Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Where are we?

NGS pipeline

Sequence preprocessing

Mapping

Variant Calling

Variant prioritization

Functional annotation

GWAS Analysis

Gene-Set Analysis

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Why IGV?

- IGV is an integrated visualization tool of large data types

Microarrays

Epigenomics

RNA-Seq

NGS alignments
Why IGV?

- Integrate different data types simultaneously
- View large datasets easily
- Fast navigation
- Run it locally on desktop
- Easy to use interface
Why IGV?

- Large-Scale projects using IGV
  - The Cancer Genome Atlas
    http://cancergenome.nih.gov/
  - Multiple Myeloma Research Consortium
    http://themmrc.org/
  - 1000 Genomes Project
    http://www.1000genomes.org/
Getting IGV

http://www.broadinstitute.org/igv/
History and Usage

- First release was in August 2008
- Current version: 2.2
- Open source and freely available
Overview

Interface

Loading Data

Browsing the Data

Sessions

Exercises

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Main Window

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)

Fco. Javier López
## Tool Bar

<table>
<thead>
<tr>
<th>Feature</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome drop-down box</td>
<td>Loads a genome</td>
</tr>
<tr>
<td>Chromosome drop-down box</td>
<td>Zooms to a chromosome</td>
</tr>
<tr>
<td>Search box</td>
<td>Displays a genomic location</td>
</tr>
<tr>
<td>Whole genome view</td>
<td>Zooms to whole genome view</td>
</tr>
<tr>
<td>Move backward and forward</td>
<td>Moves backward and forward</td>
</tr>
<tr>
<td>Refresh</td>
<td>Refreshes the display</td>
</tr>
<tr>
<td>Define a region</td>
<td>Defines a region of interest</td>
</tr>
<tr>
<td>Zoom slider</td>
<td>Fits all tracks to the window</td>
</tr>
<tr>
<td>Toggles on/off pop-up information</td>
<td>Toggles on/off pop-up information</td>
</tr>
</tbody>
</table>

---

**Fco. Javier López**

**Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)**
Available Genomes

- Human, Mouse, S. cerevisiae, C. Elegans, D. melanogaster, and others

http://www.broadinstitute.org/software/igv/Genomes
Overview

Interface

Loading Data

Browsing the Data

Sessions

Exercises
Data types

- General characteristics
  - Any data related to genome coordinates
  - Sample annotation/attributes
  - Genome annotations
# Data types

- IGV supports multiple file formats

<table>
<thead>
<tr>
<th>Source Data</th>
<th>Recommended File Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence alignment data</td>
<td>SAM (must be sorted/indexed)</td>
</tr>
<tr>
<td></td>
<td>BAM format (must be indexed)</td>
</tr>
<tr>
<td>Genome annotations</td>
<td>GFF or GFF3 format, BED format</td>
</tr>
<tr>
<td>ChIP-Seq, RNA-Seq</td>
<td>TDF format. Use igvtools to generate a binary read count. Load the resulting TDF file into IGV.</td>
</tr>
<tr>
<td>Any numeric data</td>
<td>IGV format, TAB format</td>
</tr>
<tr>
<td></td>
<td>WIG format</td>
</tr>
<tr>
<td>Gene expression data</td>
<td>GCT format</td>
</tr>
<tr>
<td></td>
<td>RES Format</td>
</tr>
</tbody>
</table>
Loading a BAM file

- **BAM format**: SAM binary. Reduces disk space and access time.
  - For each read, provides the position(s) where it maps and information about the alignment.
  - BAM files need to be indexed (samtools). SAM files need to be sorted by start position and indexed.

**Index an example bam file**

```
samtools index /home/biouser/mda13/mqc-igv/igv1.bam
ls -la /home/biouser/mda13/mqc-igv/igv1.*
```
Loading a BAM file

Open IGV

igv

Load an example bam file

▶ File → Load from file →
/home/biouser/mda13/mqc-igv/igv1.bam
Overview

Interface

Loading Data

Browsing the Data

Sessions

Exercises

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Moving around

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Moving around

Type genomic coordinates 21:32508271 in the search box

- At low resolution only coverage is shown
- At higher resolution, reads are shown including where bases differ

Leave the cursor on a read
Track options

Right click on track
RefSeq track

Type GART in the search box

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
RefSeq track

Zoom in to focus on an exon

- Reference nucleotide sequence
- Predicted aminoacid sequence
RefSeq track

Right click on the RefSeq track → Expanded

Zoom out to get a general view of the mapping profile

- Reads are grouped around exons. Is that a coincidence?
  No. This bam is part of an exome sequencing experiment.
Loading a BED file

- Let’s load capture regions
- BED format: to store a list of genomic regions. Text file with the list of regions. Each line contains one region with three required fields separated by tabs: chromosome, start coordinate, end coordinate

  http://genome.ucsc.edu/FAQ/FAQformat.html

  less /home/biouser/mda13/mqc-igv/igv.bed

Load a bed file

- File → Load from file → igv.bed
  Another track appears: blue boxes indicate target regions
Visualizing variants

Move to 21:48022375

- Variant: ref:C, alt:A, heterozygosis

Leave the cursor over that position on the histogram (top of the track)
Visualizing variants

Right click over that position on the histogram

Move to 21:47786817

- Variant: ref:C, alt:G, homozygosis
Visualizing variants

- Files containing a variant list can also be loaded: VCF files (Variant Calling Format).

  less /home/biouser/mda13/mqc-igv/igv.vcf

- VCF format:

  http://genome.ucsc.edu/FAQ/FAQformat.html

  - Text file
  - Header including information about the mapping and variant calling processes: set of lines beginning with ##
  - An additional header line beginning with #: contains a table header with column identifiers
  - One line for each variant: chromosome, genomic position, reference and alternative bases
Visualizing variants

- VCF files need to be indexed before being loaded:

```
igvtools index /home/biouser/mda13/mqc-igv/igv.vcf
ls -la /home/biouser/mda13/mqc-igv/igv.vcf*
```

Load a VCF file

- File → Load from file → igv.vcf A new track is added

Move to 21:47786817

- A peak appears indicating the variant position.
Visualizing variants

- Leave the cursor over that position on the vcf track:
Overview

Interface

Loading Data

Browsing the Data

Sessions

Exercises

Visualization of Mapped Reads: Integrative Genomics Viewer (IGV)
Create/Open/Save Sessions

- Your current session can be saved
Exercise 1

- Load /home/biouser/mda13/mqc-igv/igv2.bam
- Load /home/biouser/mda13/mqc-igv/igv1.bam
- Move to chr11:1,016,174-1,018,316
- Can you see any difference between both tracks?
- Is the igv2.bam mapping a good or a bad result?
- Can you think of the reasons that lead to such result?
- Can you think of any way of improving it?
Exercise 2

- Load `/home/biouser/mda13/mqc-igv/igv1.bam`
- Move to 21:47917047
- What is happening in our sample in that position?
Exercise 3

- Load /home/biouser/mda13/mqc-igv/igv1.bam
- Move to 21:26973663
- What is happening in our sample in that position?
Exercise 4

- Load /home/biouser/mda13/mqc-igv/igv1.bam

- Which of these variants would you trust? Why? Which is the sequence change in each case?
  1. 21:47821726
  2. 21:46596230
  3. 21:42848560
  4. 21:47917170
Exercise 5

Considering that /home/biouser/MDA13/mqc-igv/igv.bed contains the target regions of a given sequencing experiment and that /home/biouser/MDA13/mqc-igv/igv1.bam is the mapping result, how well are the target regions of the gene MX2 covered?

Focus on a particular target region and have a look at the histogram on the top of the igv1.bam track.

- Which is the shape of the histogram? Can you explain this?
- Which would be the desired shape?
Exercise 6

- Considering that /home/biouser/mda13/mqc-igv/igv.bed contains the target regions of a given sequencing experiment and that /home/biouser/mda13/mqc-igv/igv1.bam is the mapping result, look for unsequenced target regions of the PRMT2 gene.