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

Desktop application for the interactive visual exploration of integrated genomic datasets

Epigenomics
Microarrays
NGS alignments
RNA-Seq

http://www.broadinstitute.org/igv
65,000 registrations
Features

With IGV you can…

• Explore large genomic datasets with an intuitive, easy-to-use interface.

• Integrate multiple data types with clinical and other sample information.

• View data from multiple sources:
  - local, remote, and “cloud-based”.
IGV data sources

- View **local** files without uploading.
- View **remote** files without downloading the whole dataset.
Using IGV: The Basics
Using IGV: the basics

Hands-on exercise

- Launch IGV
- Select a reference genome
- Load data
- Navigate through the data
Launch IGV

http://www.broadinstitute.org/igv
Launch IGV

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

Agree  Cancel
Launch IGV

Integrative Genomics Viewer (Version 2.3)

Mac Users: Apple has pushed out an update that blocks all but the latest versions of Java. See this article for details. To run IGV from the web launch buttons below, you need the latest version of Java. Another option which avoids Mac security issues is to use the “zip” distribution below. After unzipping double-click the “igv.command” file to launch IGV.

Java: IGV 2.3 requires Java 6 or greater. To use the launch buttons below on MacOS Java 7 is required.

Chrome: Chrome does not launch java webstart files by default. Instead, the launch buttons below will download a “.jnil” file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run.

Windows users: To run with more than 1.2 GB you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error ‘could not create virtual machine’.

Launch IGV

Launch with 750 MB
Launch with 1.2 GB
Launch with 2 GB
Launch with 10 GB

Nightly Build Latest development build.
Archived Versions

Igvtools
Utilities for preprocessing data files.
- igvtools_2.3.20.zip

Download
A downloadable version that does not require launching from the web. For Windows, Mac OS X, and Linux.
- IGV_2.3.20.zip

Source Code
Source distribution archive:
- v2.3.20.zip
Source code repository is hosted at github:
- https://github.com/broadinstitute/IGV/
Launch IGV

Integrative Genomics Viewer (Version 2.3)

Mac Users: Apple has pushed out an update that blocks all but the latest versions of Java. See this article for details. To run IGV from the web launch buttons below, you need the latest version of Java. Another option which avoids Mac security issues is to use the "zip" distribution below. After unzipping double-click the "igv.command" file to launch IGV.

Java: IGV 2.3 requires Java 6 or greater. To use the launch buttons below on MacOS Java 7 is required.

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

Windows users: To run with more than 1.2 GB you must install 64-bit Java. This is often not installed by default even with the latest Windows 7 machines with many GB of memory. In general trying to launch with more memory than your OS/Java combination supports will result in the obscure error "could not create virtual machine."

Launch with 750 MB
Launch with 1.2 GB
Maximum usable memory for Windows OS with 32-bit Java.

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

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

Nightly Build
Latest development build.

Archived Versions

Igtools
Utilities for preprocessing data files.
- igtools_2.3.20.zip

Download
A downloadable version that does not require launching from the web. For Windows, Mac OS X, and Linux.
- IGV_2.3.20.zip

Source Code
Source distribution archive:
- v2.3.20.zip
Source code repository is hosted at github:
- https://github.com/broadinstitute/IGV/
Select the reference genome

Select genome from the drop-down menu
Select the reference genome

Today, we will use both Human hg18 and hg19

If Human hg19 is not in the menu, then click on More...
Select the reference genome

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

- Select Human hg19 from the list.

- Click the 'Genomes' button to see the available genomes.

- The selected genomes will be added to the genome dropdown list.
Select the reference genome

Select Human hg18
Load data

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

Open the Tutorials menu

Select UI Basics (Encode)
Screen layout
Screen layout

Click the **Home** button for whole-genome view
Screen layout

IGV

Human hg18

Tracks

GM12878 H3K27ac
GM12878 H3K27me3
GM12878 H3K36me3
GM12878 H3K4me1

RefSeq genes

6 tracks
Screen layout

- **menus**
- **toolbar**
- **genome ruler**
- **data panel**
- **tracks**
- **genome features**
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
- broadPeak
- CBS
- CN
- Cufflinks Files
- Custom File Formats
- Cytoband
- FASTA
- GCT
- genePred
- GFF
- GISTIC
- Goby
- GWAS
- IGV
- LOH
- MAF (Multiple Alignment Format)
- MAF (Mutation Annotation Format)
- Merged BAM File
- MUT
- narrowPeak
- PSL
- RES
- SAM
- Sample Information
- SEG
- SNP
- TAB
- TDF
- Track Line
- Type Line
- VCF
- WIG

- For current list see: [www.broadinstitute.org/igv/FileFormats](http://www.broadinstitute.org/igv/FileFormats)
Navigate

Click on the 1 in the genome ruler to view chromosome 1
Navigate
Navigate

Click and drag from 40 mb to 60 mb on the genomic ruler
Navigate
Navigate

Double-click to zoom in for closer view of peak
Navigate
Navigate

Click anywhere in the data panel and drag tracks left and right.
Navigate

Type gene name or other RefSeq annotation into the Search Box and click Go
Navigate

Type gene name or other RefSeq annotation into the **Search Box** and click **Go**.
Navigate
Navigate

Click on the last tick on the "railroad track" to zoom in to maximum resolution.
Navigate

Maximum zoom. We’ve moved from whole genome to base pair resolution.

Reference sequence
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

5’ UTR  |  Intron  |  Exons  |  3’ UTR

Zoomed in views

Zoomed out views
Annotation display mode

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

2. Expand the track using the popup menu
Annotation display mode

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

- **Search box**
  Enter multiple loci or features in the search box

- **Regions > Gene Lists...**
  Select from a number of pre-defined gene lists, or
  Create your own persistent list
Viewing multiple regions

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

• double-click on a region label — or —
• right-click and select “Switch to standard view”
Viewing NGS Data
Viewing alignments

Whole chromosome view
Viewing alignments

Zoom in to view alignments
Viewing alignments

Coverage track now has more detail
Viewing alignments

Zoom in to see more detail

Bases that do not match the reference sequence are highlighted by color.
Viewing alignments
Zoom in to see more detail
Viewing alignments

Zoom in to see more detail

Low-quality base calls are faint, semi-transparent.
Viewing alignments

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

- How far do you need to zoom in to see the alignments?
  - 30 kb
  - or set a different threshold in preferences

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

Downsampling:
- limits displayed read depth
- uses less memory
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

Open the Tutorials menu

Select File > Load from Server...

Select SNP Validation
Viewing SNPs

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

If necessary, click and drag the window divider for a larger data panel.
Viewing SNPs
Viewing SNPs

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

Click & drag to position mismatched bases between the center guidelines
Viewing SNPs

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

On Mac: Right-click = ⌘-click
Viewing SNPs
Viewing SNPs

Mouse over red & blue bar in coverage track. Note allele counts and frequencies.
Viewing SNPs

Type “snp2” in the Search Box and click Go

Note:
Large % of low quality base calls, and scatter of “C” mismatches
Viewing SNPs

Click & drag to position locus with 5 blue C’s between center guidelines
Viewing SNPs

Right-click on alignments and select Shade base by quality
Viewing SNPs
Viewing SNPs

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

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

DNA or cDNA

Fragment

Read from each end

insert size
Paired-end sequencing

DNA or cDNA

Fragment

Read from each end

Align to Reference

insert size

inferred insert size
Interpreting Insert Size
Interpreting inferred insert size

The “inferred insert size” can be used to detect structural variants, including:

• Deletions
• Insertions
• Inter-chromosomal rearrangements: (Undefined insert size)
Deletion

What is the effect of a deletion on inferred insert size?
Deletion

Reference Genome
Deletion

Reference Genome

Subject
Deletion

Reference Genome

Subject
Deletion

Reference Genome

Subject
Deletion

Reference Genome

Subject
Deletion

Reference Genome

Subject
Deletion

Reference Genome

Subject
Deletion

Reference Genome

inferred insert size

Subject
Deletion

Reference Genome

inferred insert size

Subject

expected insert size
Deletion

Inferred insert size is > expected value

Reference Genome

inferred insert size

Subject

expected insert size
Deletion

Pairs with larger than expected insert size are colored red.
Deletion

Note drop in coverage
Insert size color scheme

- Smaller than expected insert size:  
  ![Blue Bar]

- Larger than expected insert size:  
  ![Red Bar]

- Pairs on different chromosomes

*Each end colored by chromosome of its mate*
Rearrangement

CHRs:
- TUMOR
- NORMAL

Comparison of CHRs 1 and 6 between TUMOR and NORMAL samples.
Rearrangement

Color indicates mate is on chromosome 6
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

A

B
Inversion

Reference Genome

Subject

A
B

B
A
Inversion
Inversion

Reference Genome

Subject
Inversion

Reference Genome

Subject
Inversion

Reference Genome

Subject
Inversion

Reference Genome

Subject

A

B

B

A
Inversion

Reference Genome

Subject
Inversion

Reference Genome

A

B
Inversion

Anomaly –
Expected pair orientation is inward facing (→ ←)
Inversion

“Left” side pair
Inversion

“Right” side pair
Color by pair orientation
Inversion
Inversion

Note drop in coverage at breakpoints
### Interpretation of read pair orientations

<table>
<thead>
<tr>
<th>Category</th>
<th>Illumina</th>
<th>SOLiD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td><img src="image" alt="LR" /></td>
<td><img src="image" alt="LR" /></td>
</tr>
<tr>
<td>LL</td>
<td><img src="image" alt="LL" /></td>
<td><img src="image" alt="LL" /></td>
</tr>
<tr>
<td>RR</td>
<td><img src="image" alt="RR" /></td>
<td><img src="image" alt="RR" /></td>
</tr>
<tr>
<td>RL</td>
<td><img src="image" alt="RL" /></td>
<td><img src="image" alt="RL" /></td>
</tr>
</tbody>
</table>

**LR**  
Normal reads. The reads are left and right (respectively) of the unsequenced part of the sequenced DNA fragment when aligned back to the reference genome.

**LL,RR**  
Implies inversion in sequenced DNA with respect to reference.

**RL**  
Implies duplication or translocation with respect to reference.

These categories only apply to reads where both mates map to the same chromosome.

*Figure courtesy of Bob Handsaker*
RNA-Seq

Hands-on exercise

- Examine tissue-specific alternative splicing.

- Data: Illumina BodyMap 2.0

  http://www.illumina.com/science/data_library.ilmn
Before we start: Select **File > New Session** to clear IGV window
RNA-Seq Setup

• Step 1: Tune settings for RNA.
RNA-seq alignments

Select **View > Preferences...**
Click **Alignments** tab
RNA-seq alignments

Select Show junction track
RNA-seq alignments

Click OK to save changes
RNA-seq alignments

Select Human hg19 from genome menu
RNA-seq alignments

Select:
File > Load from Server...
RNA-seq alignments

- Open the **Tutorials** menu
- Select **RNA-Seq (Body Map)**
RNA-seq alignments

Zoom in to see alignments.
RNA-seq alignments

Type SLC25A3
RNA-seq alignments

![RNA-seq alignments](image)

Click Go
RNA-seq alignments
RNA-seq alignments

If reads are still blue & red from the settings for the last exercise, then right-click and select **Color alignments by > no color**
RNA-seq alignments
RNA-seq alignments
RNA-seq alignments

Junction Coverage
RNA-seq alignments
RNA-seq alignments

Right-click over RefSeq Genes track
RNA-seq alignments

Select **Squished**
RNA-seq alignments

Isoforms now displayed
RNA-seq alignments

Click & drag on ruler to zoom in on first 2 exons
RNA-seq alignments
RNA-seq alignments

Evidence of alternative splicing
Sashimi plot

Viewing RNA splicing with Sashimi Plots

{


}
RNA-seq alignments

Right-click over alignments
RNA-seq alignments

Select Sashimi Plot
RNA-seq alignments

Select RefSeq Genes
RNA-seq alignments

Select both Heart and Liver
RNA-seq alignments
RNA-seq alignments

A little busy. Let’s filter out low-count events
RNA-seq alignments

Right-click anywhere over plot
RNA-seq alignments
RNA-seq alignments

Enter 20
igvtools
## igvtools

A set of utilities for preparing files for efficient display.

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

• Can be launched from the IGV user interface
  *File > Run igvtools…*

• Or run from the command line
The **toTDF** utility converts large ASCII data files into tiled data format (.tdf) files.

TDF files have the following advantages:

- Data is indexed for efficient retrieval.
- Data is preprocessed for zoomed out views.
- TDF files are web friendly – large data files can be shared over the web. Only small slices of the file are actually transferred as needed.
The **count** command is used to transform alignment files to read density TDF files, e.g. for ChIP-Seq, RNA-Seq, and similar alignment counting experiments.

**Alignments**

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

**Read Density**

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

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

Example:

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

- Uses combination of memory and disk to handle large files.

  - `m = maximum # of lines to hold in memory`. When this number is exceeded a temporary file is created.

  - `t = directory used to create temporary files during sorting.`
igvtools index

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

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

Example:  igvtools index myFeatures.bed

The index file must remain in the same directory as the input file
Computing coverage: igvtools

Hands-on exercise

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

Data source

Illumina BodyMap
Download data files required for this exercise from: 

Files included in the zip:
heart.bodyMap.bam
heart.bodyMap.bam.bai
sacCer3.fa (used in next exercise)
Computing coverage: igvtools

Select Tools > Run igvtools...
Computing coverage: igvtools

Select **Count** command
Computing coverage: igvtools

Select input file `heart.bodyMap.bam`
Output filename will be filled in automatically.
Computing coverage: igvtools

Click Run
Computing coverage: igvtools

Wait for “Done” message (should be fast).
Computing coverage: igvtools

Click Close
Computing coverage: igvtools

Select
File > Load from File...
Computing coverage: igvtools

Select heartBodyMap.bam.tdf (output from igvtools)
Computing coverage: igvtools

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

Enter SLC25A3 in the search box and click Go
Computing coverage: igvtools
More about reference genomes

IGV doesn’t host the genome you need?

Use any genome you want, if you have the sequence in FASTA format.

Optionally, package genome annotations with the sequence.
Loading a genome

Hands-on exercise
Loading a genome

Select Genomes > Load Genome from File...
Loading a genome

Select *sacCer3.fa* (FASTA file for *S. cerevisiae*)
Loading a genome
Loading a genome

Click on a chromosome name or use the pulldown menu to jump to the chromosome view.
Loading a genome
Loading a genome

Click the rightmost “tick” to jump to base pair resolution
Loading a genome

Confirm that you see sequence
Acknowledgments

IGV Team
Jim Robinson, Jacob Silterra, Helga Thorvaldsdóttir, Jill Mesirov (PI)

Funding
IGV development has been made possible with funding from:
• National Cancer Institute (NCI) http://cancer.gov/
• Starr Cancer Consortium http://www.starrcancer.org/
• National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health http://www.nigms.nih.gov/
• IGV participates in GenomeSpace http://genomespace.org/, which is funded by the the National Human Genome Research Institute (NHGRI) http://www.genome.gov/
For further information and help:

http://www.broadinstitute.org/igv
http://groups.google.com/group/igv-help

Cite:

Robinson et al.  
*Integrative Genomics Viewer.*  

Thorvaldsdóttir, Robinson, and Mesirov.  
*Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.*  
Briefings in Bioinformatics (2012).