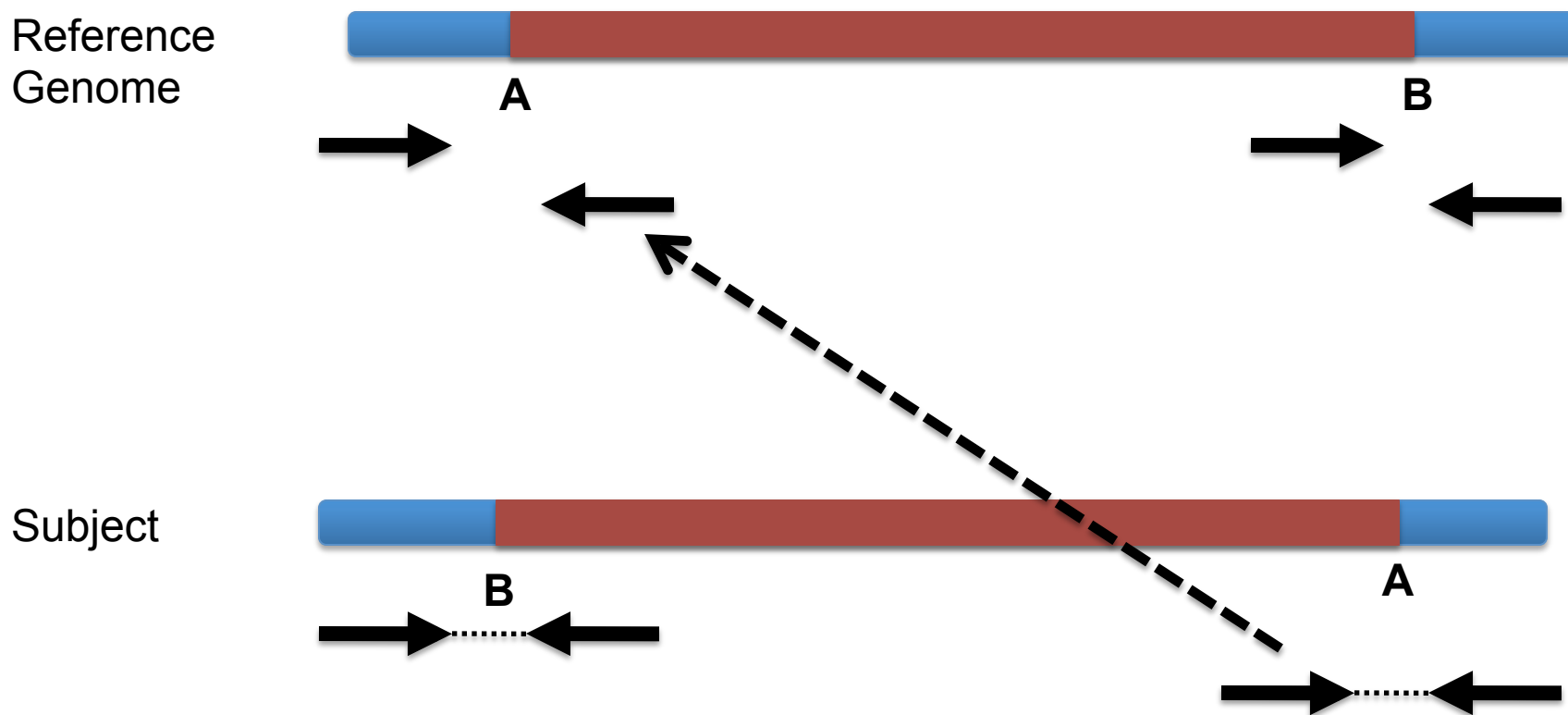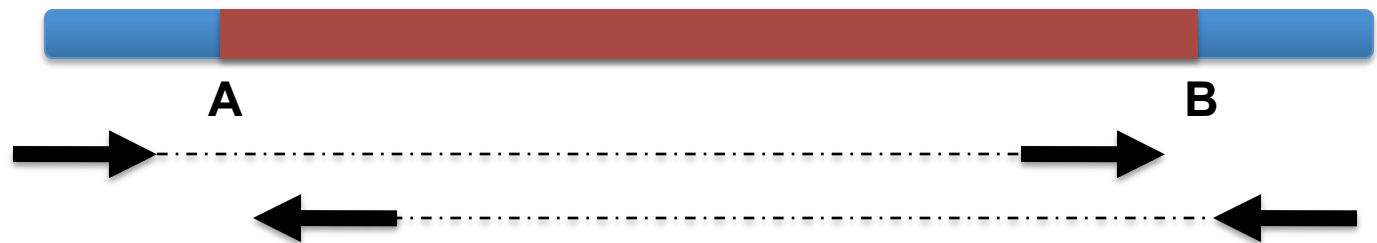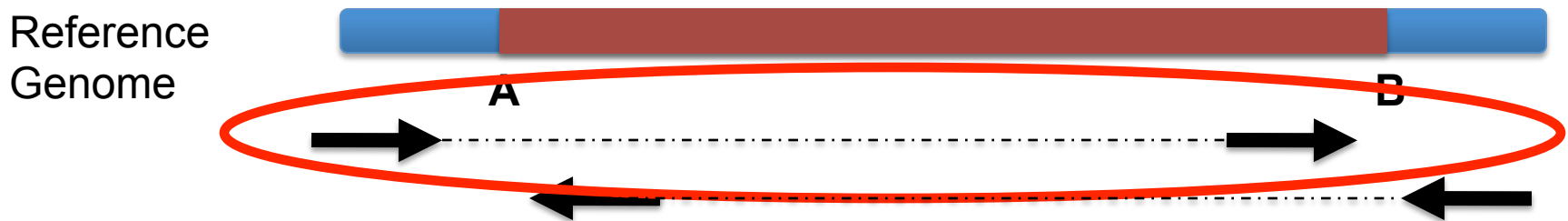B   A

Inversion

Inversion
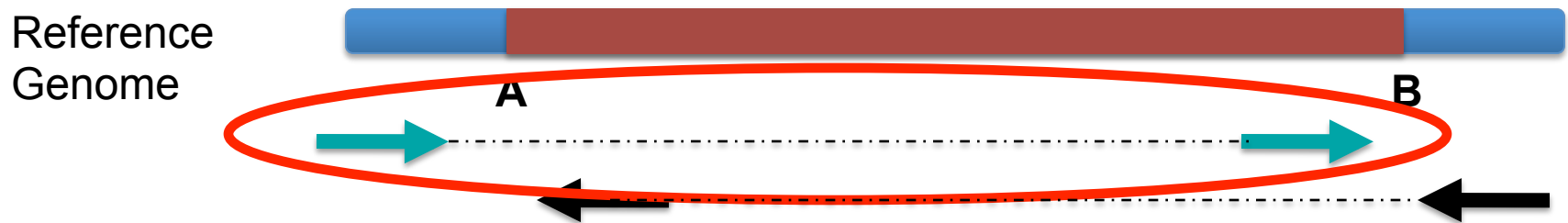
Inversion

Reference Genome

A   B

Inversion

Reference Genome

A   B
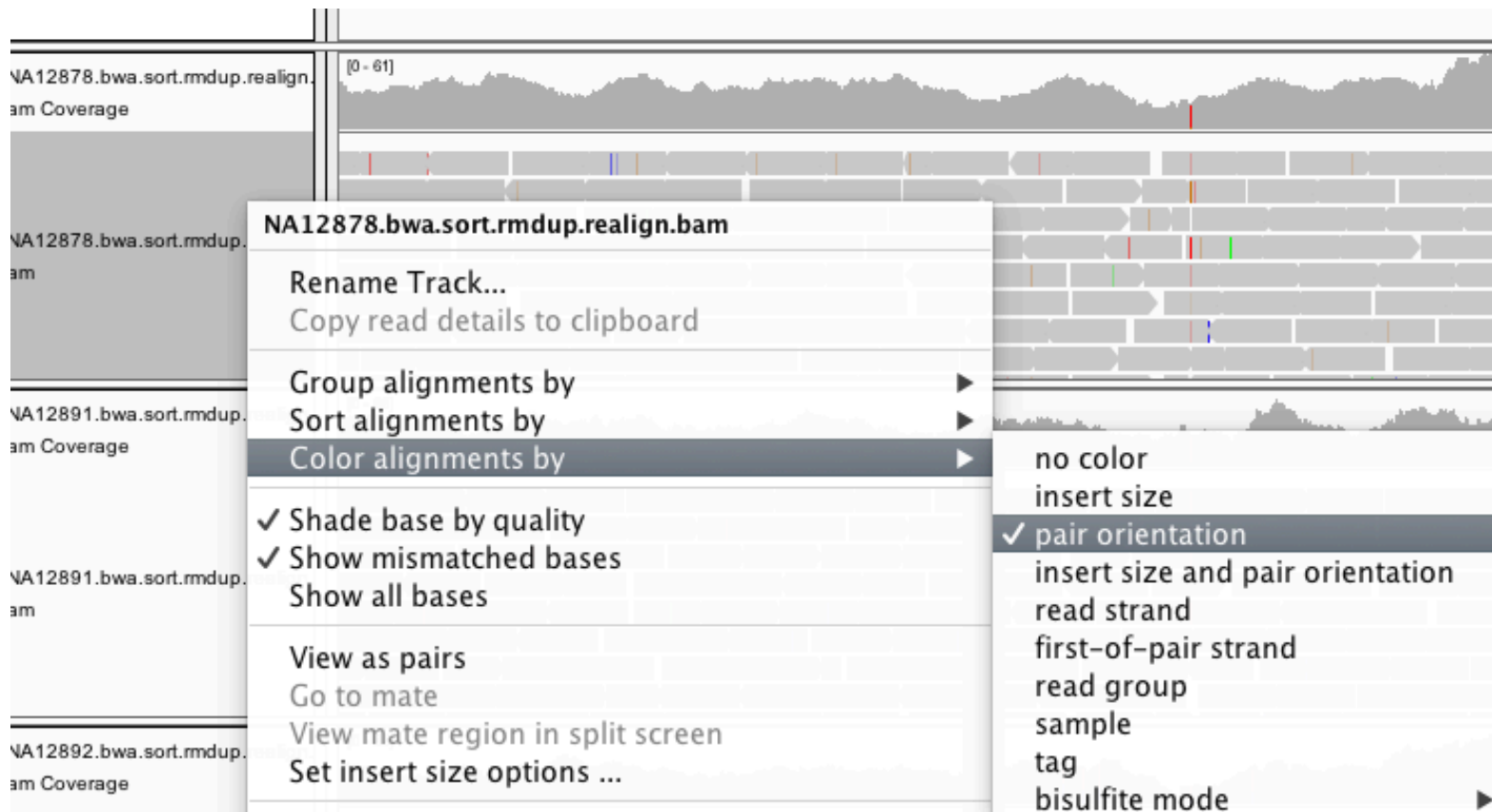
Anomaly: expected orientation of pair is inward facing (➡ ⬅)

Inversion

"Left" side pair

Inversion
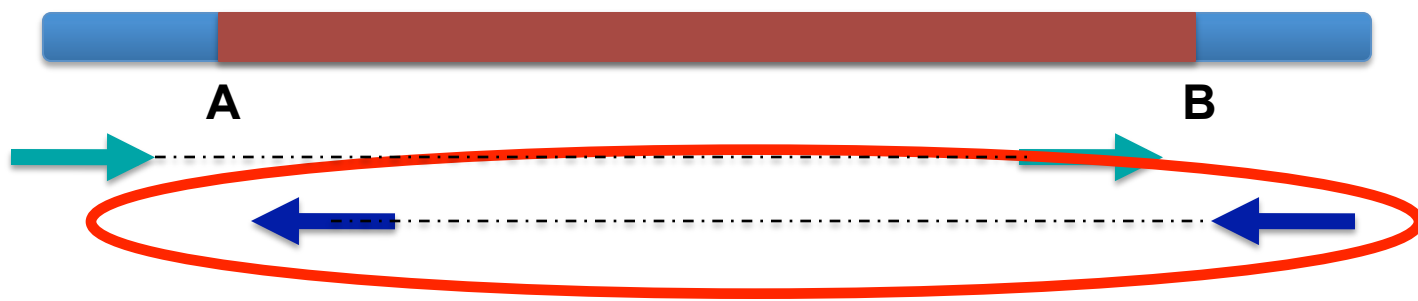
Color by pair orientation

Inversion

Note drop in coverage at breakpoints

Interpretation of read pair orientations

Category	Illumina	SOLiD

LR — Normal reads, shown with light blue bars, black arrows pointing inward for Illumina (→ ←), and SOLiD arrows labeled 1 and 2 (→ / → OR ← / ←).

LL — teal bars, both arrows pointing right for Illumina (→ →), SOLiD (1→ / 2← OR 2← / 1→).

RR — dark blue bars, both arrows pointing left for Illumina (← ←), SOLiD (1← / 2→ OR 2→ / 1←).

RL — green bars, arrows pointing outward for Illumina (← →), SOLiD (1← / 2→ OR 2→ / 1→).
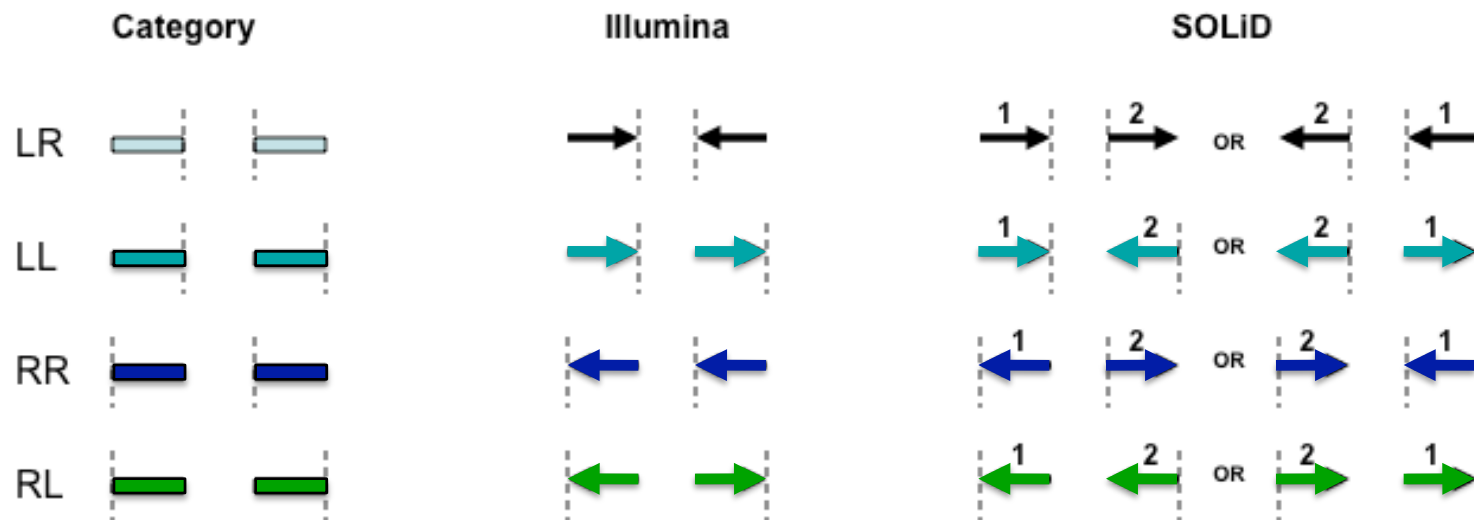
LR   Normal reads.
The reads are left and right (respectively) of the unsequenced part of the sequenced DNA fragment when aligned back to the reference genome.

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

IGV hands-on tutorial

[https://github.com/griffithlab/](https://github.com/griffithlab/rnaseq_tutorial/wiki/IGV-Tutorial)[rnaseq_tutorial/wiki/IGV-Tutorial](https://github.com/griffithlab/rnaseq_tutorial/wiki/IGV-Tutorial)

We are on a Coffee Break & Networking Session