GeneProteinViz (GPViz)

Version 1.2.0

Quick Tour

Rene Snajder

04.02.2013

Table of Contents

Introduction

<TODO: talk about what GPViz does and why it does it – maybe copy from publication?>

Purpose of this document

This document is designed to give you a **quick run** through the most important features of GPViz. It's not a complete manual. For more detailed description, go to http://icbi.at/software/gpviz/gpviz.shtml and download the GPViz **User manual**.

System requirements

GPViz runs on any system that supports the Java 7 runtime. If you do not have Java 7 installed, please get it from http://www.java.com.

Mac users please visit http://www.java.com/en/download/faq/java_mac.xml to learn about how to install Java 7 on OSX.

Installation

Downloading GPViz

To download GPViz, please visit http://icbi.at/software/gpviz and click the “Download” tab. There, click “download” next to the newest version of the client package.

This should download a ZIP file, containing all the files you need to run GPViz. After downloading, extract the contents of the ZIP file wherever you want.

Downloading Demo Data

We provide a demo data set that will help you to run the “quick tour” section of this manual. Not only that, but they contain the complete data you will need to run GPViz on the human genome (hg19). If you work with different organisms, or want to use data from different sources, the “Use your own data” section of this manual is where you'll find all the necessary information.

To download the human test data, again visit http://icbi.at/software/gpviz and click the “Download” tab. There is a section called “Homo Sapiens test data”. It contains test files from several different sources (mainly Refseq and Ensembl). Although you won't need all of them for the demo, I recommend to just download them all and put them in one folder.

Running GPViz

On Windows

Unpack the ZIP file you downloaded and double-click “GeneProteinViz.exe”. That should be it. If Java Runtime 7 (or newer) can't be found, there will be a message telling you to install it (see System requirements).

On Mac/Linux

All you need to do is run the JAR file. If Java Runtime 7 is installed (and your system is configured to run JAR files properly, all you might need to do is double-click “GeneProteinViz.jar”.

If that doesn't work, make sure Java 7 is installed. Then, open a shell/terminal and navigate to the folder you extracted GPViz to and run:

JAVA_HOME/bin/java -jar GPViz.jar
(replace “JAVA_HOME” with your Java installation directory)

Note: The first time you start GPViz it will inform you that you haven't defined any annotation file. If you want to work purely with Ensembl data, or have any other data source with consistent IDs, you won't need that. See the “Annotation Files” section for more information.

Adjusting memory settings

If you plan to load large input files, you might need to adjust the memory settings to allow for more RAM.

On Windows

Open the file GeneProteinViz.l4j.ini in a text editor and change the line **“-Xmx1024m”** to a higher value. For example **“-Xmx1536m”** for 1.5GB of memory.

On Mac/Linux

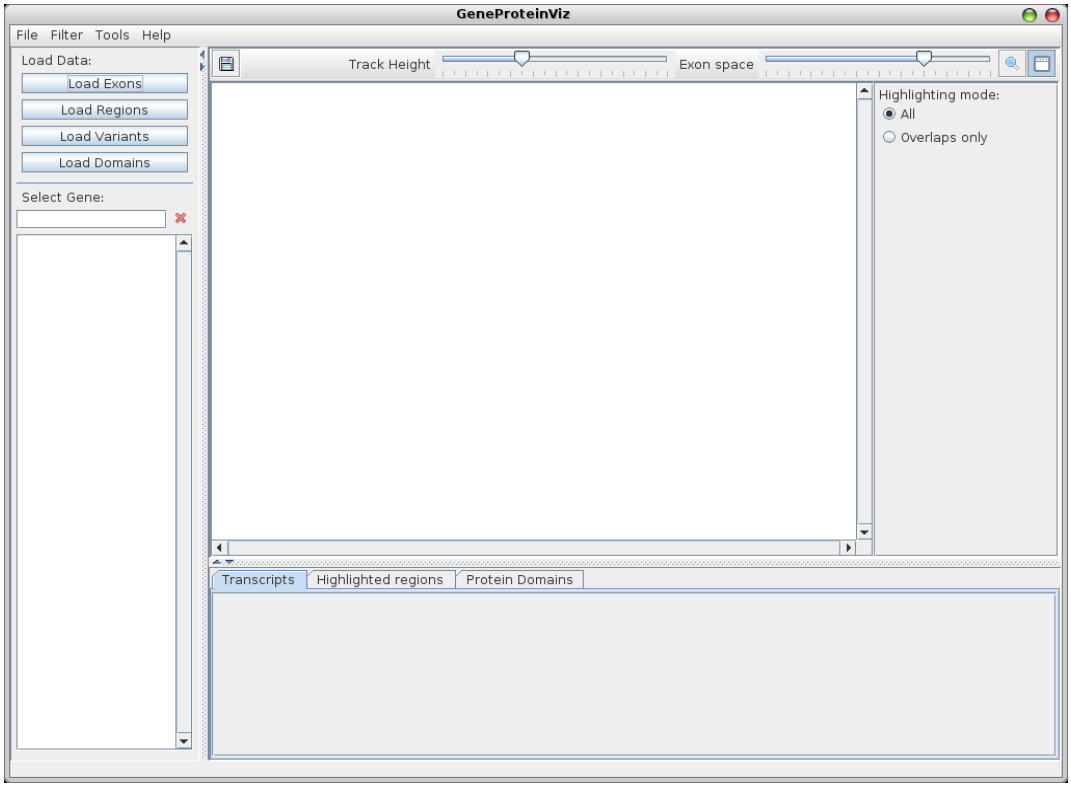
Simply add the memory parameter (for example **“-Xmx1536m”**) to the java command when launching the program. For example:

JAVA_HOME/bin/java -jar GPViz.jar -Xmx1536m

Note: If you run a **32bit version** of Java, the amount of memory you can use will probably be limited to around **1,5GB**. If you need more, you'll have to install a 64bit version of Java on a 64bit operating system.

Quick tour

After you successfully downloaded GPViz and know how to start it, this is what you should see:

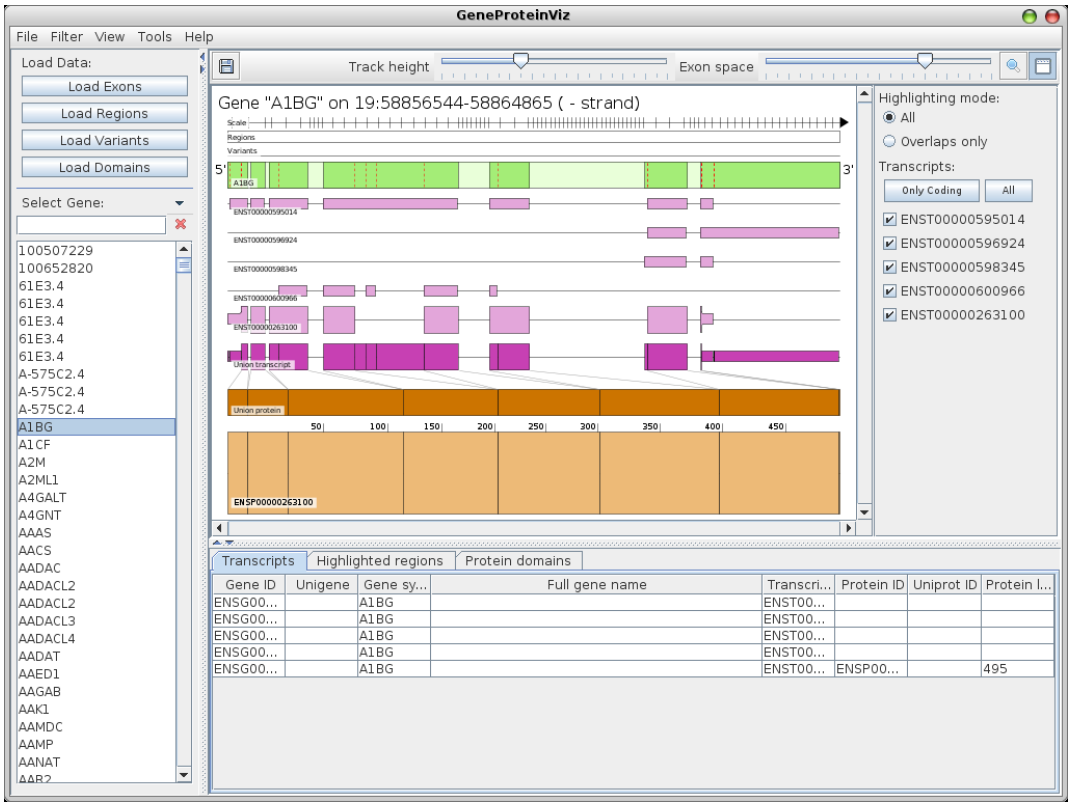


Make sure you have the demo data downloaded before continuing with this section.

Load a GTF file

First, we will load a GTF file. To get the information about the genes, exon positions, and transcription variants. Click the **Load Exons** button and select the **“Ensembl_GRCh37.gtf”** file from the demo data. This will take some time, but after it's done you can see on the left side the list of genes has been filled with the gene symbols loaded from the GTF file. Select **A1BG** for now.

GPViz should now look like this:



Below you can click through the table views, and on the right side you can hide and un-hide transcripts of the selected gene.

In the center you see the main visualization. The first of the main bars (here in green) shows the exons and introns in 5' to 3' direction. The dark green parts are exons, the light green are introns.

Under the green bar, you see the pink bars. They represent the the transcripts created by this gene.

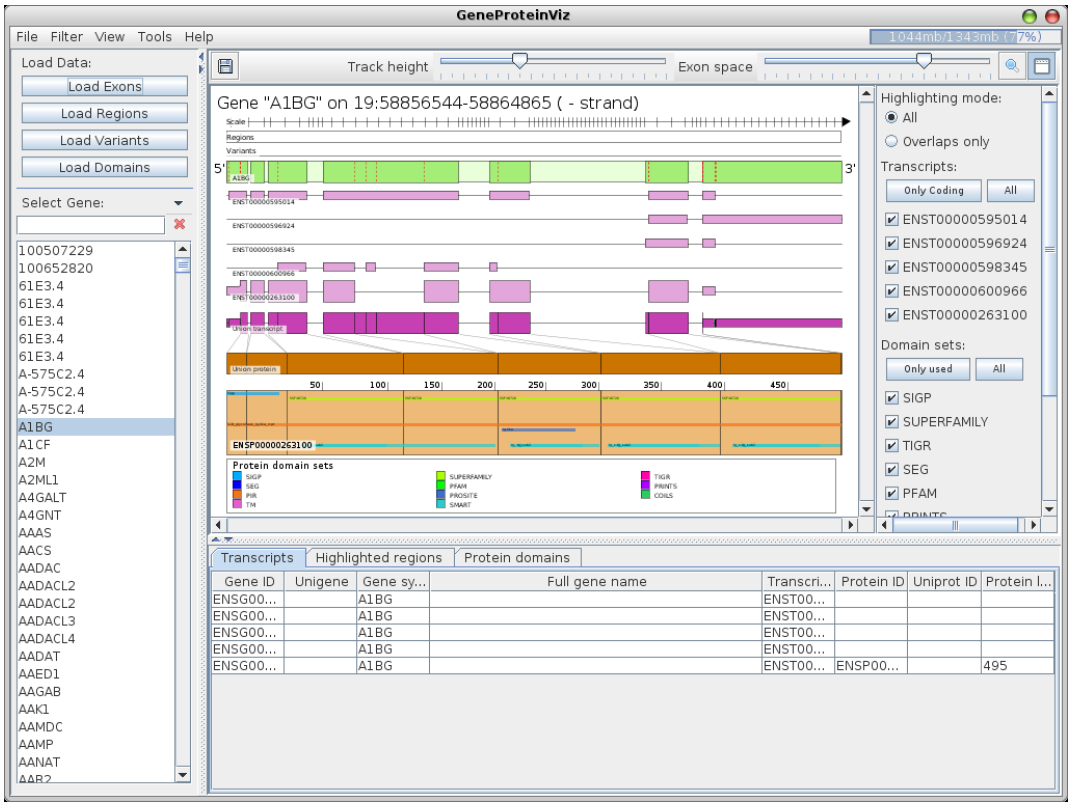
The magenta colored bar is what we call the **Union transcript**. It is basically the union of all transcripts, with each exon segment displayed as a separate block.

Same goes with the **Union protein**, which is the dark brown bar. It is a union of all the possible proteins created by this gene.

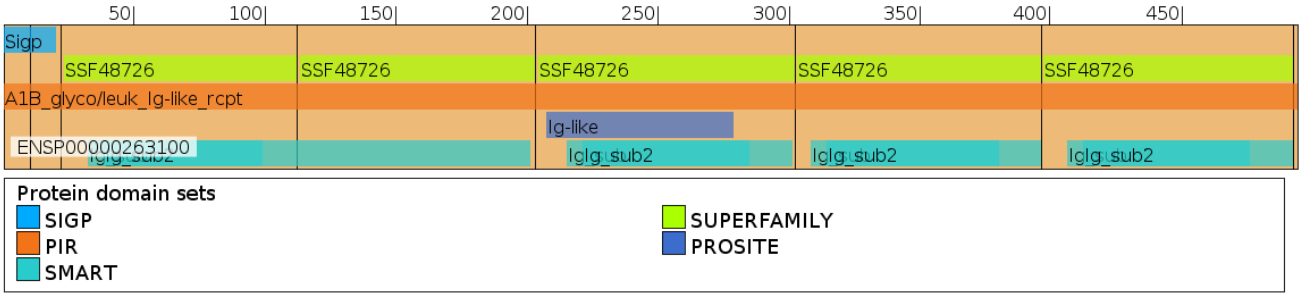
Under the Union protein you will then see one bar for each possible protein. In this case there is **only one**, as 4 of the 5 transcripts are **non-coding**. If there is more than one protein, the different segments will be aligned to the union protein.

Load protein domains

Next, we'll load the protein domains. Click the **Load Domains** button on the left side and select the Ensembl_IPR.txt file from the demo data.



Now the protein domains for all proteins have been loaded. You can see them highlighted in the protein block. Protein domains are divided in categories, which depend on the input file they have been loaded from. On the right side, in the display properties panel, you can see all the categories that have been found. Just like the transcripts, you can disable categories that you don't want displayed to make more space for the others. Use the “Only used” button above the domain categories to automatically hide all categories that are not present in the gene you are currently looking at. This will free up space in the legend and in the protein.



Once you clicked the “Only used” button, the figure will change to look like this.

You can see the separate categories now distribute vertically over the protein with no unused space in between. The surplus categories have also been removed from the legend below.

Customizing your figure

At this point you might want to familiarize yourself a bit with the various options available to change the appearance of figure. While we will go into further detail later in this manual, it's probably the best to have a quick look at them now.

Above the figure you will see two horizontal sliders called “**Track height**” and “**Exon space**”:

Track height changes the ratio between the protein tracks and all the other tracks in the figure. This is especially useful when you have many different proteins and/or many different domain categories. It allows you to increase the size of the proteins in order to have more space for the domain highlightings.

Exon space controls the ratio between exons and introns. In GPViz exons and introns aren't displayed to scale, but rather all exons are inflated in order to span over a certain percentage of the image. The Exon space slider can be used to increase or decrease that percentage. When looking at genes with a large amount of small exons, for example, it can be helpful to move the slider to the right to see them in more detail.

On the right side of the image you can see the “**Display properties panel**”. You can hide and show it with the little  icon above it. Here you can show/hide transcripts for the gene you are currently looking at, as well as protein domain categories.

Additionally, the **View** menu, which can be accessed in the top menu bar, or by right-clicking anywhere in the figure, gives you a couple of basic options, which mainly enable or disable elements in the figure.

The **Color schemes** dialog in the Options menu (Tools -> Options) allows you to create your own color scheme, different from the default scheme that comes with GPViz. You can even save and switch between them later on.

In the following screenshots some of these options have been used to make the figures easier to read. If you wonder why your results look different from the ones in here, try playing with these settings until you get a similar picture.

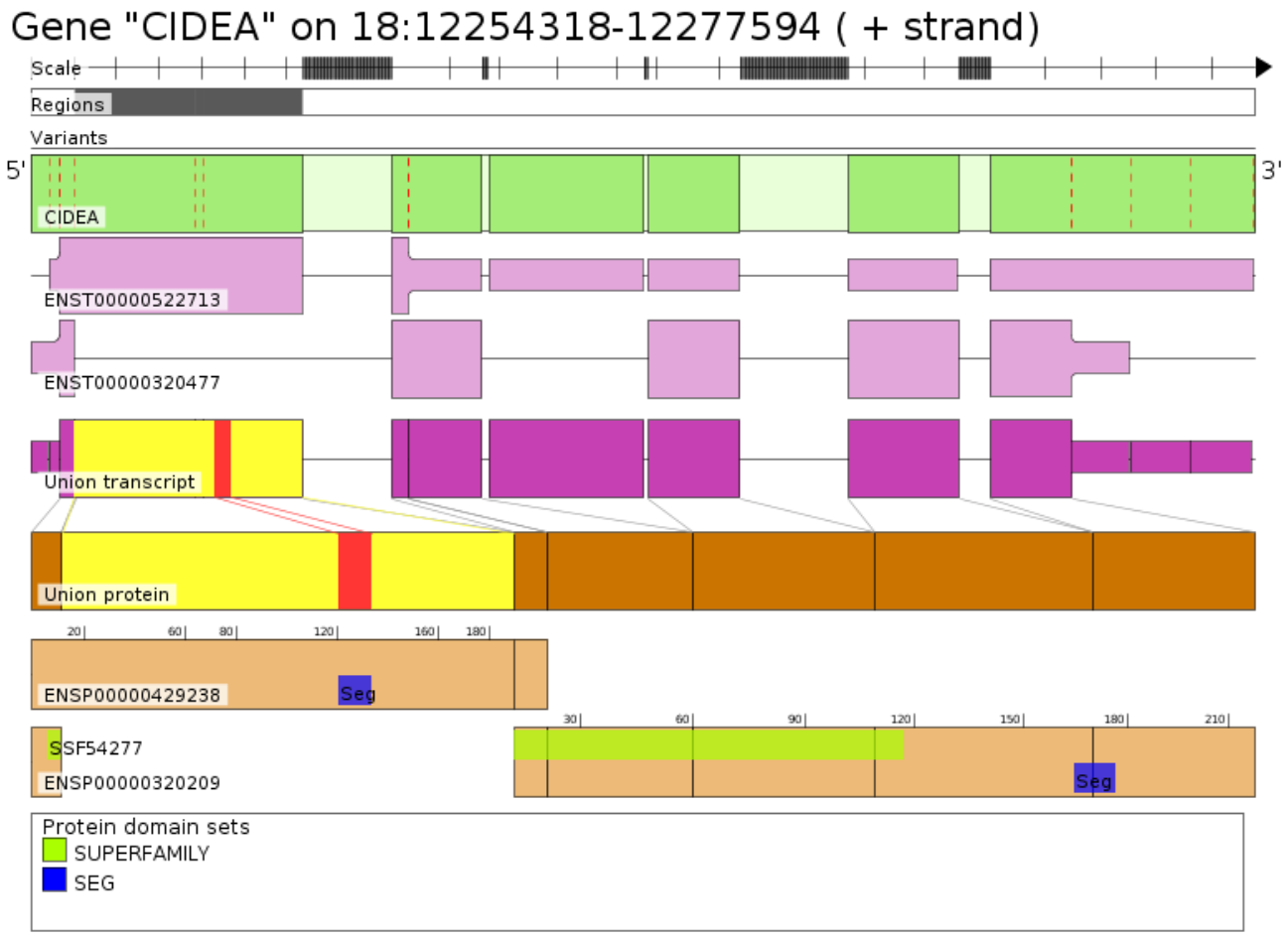
Load region highlighting

Now that we have all the data on our exons and proteins loaded, we'll want to highlight certain areas on our genome – in our case differentially expressed regions from a BED file. To do so, click “Load Regions” on the left side and select the file “Exons.bed” from the sample data. GPViz will then apply highlighting to all the areas provided.

At first you won't notice any difference, but that's because none of the regions apply to the gene you are currently looking at. So what we do now is **apply a filter** in order to find the genes that are affected.

In the top menu bar, click **Filter** and then **Genes with regions/variants**. It may take a couple of seconds to apply the filter, but once it's done you'll see that the gene list on the left side is reduced to contain only the genes that have any region highlights present.

Now, let's have a look at gene **CIDEA**.



Now we see that this gene contains one of the regions loaded from the BED file. The regions are marked on the top bar labeled as “Regions”, as well as a yellow/red highlighter on the Union transcript and Union protein.

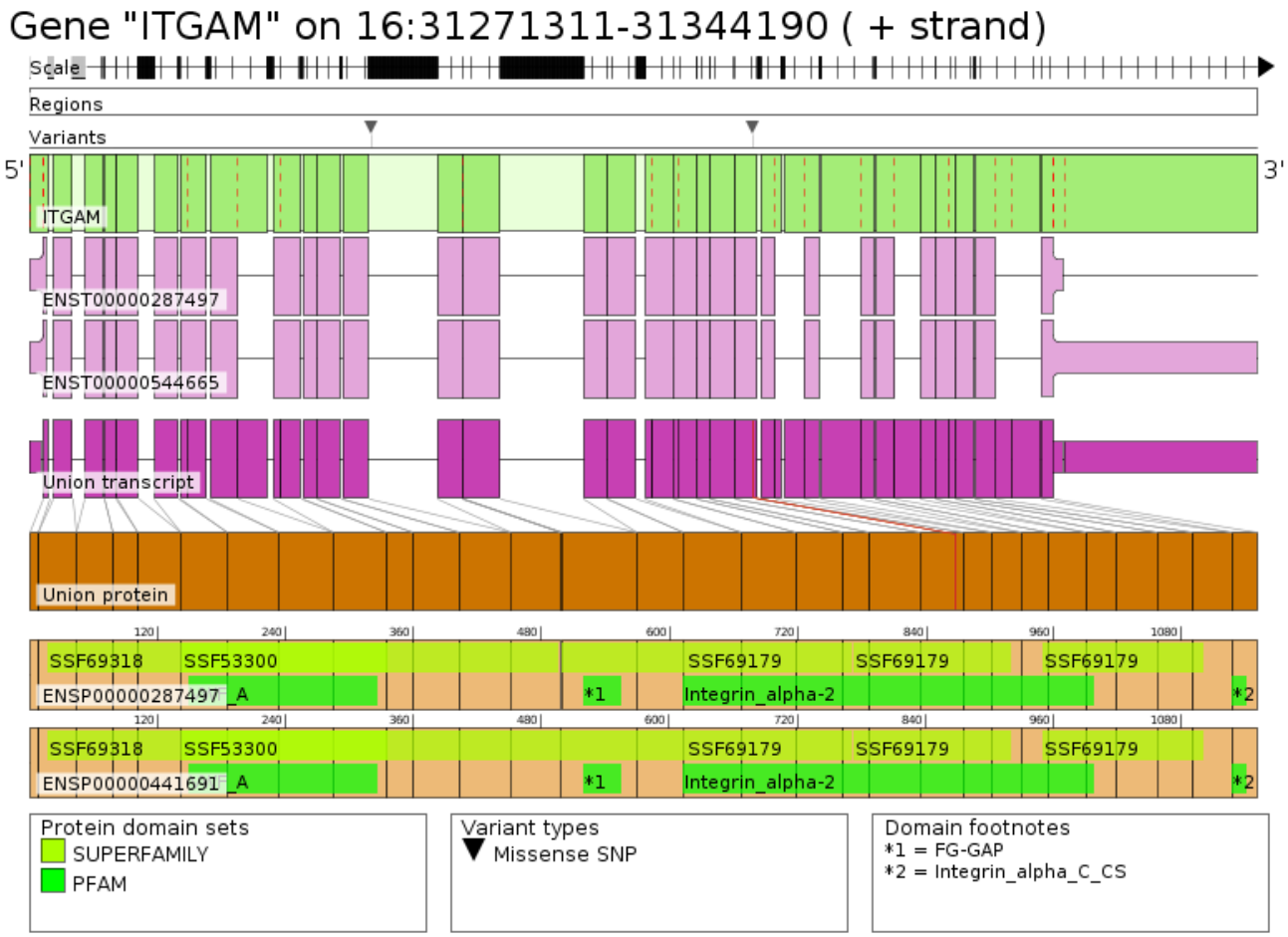
In here the **yellow** parts represent where this differentially expression region **affects the protein**, while the **red** parts show where such a change might also affect a protein domain, and thus might cause a functional change in the protein.

Load variants

Displaying variants in GPViz is very similar to displaying differentially expressed regions. Click the "**Load Variants**" button and select the "**Somatic_mutations.maf**" file from the sample data. If you still have the filter enabled from previously, the filter results will automatically be refreshed after loading the file. This may take a couple of seconds. If not, please select "**Filter**" and "**Genes with regions/variants**".

Now, just like with the regions, GPViz now displays all the Genes that are affected by the regions and variants from the input files.

If we have a look at gene ITGAM, remove all the non coding transcripts and select only the domain sets SUPERFAMILY and PFAM, we will see this picture:



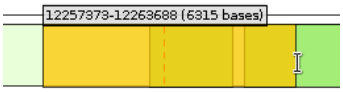
On the top in the "**Variants**" bar we can see two triangles, which we with the help of the legend can identify as Missense SNP. One of them seems to be located in an intron, and thus doesn't effect the resulting proteins, but the other one is within an exon, and if we trace the red line in the union transcript and union protein we can see that it overlaps with 2 of the protein domains.

Regions and variants from multiple files

You will have noticed that for every file you load with “Load regions” and “Load variants” a new “**Sample**” will be visible in the right hand side display properties panel. There, you can hide and unhide samples from being displayed in the figure.

Manually insert regions and variants

You can highlight regions by dragging your mouse across the gene track or one of the transcript tracks. You can also click a single exon to mark the entire exon. To apply that selection and save it as a region highlighting, simply **double click** your selection or right click and click “**Highlight this region**”.

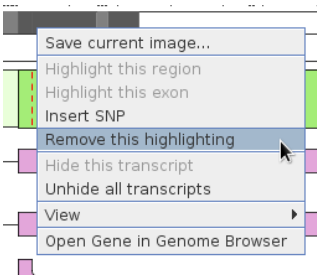


You can do the same with adding variants by simply double clicking in the “**Variants** bar above the gene, or by holding the **Ctrl** key while double clicking any area in the gene or transcript track.

The manually added regions will be saved as sample “Custom”.

Manually removing highlighted regions

Regions can be added and removed by right clicking the highlighter either in the “**Regions**” bar or in the union transcript or union protein, and then selecting “**Remove this highlighting**”.



Saving images

Finally, you'll want to be able to save your figure. You can simply save a single image by clicking the icon in the upper right corner, or by clicking “File” -> “Save current image...”.

You can also save **multiple figures at once.** To do so, use the Shift and/or Ctrl keys to select multiple genes from the list, and then click “File” -> “Save images for selected genes”.

This will open the image save dialog. Here you can choose the output directory, file format and size/resolution for your figures. You can also choose to write them all to one **multipage document**. If you choose this, every gene will be saved as a page in a multipage PDF or TIFF file.