Part 1: How to Install IGV

You have two main options: downloading the desktop application (recommended for heavy, routine use) or using the web app (great for quick checks).

Option A: The Desktop Application

1. Go to the official Broad Institute IGV download page: `software.broadinstitute.org/software/igv/download`
2. **Choose Your Operating System:** * **Windows/MacOS:** Download the standard bundle for your OS. It comes with Java pre-packaged, so you don't need to worry about installing Java separately.
    - **Linux:** Download the Linux zip file and extract it.
3. **Install and Launch:** Run the installer (Windows) or drag the app to your Applications folder (MacOS). Open the application.

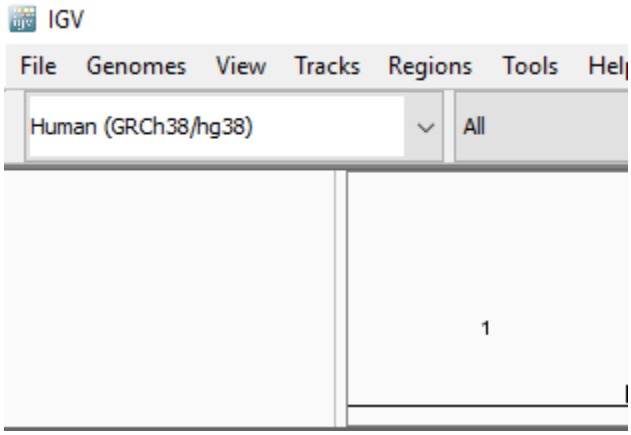
Option B: The Web Application (IGV-Web) If you can't install software on your machine, you can use IGV directly in your browser by visiting `igv.org/app`. It has a slightly streamlined interface but works great for basic viewing.

Part 2: Basic Uses for Viewing Genome Data

1. Select Your Reference Genome

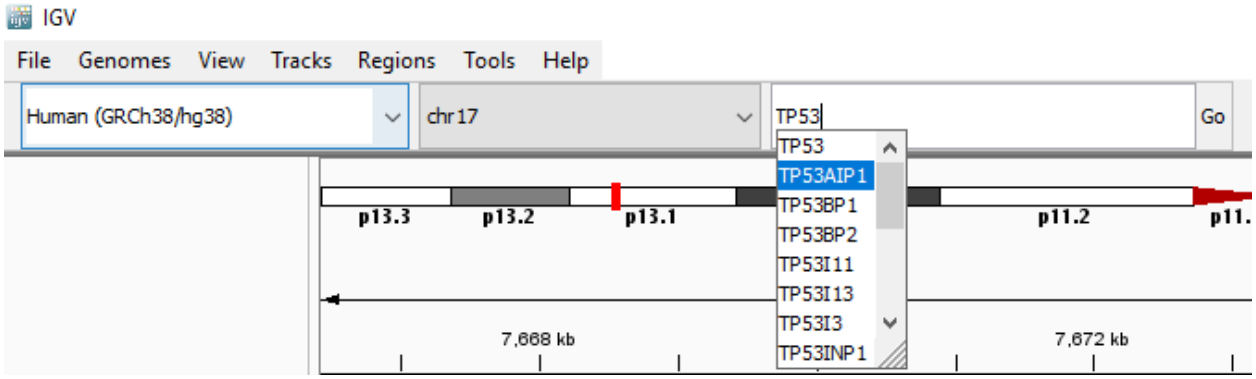
Before you load any data, IGV needs the "map" of the genome you are studying.

- Look at the top-left dropdown menu.
- By default, it usually says **Human hg19** or **Human hg38**.
- If you are studying a different organism (like a mouse or a specific bacteria), click the dropdown, select **More**, and search for your organism's assembly.

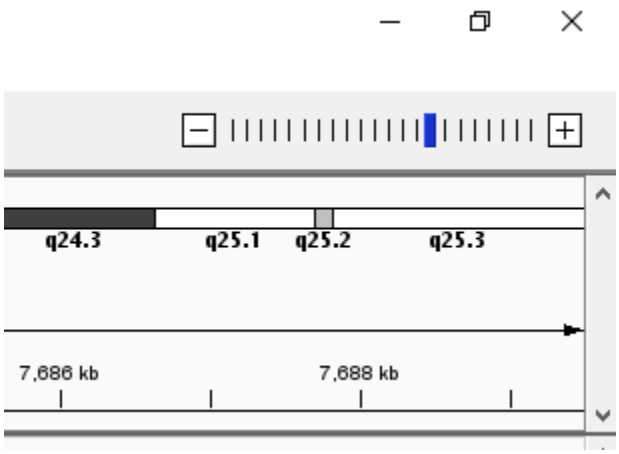


2. Navigate the Genome

- **The Search Bar:** Next to the genome dropdown is a search box. You can type in a specific gene name (e.g., *BRCA1* or *TP53*) or a specific genomic coordinate (e.g., `chr1:10,000-20,000`) and hit Enter.



- **Zooming and Panning:** Use the zoom slider in the top right corner. You can also click and drag left or right on the main viewing window to pan across the chromosome.



3. Load Your Data

- Go to **File > Load from File...**
- Select your data files. The most common formats for basic viewing are:
    - **.BED or .GTF/.GFF:** These are annotation files. They show you where known genes, exons, and regulatory regions are located.
    - **.VCF:** Variant Call Format files. These show specific mutations (SNPs or Indels) in your sample compared to the reference.
    - **BAM:** These much larger files contain the mapped reads from your experiment.

Part 3: Viewing RNA-Seq Data

RNA-Seq data allows you to see which genes are being expressed (turned into RNA) and how they are spliced.

1. Prepare Your Files

To view RNA-Seq data, you need an alignment file. This is usually a **.BAM** file.

- **Crucial Step:** IGV *requires* an index file to read a BAM file. If your file is named `sample.bam`, you must have a file named `sample.bam.bai` in the same folder. If you don't have the `.bai` file, IGV will throw an error.

2. Load the BAM File

- Go to **File > Load from File...** and select your `.BAM` file.

3. Read the Tracks

Once loaded (and assuming you are zoomed in close enough to a specific gene), you will see two main tracks appear for your sample:

- **The Coverage Track (Top bar chart):** This grey bar chart shows the "depth" of your data. The taller the grey bar, the more RNA-seq reads mapped to that specific base pair.
    - *What it means:* High coverage over an exon means that part of the gene is highly expressed.
- **The Alignment Track (Bottom pile of rectangles):** These are the actual individual sequenced reads (usually 50-150 base pairs long) mapped to the reference genome.
    - **Colors:** Grey reads are normal mappings. If you see vertical colored lines inside a read, that indicates a mismatch (a potential mutation or RNA editing site) compared to the reference genome.
    - **Blue lines connecting reads:** These represent spliced reads! An RNA-seq read might map half to one exon and half to the next, with a thin blue line connecting them over the intron space.

4. Pro-Tip: The Sashimi Plot

For RNA-Seq, the default alignment track can get messy when trying to look at alternative splicing.

- Right-click anywhere on the alignment track.
- Select **Sashimi Plot**.
- This opens a new window that draws quantifiable arcs connecting exons, showing you exactly how the RNA was spliced and how many reads support that specific splice junction.

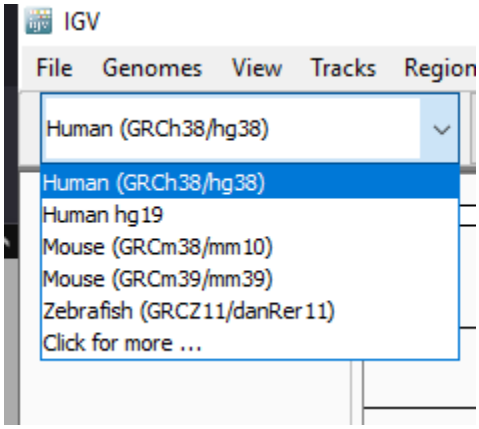
Part 4: Specifics for Human and Mouse Data

When working with human and mouse samples, the most frequent pitfall is a mismatch between the reference genome used for alignment and the reference genome selected in IGV. If these do not match exactly, your data will either look completely empty, heavily mutated, or IGV will throw an error.

1. Selecting the Correct Reference Build

Always check the upstream bioinformatics pipeline (e.g., DRAGEN, Cell Ranger) to confirm which genome build was used before loading your BAM or VCF files.

- **Human Genomes:** * Older datasets are often aligned to **hg19** (GRCh37).
    - Modern RNA-Seq and single-cell pipelines almost exclusively use **hg38** (GRCh38).
    - *How to find them:* These are usually available right in the default genome dropdown at the top-left of the screen.
- **Mouse Genomes:**
    - Common builds are **mm10** (GRCm38) and the newer **mm39** (GRCm39).
    - Mouse builds are not always in the default quick-list. You will likely need to click the genome dropdown, select **More...**, search for `Mouse` (or `mm10`/`mm39`), and check the box to download the sequence dictionary.



2. Exploring Gene Annotations

When you switch to a standard human or mouse genome in IGV, a **RefSeq Genes** track automatically loads at the bottom of the screen.

- **The Structure:** Thick blue bars are coding exons, thinner bars are untranslated regions (UTRs), and the thin lines connecting them are introns.
- **Directionality:** Look closely at the intron lines; you will see tiny arrows (`>` or `<`) indicating whether the gene is transcribed on the positive (forward) or negative (reverse) DNA strand.

- **Viewing Splice Variants:** By default, RefSeq collapses all known transcripts for a gene into one line. Right-click on the "RefSeq Genes" track label on the left side of the screen and select **Expanded**. This will unstack the gene and show you every known splicing isoform for that human or mouse gene, which is incredibly helpful when comparing against your RNA-Seq Sashimi plots.

3. Using the BLAT Tool for Unknown Sequences

If you find a strange, heavily mismatched read in your human or mouse RNA-Seq data and suspect it might be contamination or a severe structural variant:

1. Right-click the specific read in the alignment track.
2. Select **Blat read sequence**.
3. IGV will send that read's sequence to the UCSC Genome Browser database and tell you where else in the human/mouse genome that sequence belongs.

Part 5: Pro-Tips and Troubleshooting

1. The Golden Rule: Sort Before You Index

Before a BAM file can be indexed (creating that required `.bai` file), it **must be coordinate-sorted**.

- IGV will throw an error or crash.
- **The Fix:** Remember to run `samtools sort input.bam -o sorted.bam` before running `samtools index sorted.bam`.

2. Memory Management (Avoiding the "Out of Memory" Error)

Genomic datasets, especially output from robust pipelines like DRAGEN or Cell Ranger, are massive. By default, the IGV desktop app restricts how much of the computer's RAM it uses. If users load too many tracks or zoom out too far on a deeply sequenced BAM file, IGV will freeze or crash.

- **The Fix:** Go to **View > Preferences > Advanced**. Increase the **Maximum Memory** allocation to at least 4000 MB (4 GB) or 8000 MB (8 GB) if your machine can handle it. You will need to restart IGV for this to take effect.

3. Saving Sessions (The Biggest Time Saver)

If you are repeatedly looking at the same 5 BAM files and 3 VCF files—especially if you are pulling these large files over the network, reloading them one by one every time is tedious.

- **The Fix:** Once everything is loaded and colored exactly how you like it, go to **File > Save Session**.

- This creates a tiny `.xml` file. The next time you open IGV, just click **File > Open Session**, select the `.xml`, and all the tracks, file paths, and customized views will instantly reload exactly as you left them.

4. "Squish" vs. "Expand" vs. "Collapse"

When looking at RNA-Seq alignments, a deep pileup of reads will quickly take up the whole screen, hiding the annotations at the bottom.

- Right-click the track name on the left side of the screen.
- **Collapse:** Smushes all reads into a single visual line (great for just checking overall coverage).
- **Squish:** Makes the reads very thin, allowing you to see thousands of reads in a small space while still spotting vertical mismatch lines.
- **Expand:** Shows every read at full thickness (best for closely examining individual reads or short insertions/deletions).

5. Color Alignments by Read Strand

In RNASeq, it is often critical to know which strand of DNA the RNA was transcribed from (especially for overlapping genes).

- Right-click the alignment track and select **Color alignments by > Read strand**.
- Reads will turn red and blue, allowing users to instantly visually separate transcripts coming from the positive strand versus the negative strand.