Name: Preet Gami

Date: December 18, 2020

PROJECT TITLE

Viewing and analyzing BAM (Binary Alignment Map) file and VCF (variant call format) file in IGV (Integrative Genomics Viewer).

BACKGROUND

- VCF and BAM files

VCF stands for Variant Call Format. This file stores gene sequence variants produced while comparing two different genomes. BAM stands for Binary Alignment Map. BAM file stores aligned sequence data.

- Importance of Analyzing variants

In biology, variant calling is important for various reasons.

1) At molecular level, helps in distinguishing between various organisms.
2) Helps in better diagnostics of specific diseases.
3) Genetic variants highly influence phenotype.
4) Necessary for studying evolution.
5) Helps to identify genes that are conserved among various organisms.
6) Understand genes that give rise to unique characteristics.

This video shows how to analyze these vcf and bam file into igv: **https://youtu.be/nochng9pJas**

- Tools used

1) IGV (Integrative Genomics Viewer) version 2.8.9

   IGV Web Application: https://igv.org/app/

   Download IGV: http://software.broadinstitute.org/software/igv/

- Advantages of using IGV over other tools

Using IGV (Integrative Genomics Viewer) over other genomics viewer such as Tablet, Bam View, Savant, Artemis is beneficial because of the following reason:

1) IGV allows to view data in more than once region of the genome at once in different panels (multi-locus view).
2) IGV allows visualization of many data types, such as NGS alignments, genomic annotations, expression data, genetic variations, etc.
3) IGV allows to customize your data the way you want to.

- Dataset used

Reference genome: Human (Homo Sapiens) hg38

VCF file:

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.chr22.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz

BAM file:

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/HG00100/alignment/HG00100.chrom20.ILLUMINA.bwa.GBR.low_coverage.20130415.bam

- The BAM file and VCF file was downloaded from the available datasets from the 1000 genome project phase 3.
- The BAM file belongs to a female (HG0010) from British in England and Scotland population, having a European ancestry. The reads align to chromosome 20.
- The VCF file contains variants for chromosome 21.
- Studying the similarities and differences between the human genome and individuals with variants help in exploring the molecular basis of the trait that makes the difference between that particular individual and the reference genome used.
- It also helps in understanding the evolutionary steps at the molecular level, mutations that might have taken place, selective constraint that led to the formation of the human species.

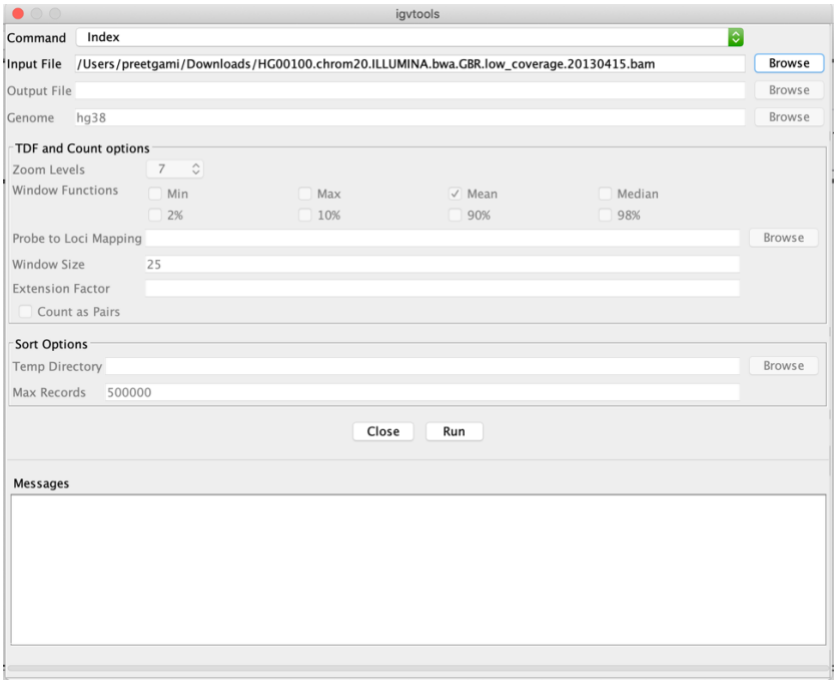
LOADING BAM FILES IN IGV

1) INDEX BAM FILES

To upload a BAM file in IGV, it is necessary to have an index file in the same folder, the BAM file is stored in. A BAM file has ". bam" extension, whereas an index file for BAM file has ".bai" extension. BAM file can be indexed using UNIX commands. However, even IGV can do create an index file for the BAM file. To do that:

Open IGV⇒Tools ⇒Run igvtools.

- Choose the option "index" from the drop-down menu from "command".
- Browse the input BAM file into the "input file" option.
- Hit run. Once the index file will be created, the "message" box will say "done".
- Look for the index file (.bai) in the folder where the BAM file is stored.



2) ADD FILES

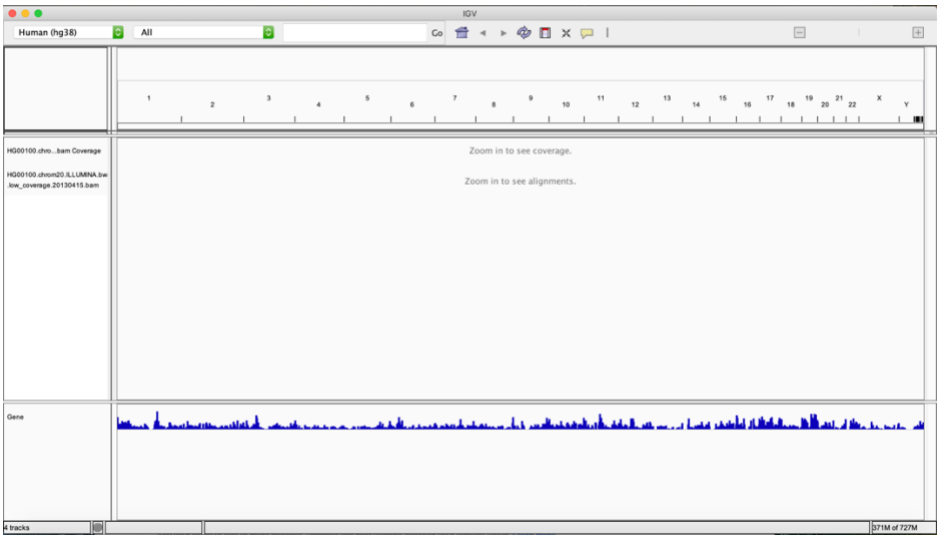
Files can be imported in various ways in IGV. Files can be directly imported from your computer/URL/IGV server. In this tutorial, the BAM files is saved on the computer, therefore, the BAM file will be uploaded from:

File ⇒ Load from file.

- Choose the original BAM file to upload. Make sure the index file (.bai) in the same folder. Do not upload the index file.

3) IGV SCREEN BASICS

Once, the file is uploaded, the IGV screen will look like shown in the picture below:



- From the drop-down menu on top-most left corner, reference genome can be selected. IGV server has already uploaded several reference genomes which can be used.
- Next to that menu, is another drop-down which is used to select the chromosome to which the reads align.
- On the white bar, next to it, specific locus on chromosome/gene name can be typed to find it on the chromosome.
- The topmost track in the middle of the screen represents the chromosome track. The number of the specific chromosome can be clicked to view only that particular chromosome.
- The track underneath it the coverage track, which will reflect the coverage produced by the aligned reads, when they are visible.
- In the track below it, the aligned reads will be visible.
- The last track below it represents the genes on the reference genome.

4) DISPLAY OF BAM FILE IN IGV

Once, the BAM file is uploaded, specific chromosome is selected, reference genome is selected. The IGV screen can be zoomed which will make the aligned reads visible. In a lot of the cases, when the IGV screen is not zoomed enough, the aligned reads will not be visible. The reads from the BAM file in IGV looks like the picture below:

LOADING VCF FILES IN IGV

1) INDEX VCF FILES

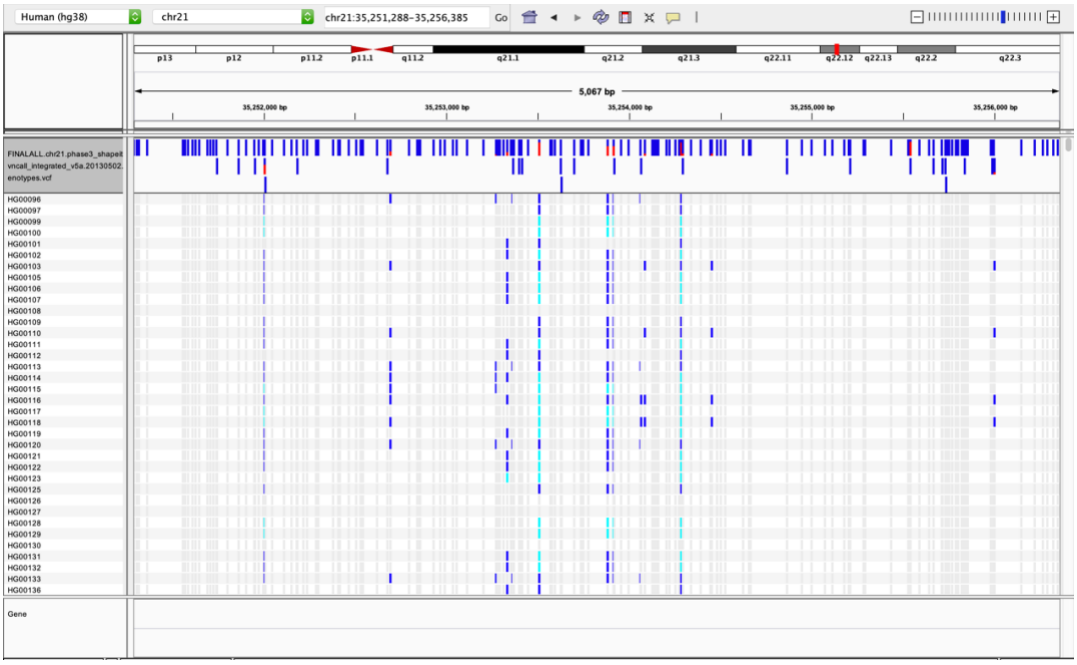
To upload a VCF file in IGV, it is necessary to have an index file in the same folder, the VCF file is stored in. A VCF file has ". vcf" extension, whereas an index file for VCF file has ". vcf.idx" extension. VCF file can be indexed using UNIX commands. However, even IGV can do create an index file for the VCF file. To do that:

Open IGV⇒Tools ⇒Run igvtools.

- Choose the option "index" from the drop-down menu from "command".
- Browse the input VCF file into the "input file" option.
- Hit run. Once the index file will be created, the "message" box will say "done".
- Look for the index file (. vcf.idx) in the folder where the VCF file is stored.

2) DISPLAY OF VCF FILE IN IGV

Once, the VCF file is uploaded, specific chromosome is selected, reference genome is selected. The IGV screen can be zoomed which will make the variants visible. In a lot of the cases, when the IGV screen is not zoomed enough, the variants will not be visible. The variants from the VCF file in IGV looks like shown in the picture below:



ANALYZING THE BAM AND VCF FILE IN IGV

This video shows how to analyze these vcf and bam file into igv: https://youtu.be/nochng9pJas