

 vcf_tutorial.md

Downloading a VCF file

1. Navigate to the DGRP website and download the dgrp2 VCF file

Follow this link to the data page of the DGRP website: <http://dgrp2.gnets.ncsu.edu/data.html>

At the top of the page you will see the following links:

Genotype files

1. [VCF file for the DGRP Freeze 2.0 calls](#)
2. Plink formatted genotype ([BED/BIM/FAM](#))
3. [Tabular formatted genotype](#) (space delimited, 0 = ref allele, 2 = alt allele (not necessarily minor), - = missing)
4. [Relationship matrix](#) (tab delimited, to conform with FastLMM format, the line IDs are repeated (family ID and individual ID, such as in PLINK)

Click on the first link to download the VCF file. The download will start automatically. The file (called dgrp2.vcf) is 3.7 GB so it may take several minutes for the download to complete.

2. Make a project folder

In your desired location, make a new folder called "dgrp_vcf". Move dgrp2.vcf from your downloads folder into the dgrp_vcf folder.

Downloading Integrative Genomics Viewer (IGV)

Integrative Genomics Viewer (IGV) is a free desktop application for exploring next-generation sequencing data, including VCF files.

Follow this link to the IGV downloads page: <https://software.broadinstitute.org/software/igv/download>

There you will see the following download options:

Install IGV 2.8.13

See the [Release Notes](#) for what's new in each IGV release.



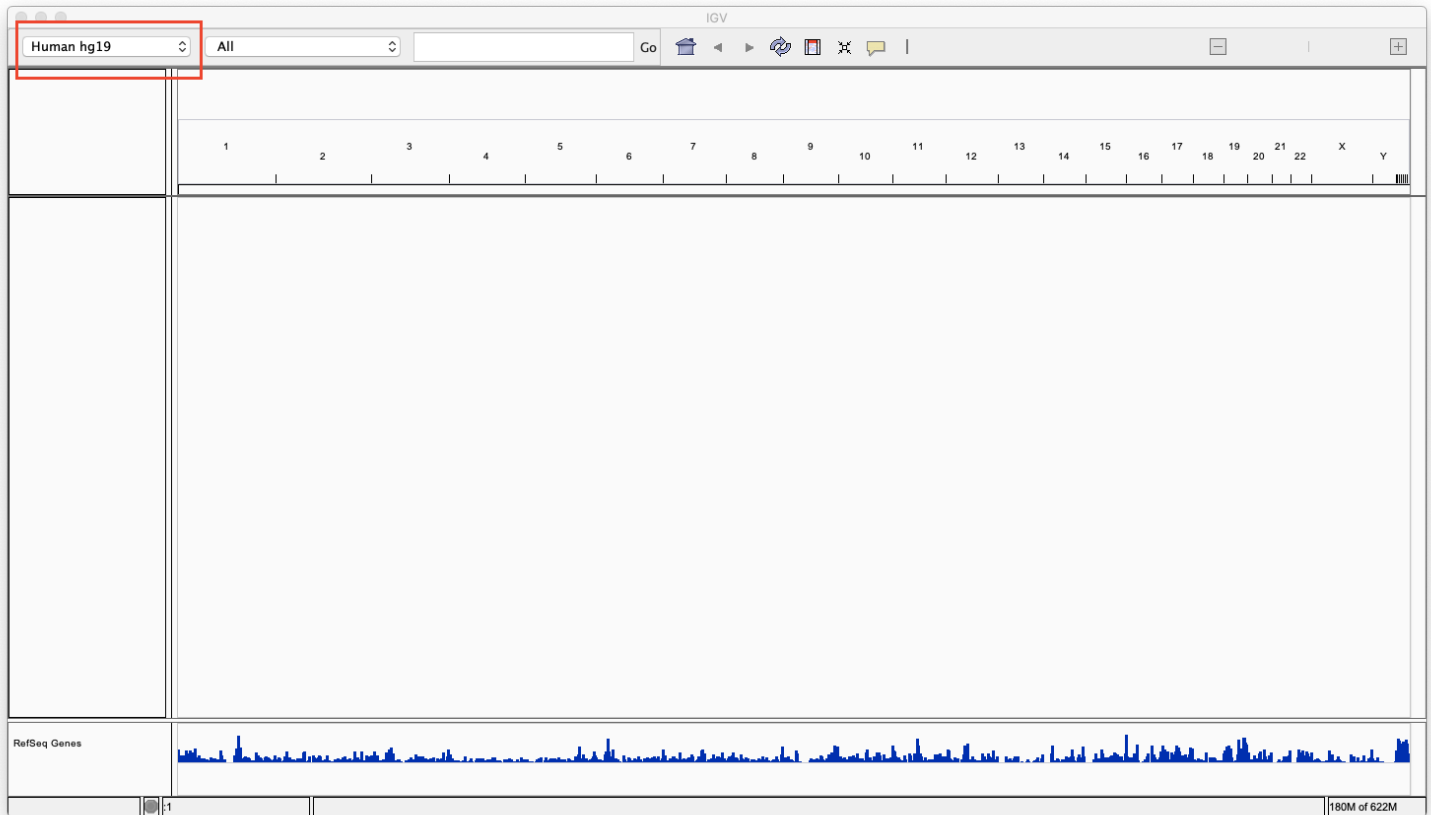
If you use a MacOS machine, click on: **IGV MacOS App (Java included)**

If you use a Windows machine, click on: **IGV for Windows (Java included)**

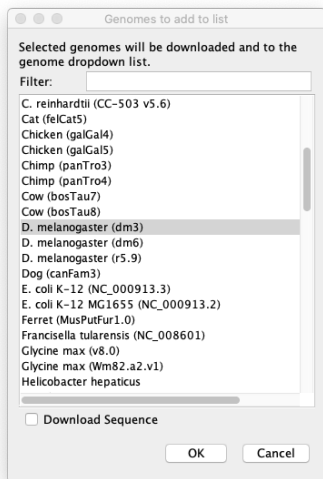
Using IGV to view VCFs

1. Open IGV

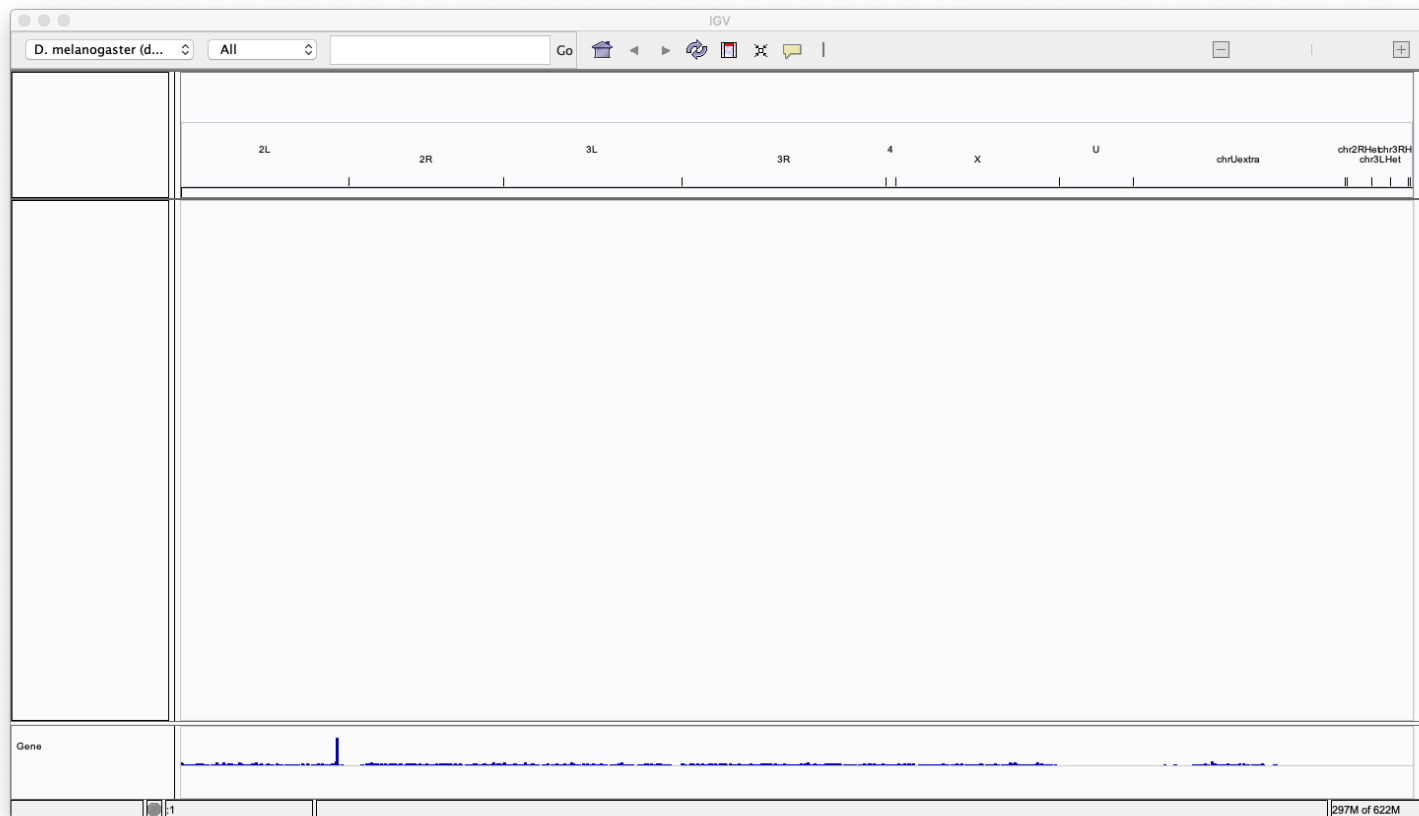
2. The human genome is set as the default genome upon opening, so we need to set the genome to the *Drosophila melanogaster* genome. To do this click on the drop-down menu in the top left hand corner, and select "More..."



3. You should see a pop-up that looks like the one below. Scroll down and select "D. melanogaster (dm3)" and press OK.

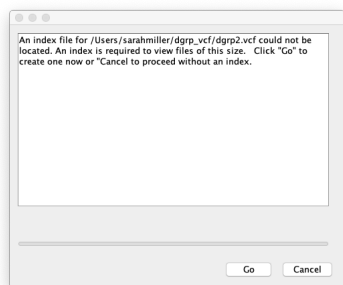


4. Your IGV window should now look like the image below.

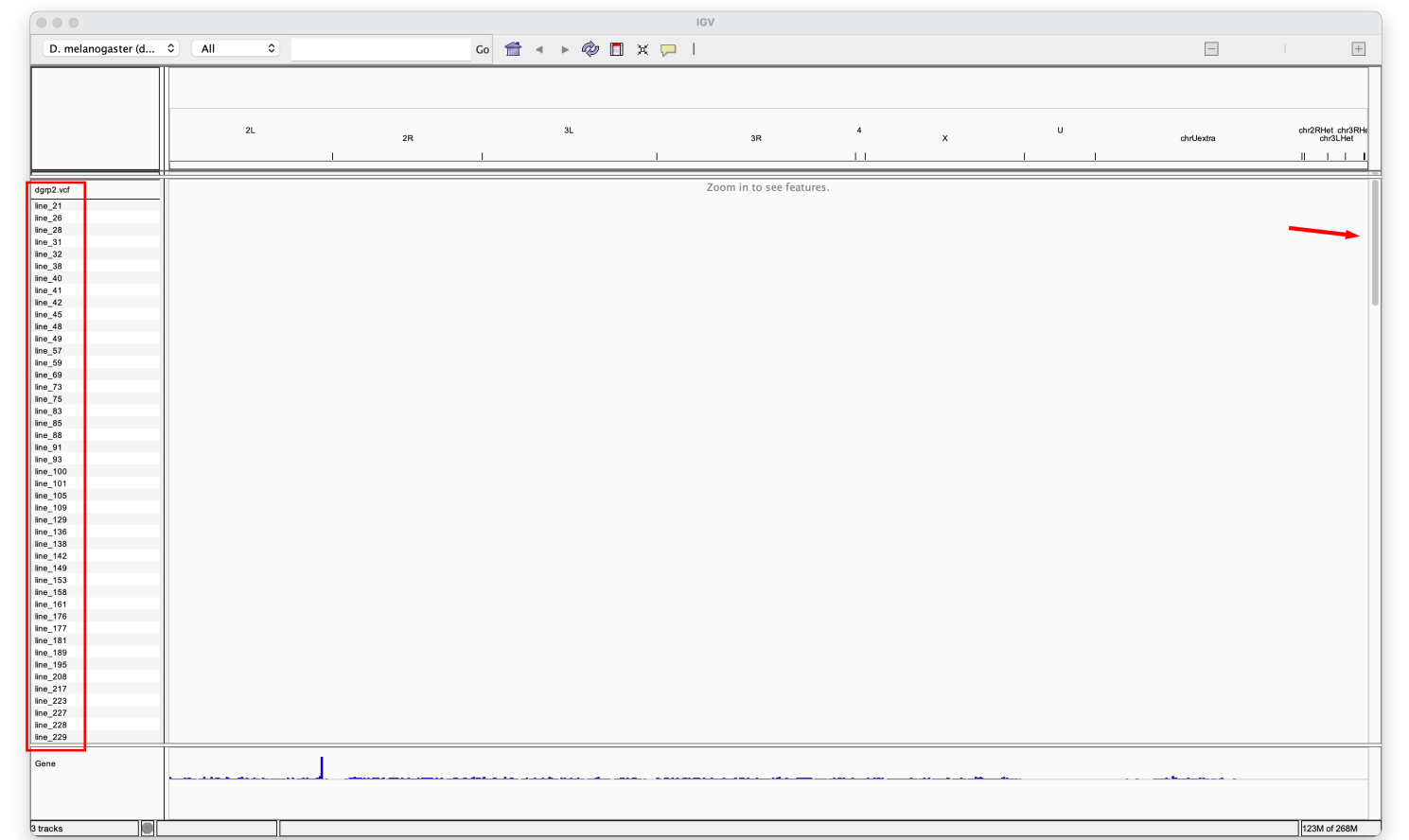


5. Now we are ready to load our VCF file.

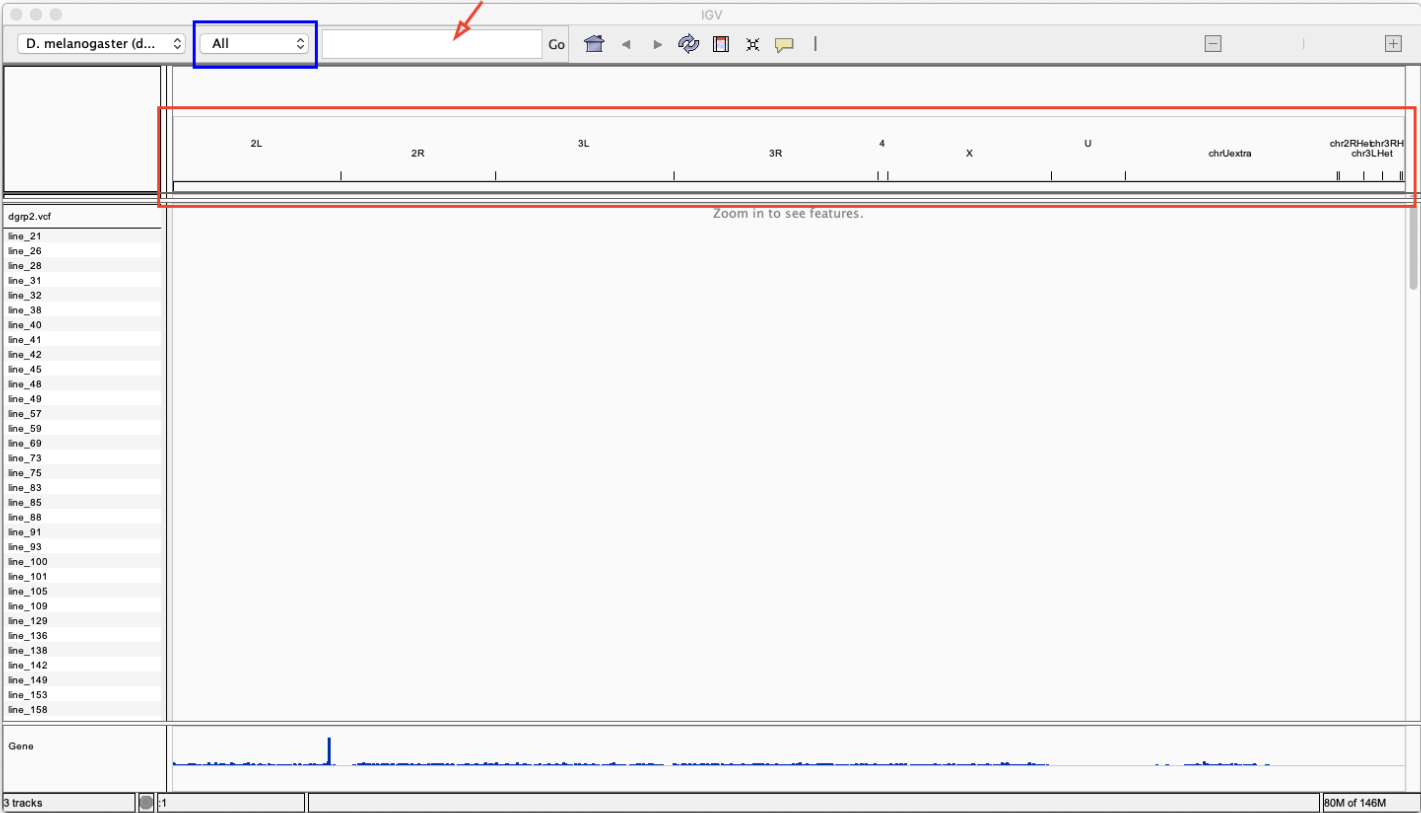
- Click on "File", then "Load from File..." (this will open a file browser)
- In the file browser, navigate to your newly-created dgrp_vcf folder, select `dgrp2.vcf` , and press open.
- You will see a pop-up that looks like the one below. Press "Go". This will make a file called `dgrp2.vcf.idx` in the dgrp_vcf folder (it might take a few minutes). You don't need to do anything with this file, but IGV requires it for loading the VCF.



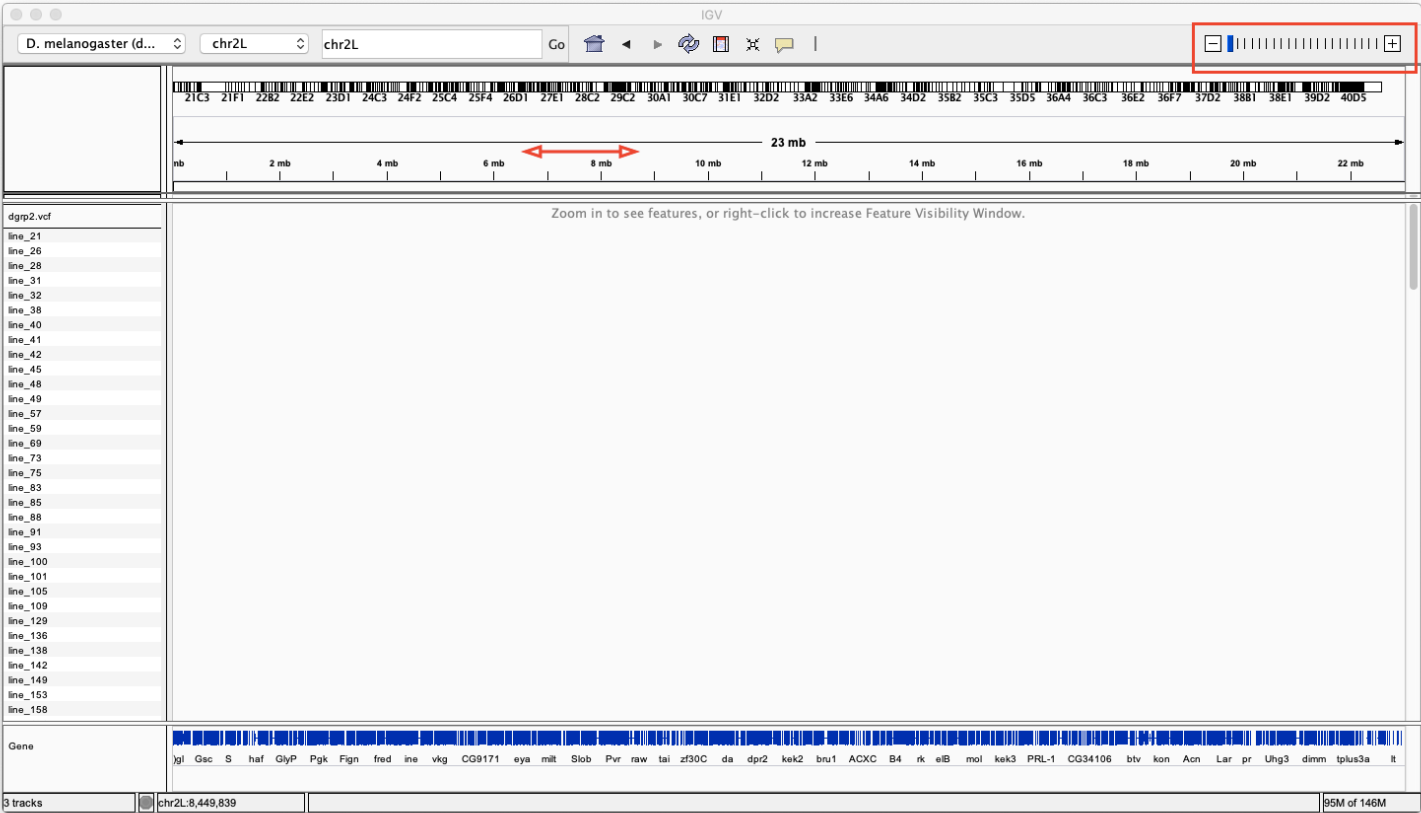
6. Once IGV has finished making `dgrp2.vcf.idx` , reload the VCF (same as the first two bullet-points from step 5). You will then see each of the 205 DGRP lines on the far left side of the window (red box in the image below). You can scroll through these using the scroll bar on the far right side of the window (red arrow in the image below).



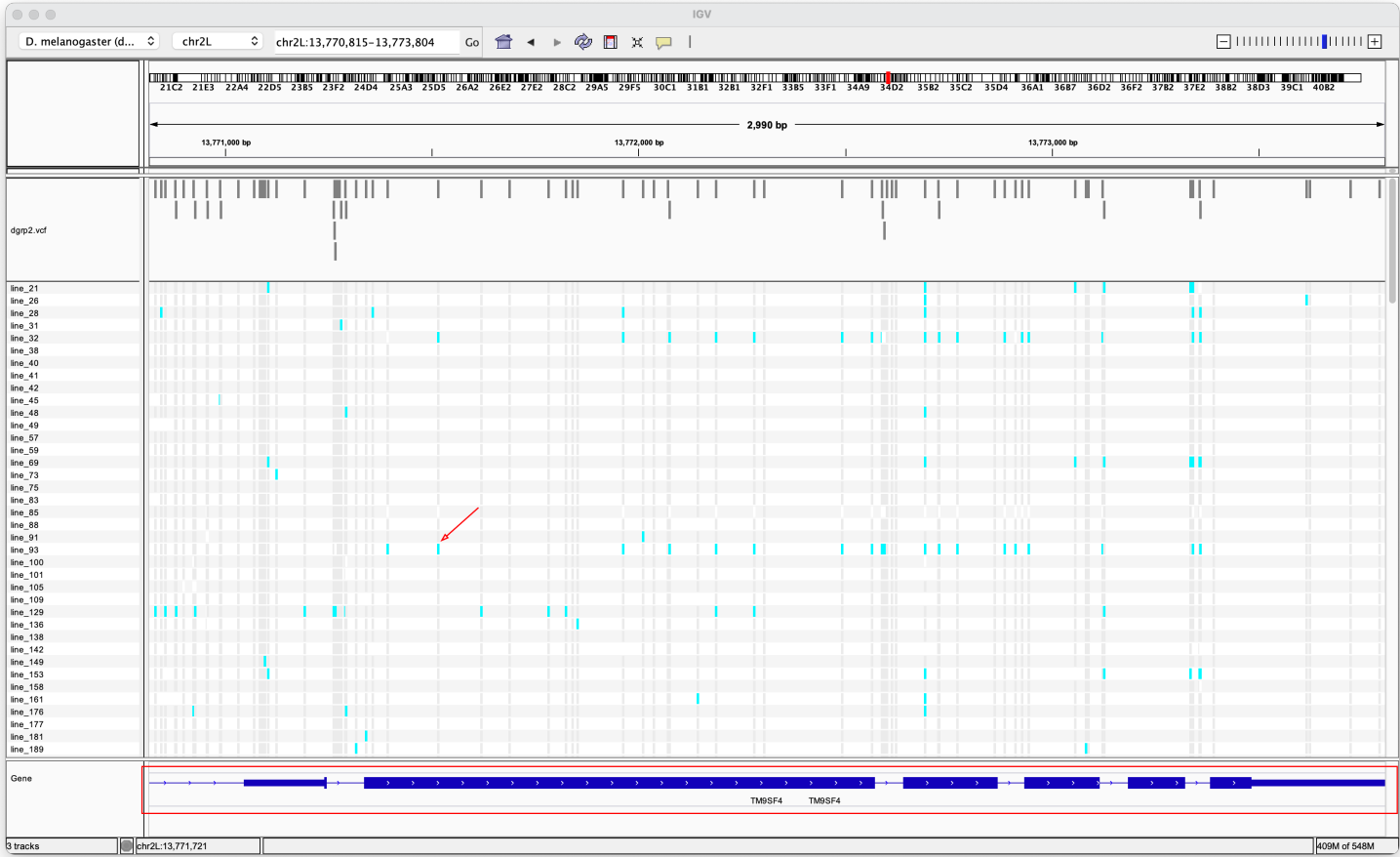
7. To explore genetic variation on a given chromosome , click on any of the chromosomes in the top panel (red box in the image below), or use the drop-down menu (blue box in the image below). Alternatively, you can jump to a specified gene or chromosome region using the text box in the toolbar (red arrow in the image below).



8. Shown below is the view after having clicked on chromosome 2L. You can zoom in and out using the toggle in the top right corner of the window (red box in the image below). Alternatively, you can zoom-in on a section of the chromosome by dragging your mouse along the desired chromosome section (red arrow in the image below).



9. When you zoom-in far enough, IGV will display the variants present in each of the DGRP lines. The red arrow in the image below points at a single variant. Genes, along with their exons and introns, are shown in the bottom panel (red box in the image below). To scroll left and right along the chromosome, swipe with your mouse in the main panel of the window (i.e., the panel displaying the variants).



10. After zooming-in even further, IGV will display the nucleotide sequence of the *D. melanogaster* genome in the bottom panel (red box in the image below). If you are zoomed-in on the coding-region of a gene, the amino acid sequence of that gene will also be shown.

