Integrated Genome Browser
User's Guide

10/11/2004

Contents

Getting started

Introduction

IGB genome browser is a powerful tool to help genome researchers:

Visualize and compare multiple genomic annotation sets from a variety of public and private sources

Target areas of interest to explore further via other tools and/or experimentation

Compare experimental results to computational results

Provide guidance for fine-tuning or modification of experiments

Overview of IGB functionality

Generally, use IGB to:

Select and load genomic annotations to analyze from a variety of sources including your own experimental results

Drill down to an area of interest

Compare annotation variations in different data sets

Use a variety of tools in IGB to seek and explore interesting possibilities

View textual information about items of interest. Hyperlink out to external information sources.

Create curations of notable annotations or hypotheses

Unique features and benefits of IGB

IGB combines in one viewer your own experimental or computational results, common reference information, and access to public and private data banks.

Other features unique to IGB include:

Slicing: a tool that facilitates examination and manipulation of alternate splicings, intron and exon boundaries,

Create, edit, and save curations

Merge graphs of scores at specified points along the genome

IGB is optimized to handle large amounts of genomic annotations with speed and efficiency

Documentation conventions

Abbreviations of commands

Onscreen text (i.e., menus, button labels, folder names, etc.) is indicated by **Arial** font.

An item followed by a “greater than” symbol followed by another item means “click the first item; the second item will become visible; then click the second item.”

Examples:

File menu > **Open File** means “click on the **File** menu and choose **Open File**.”

QuickLoad tab> **Load Sequence Residues** means “click the **QuickLoad** tab and then click the **Load Sequence Residues** button”

Installation and setup

System requirements

Hardware/ OS

IGB has been tested on Windows 2000 and Linux but should work on any Java environment.

The minimum required RAM is 32 MB, but for work involving many annotations and graphs 256 MB of dedicated RAM is recommended.

Note for Macintosh users: Operations that require right-clicking can be performed on Mac by holding down the control key while clicking the mouse.

Software

Java Runtime Engine 1.4.2 or higher must be installed before IGB can run.

Where to find IGB

The preferred method of installation uses Java Web Start, freeware from Sun Microsystems that helps ensure that you always have the most current IGB version; see Appendix 1 Using Java Web Start.

Installing

If you use Java Web Start, follow the instructions in Appendix 1 Using Java Web Start. If you were given a zipped package containing IGB, simply unzip the package in the location you prefer. No other installation is needed.

Uninstalling

To uninstall IGB from your computer:

If you used Java Web Start to install it, see Appendix 1 Using Java Web Start.

If you did not use Java Web Start to install it:

1. Move any data files you want to keep, if they are in the **IGB** folder.

2. Delete any cached data. You can find out where IGB has stored cache data by accessing the **Help** > **About** menu while IGB is running.

Alternatively, clear the cache with **QuickLoad** tab **> QuickLoad Options > Clear Cache**.

3. Delete the unwanted **IGB** folder with its contents. No other files are installed on your computer.

Setting preferences

See Appendix 3 Preferences.

Launching IGB

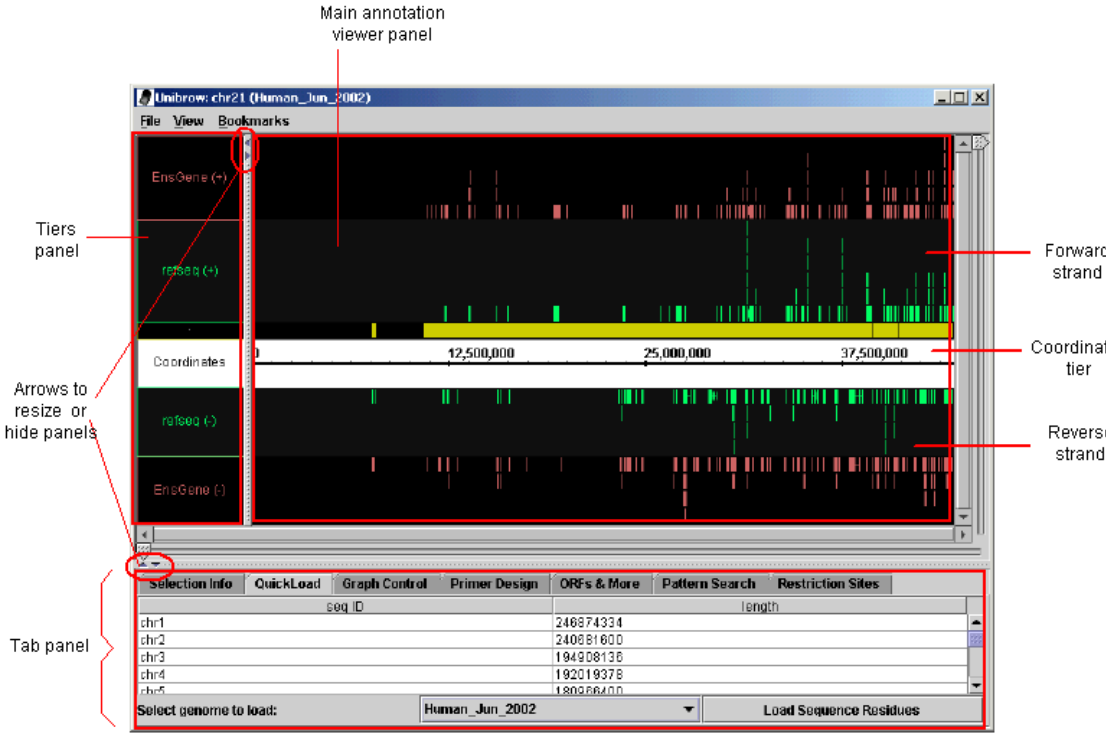
If you use Java Web Start, see Appendix 1 Using Java Web Start.

If not, double-click the **IGB.jar** file. To launch multiple copies of IGB, double-click **IGB.jar** again for each instance desired.

Simply double-clicking the **IGB.jar** file will open the program with default settings. If you need to alter any parameters, such as raising the amount of memory available, you can create a simple script to run the program. For example, the following script will allow the program to use 256 Mb of memory and will use preferences from the file "my_prefs.xml".

"java -Xmx256m -jar IGB.jar -prefs my_prefs.xml"

Overview of the IGB window



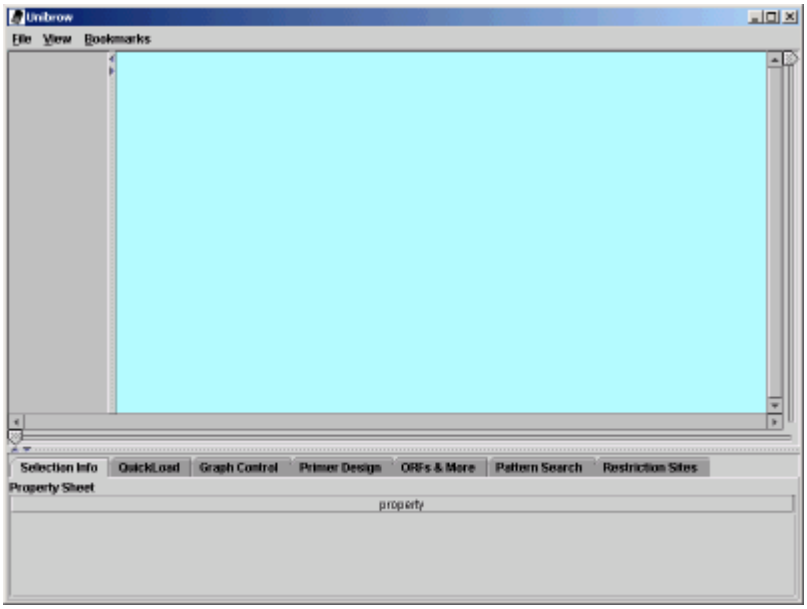
Drag the arrows to hide or resize panels in the IGB window. Resize the entire window by dragging any corner of the window. Tiers in IGB are comparable to "annotation tracks" in other genome browsers.

Information in each tab panel can be displayed in a separate window by clicking that tab, then choosing **View** menu > **Open tab in new window**. To return the window to its tab location, click the box in the upper right corner of the window to close it. You can also hide tab panels you don't use; see Appendix 3 Preferences.

Keyboard shortcuts

Some keyboard commands (keyboard shortcuts) are shown on the menus; all are contained in the preferences file. You can change them or add new ones; see Appendix 3 Preferences.

Choosing and opening data sources



IGB opens with a blank viewer. Load genomic information from any source(s) below to compare annotations.

Data sources that IGB can work with include:

- A quick-loading set of standard data sets such as RefSeq, EnsGene, and KnownGene
- Publicly available data sets from sources such as UCSC
- Data files of various formats
- Graphs

For more about each of these, see the appropriate sections below.

Fast-loading standard reference data

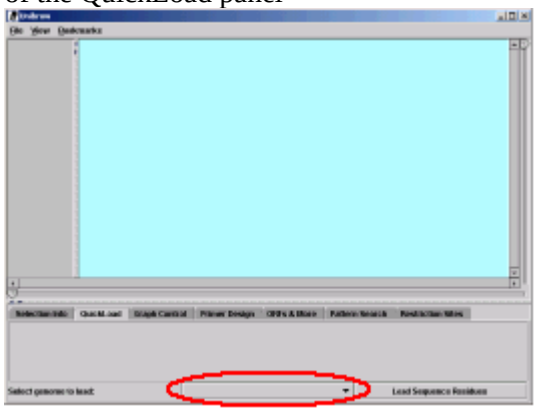
"QuickLoad" standard annotations, predictions, and/or contigs for a whole genome version you choose. Do this before you do anything else; if you do it later, it will obliterate anything else that has been loaded.

1. Make sure you're connected to the network (this may not work on slow remote connections)

2. Click the **QuickLoad** tab

3. Click to select the checkbox(es) for one or more reference data sets from the list on the right. RefSeq will automatically be checked and loaded by default.

4. Select a genome assembly to load from the drop-down list at the bottom of the QuickLoad panel



The drop-down list contracts again when loading is complete.

5. Wait a few moments for the table to load.

6. Initially, IGB displays no annotations; click a chromosome in the first column of the table to view the annotations for that chromosome.

7. Additionally load any of the reference data sets on the right by clicking to select the associated checkbox. Available data sets will vary depending on the genome you select. You may need to click the referenced chromosome to view the loaded data.

8. Load sequence residues for the selected chromosome, if desired.

If the desired sequence does not appear, it may not be available for the assembly you have selected.

QuickLoad data is read-only; to make and save changes, create a curation.

If you need to store QuickLoad data on your own computer for use while away from the network, see Appendix 3 Preferences.

QuickLoad may cache some data for efficiency. Click the **QuickLoad Options** button to set preferences for caching, or to clear the cache.

To load data from a DAS server:

1. If you are already viewing other data, make sure the viewer is displaying the genome and chromosome to be loaded from DAS.

2. If you are already viewing other data, and depending on the density of the annotations you expect to be loaded (EST's, for example, are dense) you may want to zoom in so the range displayed in the viewer is about 2 Mb.

3. Choose **File** menu > **Load DAS Features**

4. Select a DAS Server from the drop-down list; the list will contract when access is established. This may take some time.

5. If not automatically entered, specify genome and chromosome from the drop-down lists. These must match any data you are viewing already.

6. Set the range min and max if you did not zoom in on the region of interest in step 2 above. The recommended maximum range is about 2 Mb.

7. Click the checkboxes for the specific info you seek. If none are available, there may not be data for the sequence you are investigating in the data source you have chosen.

8. For information about the resources listed under the UCSC option, visit http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#IndivTracks

9. Click **OK**.

A file of data

IGB can view data from one or more files. A variety of file formats is supported.

All filenames must include the appropriate filename extension, for example ".psl". Compressed files can be read if the filename extension is part of the filename, for example "myfile.psl.gz" or "myfile.psl.zip".

Examples of these files are included with the IGB application.

Viewable file formats

IGB can display several types of standard and Affymetrix-internal genomics file formats.

An introduction to standard genomics file formats is at:
http://genome.ucsc.edu/goldenPath/help/customTrack.html
Read the introductory material, then scroll down past the UCSC-specific info to the discussions for each file type, or search the page for a file type name.

There are two types of genomic file formats; some describe annotations, others give sequence residues.

Graph file formats are addressed separately.

Annotation file formats

These files give sequence coordinate locations of exons, genes, transcripts, etc. on the genome. Some accepted formats are described in the table below.

.gff	More info is available at: http://www.sanger.ac.uk/Software/formats/GFF/index.shtml or http://genome.ucsc.edu/goldenPath/help/customTrack.html#GFF
.gtf	For information, visit http://genome.ucsc.edu/goldenPath/help/customTrack.html#GTF For IGB use, the tier name is the **source name**. **Score** and **frame** will be ignored. Coordinates will be transformed from Base1 to Interbase 0.
.psl	Displays annotations with **target name** matching the loaded sequence. For additional information about .psl files and how to use them, see http://www.soe.ucsc.edu/~kent/exe/doc/psLayout.doc or http://genome.ucsc.edu/goldenPath/help/customTrack.html#PSL. Files of interest are available via the DAS server. When you load them, you will be prompted to specify whether to use the Query or Target sequence. See Comparing genome to genome alignments.
.psl3	Affymetrix custom format, similar to psl files, described above. Specify whether to use the Query, Target, or Other sequence when prompted. See Comparing genome to genome alignments
.bed	Standard genomic file format. For information, visit http://genome.ucsc.edu/goldenPath/help/customTrack.html#BED
.axml	Affymetrix custom XML format; may or may not have sequence residues

Any of the above that are compressed with any of these extensions:

.zip

.Z (case-sensitive)

.gz

.gzip

The filename must include both the filetype extension and the compression extension, such as “myfile.psl.gz”. Zip files must contain only a single annotation file.

Sequence file formats

These files contain sequence residues only. If an annotation file you use lacks sequence residues, you can merge a sequence file for the same region.

.fasta

.fa

.bnib (Affymetrix internal)

Any of these compressed as above.

Viewing files

You can view a single file, view multiple files or sources of data in a single instance of IGB, or launch separate instances of IGB to view multiple data files.

Viewing a single file

Use this option to view a file that you don't intend to compare with any other annotation tracks. Anything already displayed in the viewer will be obliterated when you use this option.

To view a single file of data:

Choose **File** menu > **Open File.** Make sure that the checkbox "**Merge with currently-loaded data**" is not selected.

Viewing multiple files in a single instance of IGB

To view and compare annotations from multiple files in a single IGB window, or if annotations from other sources are currently loaded:

1. Make sure the genome version and chromosome of files you want to open match those of the data already displayed.

   IGB recognizes standard terms for version and sequence (For specifics, see the Synonyms file on the QuickLoad server: (http://147.208.165.250/quickload_data/synonyms.txt).

   If in doubt, check the text at the beginning of the file to be opened. If genome version and chromosome cannot be matched, you may still be able to open a file if the sequence name matches.

2. Choose **File** menu > **Open File.** Make sure that the checkbox "**Merge with currently-loaded data**" is selected.

Annotations from merged files will automatically become additional tracks alongside tracks already in the viewer.

If merging brings in the file as a narrow band at the top of the viewer panel, use the vertical slider to make the annotations easier to see and view the tier name. In the Working with tiers section, see Expanding vertically

Viewing multiple files in separate instances of IGB

To view files in separate IGB windows:

Launch two instances of IGB and open the desired sequences or files in each. You can compare annotations between the two windows. See Comparing data sets.

Graphs

Graphs contain numerical data associated with base pair positions along the genomic sequence axis, such as sets of scores in specific positions along the genome.

Examples of suitable data include:

- Expression values from whole-genome tiling arrays.
- Density of EST's across a chromosome
- GC content along the genome
- Measures of conservation between two genomes

Data should be formatted as described under one of the graph file formats below.

Graph file formats

IGB can display graphs in several file formats, all developed at Affymetrix.

Examples of the following graph file formats are included in the folder with the IGB application or can be downloaded from the Affymetrix Web site.

All files must include the filename extension, for example ".gr".

Graph files may be in the following formats:

.gr	The format of these text files is: no header; two columns of numbers separated by a single space or a tab. The first number is the base position; the second number is the score. The name of the file is the name of the graph.
.bgr	Similar to .gr files, but they are more compact and load faster. They also contain additional information, such as how the graph was created, parameters, and the genome version and sequence.
.bar	Generated by Gtrans

Opening graph files in IGB

To open graph files in any of the supported formats:

1. Copy the selected graph to your hard drive

2. Make sure the coordinates of the graph file correspond to those currently displayed in IGB

3. Choose **File** menu > **Add Graph**

Important: .gr files may display even if they don't match up with the coordinates currently loaded, but any comparisons will be meaningless.

Other sources of data

Other sources of loadable data include:

- Bookmarked data from IGB, imported
- Bookmarked data from IGB, via web browser
- Saved curations
- Saved primers

Secure data sources

QuickLoad and DAS data can be made available via secure https connections. Set your IGB preferences for that data source to point to the URL of the secure server. For information, see Appendix 3 – Preferences.

Viewing, navigating, selecting, and finding data

This section describes the different ways to:

determine what you're looking at

view annotations and information about them

navigate around a genome assembly or other loaded data set

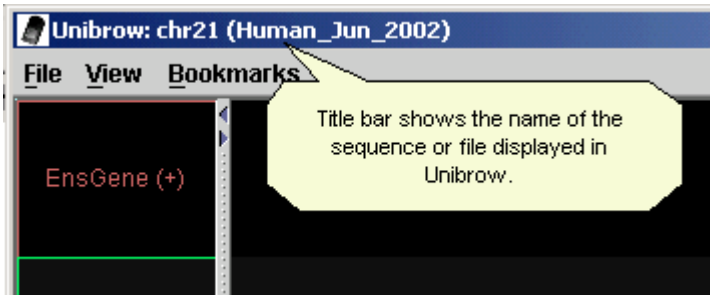
zoom in on areas of interest

find annotations, sequences, or patterns

select annotations for use with various features of IGB

Identifying the sequence displayed in the viewer

If you are viewing a standard data set, the genome version and chromosome or sequence being viewed are identified in the window title. If you are viewing a file, the filename will be displayed.



Clearing the IGB viewer

To clear the contents in the viewer in order to work with a different data set:

Choose **File** menu > **Clear All**

QuickLoad annotations and any loaded sequence residues remain loaded; click a chromosome to display it. Curation and primer tiers you have created will also remain if you return to the relevant chromosome. Bookmarks remain in the Bookmarks menu.

Loading and viewing sequence residues

Many IGB features require sequence residues to be loaded, so you may want to load them initially during each IGB session.

Loading sequence residues

For QuickLoad data, load sequence residues separately for each chromosome:

1. **QuickLoad** tab > click a chromosome in the list > **Load sequence residues** button.

2. Wait until button returns to light gray from dark gray color.

Sequence residues contain a large amount of data. If loading over the internet is too slow, consider copying the sequence residues from the QuickLoad server onto your local network. If you do this, you will need to modify your preferences file to show the new Quickload server location.

To work with large amounts of sequence data, you may need to increase the amount of memory available to the program. See Launching IGB.

For file data, if sequence residues are not included in your file, merge a .fasta sequence file for the identical sequence.

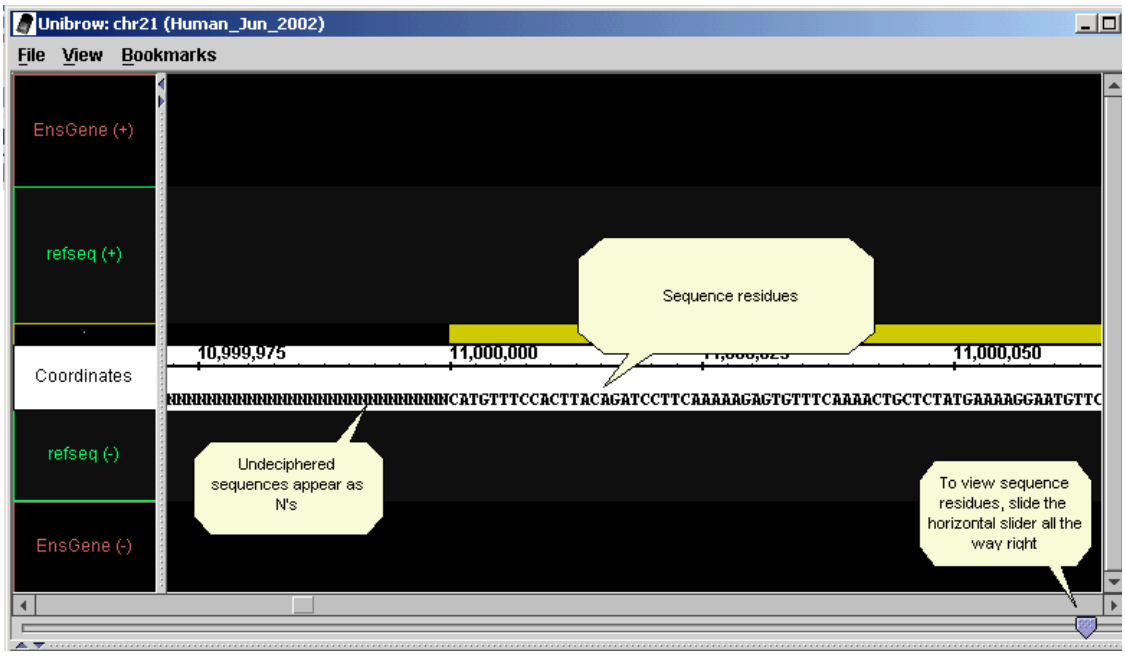
Features requiring sequence residues

The following functions require that sequence residues be loaded:

- Selection of a region or sequence for making curations, copying to other application, etc.
- ORF analysis
- Pattern searches on sequences
- Finding restriction sites
- Primer design

Viewing sequence residues

Sequence residues, when loaded, are visible in the Coordinates tier when zoomed all the way in (see picture).



The Coordinates axis appears slightly above the center of the Coordinates tier when sequence residues are loaded.

Unsequenced regions appear as Ns.

Navigating around the genome

In addition to the following navigation methods, see also Showing restriction sites and Finding instances of short sequences.

Going to chromosome (standard reference data)

QuickLoad tab > click a table row

Going to an annotation

There are several ways to go to a particular annotation:

Click the **Annotation Browser** tab, then click the annotation ID you want to go to. Wait a moment.

Click the **Pattern Search** tab, then enter the (case-sensitive) identifier for the annotation you want to jump to in the second box (**enter id of annotation to find**), then press the <Enter> key. (Ignore the other options in this panel.)

IGB displays the region of the annotation, if it is present in the data set being viewed. If QuickLoad data is displayed, the window switches to the chromosome the annotation is on.

See also Zooming in on an annotation and Select parent.

Going to a specified coordinate

To go to the coordinate you specify and focus the zoom on it:

1. Click the **Pattern Search** tab

2. Enter the coordinate number you want to jump to in the first box (**enter coordinate to center on**) then press the <Enter> key.

If you have clamped to a view or are looking at a file that does not include the coordinate you want to jump to (and QuickLoad is not loaded), this feature does not work.

Scrolling along the chromosome/ file data

When you are zoomed in, you can browse the view in several ways:

- Use the scroll box at the bottom of the viewer panel to scroll along the sequence. Box size and speed of movement will vary according to the position of the horizontal zoom slider (see screenshot).
- Click in the track on either side of the slider to move it in the direction of your click.
- For finest precision when zoomed in, use the arrows at the ends of the scroll box track.

When you are zoomed all the way out (zoom arrow is at the far left), you are already viewing the entire range of the loaded data and the scroll box is not available.

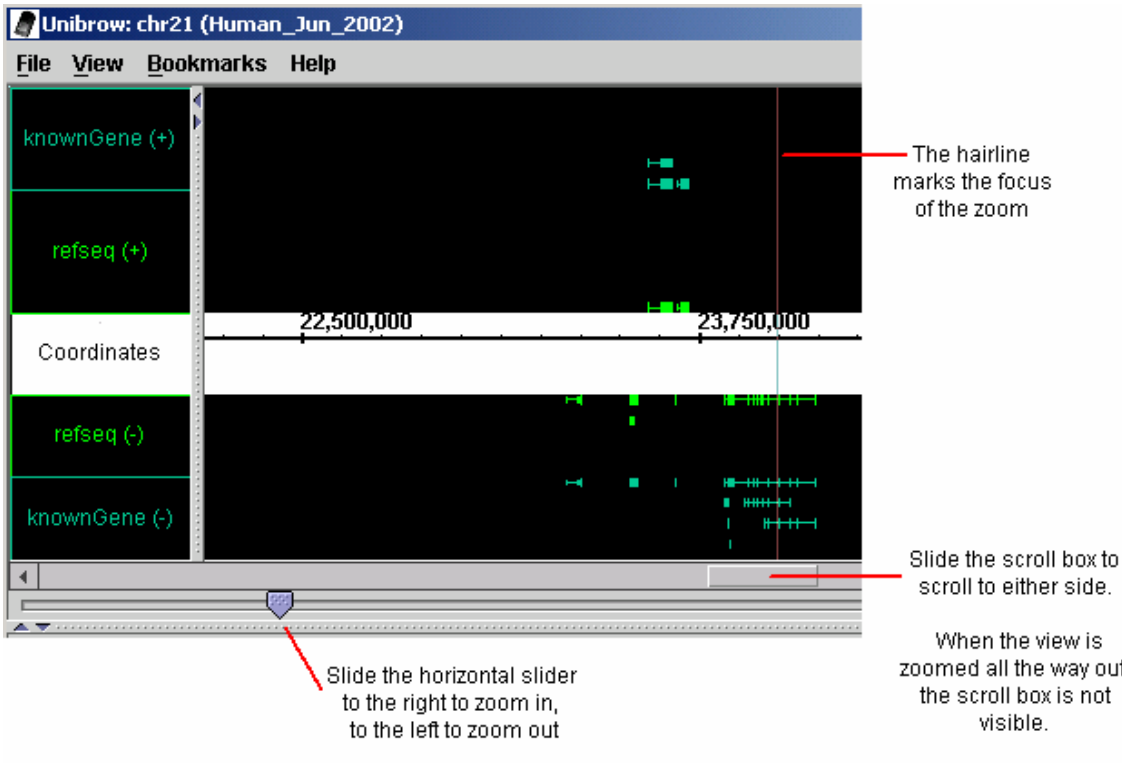
Zooming and locking

Zoom in on an area of interest, and then lock on it to protect your view. Slicing is a separate feature that shares some similarities with zooming.

Zooming in on a coordinate or region

To select a point to zoom in on, click anywhere in the viewer to set the focus of the zoom. A hairline will mark the zoom focus. Move the arrow-shaped slider to the left to zoom out, to the right to zoom in.

To see the exact number of the coordinate marked by the hairline, choose **View** menu > **Toggle hairline label.** The hairline will display the coordinate number with each subsequent click. To turn off the label, choose **View** menu > **Toggle hairline label** again.



See also Going to a specified coordinate

Zooming in on an annotation

To center on and fill the viewer with the selected object or rubber banded region:

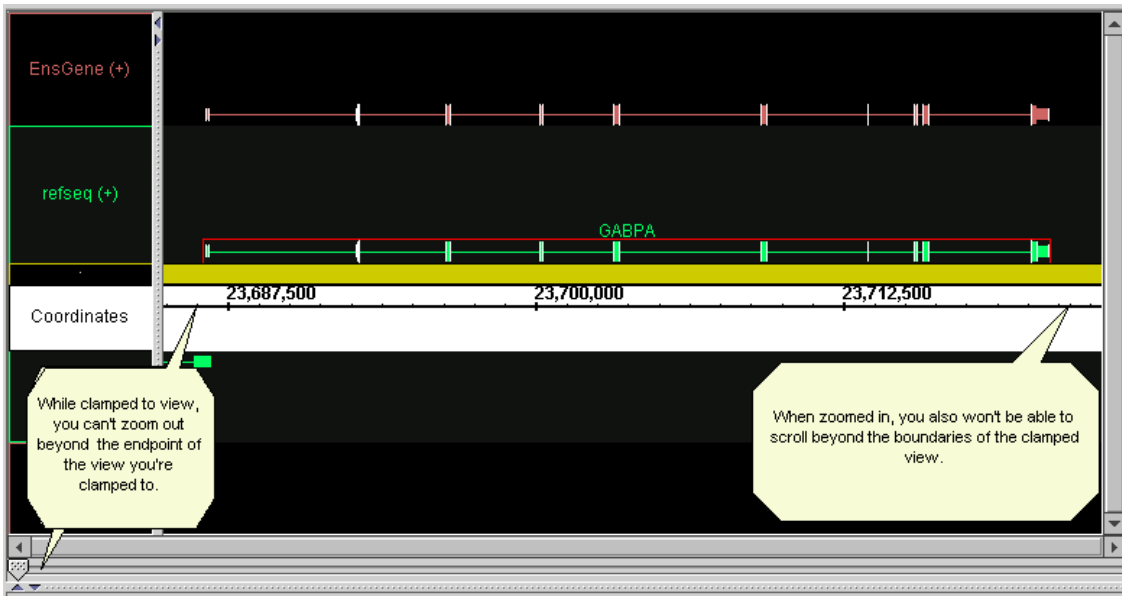
Right mouse click on an item > **Zoom to selected**

To unzoom, slide the horizontal slider back to the far left.

See also Going to an annotation.

Clamping on a view

Normally, if you are zoomed in on a small region, it is very easy to scroll away from the target region and lose it. Clamp to view so you don't lose what you're examining as you browse the region using the zoom and scroll sliders.



Clamping restricts the max and min coordinates of the view to the boundaries of the clamped view, so that you cannot zoom out or scroll beyond the range you've clamped to.

For example, if you want to examine the gene GABPA more closely, zoom in so the gene is the width of the window, then clamp to view. When you then zoom in further for a closer look, the scrollbar now will not scroll beyond the begin- and end-points of the gene. Clamping also allows finer control when scrolling by moving the horizontal scroll box.

To clamp to a view:

1. Zoom in on a region of interest. If you're zoomed all the way in or out, you can't clamp.

2. Choose **View** menu > **Clamp to view**

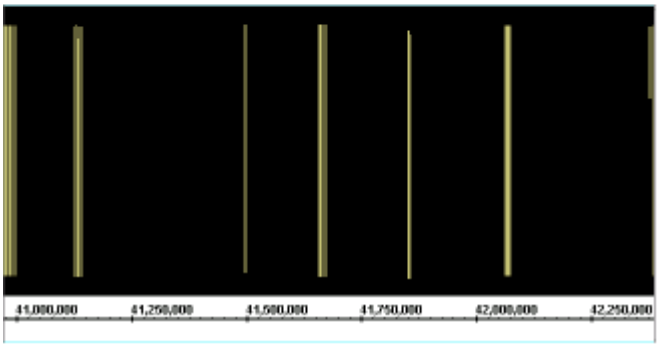
3. To unclamp and zoom out to or scroll along the entire chromosome or file content, choose **View** menu > **Unclamp**.

Zooming in and lock on a small region

When working with a file that contains annotations that only cover a small region of the sequence (for example, curations in a small region saved as a BED file), you can tell IGB to ignore most of the sequence range and zoom in on the region of interest:

Before shrink wrapping:

After shrink wrapping:



To zoom in on a focused region in a file:

Choose **View** menu > **Toggle Shrink Wrapping on**.

Note that this will not work if you have also loaded other data that has no empty black space to eliminate.

Selecting annotations or sequences to work with

Select annotations in order to perform certain operations on them. For context-sensitive menus, first select the annotation or item to work with, then right-click to bring up the menu. The following are ways to select items or sequences in the graphical window.

To select:

A single annotation:

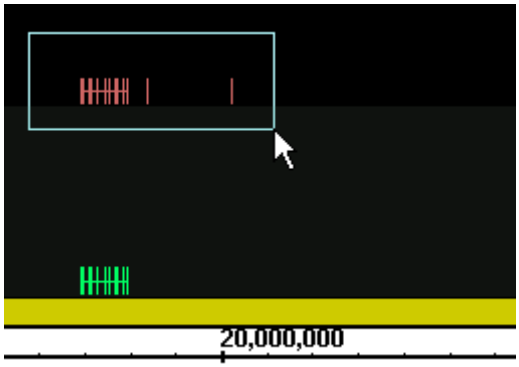
Click on it.

Multiple annotations:

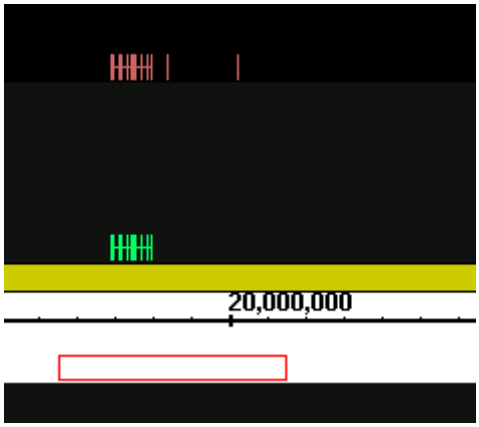
Shift-click on each or shift-drag to select multiple adjacent items (the latter selects all items touched by or enclosed by the drag boundaries.)

A region of sequence:

In the black area in the viewer window, click-drag a box around an area:



This is called rubberbanding. A red box in the Coordinates tier indicates the captured sequence/ range:



For precision rubberbanding, zoom in all the way so sequence residues appear.

Deselect everything:

Click the empty black space.

Select the gene containing the selected exon:

Right mouse click a selected annotation > **Select parent**

Or click the line connecting the exons.

Mistakenly selecting the exon rather than the gene may cause some functions to appear not to work. Use the select parent function to select the gene.

If selecting seems slow, freeze the sliced view:

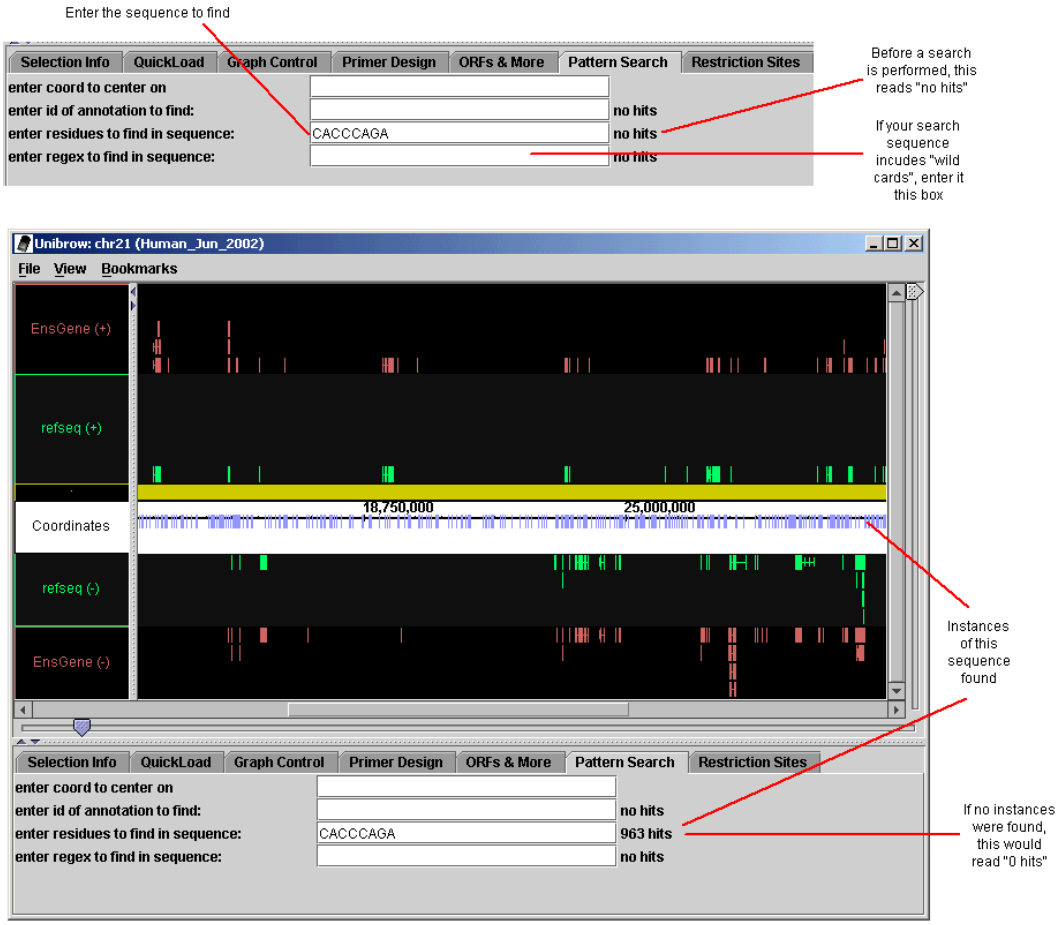
Choose **Sliced View** tab > uncheck **Slice based on selection**.

Finding sequences

For short sequences, use the Pattern Search feature described in the first sections.

Finding instances of short sequences

Use the Pattern Search feature to locate each instance of a short sequence you specify, either exact bases or including placeholders for unknown bases. All instances of that sequence on the chromosome will be marked on the Coordinates tier.



These functions work only for data loaded as files or QuickLoaded, not data loaded from remote servers (i.e., via DAS servers). Sequence residues must be loaded.

These searches are case-sensitive for efficiency. Fasta files may contain both uppercase and lowercase sequences (lowercase is for repeating sequences); if you are working with data from a .fasta or .fa file, search for upper and lower case separately as desired. It is possible to make regular-expression searches be case-insensitive by using the flag "(?i)" at the beginning of the regular expression.

Found instances of your search sequence will be marked in color on the Coordinates tier (see screenshot above). These may be difficult to distinguish from Restriction Sites you may have marked.

The number of instances found appears in the tab window next to the field in which you entered the sequence string.

No hits means no search was performed for that function. **0 hits** means no matches were found. (See the screenshot.) Watch this indicator to know when IGB is done processing your search. The cursor will also begin to blink in the text-entry field when processing is done.

Finding instances of a known sequence

In the chromosome being viewed, identify all instances of an exact base sequence you enter:

1. See the information in the previous section.

2. Click the **Pattern Search** tab and enter a case-sensitive sequence or (rubberband and) paste bases in the third text-entry field (**enter residues to find in sequence**). Press the <Enter> key. Wait a few moments.

Deleting pattern markers from the Coordinates tier

To delete pattern markers from the Coordinates tier, press **Clear**.

Finding instances of a sequence containing unknowns

In the chromosome being viewed, identify all sequences matching a regular expression (a sequence containing "wild card" nucleotides):

1. See information about finding instances of short sequences above.
2. Click the **Pattern Search** tab and enter a case-sensitive string of bases and wild cards (see table below for common ones) in the text-entry field called (**enter regex to find in sequence**). Press the <Enter> key.
3. Wait a few moments; this may take some time.

Example wildcards:

Wild card	**Represents**	**Example entry**	**Finds sequences**
.	any single nucleotide	ACCT.T	ACCTTT, ACCTAT, ACCTGT, and ACCTCT
…	any three nucleotides	ACCT…T	ACCTGAGT, Etc.
[CG]	a C or a G	ACCT[CG]TC	ACCTCTC and ACCTGTC
T{1,n}	1 to n T's	ACGGT{1,3}C	ACGGTC, ACGGTTC, and ACGGTTTC
T*	Zero or more T's	ACGGT*C	ACGGC, ACGGTTTTTC, Etc.

.*?	a string of any length containing any nucleotides	TCGGGGTTAA.*?CTGGACTC	Many possibilities.
.*	the longest possible string of any length containing any nucleotides	TCGGGGTTAA.*CTGGACTC	Differs from the search above in that the longest possible result(s) will be found
(?i)	use case-insensitive matching	(?i)A[CG]T	ACT, agt, acT, AGt, Etc.

The full list of regular expression syntax usable in IGB is available at http://java.sun.com/j2se/1.4.2/docs/api/index.html.

For more detailed information about wildcard searches in general, see Friedl's book *Mastering Regular Expressions*.

To delete pattern markers from the Coordinates tier, press **Clear**.

Getting further details about annotations

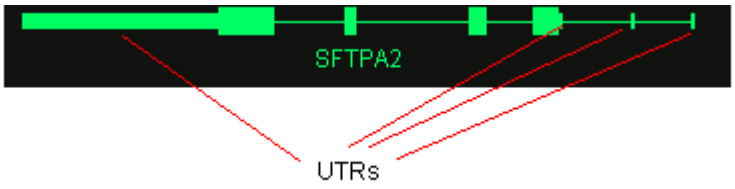
IGB can provide specific information about annotations in various ways.

Annotation ID's

To make annotation names visible, use the vertical slider at the right side of the viewer window to increase the annotation size until the text is visible.

Untranslated Regions (UTRs)

UTRs are indicated graphically by a "tongue" at either end of a gene, or a thinner-than-usual exon, as in this example:



Only some data sets and files will indicate UTRs.

Item details as text

View start, end, chromosome, and strand for all annotations in the loaded sequence that have ID's

1. Click the **Annotation Browser** tab
2. Scroll down to the annotation of interest.

Details for currently selected features

To get details about annotations, graphs, or primers:

1. Click the **Selection Info** tab
2. Click to select the item of interest in the viewer. This can be an annotation, primer, or graph.
3. Shift-click to select more than one item.
4. If you click an item and less information than you expect appears in the Selection Info panel, you may have selected an exon rather than a gene. Select the parent of the annotation to see more information.

The kinds of information displayed vary by item selected and source.

Selection Info shows one column for each gene (up to 10) that you shift-click to select in the viewer.

More information via web browser

The features below will open the results in a web browser. On Windows, your default browser will be used. On Linux/Unix, IGB will attempt to use "netscape" by default, which requires that "netscape" be in your PATH. On Linux/Unix, you can change the browser used by editing the preferences file. See Appendix 3 – Preferences.

Viewing a region in the UCSC web genome browser

The UCSC web genome browser is a rich source of genomic information.

For more information on what's available there, visit http://genome.ucsc.edu/.

To view in the UCSC browser the same region currently displayed in IGB:

1. Select an item in the viewer window

2. Choose **View** menu > **View region in UCSC browser**

Results may be slow to appear. (You may want to set your web browser preferences to show loading progress -- for example, IE has a bar on the bottom of the web window. See the documentation for your web browser.)

See also Data from DAS servers to load annotations from UCSC and other sources into the IGB viewer.

Viewing information on public web sites

For many annotations, more information about an annotation is available directly from the source. This feature automatically takes you to NCBI (for RefSeq annotations), Ensemble (for EnsGene annotations), UCSC (many DAS annotations), or other web site depending on the source of the annotation you seek information on. Not all annotation sources contain the necessary link information for this feature; if the menu option is unavailable for the annotations you are examining, this means that the data source has not been associated with a website. You can configure IGB to provide links to any website of your choice for each data track. See Appendix 3 – Preferences.

To view information for an annotation from a public web site:

1. Make sure you're selecting the gene, not just an exon. Select the parent of the annotation if necessary.

2. Right mouse-click an annotation in the viewer > **Get more info**

Comparing data sets

You can view and compare information from different sources, customize the tiers, and use unique IGB tools for analyzing and comparing.

View and compare information from various sources:

- QuickLoad data
- Data from DAS servers
- Files (Merged)
- Curated/ Saved annotations (via DAS or BED)
- Graphs
- Bookmarks (imported or via Web)
- IGB tools such as ORFs, restriction sites, pattern repeats, etc.
- Two different genomes

Tools for comparing graphs are discussed in the Working with graphs section

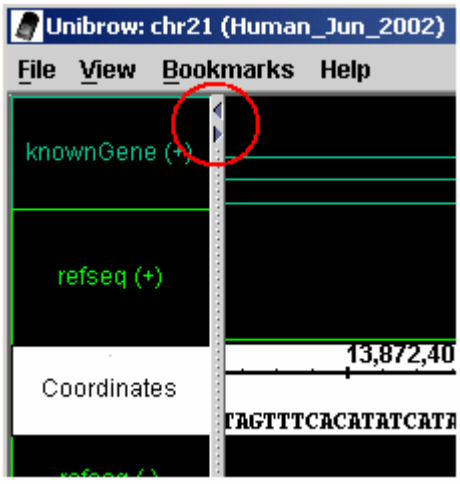
Working with tiers

Each data set/ source you load into a single IGB appears as its own tier. You can position the data sets relative to each other vertically, and temporarily hide them from view, to facilitate viewing and comparison.

Resize the tiers panel

The name of the tier may be too long to fit into the default tier size.

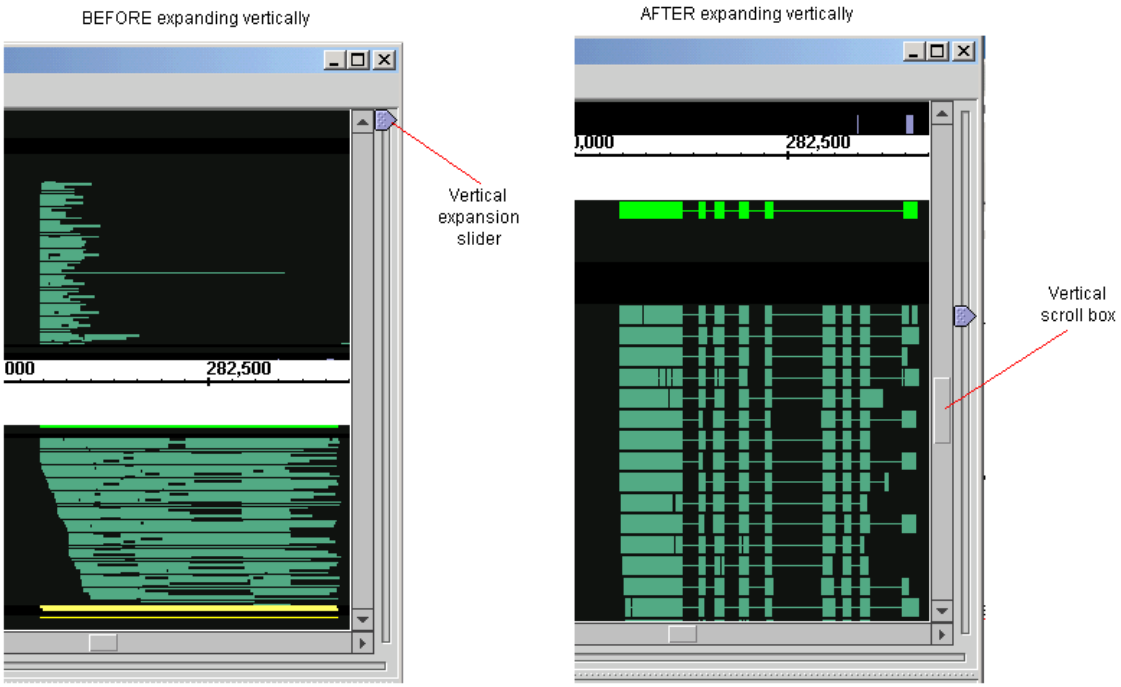
To resize the tiers column, drag the separating bar or click the tiny arrows:



Expanding vertically

This feature is especially useful for viewing ESTs.

To vertically expand the view, drag the vertical slider:



After expanding vertically, use the vertical scroll box to view all annotations.

Re-ordering tiers top-to-bottom

To re-order the tiers in the tiers column:

Click a tier name (in the panel on the left) and drag the tier up or down to the desired location.

Deleting tiers

Tiers can't be deleted; hide them instead.

Hiding and Showing (Unhiding) tiers

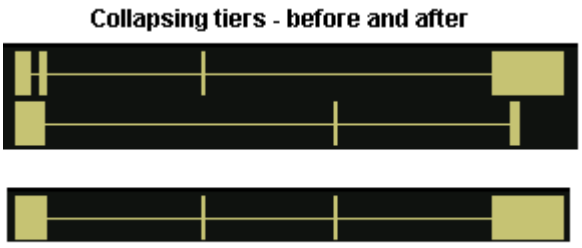
1. To hide a tier: Right-click on the tier or tiers you want to hide > **Hide**

2. To show a hidden tier: Right-click on a tier > **Show** > choose the tier you want to show. If no tiers are hidden, none will be listed for "Show".

Collapsing and Uncollapsing tiers

For tiers containing multiple rows of annotations, collapsing tiers consolidates all rows within a tier into a single row. Annotations on any row will be included in the collapsed tier. Larger annotations may obscure

smaller ones; annotations with introns may be obscured by annotations that don't show the intron.



To collapse or expand one or a few tiers:

Right-click on tier name(s) > **Collapse** or **Expand**

To collapse or expand all tiers at once:

Right-click on tier name > **Collapse All** or **Expand All**

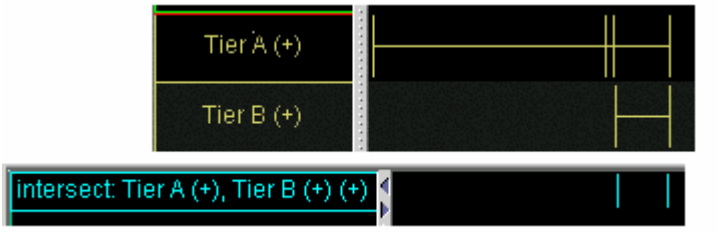
Working with multiple tiers

To select more than one tier, hold down the Shift key. To perform right-mouse operations on multiple tiers simultaneously, continue holding the Shift key while right-clicking to bring up the menu of tier options.

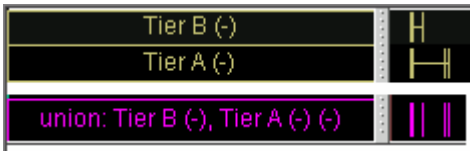
Combining tiers

You can highlight differences or commonalities between different annotations or graph thresholds on two tiers by combining the tiers in one of several ways:

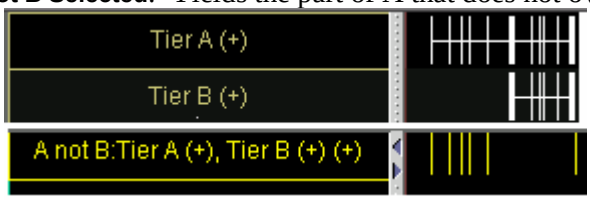
Intersect Selected: Yields only the overlapping portions of A and B



Union Selected: Yields all bases that are in either A or B



A not B Selected: Yields the part of A that does not overlap with B



B not A Selected: Yields the part of B that does not overlap with A

XOR Selected: the union of the yields of (A not B) and (B not A); i.e., all of the material that is in A or B but not both.

Not Selected: Yields all coordinates that are not in the selected tier

For example, if you are comparing graphs, you can capture annotation-like bars resulting from different threshold settings, then compare the snapshots, tweak the threshold settings, and compare them again:

1. Snapshot the graph thresholds if you are working with graphs
2. Select the tiers to compare
3. While continuing to hold down the Shift key, right-click a tier name > **Combine Tiers** > choose an expression above for comparing tiers

Tools for highlighting patterns

Highlighting matching endpoints

To see if different sources show the same start and endpoints for an annotation:

Select an item; matching edges of corresponding items in all tiers being viewed will be highlighted in white.

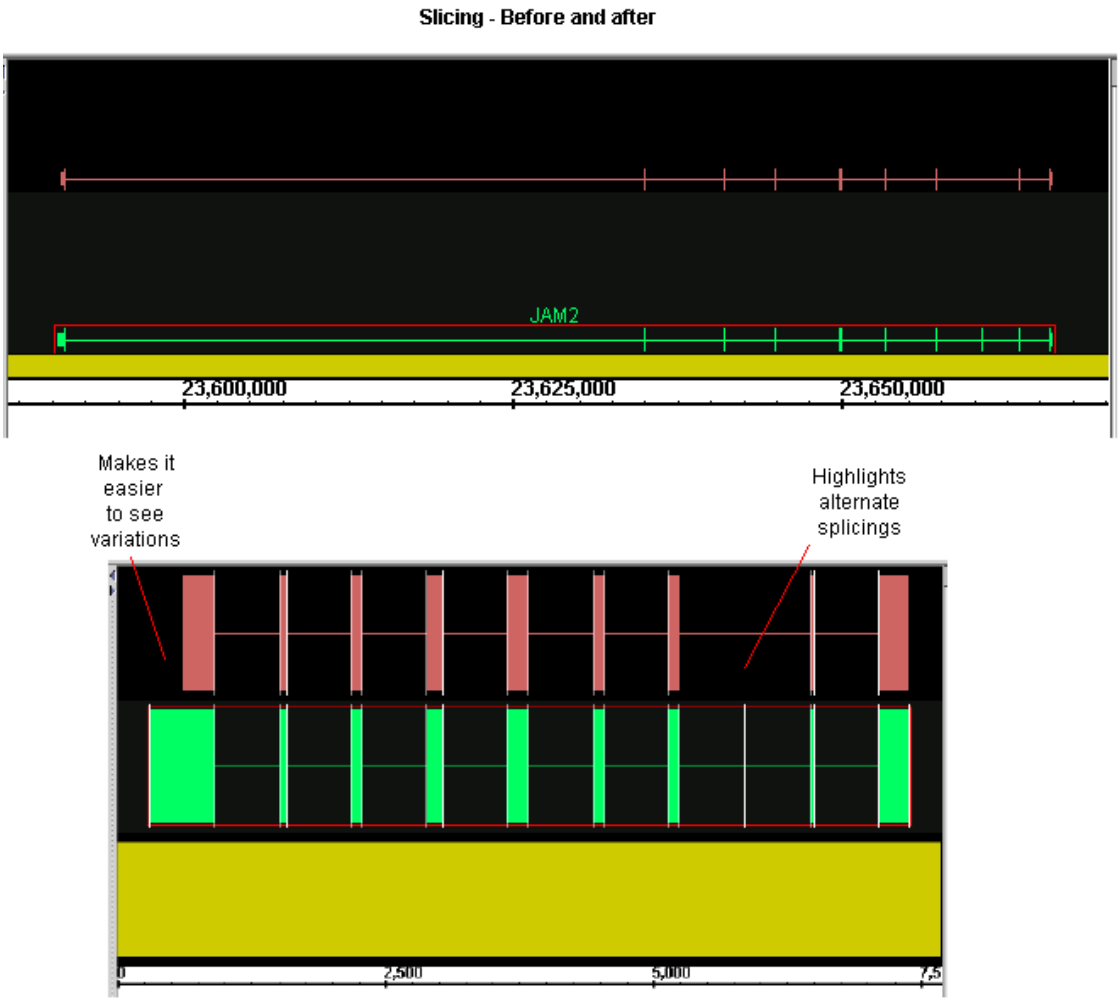
To specify the degree of closeness required for a match (how many base pairs difference is still considered a match):

Choose **View** menu > **Adjust Edge Match fuzziness**

By default, if edge match fuzziness is adjusted, all matching edges are highlighted in gray rather than in white. These colors can be modified in Preferences.

Slicing

Slicing makes slight variations more visible when comparing multiple annotations on the same sequence. Slicing cuts out long introns and realigns the exons of a selected annotation or set of annotations, making pattern irregularities and possible alternate splicings more obvious.



To slice a selection:

1. Click the **Sliced View** tab.

2. Make sure the checkbox **slice based on selection** is checked.

3. Select annotations to slice in the main view.

It may take a few moments for IGB to display the sliced region. For easy comparison, the sliced view displays all annotations and graphs in the region being viewed. The numbers in the Coordinates tier in the sliced view indicate scale; they don't correspond to genomic coordinates. Note that endpoint matching in the upper and lower windows are independent of each other.

You can view the sliced region in a separate window by opening the tab in its own window.

Each sliced region includes a buffer, in order to allow examination of the bases in the intron-exon boundary. The buffer is initially 100 bps on each side of each exon; if an exon in the same range that was not selected for slicing extends beyond the buffer, it may be cut off. To avoid this problem, increase the buffer enough to accommodate the entire exon by changing the number of base pairs before and after each exon.

To adjust the buffering number of base pairs before and after each exon:

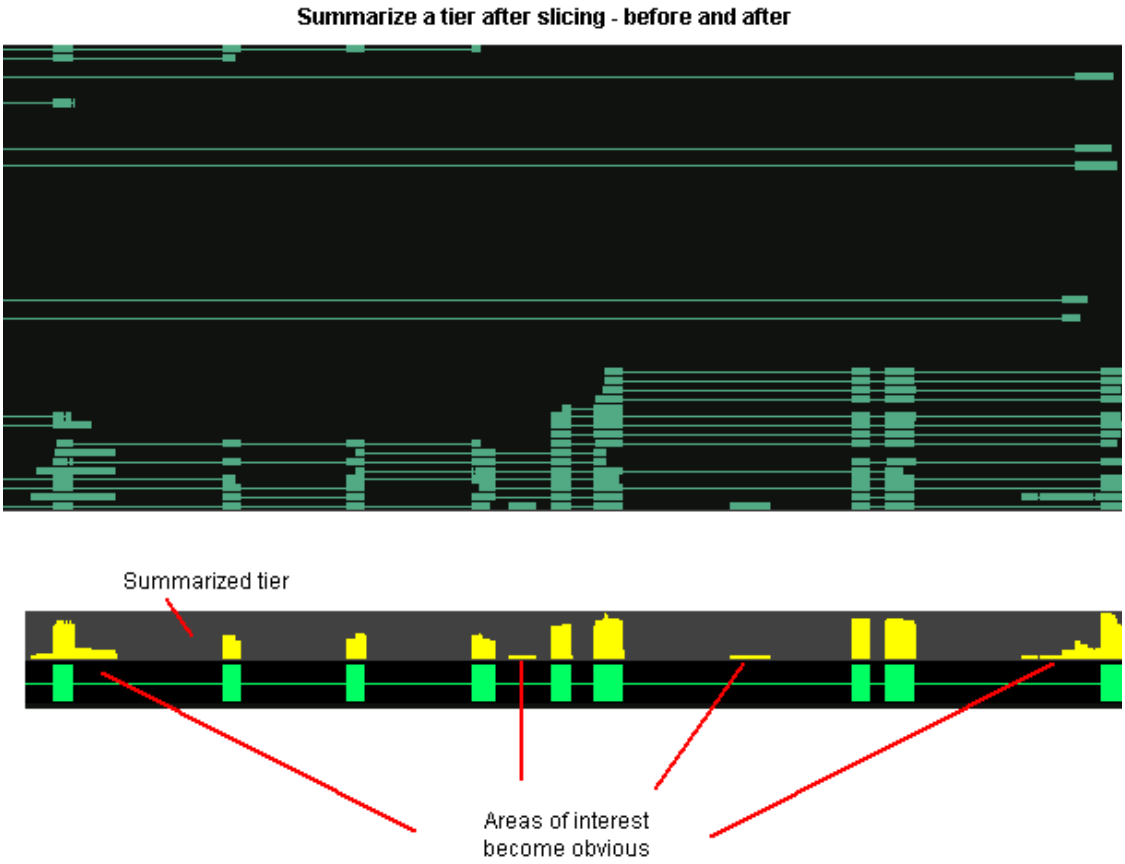
1. Visually estimate the number of base pairs required to achieve the adjustment needed

2. In the **Sliced View** tab edit the value in the field **Adjust slice buffer**

To further enhance viewing and comparison, especially of ESTs, summarize the tier after slicing.

Summarizing tiers

For tiers containing multiple rows of data, summarizing collapses a stack of annotations so that anomalies are readily visible. This is especially useful after slicing and when examining ESTs.



To summarize a tier:

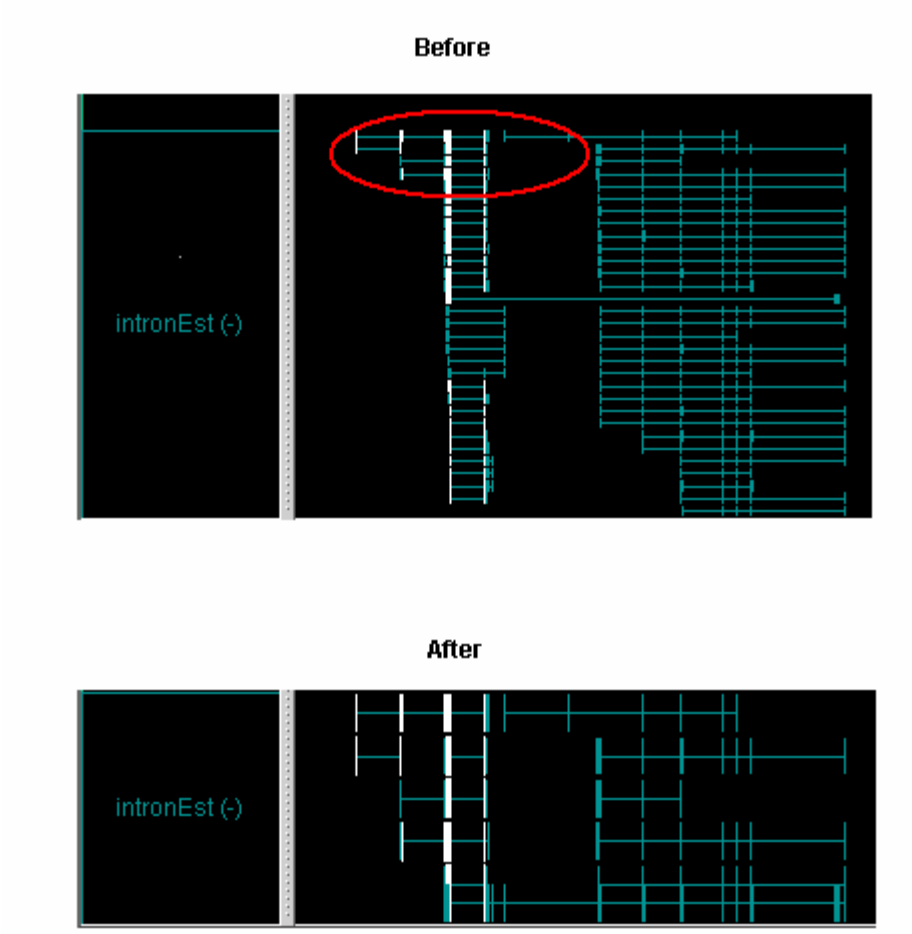
1. If you will use slicing during this analysis, slice now
2. Right-mouse click on the label tier > **Make density landscape**
3. Wait a few moments for processing.

The summary tier appears just below the tier being summarized.

Limiting the number of annotation rows in a tier

If a single tier contains many rows of annotations, some annotations may be represented by many instances while others have few representations. To focus on annotations that have fewer representations, limit the number of rows displayed.

In the example below, to make the circled annotations more visible, limit the number of annotation rows in the tier to 5:



To limit the number of annotation rows in a tier:

1. Right-click on the tier > **Adjust max expand**
2. Enter the new maximum tier height
3. Click **OK**

To apply the same limit to all tiers in the viewer:

1. Right-click on the tier > **Adjust max expand all**
2. Enter the new maximum tier height
3. Click **OK**

To return to the original view and make all annotation rows in a tier visible:

1. Right-click on a tier > **Adjust max expand**
2. Enter **1000** for the new maximum tier height

3. Click **OK**

Working with graphs

Introduction

Graphs allow you to visualize numerical data (scores) associated with base positions along the genome.

IGB graph views were originally developed to show expression data from genome tiling arrays, arrays that contain probes placed at specific positions along the genomic sequence axis. These arrays have been used to survey expression across entire chromosomes. To find out more about genome tiling arrays, visit the Affymetrix Transcriptome Project Web sites at: http://www.affymetrix.com/transcriptome/index.affx and http://transcriptome.affymetrix.com/local_index.html.

IGB provides a variety of ways to fine-tune how graphs are shown to make important features in the data easier to see. This chapter explains how to:

- Change graph appearance
- Fine-tune graph scaling
- Use graphs to create new annotations with thresholding

For information about graph data sources, file formats, and loading of graphs into IGB, see the Graphs section in the Choosing and opening data sources chapter.

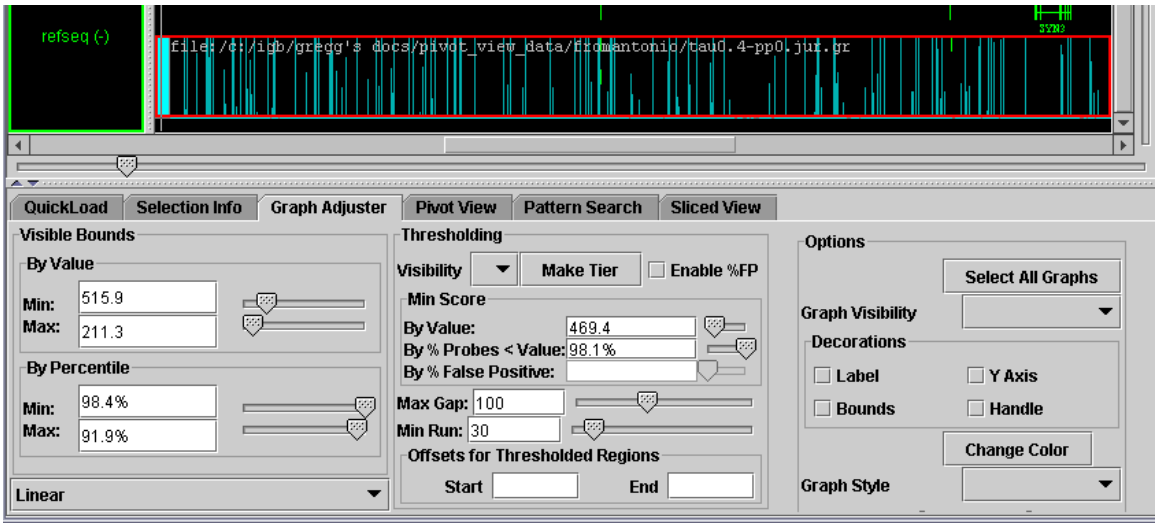
Graph adjuster tab

Settings and adjustments for graphs can be made using the **Graph Adjuster** tab. Note that earlier version of IGB used right-clicking on graphs to change settings; the current version now uses the Graph Adjuster tab instead.

To view the settings for an individual graph, you just select the graph and click the **Graph Adjuster** tab. Changes can be made to the graph settings by typing in new values or by operating sliders. You can also change the settings for a group of graphs all at once by selecting more than one graph. Any changes you then make to the values in the **Graph Adjuster** tab will apply to all currently selected graphs.

To make sure that the Graph Adjuster tab is included in the IGB display, be sure to set the "USE_GRAPH_ADJUSTER" option to "true" in your preferences file. See Appendix 3 Preferences.

When set to "true," the tabbed panel shown in the figure below will appear in the tabbed panel sections of the IGB display:



Graph appearance - customizations.

View graph as a tier

By default, graphs can be click-dragged to new locations across other tiers. To put a graph into its own tier (like the other annotation types) use the **Attach** radio button:

When a graph is attached, sometimes the title is hard to read. To view the full title of the graph in the tiers column, resize the tier column.

To detach the graph from the tier so that it can be dragged over the other tiers, select the **Float** radio button.

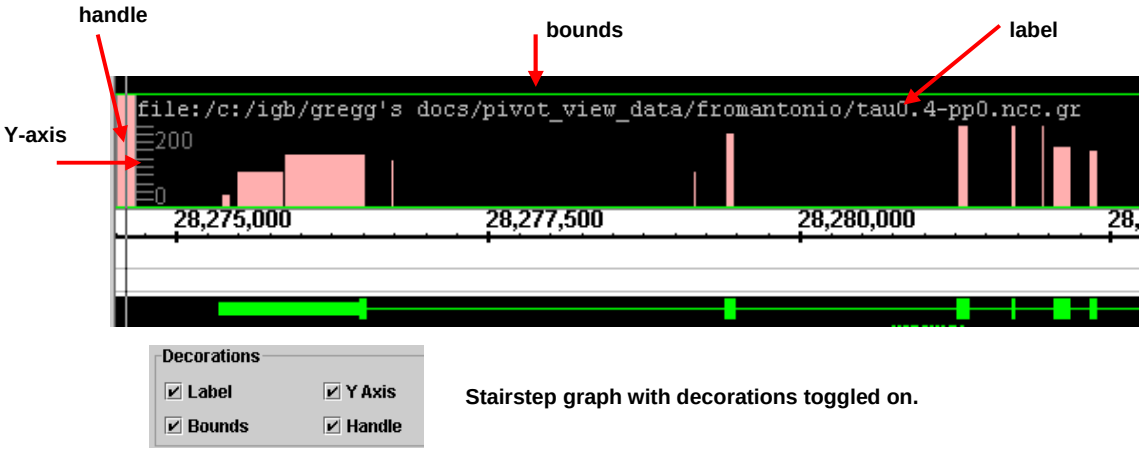
Decorations

Use the Decorations section of the Graph Adjuster panel to modify the appearance of selected graphs.

Currently, there are four types of decorations that can be shown:

- **y-axis** scale
- a **boundary** line around the graph
- the data source's **label** (typically a file name)
- the graph **handle** used for click-dragging the graph to new location

An example graph showing all four options turned on is shown below:



Stairstep graph with decorations toggled on.

Change graph color

To change graph color:

1. Select the graph to change, and then click the **Change color** button.
2. Choose the desired color, and then click **OK**.
3. Click **Reset** if you choose not to use the selected color.

Change the height of a graph

To stretch out the graph in the vertical direction:

1. If the graph is attached as a tier, detach it so it floats
2. Hold down the shift key and drag the mouse up or down on the graph bar at the left of the graph.

Hide/ show graph

To hide the data in a graph or to show a hidden graph, use the **Graph Visibility** pulldown menu.

Delete one graph from the viewer window

To delete a single graph from the viewer:

1. If the graph is attached as a tier, detach it so it floats
2. Select the graph to delete and click the **Delete graph** button.

Delete all graphs from viewer window

To delete all graphs currently displayed:

Choose **File** menu > **Clear graphs**

Change graph styles

Graphs can be shown the following representational styles. To change the style, **Graph style** menu:

MinMaxAvg – This is the default style and is usually fine for most purposes. It is especially useful for showing very densely populated graphs with data points for large numbers of positions.

When IGB is zoomed all the way in, values at individual bases resolve into bars, one for each data position/numerical value pair. When zooming out, IGB starts to summarize values. When the scale of the display reaches the point where individual pixels are associated with multiple values, IGB picks the maximum and minimum values and draws a bar between them. In addition, IBG draws lines through the centerpoints (the average) of all the bars.

Line – In this style, adjacent values are linked with a line. If the zoom level forces a single pixel to represent multiple values, the average of all them is shown.

Bar – Individual values are shown as vertical bars that are one base wide. Height is determined by the value and also the scaling. Zooming is handled as with line graphs.

Dot – Instead of bars, dots are shown. Zooming is handled as with bar and line graphs.

StairStep – This style is particularly useful for viewing ESTs or other density data. The **StairStep** style is similar to the bar graph style except that bar widths along the horizontal axis are stair-stepped.

For example, if position 100 has a value of 50 and position 200 has a value of 75 and there are no values in between, then IGB will draw a bar of height 50 that starts at position 100 and stops at position 200. Then, at position 200, IGB will draw a new bar of height 75 that terminates at the next location with a value.

Interval – This style is used for certain types of internal Affymetrix Transcriptome Project graphs and is not recommended for general use.

Change visible graph bounds

Changing the visible bounds involves changing the scale of the graph while at the same time setting the maximum and minimum values to be displayed.

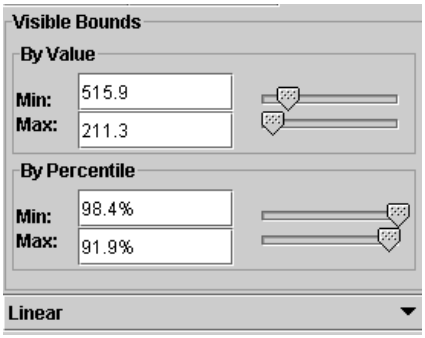
Scaling is an important issue to consider when viewing expression values because the distribution of these values across the range defined by their minimum and maximum may be highly asymmetrical.

For example, the majority of values could range between 100 and 1,000, with just one or two percent exceeding 100,000. These extremely high values (relative to the rest) would tend to dwarf the others if displayed to scale in a graph because the vertical axis would be forced to show a range of 1 to 100,000. Thus, the majority of the values in the lower range of 100 to 1,000 would appear be dwarfed or "squashed" by the minority values in the higher range. Although technically accurate, such a graph would not be very informative because most of the information – the relative differences in value between different base pair positions – would be impossible to discern visually. Thus, if you are using graphs to display numerical values, you should find out what the overall distribution pattern of your values looks like. If the distribution is asymmetrical (for example, not following a bell-shaped, normal curve) then you need to come up with good **Visible Bounds** settings.

Min/Max Visible Bounds scaling offers one solution to the asymmetrical distribution problem. In Min/Max scaling, minimum and maximum thresholds are chosen by some criteria, and these thresholds define the visual range of values displayed in the graph. Then, any expression value that is above the maximum threshold is occupies the maximum possible vertical space within the graph. Likewise, any expression value that is below the minimum threshold is shown as having a barely non-zero height.

Thus, any extreme values that are above or below the chosen thresholds appear the same visually. And the values in between are displayed across a range that allows their differences to be noted.

To set these Min/Max Visible Bounds, use the **Visible Bounds** section of the Graph Adjuster tabbed panel:



1. If it's on, toggle thresholding off before you adjust the graph bounds, then on again after.

2. Select the graph(s) to change (Shift-click to select more than one)

3. Click the Graph Adjuster tab.

4. Choose a way to adjust the visible graph bounds:

    - To make smaller changes to the graph view, choose **Adjust by Value**

- To make larger changes to the graph view, choose **Adjust by Percentile**

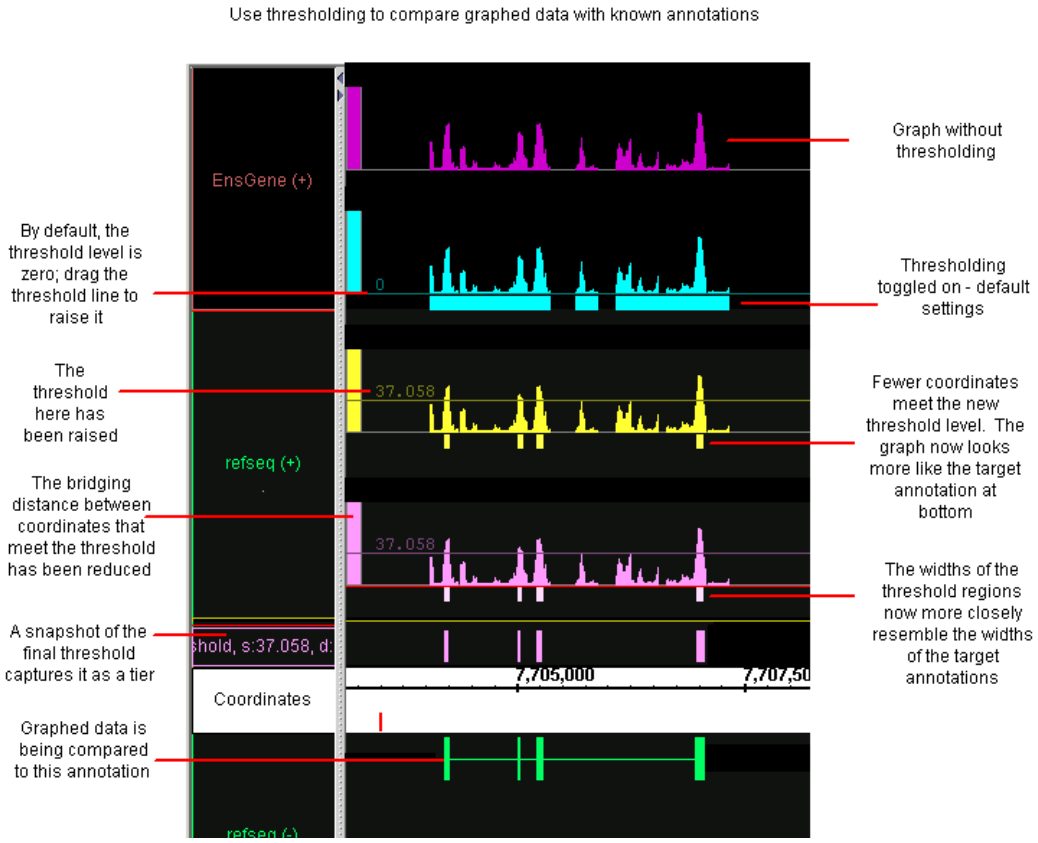
To undo an adjustment and return to the original graph boundaries:

Adjust by Percentage and reset the **Min** to 0 and **Max** to 100

View Thresholds

Graphs from genome tiling arrays represent hybridization activity, but only activity levels above a certain threshold should be deemed meaningful. To screen out non-meaningful values and to view meaningful features of the graph as annotations, adjust the graph's threshold values.

Thresholding displays graph features as an annotation-like bar when the threshold you define is met (i.e., when the level of activity at the coordinate meets your defined threshold.)



Adjust your graph's threshold so the bars correlate to existing annotations; any threshold bars visible at that point that don't have known annotation counterparts (or vice-versa) may indicate areas for further investigation.

Turn thresholding on

To turn thresholding on:

1. Select the graph to threshold
2. Select **Thresholding** > **Visibility > On**.

By arbitrary default, most graphed coordinate points meet the threshold, so it appears as a solid annotation bar wherever there is any graphed data. To obtain meaningful thresholds, adjust the threshold settings using any of the following methods:

Change the threshold level

To change the threshold so that more or fewer points meet the threshold limit:

1. Drag the horizontal line marking the threshold. Initially, it coincides with the lower graph bound.

Or:

1. Select the graph to threshold

2. Slide the **Min Score** sliders to adjust thresholding, or enter values into the boxes for each of the following parameters:

    - **Score** -- For expression results, this score represents the level of hybridization activity and is measured on the Y-axis. By default, all graphed coordinates meet the threshold. Enter a number greater than zero to raise the threshold to a meaningful level. For a hybridization activity level to be considered significant, the score must be above a certain threshold. Toggle the Y-axis on to see the scale on the Y-axis if needed.

    - **Max Gap --** Groups of coordinates that meet the threshold may be separated by gaps where the threshold is not met. By default, thresholding bridges such gaps if they are less than 100 bps long. For example, if there are 300 bps that meet the threshold, then 99 bps that don't, followed by 100 that do, the threshold bar would bridge the 99 bps to produce a single threshold annotation 499 bps long. To change the length of this bridge, enter a different value into the Max Gap text-entry field.

    - **Min run --** Min run is the minimum number of bases in a row that must meet the threshold before an annotation bar will appear. For example, by default the Min run is 30 base pairs, so a sequence of 29 base pairs that meet the threshold will not be marked by an annotation bar.

Limit the percent of false positives (%FP)

Graphs from the Transcriptome Project capture hybridization activity of the probes on the genome tiling arrays Affymetrix chip, including measurements from negative control probes from bacteria or another non-human genome

that should not hybridize with human samples. When these control probes do cross-hybridize, this is called a "false positive" reaction.

The percent false positive setting lets you use these negative control probes to set the graphs threshold. Whatever percentage is entered here causes IGB to compute a threshold that would allow this same percentage of negative control probes to be called positive. For instance, setting percent false positive to 5% sets the threshold to the negative control probe population's 95^{th} percentile, the value that marks the cutoff separating the top 5% and bottom 95% of the negative control probes.

Thus, the **Adjust threshold by % False Positive** parameter lets you specify an acceptable percentage of probes on the chip that light up that shouldn't.

Typical values are 1%, 3%, or 5% false positive. Decreasing the false-positive value improves the likelihood that values above the threshold are valid.

Capture the threshold bars

Capture the threshold bars (which look similar to annotations) in their own tier to more easily compare different threshold settings for a single graphed region. (See the thresholding screenshot, above)

To capture the threshold bars for use in this IGB session:

1. Click to select the graph
2. Click the **Make Tier** button.

The snapshot does not capture the graph ID; it does, however, give the score (the level at which the threshold is set, or the "s" value in the tier name) and distance (the length of the "bridge" joining the gaps where the threshold is not met, or the "d" value in the tier name).

To permanently capture the threshold snapshot for later use or sharing:

1. Right-click the tier and choose **Save tier as BED file**
2. Name your file. Be sure to include the .bed filename extension.

Common useful graph customizations

Some common useful graph customizations are:

Dampen the distorting effects of outliers – for example, ignore any points that are above the 99.5th percentile or below the .05th percentile.

Under **Visible Bounds,** set **By Percentile** to the desired lower and upper percentile bounds.

Change scales to match when comparing multiple graphs of different scales. For example, if one graph has a minimum value of 12 and a max value of 24, and another has a min value of 24 and a max value of 48, make them easier to compare by changing the min and max bounds so they have the same scale.

For each graph being compared, toggle the Y-axis on to view the min and max values for each

Adjust the visible graph bounds by Value to make them compatible

Display the optimal type of graph for a particular purpose, for example use a stairstep style graph for analyzing ESTs or performing other density analyses. Then slice the selection to see details.

Find areas of activity that do not fall in known exons.

Adjust your threshold bars to match the known annotations; any "extra" threshold bars visible in your graph indicate potential areas of interest.

Comparing two graphs

You can compare the difference between two graphs representing data from the same chip or otherwise having the identical genomic position on the chromosome. For example, a graph of normal cells will show what's turned on or off when compared to a graph of cancerous cells. This feature makes the differences easier to see.

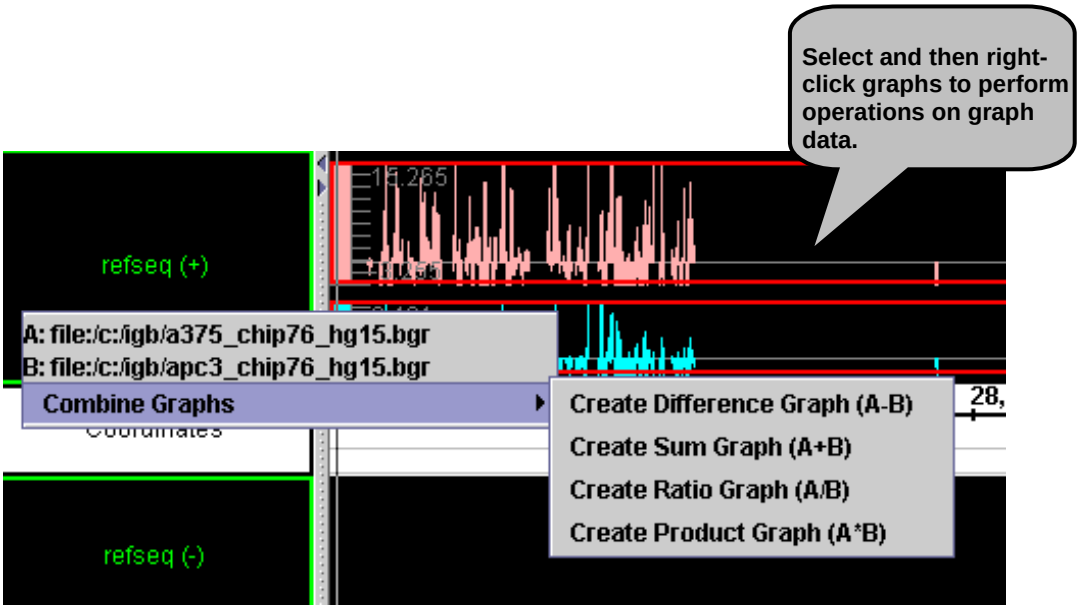
For this feature, Graph A is the first one clicked to select it, and Graph B is the second one clicked.

To compare the difference or ratio between two graphs:

1. Shift-click the graph bar of each graph to select them. The graph you select first is A, the second-clicked one is B.

2. Right-click either graph bar and choose **Create Difference Graph (A-B)** or **Create Ratio Graph (A/B)**

Results that show significantly more or less activity than a control graph for the same region may indicate an area of interest.



Getting details about the graph

Textual information about TDB graphs loaded into IGB is available in Selection Info.

Click the bar at the left of the graph (see graphs screenshot), then click the **Selection Info** tab. Resize the columns if needed.

Finding MVA images for selected graphs

MVA plots provide an assessment of the variability between probes in a pair of chips as a function of probe intensity.

For a given pair of chips with intensities (x_1,x_2,...,x_n) and (y_1,y_2,...,y_n) the values of M=log(x_i)-log(y_i) are plotted against A=0.5*(log(x_i)+log(y_i)). M is the log of the ratio of intensities and A is the log of the geometric mean of intensities. In the case of perfect reproducibility all the points would lie along the A-axis, and the extent to which points scatter around the axis gives a qualitative sense of the variability as a function of probe intensity. Note that a multiplicative difference between two chips would lead to the MVA plot being shifted along the M-axis, so any observed trend of nonzero slope is indicative of a non-linear relationship between the intensities on each chip.

To view the variance in the replications of the experiments generating the graph selected:

Click the **Find MVA images for selected Graphs** button at the bottom of the **Graph Control** tab panel.

Saving, exporting, and sharing graph views

Certain aspects of graph manipulation can be captured for future use by bookmarking temporarily for later use in the same session of IGB, or permanently for use in a later session or by another person.

Saving a graph as a file

Save a graph as a separate file for future reference or offline use. Note that this saves the generic graph, without any visualization customizations you may have made to it.

To save a graph to file:

1. Select the graph by clicking the graph handle.
2. Click the **Save Graph** button under **Options** in the Graph Adjuster panel.
3. Specify location and graph name, then add the **.gr** filename extension.
4. Click **OK**.

Clearing all graphs from the viewer

To clear all graphs from the viewing window:

Choose **File** menu > **Clear graphs.**

Pivot View – visualizing expression across many experiments

Introduction

Use the Pivot View tabbed panel to examine and compare the results from multiple gene expression experiments.

Using Pivot View is similar to using Graphs in that both show expression values in the context of genomic sequence. The difference is that in the Pivot View, the information on the x- and y-axes in the main view is flipped and individual sequence segments are separated into individual rows.

With Graphs, expression values from a single experiment appear as a graphs laid out across the horizontal sequence axis. In Pivot View, columns represent individual experiments, while rows represent individual regions.

Turning Pivot View on and off

Using Pivot View is optional. To run IGB without it, edit the "USE_PIVOT_VIEW" option in the IGB preferences file as described in Appendix 3 Preferences. If the pivot view option is set to "true," then the Pivot View will appear in a separate tab at the bottom of the IBG display.

Expressed intervals and .map files

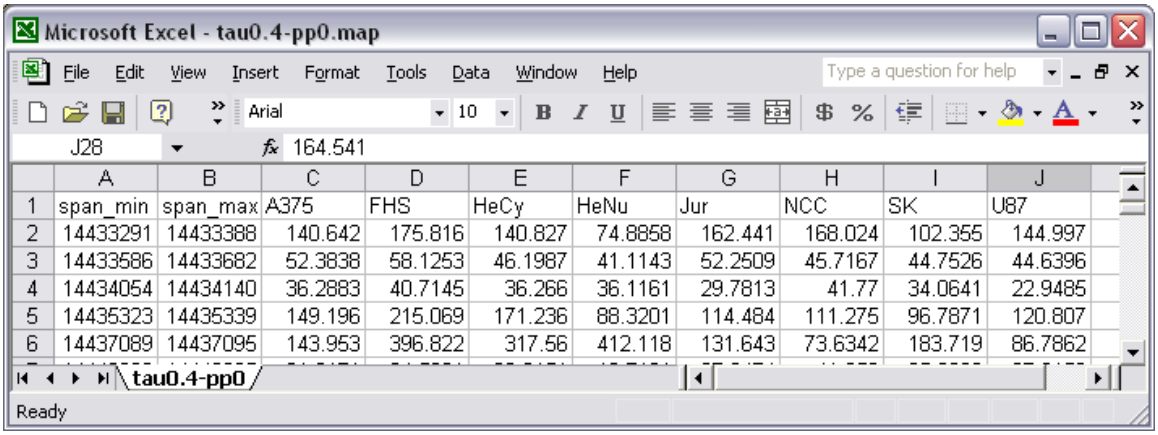
Pivot View displays data from expression map (file name suffix: .map) files, tab-delimited plain text files containing summarized expression data for expressed intervals.

Expressed intervals are segments of genomic sequence that a data analysis has determined are expressed together. For example, a data analysis algorithm might determine that an exon whose boundaries never vary is an expressed interval. Conversely, if an exon coincides with two expressed intervals, then it is possible that the exon may be alternatively spliced, depending on the tissue type or developmental state. Another way to look at expressed intervals is to regard them merely as contiguous runs of bases whose expression values can be intelligently summarized using a single value.

The image below shows an example IGB expression map file that was opened using Excel. Note that the first row of data contains the column headings and that the values are separated by tabs.

The first two columns give start and end positions for the expressed intervals described in the file. The columns after that list expression values for expressed intervals in eight different samples, including cytoplasmic and nuclear fractions from HeLa cells and samples from other cell lines, as well.

Note that as of this writing, map files do not contain information regarding which chromosome or genome release the expressed intervals come from. Future versions may contain this information, however. As with graphs, it is important to match up the map files with the appropriate segment of the genome being viewed.



	A	B	C	D	E	F	G	H	I	J
1	span_min	span_max	A375	FHS	HeCy	HeNu	Jur	NCC	SK	U87
2	14433291	14433388	140.642	175.816	140.827	74.8858	162.441	168.024	102.355	144.997
3	14433586	14433682	52.3838	58.1253	46.1987	41.1143	52.2509	45.7167	44.7526	44.6396
4	14434054	14434140	36.2883	40.7145	36.266	36.1161	29.7813	41.77	34.0641	22.9485
5	14435323	14435339	149.196	215.069	171.236	88.3201	114.484	111.275	96.7871	120.807
6	14437089	14437095	143.953	396.822	317.56	412.118	131.643	73.6342	183.719	86.7862

Loading .map files

Loading map files is similar to how other annotation files are loaded.

1. Make sure the genome version and chromosome of files you want to open match those of the data already displayed.

2. Choose **File** menu > **Merge from File**

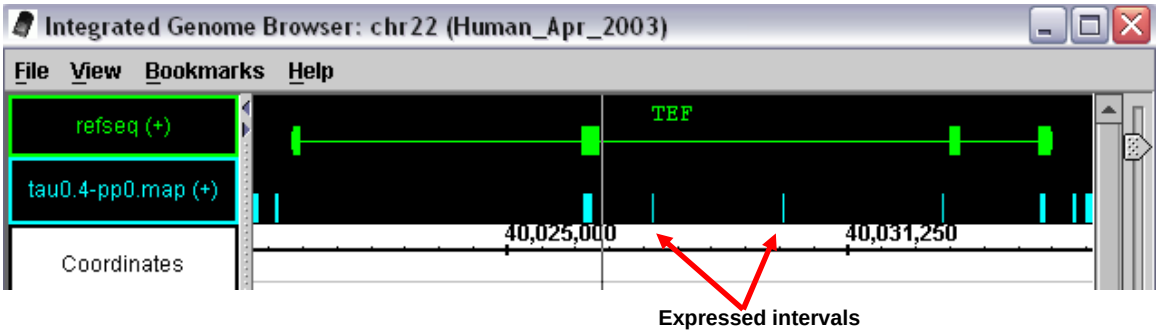
3. Select a .map file.

The data from this merged file will automatically be added to an additional tier alongside the tiers already in the viewer.

Map data in the main window

When a map file is first loaded, the data appear in a new tier named after the data file. The image below shows a new tier that appeared after loading a data file called **tau0.4-pp0.map**.

Expressed intervals in main display window.



In the picture above, the blue blocks above the genomic sequence axis represent expressed intervals occurring at the indicated positions. These purely positional data correspond to the interval information that was stored in the first two columns of the tabular expression map file shown above.

Note that in this case, some of these expressed intervals overlap with exons in the RefSeq mRNA transcript annotation shown in the tier directly above the expression interval tier. However, there are also some smaller expressed intervals which do not overlap with any RefSeq annotations. In fact, they appear to overlap with intronic regions which should not be detected on a GeneChip because they are removed from the transcript during splicing. However, it is possible that these intervals correspond with genes and RNAs that have not been entered into the RefSeq database. If so, these unreported genes or transcripts could be transcribed from either the plus (top) or minus (bottom) strands. In this case, at least, the experimental protocol used to generate the expression map data cannot distinguish strand information.

Thus the expression map data format does not include strand information. By default, expression intervals are always shown as features on the plus (top) strand of the genomic sequence.

Map data in the Pivot View

To view expression data associated with the expressed intervals, select them in the main window: Anything that is selected (outlined in red) in the main window is automatically shown in the Pivot View.

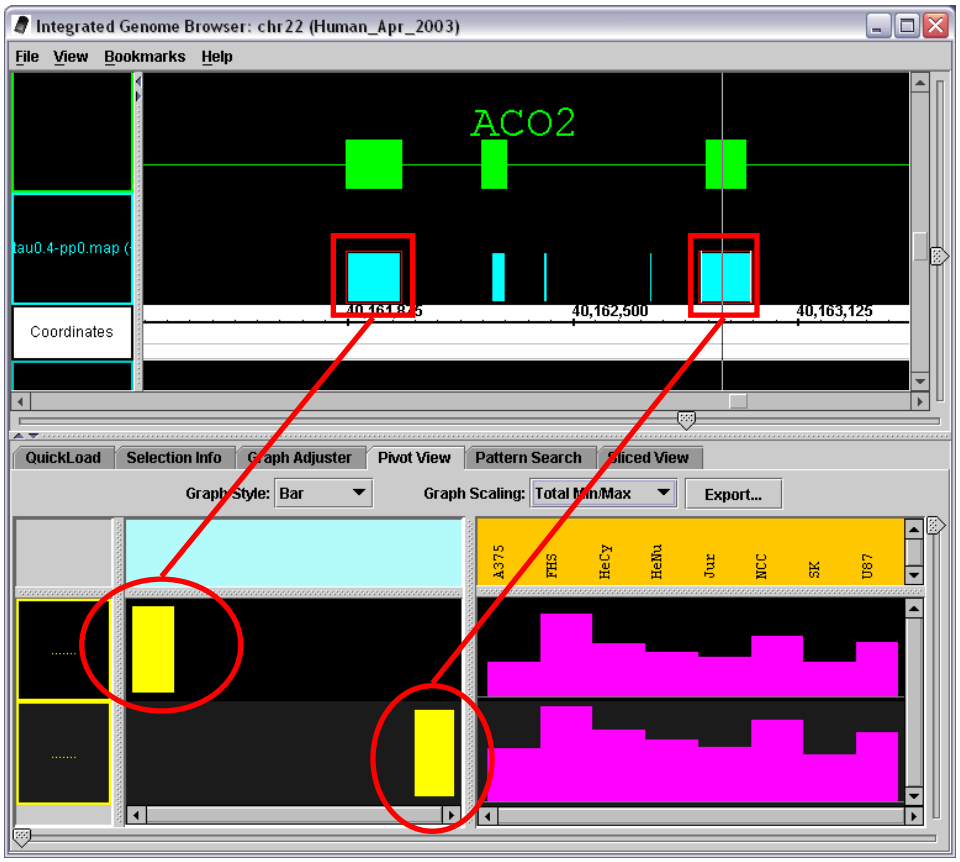
To select a region of expressed intervals for display in Pivot View, click-drag over the expressed interval blocks while pressing the **SHIFT** key. To add more expressed intervals to the selection, click or click-drag over additional blocks while pressing the **SHIFT** key.

It is also possible to **SHIFT**-click annotations to add them to the Pivot View display.

The image below shows a sample Pivot View display that was built by selecting two expressed intervals from the Aco2 region. (Aco2 encodes

aconitase, a mitochondrial protein that catalyzes the second step of the TCA cycle.)

Pivot View showing expression values for two expressed intervals



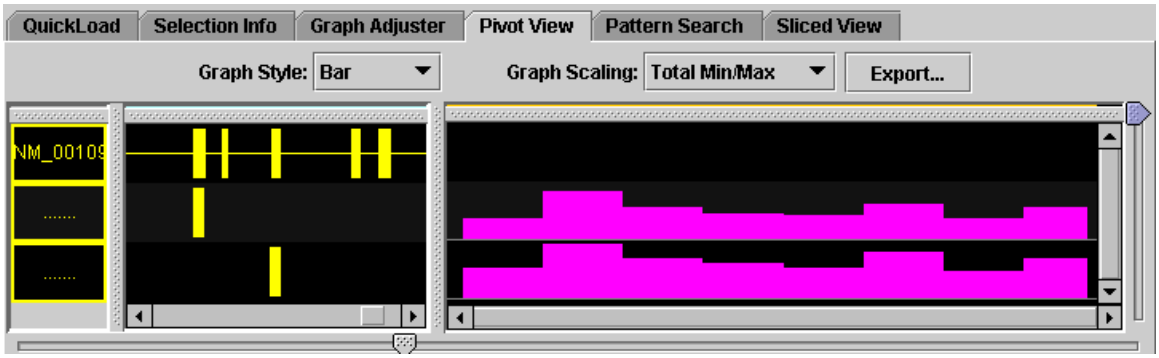
In this figure, the two expressed intervals that were selected are outlined in red in the main IGB display. They also appear as yellow blocks in the left-hand portion of the Pivot View tabbed panel. Note that individual expressed interval blocks occupy separate rows, and their positions within the row indicate where they lie with respect to the genomic sequence axis. Thus, the leftmost block represents the 5'-most interval, while the rightmost block represents the 3'-most interval.

Expression patterns for each interval appear in the expression values panel occupying the right side of the Pivot View. This panel shows the expression values for each expression interval (rows) measured across several experiments (columns).

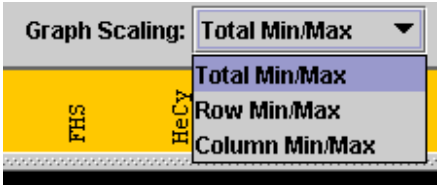
In this example, the expression panel shows expression values for as a bar graph that represents expression for an individual expression interval across a panel of experiments. Within each graph, the height of individual bars represents expression intensity. Taller bars indicate higher levels, while shorter bars indicate lower levels of expression.

What should be clear from this very simple example is that the overall pattern of expression for both blocks is similar. That is, if you compare the height of bars in adjacent cells for both blocks, you notice that when the expression for one block goes up from experiment to experiment, the expression for the other block goes up as well. Similarly, if expression goes down from experiment to experiment for one block, the expression of the other blocks also decreases. This makes perfect sense when you consider that the two blocks probably represent nearby exons in the same mRNA transcript. Thus, it is very likely that these blocks measure aconitase expression.

To compare the blocks shown in Pivot View with RefSeq annotations in the main view, just add the annotation of interest to the currently selected items by clicking the item while holding the SHIFT key. The following figure shows how this looks for our aconitase example.



Expression panel graph scaling



The **Graph Scaling** menu at the top of the Pivot View tab contains settings that control the vertical scaling and range of values displayed in the expression panel.

For example, if a row occupying 1 cm of vertical space has an expression value range of 0 to 1000, then a bar that is 0.10 cm high indicates an expression value of 1000 expression units / cm x 0.10 cm = 10 expression units.

The **Graph Scaling** menu contains three options:

Total Min/Max

Row Min/Max

Column Min/Max

Currently only the first two of these are implemented. The default setting is **Total Min/Max** for newly displayed graphs, but this default behavior may at some point become settable in the IGB preferences file.

In **Total Min/Max** scaling, each row shows the same range of values. The top end of this range is set as the largest expression value across all the rows, while the bottom end of this range is set by the minimum value across all the rows.

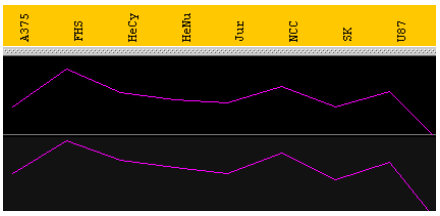
In **Row Min/Max** scaling, scaling is done on a per-row basis. That is, the maximum and minimum values within a row determine the relationship between vertical space and expression values. Use **Row Min/Max** scaling when one row has unusually high values.

Changing the graph style

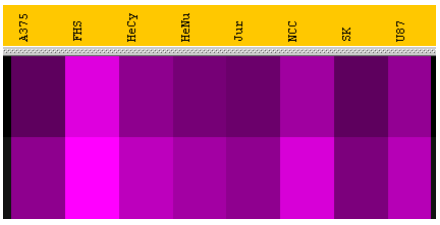
Using the **Graph Style** menu, it is possible to switch between three different graph presentation styles in the Pivot View expression panel. The three options currently available include:

Bar graph

Line Graph



Heat map



The type of graph that is most appropriate depends on the type of question being asked about the data. For example, when comparing trends, it is very useful to use the line graph display method because this method makes it

easier to compare trends and patterns. The number of expression intervals being shown also can affect the graph display choice. The user is encouraged to experiment with the different display types to find out which method works best for specific purposes.

Display Tips

When viewing data from a large number is of expression experiments and map files, it helps to put the Pivot View panel into its own separate window on the desktop. This is easily done by choosing **View** menu > **Open Tab in New Window**.

Visualizing probe sets

Introduction

Chip design procedures have varied somewhat from chip to chip, but, in general, probes on Affymetrix' standard commercial expression microarrays (such as the Human U133 chip) are designed to match the 3' regions of known or computed mRNA sequences. These probes are typically 18 to 25 bases long depending on the chip.

In standard arrays, probes are grouped conceptually into probe sets, groups of probes that are expected to measure expression for individual known or computationally-deduced mRNA molecules.

Probe sequences are typically selected from the 3' regions of these design sequences in order to maximize the amount of sample mRNA to be measured. (This so-called "3-prime bias" will most likely change as sample preparation protocols improve over time.)

These design sequences may be identical to known mRNA sequences in GenBank, or they may have been produced computationally by merging ESTs or mRNA sequences into a single sequence, sometimes called a "consensus" sequence.

IGB can be used to visualize the location of design sequences and probes within the genomic sequence. Being able to see where these sequences are located within the genome can be extremely useful when several overlapping probe sets recognize diverse mRNAs originating from the same gene. Around 60% of human genes produce multiple variants, and these diverse variants often exhibit very complex configurations of exons and introns. As a result, it often helps to be able to view a diagram of these complex structures together with probes in order to determine which individual mRNAs are being detected by a given probe set.

To make this easier, IGB shows where the design sequences and their probes align to the genome together with alignments between the genomic sequence and known or predicted mRNAs. By showing probes, known mRNAs, and design sequences in the same view, IGB makes it possible to determine quickly which known mRNA a probe set could detect.

However, it is important to note that these genome alignments can only be as good as the genomic sequence on which they are based. Genomes vary in their level of quality – for example, the human and fruit fly genomes are very high quality, having gone through several releases involving many refinements. More recently-sequenced genomes (such as mouse or rat) are not yet as reliable. For the less-refined genomes, there may be many examples of design or mRNA sequences that simply do not align anywhere to the genomic sequence or, if they do align, do so imperfectly. Readers interested in understanding more about these issues can read more about this

topic at the following link:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=12149135/.

Invoking IGB via links from NetAffx

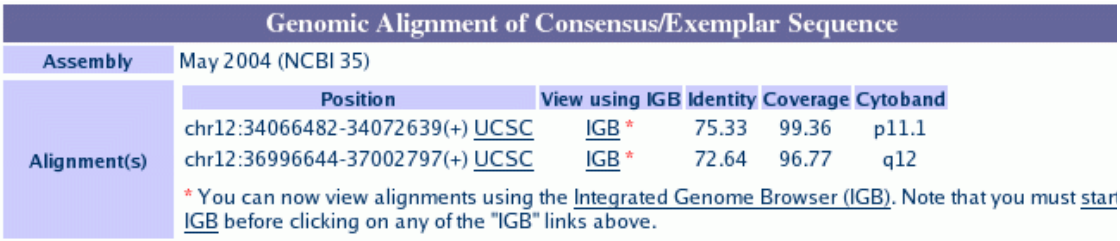
Probe set pages on the NetAffx customer support Web site contain links to the IGB browser under the “Genomic Alignments” section of the page. Clicking these links (labeled “IGB” as shown below) will tell IGB to show the region of the genome containing the probe set and its design sequence.

IGB must currently be running for these links to work. If IGB is not currently running, start it in your usual way, or by clicking the “start IGB” link provided for that purpose. If IGB is already running, then the current scene will be replaced with a view of the probe set design sequence alignment region.

Depending on the probe set, there may be more than one reasonably good alignment for the design sequence. Most genomes contain large amounts of duplicated sequence, and so it is to be expected that many design sequences will align to multiple locations.

For example, the figure below shows links to two different alignments for a single probe set and its corresponding design sequence.

IGB links in probe set page.



Genomic Alignment of Consensus/Exemplar Sequence					
Assembly	May 2004 (NCBI 35)				
Alignment(s)	Position	View using IGB	Identity	Coverage	Cytoband
	chr12:34066482-34072639(+) UCSC	IGB *	75.33	99.36	p11.1
	chr12:36996644-37002797(+) UCSC	IGB *	72.64	96.77	q12
	* You can now view alignments using the Integrated Genome Browser (IGB). Note that you must start IGB before clicking on any of the "IGB" links above.				

Visualizing probe sets: loading from data files.

Sometimes it is useful to load up an entire chip’s design sequence alignments into IGB. To do this, download the files from the Affymetrix customer support Web site.

These can be obtained on-line at the NetAffx customer support Web site. To obtain these files for individual chips, go to http://www.affymetrix.com/support and choose the array of interest. On the array’s support page, look for alignment data files in the "NetAffx Annotation Files" section. Note that alignment data files will only be available for species with well-characterized genomes.

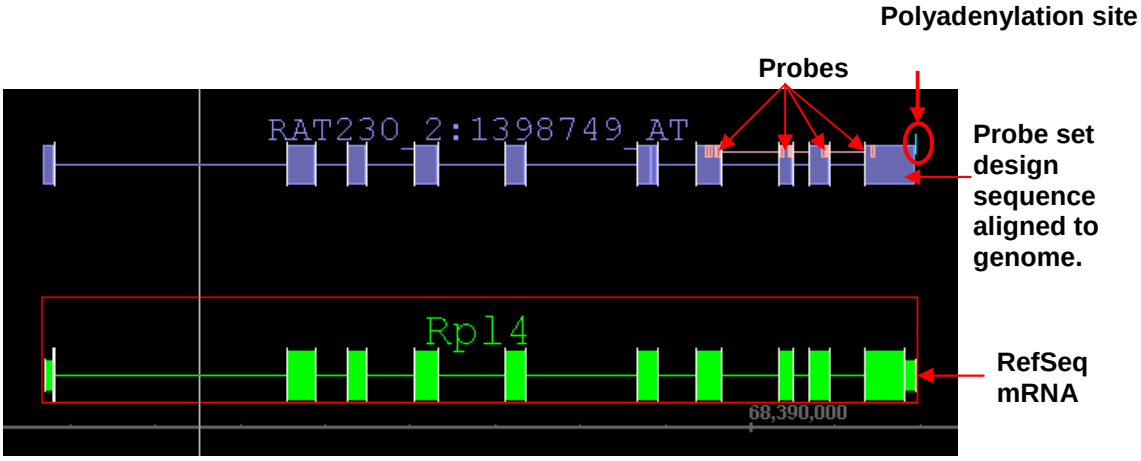
These alignment files are in a special format called “PSL” (pslayout) which was first designed by Jim Kent of U.C. Santa Cruz. For more details on this format, visit the U.C.S.C. Genome Bioinformatics Web site at http://genome.ucsc.edu.

To visualize all the probe sets and associated design sequences for a particular chip, load the genome of interest and merge a probe set data file, see Viewing multiple files in a single instance of IGB.

Probes and probe sets on display

Once the data have been loaded, the probes and design sequence alignments will appear in a separate tier labeled by chip. Within the tier, probes and their matching design sequences always appear together. The following figure shows an example from the rat 230 chip.

This image shows a probe set that detects mRNA transcripts from the rat Rp14 gene, which encodes a ribosomal protein.

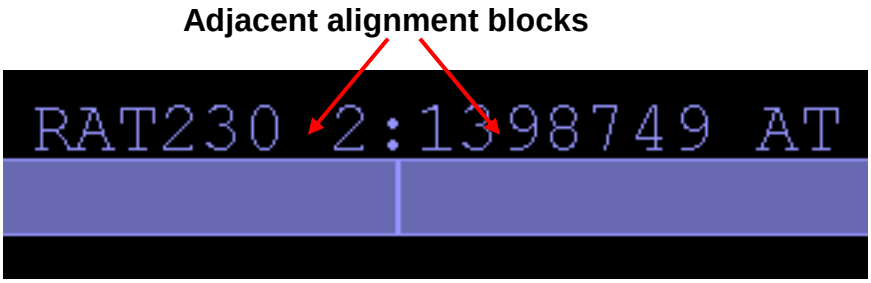


Probe set design sequence.

The alignment between the probe set design sequence and the genome is represented at the top of the figure as a series of blocks. Each block represents a block of alignment in which each base in the genome matches a corresponding base in the design sequence. Gaps between the blocks typically represent areas where the genomic sequence contains inserts relative to the aligned design sequence. Usually, these gaps are due to introns.

There are some exceptions to this, however. For example, the sixth and seventh blocks in the figure above are so close together that they almost appear as a single block at this level of zoom.

Zooming in for a closer view reveals that these two alignment blocks are immediately adjacent to each other. This indicates that these two blocks of alignment were separated by an insert in the design sequence relative to the genomic sequence. That is, the design sequence contained some bases that were not present in the genomic sequence. This may present a problem if this missing region (in the genome, that is) contains some probe sequences. In this particular case, however, the alignment irregularity occurs in a 5' region, outside the area covered by the probes. (See the discussion below.)



How can this happen? There are a number of reasons, but gaps in the genomic sequence – regions that simply have not yet been sequenced – are the most common culprit.

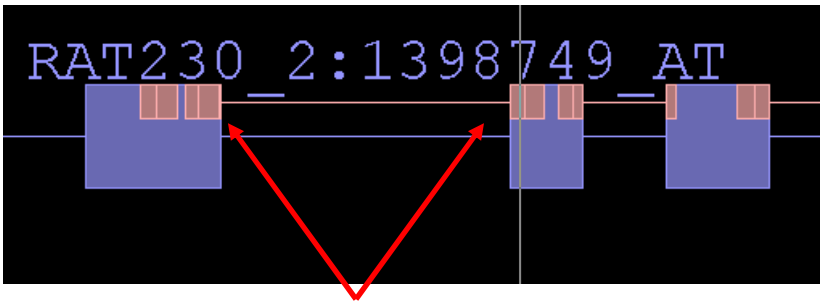
Probes and probe sets.

Each design sequence is shown with its corresponding probe set. Each probe set consists of a group of probes, which are shown superimposed on the alignment blocks of the design sequence.

The figure below shows a close-up view near the 3' end of the design sequence. The 3' end of the design sequence is annotated with blocks which represent individual probes.

Two things are important to notice about this image. First, sometimes individual probes are split across gaps in the alignment, which typically correspond to introns. When this occurs, the two halves of the probe are connected by a line.

Second, sometimes probes overlap with each other. This can be seen by clicking on probes. If a selection outline is visible in the middle of what otherwise looks like a single block, then there is really more than one block.



Probe split across genomic gap (intron)

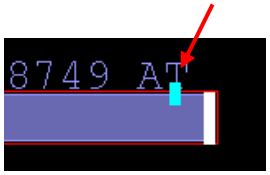
Probe set labels

Each probe set and probe set design sequence is labeled with the name of the chip (Rat 230, in this case) and the probe set identifier.

Polyadenylation sites

For many probe set design sequences, Affymetrix bioinformatics scientists have used computational methods to predict putative polyadenylation sites near the 3' end of the design sequence. These are shown as dark or light blue boxes riding on top of the terminal alignment span.

Predicted
polyA addition site



If the box is light blue, then the site was deduced from the overlap of multiple expressed sequences (usually 3' ESTs) whose genome alignments all terminate at a common location. This is sometimes called a "polyA stack."

If the box is dark blue, then the site has been deduced through sequence analysis of an individual, exemplar sequence, such an mRNA sequence record from GenBank. This is sometimes called a polyA site to distinguish it from a polyA stack.

mRNAs in the region – putative probe set targets

It is usually a good idea to load additional tiers of data besides just the probe set information. For example, in the figure above, the tier below the probe set tier shows the probe set's likely target, an mRNA from the Rp14 locus. Based on their relative alignments to the genome sequence, it appears that both sequences overlap in the region that contains the probes. Thus, it is very reasonable to assume that this probe set does indeed detect the mRNA shown below it.

Getting more information about a probe set

Right-clicking a probe set or its design sequence and selecting the **Get More Info** option will open a Web browser window showing detailed information about that probe set, including its design sequence.

About alignments

Probe mappings

A program called "blat" was used to map probe set target sequences onto various versions of various genomes. More information can be obtained about this program from the U.C. Santa Cruz Genome Web site, which also uses the blat program to map mRNA sequence onto genomic sequence. The link to this Web site is: http://genome.ucsc.edu/ .

The location of the probes within the probe set design sequence is already known from the design process. This information, in combination with information about the location of the probe set design sequences within the larger genomic sequence, was used to map probes onto the genome.

Many chip probe names use special suffixes to indicate when a probe set may detect multiple distinct target mRNAs. For details, consult the data sheet for your array of interest or contact Affymetrix technical support. (See: http://www.affymetrix.com/site/contact/index.affx for contact information.)

Viewing tips

Use the spliced view display to examine probe sets, probe set design sequences, and mRNA transcript annotations. Introns in mammalian genomes tend to be very large. The "sliced" view takes away the uninformative intronic sequence, allowing one to take in the entire structure of gene without having to scroll or zoom.

Capturing, saving, and sharing IGB data

There are several ways to save data for use in other applications, sharing, or later use in IGB.

Save or share data you from IGB in any of several ways:

- Saving graph files
- Making and exporting bookmarks
- Creating and saving curations
- Saving graph threshold snapshots, then saving them as curations
- Exporting primer details
- Copying and pasting to another location (sequences or data in IGB tables)
- Printing

Bookmarking

To return later to a region or annotation and zoom level of interest, bookmark it. These will appear under the Bookmarks menu below a few additional menu items for manipulating and managing bookmarks.

Any bookmarks you create during a session will be saved when you exit. (If you run multiple copies of IGB at the same time, the bookmarks set in one of the instances will overwrite the ones set in the other instance. To avoid this problem, you can manually export the bookmarks to a file.)

To export bookmarks to a file, choose **Bookmarks** menu > **Export Bookmarks**.

If you would like to edit these bookmarks and inspect their properties, be sure also to set the "USE_BOOKMARK_MANAGER" option in your IGB preferences file to "true." Setting this option to true tells IGB to display the **Manage Bookmarks** tabbed panel.

Bookmark types

There are two types of bookmarks in IGB:

1. Position bookmarks.

   Position bookmarks record the currently shown genome, the location, and the current zoom level.

   To create, choose **Bookmarks** menu > **Add position Bookmark**

2. Position-and-graphs bookmarks.

   Position-and-graph bookmarks record the same things as position bookmarks, but also keep track of any graphs that have been loaded thus far. In addition to remembering the data files that have been loaded, position-and-graphs bookmarks can also remember many of your customizations for graph appearance, such as color, graph bound adjustments, thresholding, and labeling.

   To create, choose **Bookmarks** menu > **Add position & graphs bookmark**

Viewing bookmark details: Manage Bookmarks

To view details about bookmarks, such as their position, which genome they refer to, or viewing customizations for position-and-graphs bookmarks, select the **Manage Bookmarks** tabbed panel. Select a bookmark and then click the "Properties" button to get more information about the bookmark. The information will appear in tabular format in a new window. If you would like to make changes to the bookmark, make them here.

Exporting bookmarks

To save a copy of the bookmarks to a file, choose **Bookmarks** menu > **Export Bookmarks**

Bookmarks are saved in a single file in the Netscape bookmarks file format. This format resembles XML and HTML, but is not identical to either of those, so use caution if you choose to edit the file.

Importing bookmarks

You can import and use bookmarks from other files besides the bookmark file specified in your IGB preferences. To import bookmarks from a different file, such as one you receive from another IGB user:

Choose **Bookmarks** menu > **Import Bookmarks** > choose the file containing the bookmark of interest.

Multiple sets of bookmarks can be imported during the same IGB session. Imported bookmarks will be added to the end of the list of existing bookmarks. All imported bookmarks will be persistent between sessions of IGB.

Controlling IGB through a browser, using bookmarks

If you open a bookmarks file in a Web browser and then click one of the links *while IGB is running*, then IGB will respond by changing the display to show the selected bookmark.

This will only work for the first instance of IGB that you are running. For example, if you are running two copies of IGB, then the browser will only affect the IGB window that you started first.

Technical details on how this works.

If you open a bookmarks file in a text editor, you will find that the URL for each bookmark looks something like **http://localhost:7085/UnibrowControl?seqid=chr21&end=44542686&version=Human_May_2004&start=44534188**.

In this case, "localhost" simply refers to your local machine and "localhost:7085" refers to port 7085 on your local machine. IGB, by default, "listens" to port 7085 on whatever machine it happens to be running. So if a Web browser sends a request to "localhost:7085," IGB will hear it and respond. If the URL contains the right parameters (version, seqid, start, end, etc.) then IGB responds by showing the genomic "scene" described in the bookmark.

However, if you start up a new instance of IGB while one is already running, the new instance will listen to a different port – the first port above 7085 that happens to be vacant. This is why the bookmark links will only work for the first IGB instance and not for any others you start running after that.

Curation

Create curations to represent and store your own custom annotations based on your experimental results or modifications or interpretations of annotation(s) from other sources. Save curations for later reference or to share with others.

Creating new curations

You can create curations in two different ways, with different results.

Making your initial curation during a IGB session

Depending on how you plan to use a curation, you can create a curation using one of two options.

To create an initial curation:

1. Choose an annotation or sequence region to include in your curation.

2. Right mouse click on your selection > **Curation** > choose one of the following:

   - If you want to give your curation a unique name for easier identification later, or if you plan to save the curation as a distinct entity separate from all other curations you create, choose **Make new curation type**. When prompted, enter a descriptive name for this curation.
   - Otherwise, choose **Make new curation**, which automatically assigns the generic name "Curation". (Each time you launch IGB and "make new curation" the first time, it will have the same generic name: Curation. The curation is also saved to a generic name. When you use saved curations, if they are saved to DAS, all curations you have ever created with **Make new curation** will appear in the same tier. If they are saved to BED and you load multiple curation files into IGB

at the same time to compare them, they will all load into the same tier. If this behavior is not what you want, use **Make new curation type** instead.)

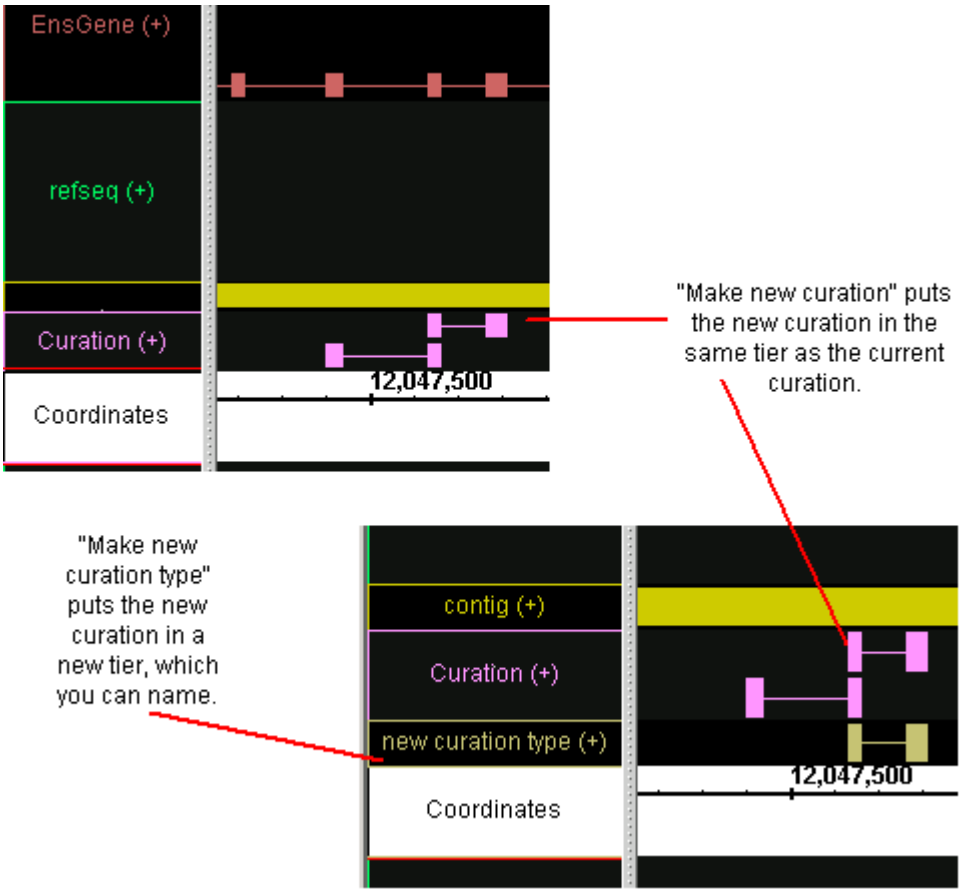
The curated annotation will appear in a new tier. If you can't see the new tier, use the vertical slider and scroll box to expand the tier and make it visible.

Currently, curations do not distinguish translated from untranslated regions.

You can now modify this curated annotation, create another new curation either in the same tier or in a different tier, or save this curation and share it.

Making discrete new curations

If you have created one curation and want to create a second one, you can create the second curation in the existing curation tier or in its own tier.



Making a discrete new curation within the same tier as another curation

To create a discrete new curation within the same tier as another curation:

1. If more than one curation tier is currently displayed, Right mouse click a curation in the tier you want to put the new curation into > **Curation** > **Set as current curation**

2. Right mouse click on the annotation > **Curation** > **Make new curation**

Making a discrete new curation in a separate tier

To create a discrete new curation in a separate tier:

1. If more than one curation tier is currently displayed, Right mouse click a curation in the tier you want to put the new curation into > **Curation** > **Set as current curation**

2. Right mouse click on the annotation to put in the new curation tier > **Curation** > **Make new curation type**

3. Enter a name for your curation when prompted.

Editing curations

You can add or subtract annotations, undo changes, redo changes, and more.

Working with multiple curations

If you have more than one curation (whether in one tier or in more than one tier) in the viewer:

Before performing an operation, specify which curation to perform the operation on. For example, if you have two curations in a tier, or two curation tiers, and you want to delete an exon from one of the curations, specify the one to delete the exon from.

To specify the curation to operate on:

Right mouse click on the curation you want to operate on > **Curation** > **Set as current curation**.

Add to the current curation

You can add an annotation or a range of coordinates.

Add an annotation

To add an annotation to a curation:

1. If you are working with more than one curation, Right mouse click the curation you want to add the new annotation to > **Curation** > **Set as current curation**

2. Right-mouse click on the annotation to add to the curation > **Curation** > **Add to current curation**

Add a range of coordinates

If you wish to add a precise range of coordinates that is longer than the maximum width of the screen, see the next procedure.

To add a range of coordinates to a curation:

1. Make sure sequence residues are loaded

2. If you're working with more than one curation, Right mouse click the curation you want to add the new annotation to > **Curation** > **Set as current curation**

3. Zoom in all the way if precision is needed

4. Rubberband to select the range to add

5. Right-mouse click on the red box outlined in the Coordinates tier > **Curation** > **Add to current curation**

To add a precise range of coordinates that is longer than the maximum width of the screen:

1. Zoom out

2. Rubberband and add an approximation of the desired range

3. Zoom in at each end respectively and delete the bases you didn't want to include.

Delete an exon or sequence from a curation

To delete an exon or sequence from a curation:

1. Right-mouse click on the curation to be deleted from > **Curation** > **Set as current curation**

2. If you want to delete a range of coordinates, rubberband the range.

3. Right mouse click on the exon or rubberbanded range to delete > **Curation** > **Delete from current curation**.

4. If you have difficulty deleting an annotation, try rubberbanding the region around it and deleting the rubberbanded region instead.

Undo or redo additions or deletions to curations

Undo/ Redo the "previous edit" applies to whichever curation you made the last change to, not necessarily the curation set as current. Therefore it's best to work with one curation at a time.

To undo or redo changes to curations:

Right mouse click on any item in the viewer > **Curation**> **Undo previous edit** or **Redo previous edit.**

You can successively undo or redo a nearly infinite number of edits; IGB remembers your every move!

Flip a curation onto the opposite strand (forward/ reverse)

To flip a forward-strand curation onto the reverse strand or vice-versa:

1. If you are displaying more than one curation, Right mouse click on the curation to flip > **Curation** > **Set as current curation**

2. Right mouse click to select it > **Curation** > **Flip current curation**

Fuse separate curations

To join separate curations into a single curation:

1. Set one curation to the current curation
2. Add the second one to it.
3. Delete the leftover copy of the curation you added.

Separate a single curation into pieces

The best way to separate a single curation into pieces is to make two copies of the curation, and then delete exons from each copy until you have the two different curations you seek.

To do this:

1. Right mouse click the curation to be separated > **Curation** > **Make a new curation**
2. Delete exons as needed from the new curation until you arrive at your first intended piece
3. Perform step 1 a second time, so that you have another new copy of the original curation
4. Delete exons as needed from the second new curation until you arrive at the second of your intended two pieces.

Saving curations

There are two options for saving curations. One method is a convenient way of saving genomic annotations, but is possible only if you are saving HG12 curations (the June 2002 version of the genome) for your own use. To save curations from any genome version or to share your curations with others, use the other method.

The Save option appears only when you Right-mouse click on a Curation tier.

Save HG12 curations to DAS

Saving to your own personal DAS server is a convenient way to save genomic annotations. For information about DAS servers, see the section about Data from a DAS server.

To save HG12 curations to your own personal DAS server:

1. Launch the **stashserver.bat** file.

   See Launching associated files for information.
2. In IGB, Right mouse click on the curation tier you want to save > **Save** > **Save to DAS**

This method saves the curation to your personal DAS server in your IGB folder.

All curations created with **Make new curation** are saved as "Curation_saved"; curations created using **Make new curation type** are saved under the name you assigned the curation tier at that time.

You will receive no feedback in IGB confirming or rejecting the save.

Save curations to a separate BED file

Use this method to save annotations from any assembly.

This method saves your curation as a .bed file, a standard genomics file format.

All curations in the tier are saved together; if you wish to save a single curation, create a separate tier with just that curation. Each saved curation tier will be a separate file.

To save a curation tier to a .bed file:

1. Right mouse click on the curation tier > **Save to BED File**
2. Name your file. Be sure to include the ".bed" filename extension.

Viewing saved curations

You can access curations you or another person have previously created and saved:

Access curations saved to DAS

If you have saved curations to your personal DAS server, to access them:

1. Launch **stashserver.bat**. Leave it running and return to IGB.

   See Launching associated files for information about this file.
2. In IGB, follow the instructions for loading DAS features; choose **Personal** as the DAS server to load.

   "Curation_saved" (which contains all curations created with "make new curation") and/or any uniquely-named curations you created with "Make new curation type" will be listed.

   If the list of available curations in the Load DAS Features dialog box doesn't include curations you expect to see, refresh the list: choose a different DAS server, let it establish access, then choose **Personal** again.
3. Select the curation type to load and click **OK**.

   All curations you have ever created with "**Make new curation**" will appear in the **Curation_saved** tier.

Access curations saved as BED files

If you or another person has saved curations as BED files:

See opening a file, opening multiple files.

If you merge multiple bed files during a single IGB session that were created using "**Make new curation**" (as opposed to **Make new curation type**), all curations will appear in the same tier, **Saved Curation**.

To zoom in on the annotation, toggle Shrink Wrapping on.

Capturing a sequence of base codes to clipboard

You can copy a sequence of bases and paste the sequence into another application for use there.

The only way to capture a sequence of bases is:

1. Make sure sequence residues are loaded.
2. Rubberband to select the desired coordinate range

   This is the only way to capture a sequence; you cannot select and capture an annotation.
3. Choose **View** menu > **Copy last RubberBand to clipboard**
4. Paste the text string into the target application or location.

Capturing data in table cells

You can copy and paste table cells from the Selection Info tab (or any other tabular format tab) and paste them into Excel or some other program.

To capture data in table cells:

Click and drag to select the cells you want.

Printing

To print the entire IGB window including the tab section:

Choose **File** menu > **Print Whole Frame**

To print the contents of the viewer window only (not including the tab section):

Choose **File** menu > **Print**

Other tools

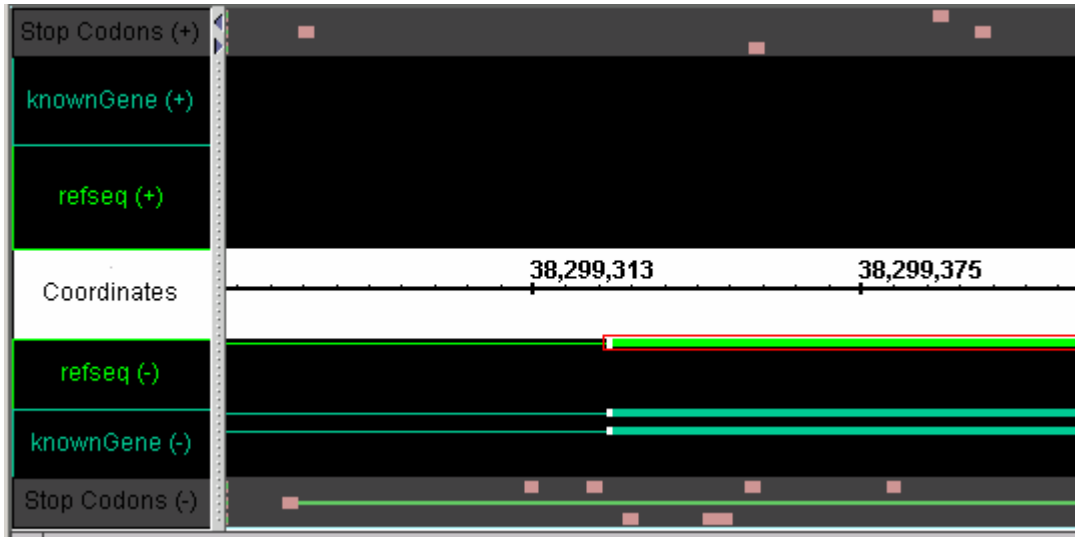
Showing ORFs and stop codons

Use ORFs to find coding regions; the longer the ORF is, the more likely it is to contain live coding regions.

To show ORFs and stop codons:

1. Make sure sequence residues are loaded.
2. Select an item of interest in the main view and click the **Sliced View** tab.
3. Adjust the minimum length of ORFs you want to see, if necessary. By default, only ORFs 300 bps or longer are shown.
4. Click the **Analyze ORFs** checkbox

   Three rows of ORFs in each direction appear, one for each start position. Stop codons are pink. ORFs of the specified length or longer appear as a green line:



   If too many green lines (ORFs) appear, increase the minimum ORF length (see step 4 above). If too few appear, decrease the length.

Clearing ORFs from the viewer

To clear ORF and stop codon markers from the viewer window:

In the **Sliced View** tab, deselect the **Analyze ORFs** checkbox.

Identifying a translation frame

To identify a translation frame:

1. Load sequence residues for the region you're viewing
2. Slice the region
3. Adjust the slicing to a buffer of 0 bases to completely remove the introns
4. Analyze the ORFs as above.
5. Look for the long, continuous green line that corresponds to your strung-together exon set. This is your translation frame.

   Currently, if a stop codon is located beyond the region being viewed, the green line is not visible; instead, the evidence of a translation frame is the lack of stop codons over a long sequence.

Showing restriction sites

Map all instances of commonly used restriction sites across the chromosome or file being viewed. Then see how far apart the restriction sites are and compare the length of the fragments that should result.

To show restriction sites:

1. Make sure sequence residues are loaded
2. Click the **Restriction Sites** tab
3. Click one or more restriction sites on the left.

   Control-click to select multiple sites or to deselect ones you've already selected.
4. Click the **Map selected Restriction sites** button

Restriction sites will appear on the Coordinates tier; the color of the marker will be the color of the restriction site as listed in the tab panel's box of selected sites.

Clearing mapped restriction sites

To clear mapped restriction sites from the Coordinates tier:

1. Click the **Restriction Sites** tab
2. Click **Clear**
3. Wait a few moments for IGB to process the request
4. Click in the Coordinates Tier

Launching IGB with Java Web Start

Once you have installed Web Start with IGB, to launch IGB when you are online, visit

http://www.affymetrix.com/support/developer/tools/download_igb.affx and click the **Launch IGB** button; the most current version of IGB will automatically download (if needed) and launch.

To launch IGB if you are offline, click the IGB icon on your desktop or from **Start** menu > **Programs**, or choose **Start** menu > **Programs** > **Java Web Start** > **Java Web Start**. You will see the Web Start Console with several default applications and IGB. Double click the IGB application to launch IGB.

You can launch multiple instances of IGB.

Appendix 1 Using Java Web Start

You can use a free utility from Sun Microsystems called Java Web Start to simplify installation and use of IGB and ensure that you always have the most current version of IGB. When you use this utility to launch IGB, it will automatically check for, download and install updates as they become available. Java Web Start is a component of the Sun J2SE Standard Edition.

Obtaining Web Start and IGB

To use IGB with Web Start, visit http://www.affymetrix.com/support/developer/tools/download_igb.affx.

If you see the **Launch IGB** button, you already have the needed software. See Launching IGB with Java Web Start.

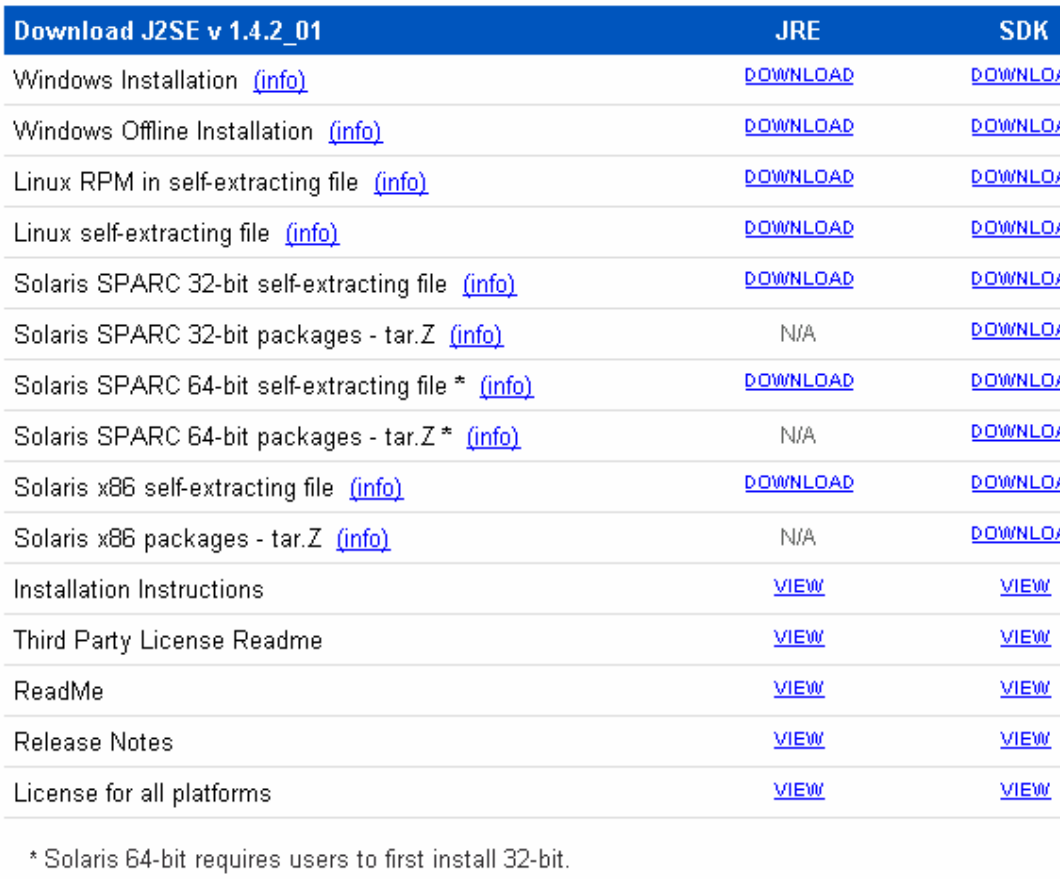
If not, go to the URL shown on that page and download the software indicated, or visit http://java.sun.com/downloads and download the Java 1.4.2 or higher JRE Standard Edition (J2SE) for your platform (Windows, UNIX, etc.) ***Many of the downloads at that site have similar names; choose carefully***, since you do ***not*** want the Enterprise Edition, or the Micro Edition, or Net Beans, or a Cobundled package, or the SDK. If you already have a version of the JRE or the JDK, this download should not interfere with that installation.

Note: the following process is subject to change; screenshots are an indication of what was current at the time this document was written, and what you will actually see on Sun's web site is subject to change.

1. On the Sun Microsystems Java downloads website, in the section labeled **Java 2 Platform, Standard Edition (J2SE)**:

2. From the drop-down list, under the item J2SE 1.4.2, select **-All Platforms**

3. Click **Go**.



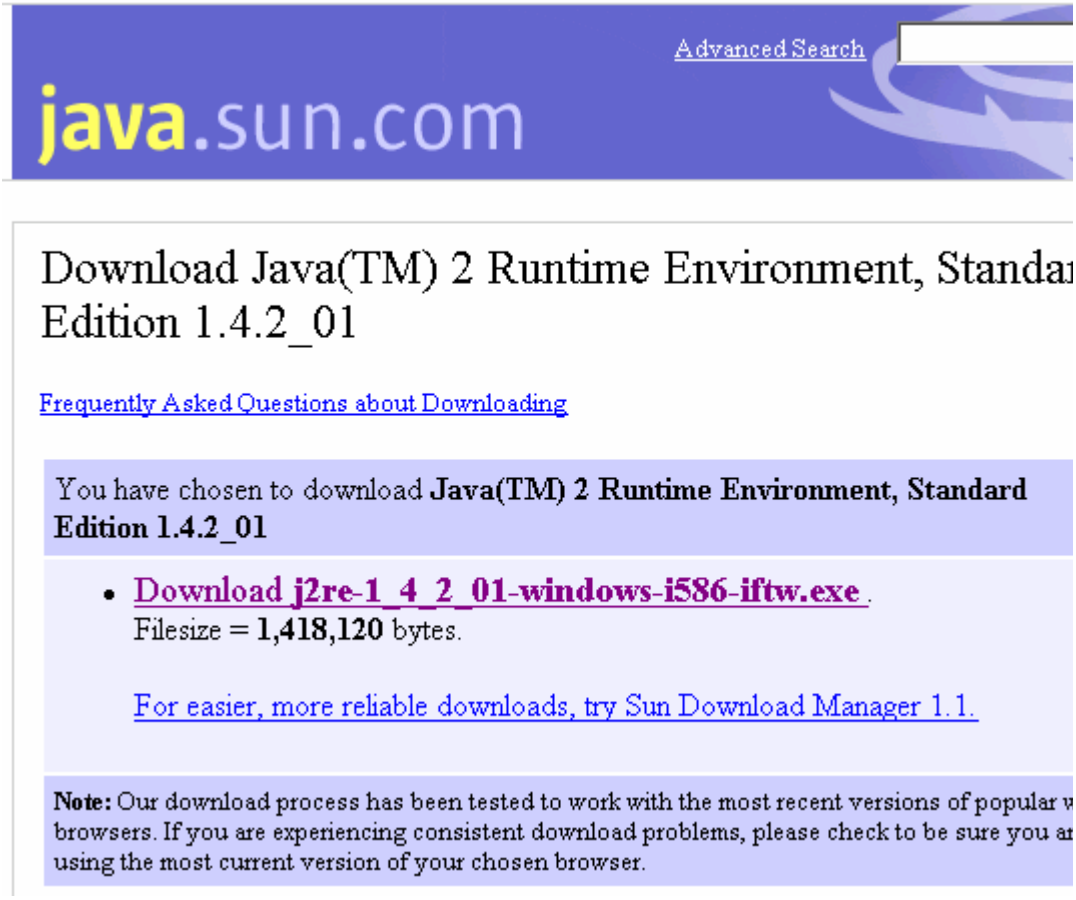
Download J2SE v 1.4.2_01	JRE	SDK
Windows Installation (info)	DOWNLOAD	DOWNLOA
Windows Offline Installation (info)	DOWNLOAD	DOWNLOA
Linux RPM in self-extracting file (info)	DOWNLOAD	DOWNLOA
Linux self-extracting file (info)	DOWNLOAD	DOWNLOA
Solaris SPARC 32-bit self-extracting file (info)	DOWNLOAD	DOWNLOA
Solaris SPARC 32-bit packages - tar.Z (info)	N/A	DOWNLOA
Solaris SPARC 64-bit self-extracting file * (info)	DOWNLOAD	DOWNLOA
Solaris SPARC 64-bit packages - tar.Z * (info)	N/A	DOWNLOA
Solaris x86 self-extracting file (info)	DOWNLOAD	DOWNLOA
Solaris x86 packages - tar.Z (info)	N/A	DOWNLOA
Installation Instructions	VIEW	VIEW
Third Party License Readme	VIEW	VIEW
ReadMe	VIEW	VIEW
Release Notes	VIEW	VIEW
License for all platforms	VIEW	VIEW

* Solaris 64-bit requires users to first install 32-bit.

4. Scroll down to the section for **Download J2SE v 1.4.2_01**

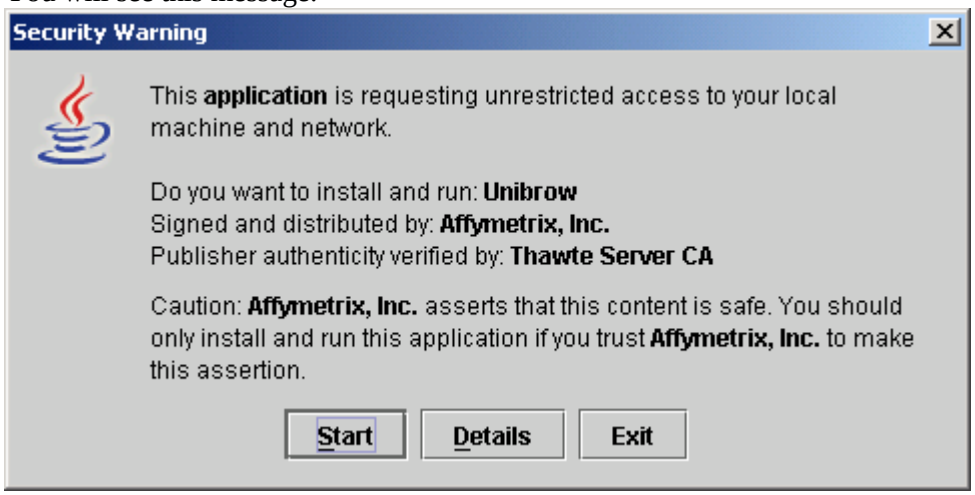
5. In the column headed JRE, click the link for your platform (Windows, UNIX, etc.)

6. Read and accept the license. Before you click the final link to perform the download, verify that you are downloading the correct software:



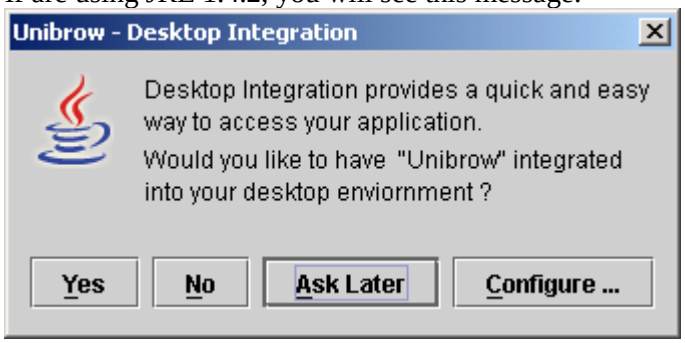
7. Click **Open** to run the installation from the Sun server, or click **Save** to save the installer to your hard disk and run the installation from there. When you run the installer, accept the defaults.

After Web Start has loaded, it will begin to download and install IGB. You will see this message:



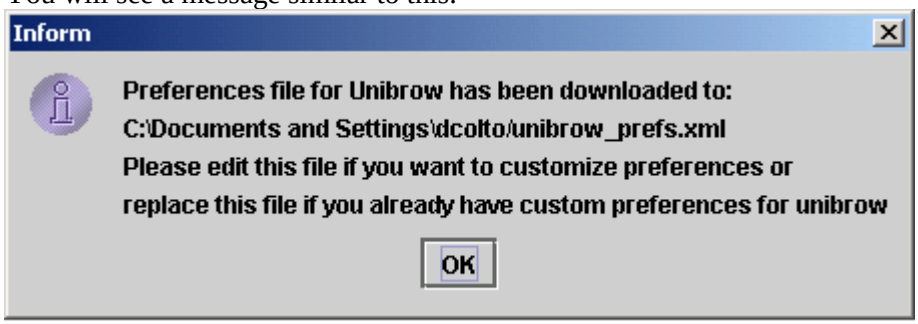
8. Click **Start**

9. If are using JRE 1.4.2, you will see this message:



10. If you wish to place a shortcut to IGB on your desktop and in your **Start** menu > **Programs**, click **Yes**. Otherwise, click **No**.

11. You will see a message similar to this:



Click **OK**. A Preferences file will be installed in your Windows home directory (in Windows 2000 and NT, that directory is C:\Documents and Settings\<your directory name>\igb_prefs.xml.) If you have been using

IGB and already have a customized preferences file, copy it to this location.

IGB will then launch.

Note: If you change web browsers, the Java Web Start process may start from the beginning and download the Web Start application again.

Appendix 2 Troubleshooting

If you have problems, first double-check the instructions for the specific task you're working on, and then try solutions below that are specific to the area of IGB you're working on, and finally try the general troubleshooting items at the end of this appendix.

Troubleshooting specific areas

Below are troubleshooting tips for some areas of IGB.

IGB seems slow or sluggish

Turn off slicing: Choose **View** menu > **Turn off slicing**.

Keyboard command doesn't work

Maybe you've changed the command in the Preferences file. Or some component other than the viewer window has the "keyboard focus". For example, your key strokes may be going to a text entry box. Click the mouse anywhere in the view to return the keyboard focus to it.

Out of memory errors

If you are working with many graphs or sequences, you may run out of memory. After you receive an out-of-memory error, most functions will become unusable and you will have to close and re-launch IGB.

It is possible to increase the amount of RAM allocated to Java to run IGB. Check the documentation for your operating system. If you are launching the program from a script, you can indicate the maximum amount of memory to use with the "-Xmx" flag. For example:

java -Xmx256m -jar IGB.jar

Tab or tab window is missing

If you have opened a tab in a new window, the tab window may have become hidden behind the main IGB or another window.

It is possible to choose which tabs will be shown with settings in a preferences file. See Controlling available functions.

General troubleshooting

If anything you're working on doesn't work, try these common fixes, even if they don't seem intuitive:

Load Sequence residues for the chromosome/ file you're using

Make sure the genome version and chromosome in the viewer match those of whatever you're trying to load/ compare.

Make sure you've selected the gene, not the exon, or vice-versa. See select parent.

Unclamp the view

Try limiting the region in the viewer to less than 2Mb

Make sure you're connected to the network and the server/databank you're connecting to is up and running…

Make sure you haven't hidden any tiers or graphs

Appendix 3 Preferences

To change preferences for data sources, colors, keyboard commands, browser, and more, you can customize the **igb_prefs.xml** file included in the IGB installation.

Changing preferences

To change the preferences, open the **igb_prefs.xml** file in a text editor, search for the function you'd like to modify, and then make changes, using the information below for reference.

To specify a non-default prefs file (or multiple prefs files), use the command-line argument "-prefs" to do this, and pass in a list of file paths to load prefs from.

For example:
java -Xmx256m -jar IGB.jar -prefs "./prefs1.xml;../../prefs2.xml"

To revert to the default preferences, exit IGB, delete the **igb_prefs.xml** file, then restart IGB.

Customizing locations of data servers

You can point IGB to different versions, copies, or locations of key data sources such as DAS or Quickload.

To do this, search in the Prefs file for the <tagval> command for the data source of interest.

For example, the default source for QuickLoad is:

<tagval tag="QuickLoadUrl" val="http://147.208.165.250/quickload_data/" />

Some data sources, for example QuickLoad, have more than one source listed; however, all but one are commented out (deactivated) by enclosing them between **<!--** and **-->**.

You can specify URLs on secure servers using https instead of http if you want more security. See your IT department to set up a secure server.

Storing QuickLoad data on your own computer

You can store QuickLoad data on your own computer for access while away from the network, and point IGB to that QuickLoad source file.

1. To obtain the QuickLoad data, look in the Preferences file for the URL given for QuickLoad data, then copy that directory into your browser's address/ location field.

2. Pick the genome assembly you want

3. Use your favorite FTP application to download the information to your hard disk. Also copy the Synonyms.txt file, then copy and edit the Contents file, which lists each genome, one per line, so that it lists only the genomes you have downloaded.

4. Modify the preferences file so that it points to this copy of QuickLoad data on your hard drive, using its pathname on your computer, for example:

   <tagval tag="QuickLoadUrl" val="C:\Program Files\IGB\QuickLoad\." />

   or

   <tagval tag="QuickLoadUrl" val="file:./data/" />

5. Comment out the default listing for QuickLoad data.

Deactivating unused locations

You can toggle back and forth between different sources of the same data by deactivating each source you are not using.

To deactivate an unused command line in the preferences file, comment it out by prefacing the command with **<!--** and following it with **-->**

For example:

<!-- <tagval tag="QuickLoadUrl" val="file:./data/" /> --> comment

To reactivate a deactivated command, remove the **<!--** and **-->**.

Controlling available functions

To make tabs visible or not visible, change the tag for that tab:

For example, to turn off the ORFs tab, change this value:

<boolean tag="USE_ORF_CONTROL" val="true" />

from “true” to “false”

Or:

To view only Untranslated regions, change this value:

<boolean tag="DrawCds" val="false" />

from “false” to “true”

Changing keyboard commands

You can set keyboard accelerators for many menu functions.

To do so, use the following format:

```
<keystroke function="Clear All" stroke="ctrl shift C" />
```

Where **function** = the exact text of the menu item (case-insensitive).

Because of the nature of XML, you will have to make substitutions if the text contains one of these special characters:

Symbol	XML code
&	&amp;
>	&gt;
<	&lt;
"	&quot;
'	&apos;

for example:

```
<keystroke function="Add Pm &amp; Mm Graph" stroke="" />
```

stroke = a key combination expressed in the format

"<modifier1> <modifier2 ...> key"

where modifier = one or more of

shift | control | meta | alt

and key = a single key, in capital letters, like "K" "C" "DELETE" "INSERT"

Notes:

If you leave the "stroke" empty, then no accelerator will be assigned to that key. You can leave these undefined keystrokes in the preferences file to remind you which menu items can be given accelerators.

If you give a meaningless keystroke, like "double click", it has the same effect as giving an empty keystroke.

If you assign the same keystroke to multiple functions, the result is unpredictable.

If you assign a keystroke that already has a meaning in your operating system or window manager, the result is unpredictable. For example, don't try using the keystroke "ctrl alt DELETE"

For letter keys, use upper-case only.

- For the non-letter keys, like "DELETE", use the codes given here: http://java.sun.com/j2se/1.4.2/docs/api/java/awt/event/KeyEvent.html, but remove the “VK_” prefix (e.g. "VK_DELETE" becomes "DELETE").

You can abbreviate "control" as "ctrl", but otherwise the system is inflexible. Shiftlock is ignored, but numlock is not.

Examples:

"control alt Z" -- a 3-key combination

"Z" -- the key 'z' all by itself (Choosing regular keys like this may cause problems and is not recommended.)

"z" -- meaningless

"shift Z" -- the shift key plus the 'z' key

"ctrl alt Z" -- same as "control alt Z"

"CTRL ALT Z" -- meaningless

"meta F4" -- the meta key plus the "F4" key

"$" -- meaningless, try "shift 4" (on most American keyboards)

"ctrl alt DELETE" -- a very bad choice for an accelerator

"NUMPAD4" -- "4" on the number pad, with "NUM LOCK" on

"KP_LEFT" -- "4" on the number pad, with "NUM LOCK" off

"LEFT" -- the LEFT key that isn't on the number pad

The description above is sufficient for most needs, but more details on the keystroke format are given here: http://java.sun.com/j2se/1.4.2/docs/api/javax/swing/KeyStroke.html#getKeyStroke(java.lang.String)

Default controls for moving the view:

```
<keystroke function="SCROLL_RIGHT" stroke="RIGHT" />
<keystroke function="SCROLL_LEFT" stroke="LEFT" />
<keystroke function="SCROLL_UP" stroke="UP" />
<keystroke function="SCROLL_DOWN" stroke="DOWN" />
<keystroke function="ZOOM_IN_X" stroke="control RIGHT" />
<keystroke function="ZOOM_OUT_X" stroke="control LEFT" />
<keystroke function="ZOOM_IN_Y" stroke="control UP" />
<keystroke function="ZOOM_OUT_Y" stroke="control DOWN" />
<keystroke function="ZOOM_OUT_FULLY" stroke="control HOME" />
```

Setting color and axis display preferences

When setting colors in the prefs file, the names are case-insensitive. Thus <curation ...> and <Curation ...> have the same effect.

To set the background color for any tier, add "_background" to the name:

```
<curation_background red="128" green="128" blue="128" />
```

The foreground and background colors for the axis are set the same way:

```
<axis red="0" green="0" blue="0" />
<axis_background red="255" green="255" blue="255" />
```

You can set the format of the numbers along the axis to either: FULL or COMMA, to see numbers like “10000” or “10,000”.

```
<tagval tag="AXIS_LABEL_FORMAT" val="COMMA" />
```

To set the color for edge-matching to change when fuzzy edge-matching is used:

```
<edgematcher red="255" green="255" blue="255" />
<fuzzyedgematcher red="200" green="200" blue="200" />
```

Changing the default Unix browser

On Windows, your default web browser will be used. On Linux or Unix systems, the default browser command is "netscape". For this to work, "netscape" must be in your execution path.

You can change the command used to open your browser. To do so, create an executable file which takes a single URL as an argument. For example, if you create a command called "/home/john/openURL.sh", then you can configure IGB to use this command like this:

```
<tagval tag="UNIX_BROWSER_CMD" val="/home/john/openURL.sh" />
```

Specifying the appearance of annotations

To customize the appearance of a particular type of annotation, use the **<annotation_style>** element with at least one of the following type attributes:

annot_type="val"

annot_type_starts_with="val"

annot_type_ends_with="val"

annot_type_regex="val"

For example, to specify the color of RefSeq annotations, you might specify any one of the following:

```
<annotation_style annot_type="refseq" red="0" green="255" blue="0" />
<annotation_style annot_type_starts_with="refseq" red="0" green="255" blue="0" />
<annotation_style annot_type_regex="(?i)refseq" red="0" green="255" blue="0"  />
```

If an annotation is loaded that has no **<annotation_style>** entry in the prefs file, or no display characteristics are specified, then a default annotation style is automatically assigned.

After specifying an annotation color, enter advanced display characteristics for it using the sections below.

All attributes and sub-elements below are optional, and are either non-essential or automatically assigned a default if not present. The advanced sections below are intended for use primarily by programmers.

Specifying the shape of annotations

"Glyphs" are the things annotations are rendered with, such as rectangles and lines. They could also be triangles, circles, etc.

The "glyph factory" is the engine in IGB that translates an annotation's definition into the visual representation you see in IGB.

Specify the factory class:

The "factory" attribute specifies a class to be instantiated as the glyph factory to "glyphify" annotations of the given annotation type.

Other attributes of an <annotation_style> element are passed to the factory in an initialization step as a Map with key/value pairs of form { attribute_name ==> attribute_value }, and it is up to the specific factory implementation to decide what to do with this information. This means that there are no restrictions on the attribute names in the <annotation_style> element, as different factories may recognize different attributes.

Example:

```
<annotation_style annot_type="refseq"
factory="com.affymetrix.IGB.glyph.GenericSymGlyphFactory" />
```

Specifying glyph color

There are two ways to specify glyph color:

Specify a single color for this annot type by adding red, green, blue attributes:

```
<annotation_style annot_type="refseq" red="200" green="0" blue="200" />
```

Specify multiple colors to be used for this annot type by adding <color> sub-elements to the <annotation_style> element:

```
<annotation_style annot_type="test">
    <color name="parent_color" red="255" green="255" blue="255"/>
    <color name="child_color" red="200" green="255" blue="100"/>
</annotation_style>
```

The name attribute value for a color element should *not* match any attribute id in the annotation_style. Colors specified this way will be passed to the glyph factory used for this annotation style as part of the initialization hash,

as entries of the form { "name_value" ==> new Color(redval, greenval, blueval) } It is up to the glyph factory to decide how to use these colors, based on their names. If red/green/blue attributes are included in <annotation_style> element (Method 1), the resulting Color is also added to the hash with key = "color".

Default glyph factory

The usual default factory is the GenericAnnotGlyphFactory.

Attributes that GenericAnnotGlyphFactory recognizes currently include:

"child glyph": This attribute specifies what glyph to use to render the (visible) leaf spans of the annotation

"parent_glyph": This attribute specifies what glyph to use to connect the child glyphs.

Example:

<annotation_style annot_type="test2"

parent_glyph="com.affymetrix.IGB.glyph.EfficientOutlineContGlyph"

child_glyph="com.affymetrix.IGB.glyph.EfficientFillRectGlyph" />

"glyph_depth": This attribute specifies which symmetries in a symmetry hierarchy should be rendered into glyphs. For annotations like transcripts or alignments, it is usually "2", indicating that the leaf symmetry nodes (for example, exons or alignment blocks) should be rendered with objects of class "child_glyph", and that the nodes just above the leaves (for example, the genomic span of the transcript or alignment) should be rendered with objects of class "parent_glyph".

If glyph depth is "1", then only the leaf nodes are rendered as glyphs, as objects of class "child_glyph" (good for repeats and SNPs, for example). It is not currently recommended to use glyph_depths greater than 2. The default is 2.

Color names that GenericAnnotGlyphFactory currently recognizes:

"parent_color": the color used for parent glyphs

"child_color": the color used for child glyphs

"color": if no "parent_color" or "child_color", uses this for both

Specifying URLs for annotation types

It is possible to associate links to any URL for items in a given tier. To do so, use an **<annotation_url>** element, like this:

<annotation_url annot_type_regex=" (?i)refseq"
url="http://my.server.com/annot?type=refseq&id=$$" />

In this example, any annotation tier matching the type "refseq" -- in a case-insensitive manner due to the "(?i)" flag -- will create links to our hypothetical server at “my.server.com”. Every instance of "**$$**" in the url parameter will be replaced with the id of the annotation. For example, the id "NM_001704" would be mapped to the url "http://my.server.com/annot?type=refseq&id=NM_001704". Note that to encode a "**&**" symbol in the url, it is necessary to use "**&**".

License

Source code for IGB is released under the Common Public License, v1.0, an OSI approved open source license. IGB uses other open source software packages that are not distributed with the IGB source code release, including Xerces and Jetspeed from Apache, and Jetty from Mortbay Consulting, which are covered by their own open source licenses. IGB is Copyright (c) 2000-2004 Affymetrix, Inc. Research and development of IGB is supported in part by NIH grant R01HG003040.

Common Public License Version 1.0

THE ACCOMPANYING PROGRAM IS PROVIDED UNDER THE TERMS OF THIS COMMON PUBLIC LICENSE ("AGREEMENT"). ANY USE, REPRODUCTION OR DISTRIBUTION OF THE PROGRAM CONSTITUTES RECIPIENT'S ACCEPTANCE OF THIS AGREEMENT.

1. DEFINITIONS

"Contribution" means:

a) in the case of the initial Contributor, the initial code and documentation distributed under this Agreement, and

b) in the case of each subsequent Contributor:

i) changes to the Program, and

ii) additions to the Program;

where such changes and/or additions to the Program originate from and are distributed by that particular Contributor. A Contribution 'originates' from a Contributor if it was added to the Program by such Contributor itself or anyone acting on such Contributor's behalf. Contributions do not include additions to the Program which: (i) are separate modules of software distributed in conjunction with the Program under their own license agreement, and (ii) are not derivative works of the Program.

"Contributor" means any person or entity that distributes the Program.

"Licensed Patents " mean patent claims licensable by a Contributor which are necessarily infringed by the use or sale of its Contribution alone or when combined with the Program.

"Program" means the Contributions distributed in accordance with this Agreement.

"Recipient" means anyone who receives the Program under this Agreement, including all Contributors.

2. GRANT OF RIGHTS

a) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free copyright license to reproduce, prepare derivative works of, publicly display, publicly perform, distribute and sublicense the Contribution of such Contributor, if any, and such derivative works, in source code and object code form.

b) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free patent license under Licensed Patents to make, use, sell, offer to sell, import and otherwise transfer the Contribution of such Contributor, if any, in source code and object code form. This patent license shall apply to the combination of the Contribution and the Program if, at the time the Contribution is added by the Contributor, such addition of the Contribution causes such combination to be covered by the Licensed Patents. The patent license shall not apply to any other combinations which include the Contribution. No hardware per se is licensed hereunder.

c) Recipient understands that although each Contributor grants the licenses to its Contributions set forth herein, no assurances are provided by any Contributor that the Program does not infringe the patent or other intellectual property rights of any other entity. Each Contributor disclaims any liability to Recipient for claims brought by any other entity based on infringement of intellectual property rights or otherwise. As a condition to exercising the rights and licenses granted hereunder, each Recipient hereby assumes sole responsibility to secure any other intellectual property rights needed, if any. For example, if a third party patent license is required to allow Recipient to distribute the Program, it is Recipient's responsibility to acquire that license before distributing the Program.

d) Each Contributor represents that to its knowledge it has sufficient copyright rights in its Contribution, if any, to grant the copyright license set forth in this Agreement.

3. REQUIREMENTS

A Contributor may choose to distribute the Program in object code form under its own license agreement, provided that:

a) it complies with the terms and conditions of this Agreement; and

b) its license agreement:

i) effectively disclaims on behalf of all Contributors all warranties and conditions, express and implied, including warranties or conditions of title and non-infringement, and implied warranties or conditions of merchantability and fitness for a particular purpose;

ii) effectively excludes on behalf of all Contributors all liability for damages, including direct, indirect, special, incidental and consequential damages, such as lost profits;

iii) states that any provisions which differ from this Agreement are offered by that Contributor alone and not by any other party; and

iv) states that source code for the Program is available from such Contributor, and informs licensees how to obtain it in a reasonable manner on or through a medium customarily used for software exchange.

When the Program is made available in source code form:

a) it must be made available under this Agreement; and

b) a copy of this Agreement must be included with each copy of the Program.

Contributors may not remove or alter any copyright notices contained within the Program.

Each Contributor must identify itself as the originator of its Contribution, if any, in a manner that reasonably allows subsequent Recipients to identify the originator of the Contribution.

4. COMMERCIAL DISTRIBUTION

Commercial distributors of software may accept certain responsibilities with respect to end users, business partners and the like. While this license is intended to facilitate the commercial use of the Program, the Contributor who includes the Program in a commercial product offering should do so in a manner which does not create potential liability for other Contributors. Therefore, if a Contributor includes the Program in a commercial product offering, such Contributor ("Commercial Contributor") hereby agrees to defend and indemnify every other Contributor ("Indemnified Contributor") against any losses, damages and costs (collectively "Losses") arising from claims, lawsuits and other legal actions brought by a third party against the Indemnified Contributor to the extent caused by the acts or omissions of such Commercial Contributor in connection with its distribution of the Program in a commercial product offering. The obligations in this section do not apply to any claims or Losses relating to any actual or alleged intellectual property infringement. In order to qualify, an Indemnified Contributor must: a) promptly notify the Commercial Contributor in writing of such claim, and b) allow the Commercial Contributor to control, and cooperate with the Commercial Contributor in, the defense and any related settlement negotiations. The Indemnified Contributor may participate in any such claim at its own expense.

For example, a Contributor might include the Program in a commercial product offering, Product X. That Contributor is then a Commercial Contributor. If that Commercial Contributor then makes performance claims, or offers warranties related to Product X, those performance claims and warranties are such Commercial Contributor's responsibility alone. Under this section, the Commercial Contributor would have to defend claims against the other Contributors related to those performance claims and warranties, and if a court requires any other Contributor to pay any damages as a result, the Commercial Contributor must pay those damages.

5. NO WARRANTY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, THE PROGRAM IS PROVIDED ON AN "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Each Recipient is solely responsible for determining the appropriateness of using and distributing the Program and assumes all risks associated with its exercise of rights under this Agreement, including but not limited to the risks and costs of program errors, compliance with applicable laws, damage to or loss of data, programs or equipment, and unavailability or interruption of operations.

6. DISCLAIMER OF LIABILITY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, NEITHER RECIPIENT NOR ANY CONTRIBUTORS SHALL HAVE ANY LIABILITY FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OR DISTRIBUTION OF THE PROGRAM OR THE EXERCISE OF ANY RIGHTS GRANTED HEREUNDER, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. GENERAL

If any provision of this Agreement is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Agreement, and without further action by the parties hereto, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

If Recipient institutes patent litigation against a Contributor with respect to a patent applicable to software (including a cross-claim or counterclaim in a lawsuit), then any patent licenses granted by that Contributor to such Recipient under this Agreement shall terminate as of the date such litigation is filed. In addition, if Recipient institutes patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Program itself (excluding combinations of the Program with other software or hardware) infringes such Recipient's patent(s), then such Recipient's rights granted under Section 2(b) shall terminate as of the date such litigation is filed.

All Recipient's rights under this Agreement shall terminate if it fails to comply with any of the material terms or conditions of this Agreement and does not cure such failure in a reasonable period of time after becoming aware of such noncompliance. If all Recipient's rights under this Agreement terminate, Recipient agrees to cease use and distribution of the Program as soon as reasonably practicable. However, Recipient's obligations under this Agreement and any licenses granted by Recipient relating to the Program shall continue and survive.

Everyone is permitted to copy and distribute copies of this Agreement, but in order to avoid inconsistency the Agreement is copyrighted and may only be modified in the following manner. The Agreement Steward reserves the right to publish new versions (including revisions) of this Agreement from time to time. No one other than the Agreement Steward has the right to modify this Agreement. IBM is the initial Agreement Steward. IBM may assign the responsibility to serve as the Agreement Steward to a suitable separate entity. Each new version of the Agreement will be given a distinguishing version number. The Program (including Contributions) may always be distributed subject to the version of the Agreement under which it was received. In addition, after a new version of the Agreement is published, Contributor may elect to distribute the Program (including its Contributions) under the new version. Except as expressly stated in Sections 2(a) and 2(b) above, Recipient receives no rights or licenses to the intellectual property of any Contributor under this Agreement, whether expressly, by implication, estoppel or otherwise. All rights in the Program not expressly granted under this Agreement are reserved.

This Agreement is governed by the laws of the State of New York and the intellectual property laws of the United States of America. No party to this Agreement will bring a legal action under this Agreement more than one year after the cause of action arose. Each party waives its rights to a jury trial in any resulting litigation.

Index