Visualization of mapped reads

Integrative Genomics Viewer (IGV)

University of Cambridge

Cambridge, UK

4th March 2014

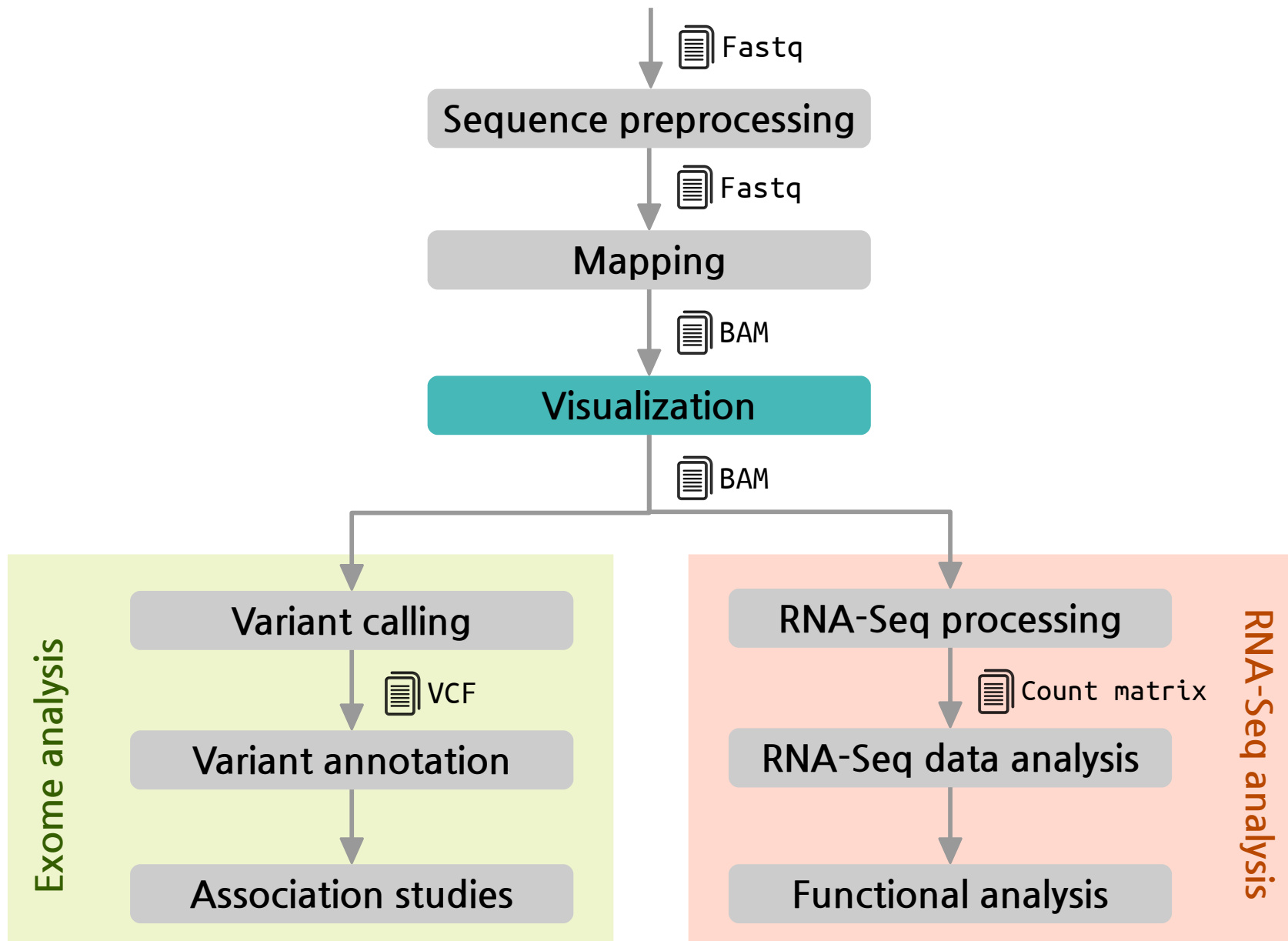**Marta Bleda Latorre**

mbleda@cipf.es

PhD Student at the Computational Genomics Institute
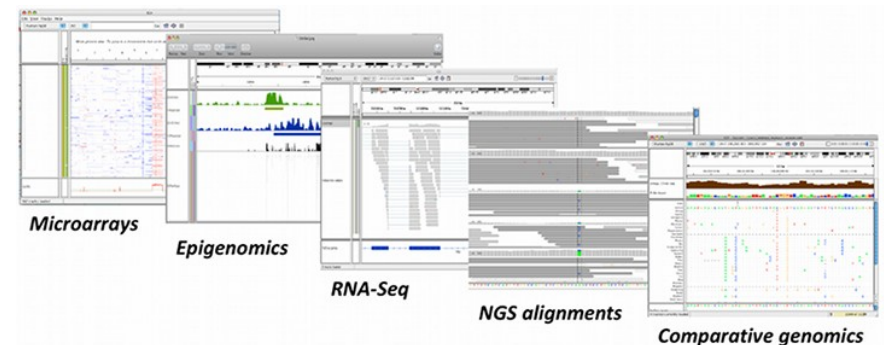
Centro de Investigación Príncipe Felipe (CIPF)

Valencia, Spain

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

ciberer
CENTRO DE INVESTIGACIÓN BIOMÉDICA EN RED
DE ENFERMEDADES RARAS

cibererBIER
PLATAFORMA BIOINFORMÁTICA DE ENFERMEDADES RARAS

The pipeline

Introduction

- **Large** quantities of genomic **data** (NGS, array based methods, etc)

- **Human interpretation** and judgment using visualization can help complex biological relationships

- Integrative Genomics Viewer (**IGV**)

  - **Integrate** different data types simultaneously

  - View **large datasets** easily

  - Faster navigation or browsing

  - Runs **locally** on your desktop

  - Used by large-scale projects

  - Open source and **freely available**

Microarrays · Epigenomics · RNA-Seq · NGS alignments · Comparative genomics

THE CANCER GENOME ATLAS

1000 Genomes
A Deep Catalog of Human Genetic Variation

MMRC®
Multiple Myeloma
Research Consortium
Powerful Collaboration Accelerates Results℠

Helga Thorvaldsdóttir, James T. Robinson, and Jill P. Mesirov
Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration
Brief Bioinform (2013) 14 (2): 178-192

Data types

- Any data related to **genome coordinates**

- **Sample** annotations or attributes

- **Genome** annotations

Recommended file formats

Source data	Recommended File Formats
Sequence alignment data	SAM (must be sorted/indexed) BAM (must be indexed)
Genome annotations	GFF or GFF3 format BED format
ChIP-Seq, RNA-Seq	TDF format (can be generated using igvtools)
Any numeric data	IGV format, TAB format WIG format
Gene expression data	GCT format RES format

Indexing a BAM file

- BAM format: Binary **SAM** file → Reduces disk space and time

  - For each read, provides the position(s) where it maps and information about the alignment

  - BAM files need to be **indexed** (using **samtools**) → SAM files will be sorted by start position and indexed

Index an example BAM file

```
samtools index NA12878_child.bam
samtools index NA12891_dad.bam
samtools index NA12892_mom.bam
```

Take a look at the file size

```
ls -lh
```

Registration and download

1. Be sure that **Java 6 or later** is installed on your machine

2. Go to the IGV website:

   http://www.broadinstitute.org/igv/home

3. Click **Downloads** at the left panel

4. Click to register and fill the form

5. Download the most suitable file for your system

6. Run IGV
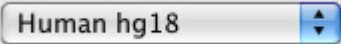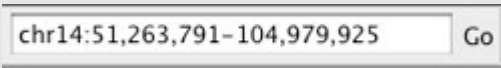
```
./igv.sh
```
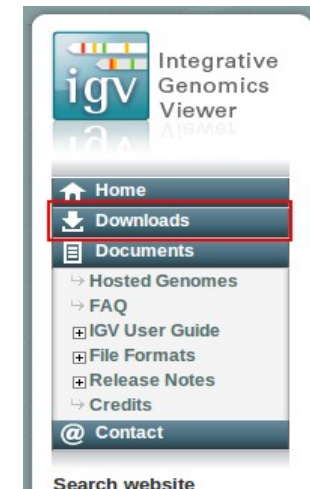
IGV interface
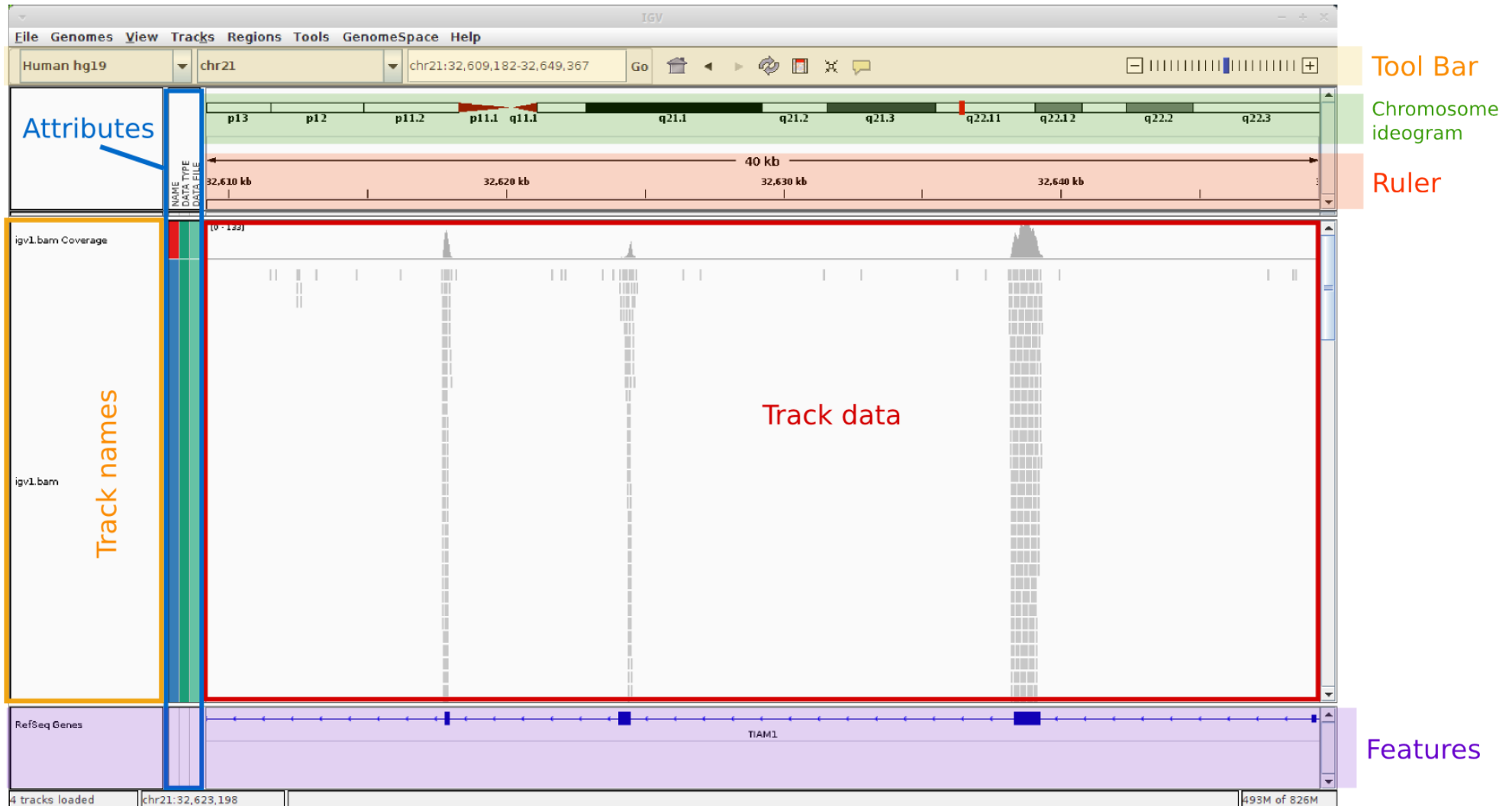
IGV Tool Bar

|	
Human hg18	Genome drop-down box
chr1	Chromosome drop-down box
chr14:51,263,791–104,979,925 Go	Search box
🏠	Whole genome view
◀ ▶	Moves backward and forward through views of the genome
♻	Refreshes the display
▣	Defines a region of interest on the chromosome
✸	Reduces the row height on all tracks to fit all data
💬	Toggles the pop-up information windows in IGV on or off
⊟ ❙❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘❘ ⊞	Zooms in and out on a chromosome

Download genome

Initially, the genome drop-down list contains a single item, "**Human hg18**"

IGV provides **a number of genomes that are hosted** on a server at the Broad Institute

List of genomes hosted: http://www.broadinstitute.org/software/igv/Genomes
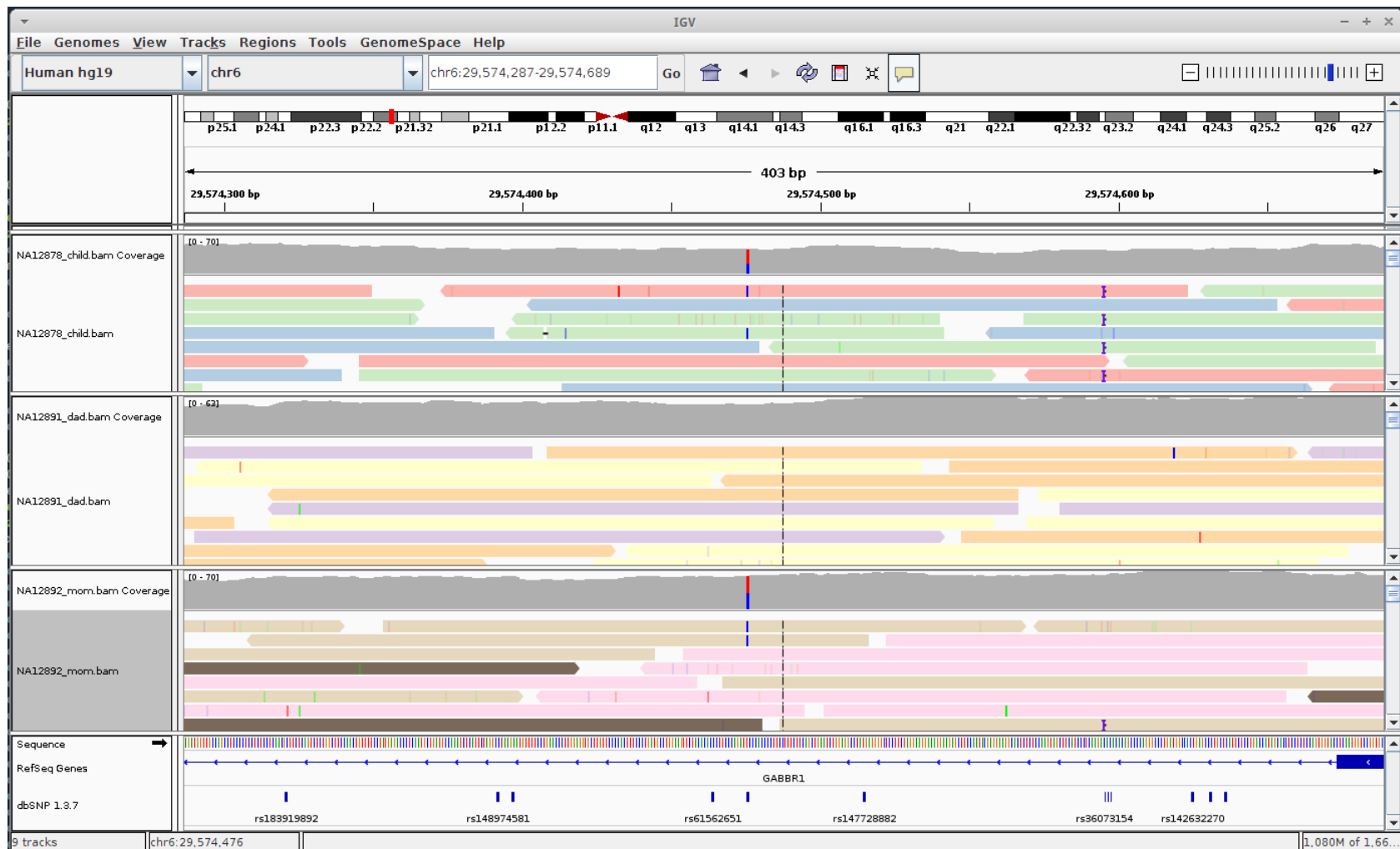
- Genomes → Load genome from server…

  Select **Human hg19**
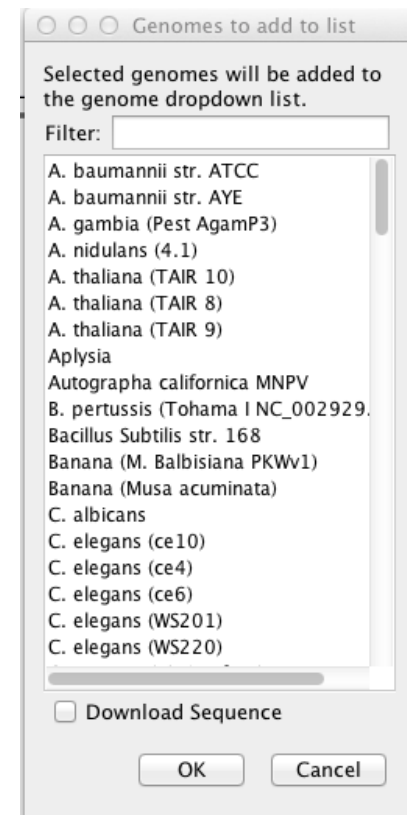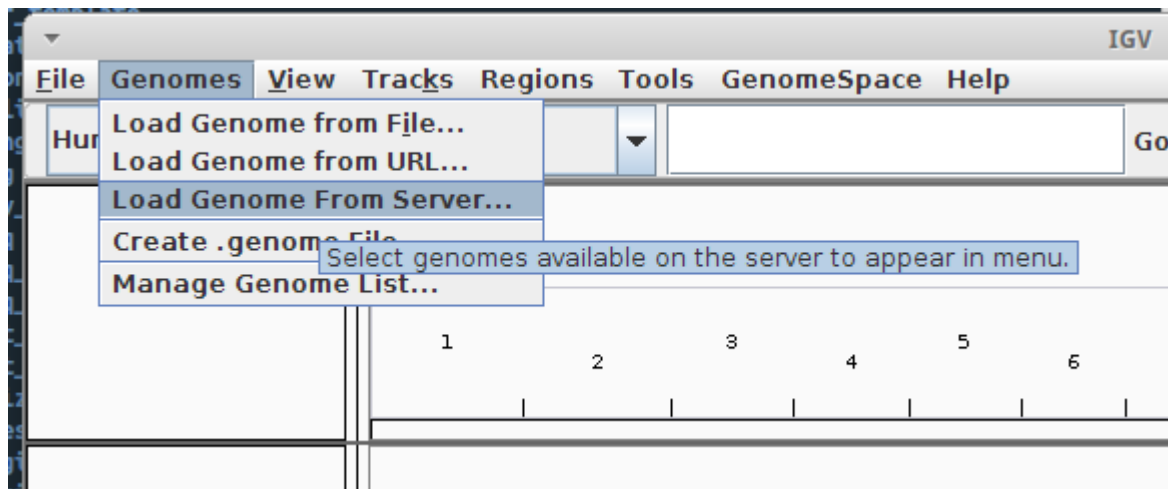
Loading and browsing files

- File → Load from file…
  Select `NA12878_child.bam`, `NA12891_dad.bam` and `NA12892_mom.bam`

Hands on!