Integrative Genomics Viewer

The Rockefeller University

Bioinformatics Resource Centre

Overview

- Introduction to IGV.
  - What is IGV.
  - How to run IGV.
- Navigating IGV.
  - The IGV user interface.
  - Moving around genomes.
- Loading and visualising data.
  - Genome information and annotation.
  - User supplied data.
  - Sample information.
  - External data.
- Displaying genomics data
  - Basic visualisation.
  - Data dependent visualisation.

What is IGV?

- Created by the Broad institute.
- Genome browser.
  - Visualises genomic data (expression, ChIP, resequencing, multiple alignment, shRNA)
  - Handles most common genomic data types.
- Java Desktop application
  - No dependence on server
  - Loads data locally or from URL, consumes memory and CPU.

How to run IGV?

- Requires Java
- IGV available from Broad

http://www.broadinstitute.org/software/igv/download

How to run IGV? (Binary version)

3. Binary Distribution

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 launch igv on Mac or Linux platforms use the shell script "igv.sh". On Windows use "igv.bat".

Download Binary Distribution

igvtools

Utilities for preprocessing data files.

- igvtools_2.3.40.zip

Other IGV Versions

Development Snapshot Build *Latest development snapshot; built at least nightly.*

Archived Versions

- Download to computer.
- Runs locally.
- Archived versions available

How to run IGV? (Webstart)

2. Java Web Start

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

***Mac Users**: The Java Web Start 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 Launch with 750 MB	Launch Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	Launch Launch with 2 GB Maximum usable memory for 32-bit MacOS.	Launch Launch with 10 GB For large memory machines with 64-bit Java.

- Runs from webstart.
- Always runs latest version of IGV.

IGV GUI

IGV 2.3
File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
100%
Tue 09:56
Thomas Carroll
IGV
Human hg19
All
Go
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
RefSeq Genes
2 tracks loaded
194M of 415M

IGV GUI

- Sample information panel <1>
- Genome Navigation panel <2>
- Data panel <3>
- Attribute panel <4>

Menu Bar

- **File -** Load data/sample information.
- **Genome -** Load and manage genomes.
- **View -** Display preferences.
- **Tracks -** Group/sort/filter data tracks.
- **Regions -** Create region/gene lists.
- **Tools -** Access to Integrated tools (IGVtools/Bedtools).
- **GenomeSpace -** Export/import from Genomespace

Moving around genomes

- Cytoband selection and zooming
- Scrolling
- Selection of region of interest

Whole genome view

Zooming

IGV
Mouse mm9
All
Go
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Refseq genes
2 tracks loaded
chr1:40,216,379
219M of 396M

IGV
Mouse mm9
chr1
chr1
Go
qA2 qA4 qB qC1.1 qC2 qC4 qD qE1.2 qE2.3 qE4 qF qG2 qH2.2 qH4 qH6
162 mb
20 mb 40 mb 60 mb 80 mb 100 mb 120 mb 140 mb 160 mb
Refseq genes
8L16Rik Gm597 Aff3 Tmeff2 Abi2 Ikzf2 Pax3 Sp100 Slco4c1 Bcl2 Tsn Mgat5 Ptprc Pla2g4a Tnr Uck2
2 tracks loaded
chr1:26,791,616
216M of 396M

Scrolling

IGV
Mouse mm9
chr1
chr1:68,125,345-197,077,344
Go
128 mb
RefSeq genes
2 tracks loaded

Jump to Region

“Bookmarking” regions of interest

- Regions may be added to “Regions of interest”
- These act as bookmarks for areas of particular interest
  - Bookmarks can be added by -
    - Bookmarking visible window
    - Selecting region within window

Bookmarking - Visible window

“Bookmarked” Region Of Interest

Bookmarking

- Bookmarks may be created from selecting "region of interest" button and edge of region

Sequences from regions of interest

- A useful feature of "regions of interest" is to retrieve of BLAT sequences.

Sequences from regions of interest

BLAT

chr	start	end	strand	score	match	mis-match	rep. match	N's	Q gap co...	Q gap ba...	T gap co...	T gap ba...
chr6	125115...	125116...	+	1000	1018	0	0	0	0	0	0	0
chr2	20341885	20341948	-	30	34	0	0	0	1	4	2	29
chr2	126905...	126905...	+	27	32	0	0	0	2	3	2	48
chr5	129894...	129894...	-	25	27	0	0	0	0	0	1	1
chr11	72432755	72432781	-	25	26	0	0	0	0	0	0	0
chr2	25195684	25195710	+	21	24	0	0	0	1	1	1	2
chr1	14007130	14007155	+	21	23	0	0	0	0	0	1	2
chr2	33115816	33115836	-	19	20	0	0	0	0	0	0	0

Viewing sequences

- At a predefined resolution, sequence information becomes available.

Viewing sequences

- Strand of sequence can be altered.
- Differing translation tables can be selected

Loading data in IGV

Loading Genome Information

- Most genomes can be selected from dropdown.

Loading Genome Information

- Genomes not included may be downloaded from repository

Loading Genome annotation

- For supported genomes, gene positions are automatically included in “feature” panel.
- Additional gene positions can be loaded into IGV in gff format.

Loading Sample data

- Most common formats can be loaded into IGV through file menu
- Acceptable data formats include:-
- BED (.bed)
- BAM and index (.bam with .bai/.bam.bai)
- BigWig (.bw)
- BedGraph/Wig (.bedGraph, .wig)
  - And many more…
  - link to IGV formats

Loading Sample Metadata

- IGV allows the inclusion of information on samples.
- Sample information is then included in sample information panel.

Example Sample information file

- http://www.broadinstitute.org/igvdata/exampleFiles/exampleSampleInfo.txt

TRACK_ID	Data_Type	PARTICIPANT_ID	SAMPLE_ID	GENDER	T/N	Tumor_type	Treated	Primary/Secondary	Hypermutated
EX-01-001	Expression	P-01-P001	P-01-S001	M	Tumor	GBM	Y	Primary	Y
CN-01-002	CopyNumber	P-01-P001	P-01-S001	M	Tumor	GBM	Y	Primary	Y
MU-01-003	Mutation	P-01-P001	P-01-S002	M	Tumor	GBM	Y	Primary	Y
EX-01-004	Expression	P-01-P002	P-01-S003	M	Normal	GBM	Y	Secondary	Y
CN-01-005	CopyNumber	P-01-P002	P-01-S004	M	Tumor	GBM	Y	Secondary	N
EX-01-006	Expression	P-01-P002	P-01-S004	M	Tumor	GBM	Y	Secondary	N
ME-01-007	Methylation	P-01-P002	P-01-S004	M	Tumor	GBM	Y	Secondary	N
EX-01-008	Expression	P-01-P003	P-01-S006	F	Tumor	GBM	N	Primary	Y
EX-01-009	Expression	P-01-P004	P-01-S009	F	Tumor	GBM	N	Primary	Y
EX-01-0010	Expression	P-01-P005	P-01-S0011	M	Control				

Tracktables package.

- R/Bioconductor packages
  - Sanjay and Tom.
- Creates HTML reports with sample information for use with IGV.

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Search:

Home | Install | Help | Developers | About

Home » Bioconductor 3.1 » Software Packages » tracktables

tracktables

Build IGV tracks and HTML reports

Bioconductor version: Release (3.1)

Methods to create complex IGV genome browser sessions and dynamic IGV reports in HTML pages.

Author: Tom Carroll, Sanjay Khadayate, Anne Pajon, Ziwei Liang

Maintainer: Tom Carroll <tc.infomatics at gmail.com>

Citation (from within R, enter `citation("tracktables")`):

Carroll T, Khadayate S, Pajon A and Liang Z (2014). *tracktables: Build IGV tracks and HTML reports*. R package version 1.2.0.

Installation

To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("tracktables")
```

Documentation

To view documentation for the version of this package installed in your system, start R and enter:

Workflows »

Common Bioconductor workflows include:

- Oligonucleotide Arrays
- High-throughput Sequencing
- Counting Reads for Differential Expression (parathyroideSE vignette)
- Annotation
- Annotating Variants
- Annotating Ranges
- Flow Cytometry and other assays
- Candidate Binding Sites for Known Transcription Factors
- Cloud-enabled cis-eQTL search and annotation
- RNA-Seq workflow: gene-level exploratory analysis and differential expression
- Changing genomic coordinate systems with rtracklayer::liftOver
- Mass spectrometry and proteomics data analysis

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the posting guide before posting!

Using sample information.

- Sample information can include discrete and continuous.
    - Can be used to “sort” and “filter” tracks.
    - Can split tracks across panels by “group”

Grouped by unique ID

Grouped related files by name

Loading external data and annotation

Load data from a URL.

- As with UCSC, IGV supports data hosted on external servers.
- Data accessible from a URL such as HTTP and FTP can be loaded using the “Load from URL”.

IGV 2.3
File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
IGV – Session: /Users/tcarroll/Documents/teste/tracktables/RNAPol2igv.xml
Mouse mm9
Go
Load from File...
Load from URL...
Load from Server...
Load from DAS...
Load from ENCODE...
New Session...
Open Session...
Save Session...
Save Image ...
Exit
EBF

IGV – Session: /Users/tcarroll/Documents/teste/tracktables/RNAPol2igv.xml
Mouse mm9
All
Go
EBF
H3K4me3
H3K9ac
RNAPol2
Enter URL (http or ftp)
https://github.com/ThomasCarroll/tracktables-Data/raw/master/H3K4me3_Example.bw
Cancel
OK

Loading external data and annotation

Load data from a server (IGV/Encode servers).

- Unlike UCSC, IGV comes with few external tracks.

- External tracks (relevant to the genome) can be loaded from the IGV server or Encode-IGV server.

IGV 2.3
File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
IGV
human hg19
Go
Load from File...
Load from URL...
Load from Server...
Load from DAS...
Load from ENCODE...
New Session...
Open Session...
Save Session...
Save Image ...
Exit
Load tracks or sample information from a server

IGV
human hg19
All
Go
Available Datasets
Available Datasets
Annotations
Genes
Phenotype and Disease Associations
Sequence and Regulation
Variation and Repeats
Comparative Genomics
The Cancer Genome Atlas
TCGA Broad CDAC
ENCODE
Txn Factor ChIP
DNase I Clusters
Regulatory Transcriptome
UW DNase
UW CTCF Binding
Broad Histone
Chromatin Regulators
Body Map 2.0 Illumina HiSeq
1000 Genomes
Tutorials
U Rochester (Pennacchio)
RNA-Seq (Body Map)
Cancel
OK
Refseq Genes
Loading...

Grabbing encode data directly

- Encode data can be downloaded from UCSC.
- http://genome.ucsc.edu/ENCODE/
- This however does not come with sample information provided through IGV interface

Viewing data

Viewing data

- IGV associates common file formats with default display methods.
- Most of the time IGV will make a sensible choice how we wish to display data.

Accepted formats and default display.

Information on accepted file formats and default display can be found at -

http://www.broadinstitute.org/software/igv/RecommendedFileFormats

Recommended File Formats

The following table lists recommended file formats for data types. **Tip**: Using igvtools to convert a source data file to a binary TDF file can reduce loading time and improve performance.

Source Data	Recommended File Formats
ChIP-Seq, RNA-Seq	TDF format. Use the igvtools package (count command) to generate a binary read count density file in TDF format. Load the resulting TDF file into IGV.
Copy number	CN format, SNP format **Log values**: If copy number data contains negative values, IGV assumes they are scaled log values and displays them without modification. If copy number data contains all positive values, IGV assumes they are unscaled. It centers the unscaled values around 2 (1 for allele specific files), log transforms them (logValue = log2(copyNumber / 2)), and displays the log values.
Gene expression data	GCT format, RES format

Bed/bigBed

- Basic
- Tab-delimited
- Chrom,Start,End
- Bed6
- bigBed (recommended)

Wig, BedGraph and BigWig

- Wig/bedGraph require high memory load
- Recommended format is bigWig

BAM alignment files

- BAM files contain alignment information.
- Require an accompanying .bai index file for display

Finer control of display

- IGV allows for customization of track display.
    - Menu bar -> View -> Preferences
        - Select track (right click)

Display preferences

General

General | Tracks | Mutations | Charts | Alignments | Probes | Proxy | Advanced | IonTorrent

☐ Distinguish missing data

Distinguish regions with zero values from regions with no data on plots (e.g. bar charts). Regions with no data are indicated with a gray background.

☐ Display all tracks in a single panel

☑ Show attribute panel

☑ Show default track attributes

Display default track attributes (NAME, DATA_TYPE, and DATA_FILE) in the attribute panel.

☐ Show region boundaries

☑ Zoom to features

This option controls the behavior of feature searchs. If true, the zoom level is changed as required to size the view to the feature size. If false the zoom level is unchanged.

Sequence resolution threshold (bp/pixel): 2

Resolution in base-pairs per pixel at which sequence track becomes visible.

Feature flanking region (bp or %): 2000

Added before and after feature locus when zooming to a feature. Also used when defining panel extents in gene/loci list views. ***A negative number is interpreted as a percentage.***

Background color click to change): Reset to default

Default font: Arial 10 Change...

OK Cancel

Display preferences

Tracks

Display preferences

Alignments

General | Tracks | Mutations | Charts | 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

☐ Filter supplementary alignments ☐ Flag unmapped pairs

☑ 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

Track display options

BAM

- Read Packing, grouping, sorting, colouring options.

Track display options

Graph/interval files

- Track colour/appearance

Track display options

Graph/interval files

- Graph type

Track display options

- Data Scaling.

- Autoscaling adjusts to track's visible signal maximum

Some cool features

IGV can display splicing information!

...and do Sashimi plots

(http://www.broadinstitute.org/igv/Sashimi)

IGV can overlay tracks.

Track overlay example

IGV can be used to combine tracks

IGV can be used to combine tracks

IGV can be used to combine tracks

IGVtools

- IGVTools can be used to post-process genomics data.
- Includes indexing, sorting and genome graph creation.

Where to get help?

- http://www.broadinstitute.org/igv/UserGuide
- https://groups.google.com/forum/#!forum/igv-help