Software & Requirements

Javier Perales-Patón
jperales@cnio.es

Fátima Al-Shahrour
[falshahrour@cnio.es]
Elena Piñeiro-Yáñez
[epineiro@cnio.es]
Pedro Fernandes
[pfern@igc.gulbenkian.pt]

cnio
Centro Nacional de Investigaciones Oncológicas

Bioinformatics Unit
CNIO. Madrid, Spain.

INSTITUTO GULBENKIAN DE CIÊNCIA

Computational Facilities

- The **genome reference** must be **indexed** only once in your computer, but it takes a long time (2-3 days). The sizes of the indexes are huge (~100 Gb). See "Before running a variant analysis for first time" at: http://rubioseq.bioinfo.cnio.es/sites/default/files/PDF/RUbioSeq-book.pdf

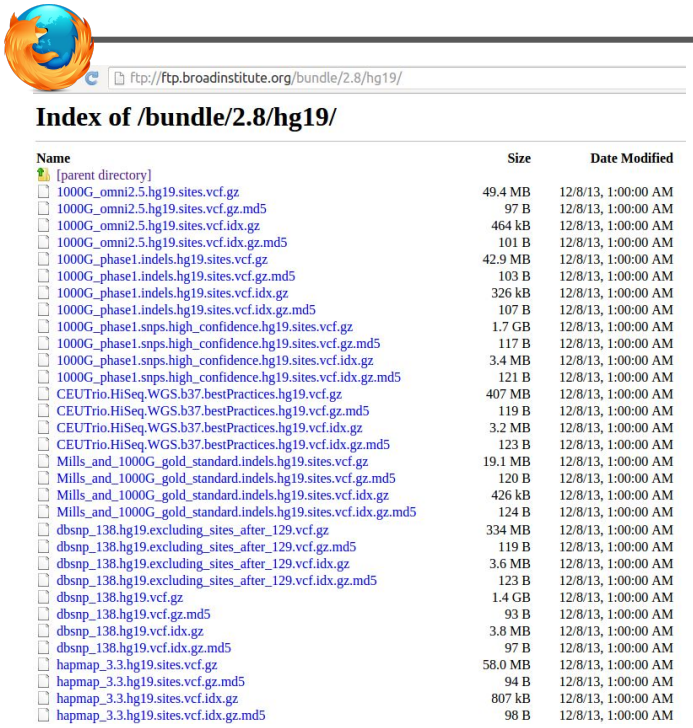
- You need a huge **hard disk capacity**:
  - **Human genome reference** (including indexes) : 150 Gb.
  - **Bundle of files for Variant Calling Analysis** (GATK) : 14 Gb.
  - **For each sample**:
    - Raw data : 7 - 15 Gb.
    - Intermediate files for each sample : 15 - 20 Gb (whole-exome seq).

WorkStation:

- **Minimal requirements:**
  - 16 Gb RAM
  - 500 Gb hard disk
  - 8 threads ~ 8 cores.

- **Recommended requirements (multi-sample analysis, samples storage):**
  - 25 Gb RAM
  - 1 Tb hard disk (e.g. Exome-seq analysis from 50 samples → 1.5 Tb).
  - 16 threads.

GATK's bundle from FTP repository (14 Gb)



You can download the different bundles from GATK's FTP (Broad Institute) visiting this URL with your Internet Browser:

ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/2.8/

1. **Genome Reference** (standard 1000 Genomes, fasta).
2. List of **Target beats or intervals** of genomic regions sequenced by the Library protocol.
3. **dbSNP** (VCF file) for a recent dbSNP release (build 138, it includes the 1000 Genomes).
4. HapMap genotypes and sites VCFs
5. **OMNI 2.5 genotypes for 1000 Genomes samples** (VCF).
6. The current best set of **known indels** to be used for local realignment); use both files:
   - 1000G_phase1.indels.b37.vcf (currently from the 1000 Genomes Phase I indel calls)
   - Mills_and_1000G_gold_standard.indels.b37.sites.vcf

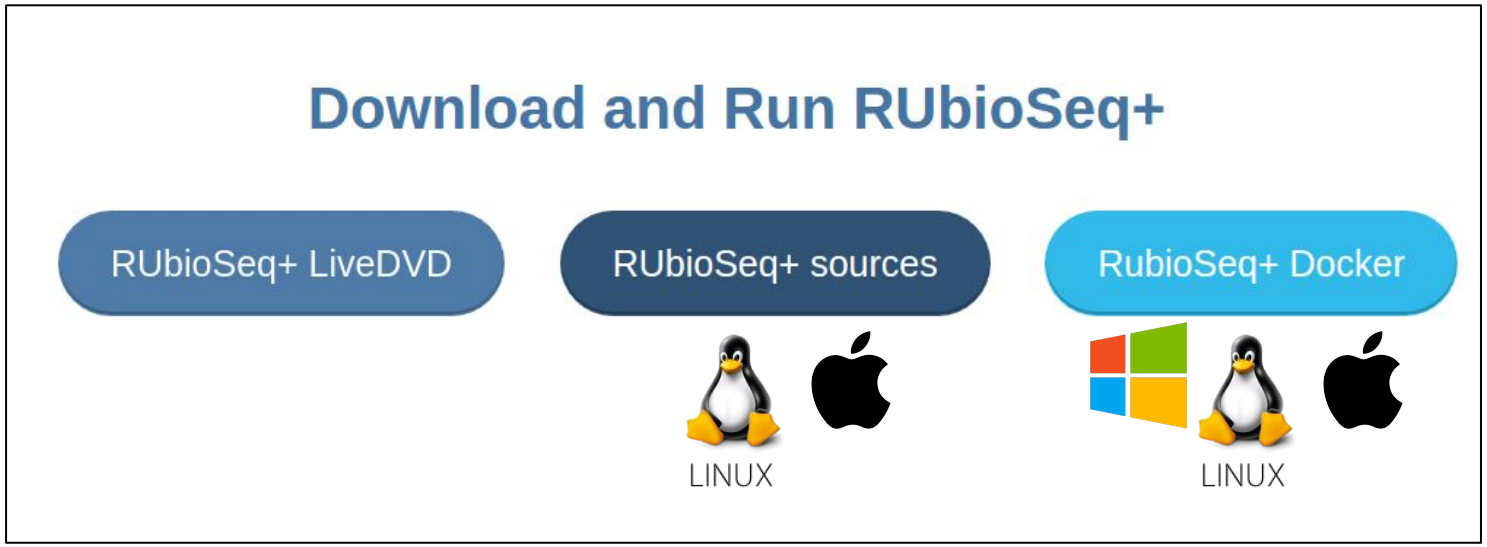
Tip for home: The following UNIX command downloads the whole bundle for **hg19** in one step (~hrs) :

```
$ wget -r -nH --cut-dirs=2 --reject-regex "NA12878|CEUTrio" \

-P /path/to/your_directory/ ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle/hg19/*
```

Installing RUbioSeq+ in your computer

http://rubioseq.bioinfo.cnio.es/



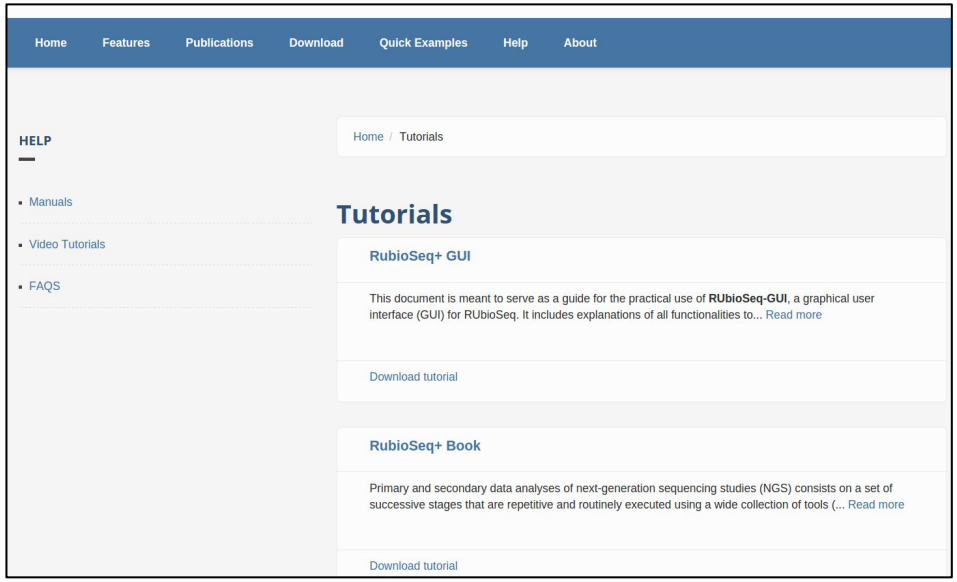
Ask for some help

RUbioSeq is under **development** by the Bioinformatics Unit at CNIO (Madrid, Spain):

- Miriam Rubio-Camarillo (mrubioc@cnio.es).
- José María Fernández (jmfernandez@cnio.es).
- Gonzalo Gómez-López (ggomez@cnio.es).

http://rubioseq.bioinfo.cnio.es/tutorials

- Video-tutorials.
- Example exercises.



RUbioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses
Miriam Rubio-Camarillo,*,Gonzalo Gómez-López, José M. Fernández, Alfonso Valencia and David G. Pisano.
2013, Bioinformatics, 29 (13): 1687-1689

RubioSeq+ Docker

Quick Manual Installation

(Virtual machine, Ubuntu system)

1. First, install Docker client (depending on your Operating System):
   Follow the corresponding guidelines:
   - Ubuntu (https://docs.docker.com/installation/ubuntulinux/)
   - Mac OS X (https://docs.docker.com/installation/mac/)
   - Windows (https://docs.docker.com/installation/windows/)

2. Second, launch the Docker Image (it will download it from the Internet):
   - Windows:
     (Start Menu): Program Files > Boot2Docker

   - Unix (in terminal):
     `docker run -ti -p 0.0.0.0:8080:8080 --name RUbioSeq ubio/rubioseq:latest /bin/bash`

3. Try it out:

```
perl /home/RUbioSeq/RUbioSeq3.7/RUbioSeq.pl -h
```

More info: http://rubioseq.bioinfo.cnio.es/rubioseq_docker

RUbioSeq+ sources

Quick Manual Installation

Dependencies (Ubuntu system)

https://cniobu.github.io/pm17/

Software Installation

Software Installation Quick Guide (root user)

RUbioSeq+ sources

Quick Manual Installation

Dependencies (Ubuntu system)

https://cniobu.github.io/pm

pm17 Precis

Instituto Gulbe

14-17 Nov

Software Installation

Software Installation Quick Guide (ro

PM17 - Software Dependencies

Javier Perales-Paton
jperales@cnio.es

Quick manual installation (root user in Ubuntu 16)

1. Install **Java (version 1.7 or 1.8) in ubuntu.**
When the installation is finished, check that Java is correctly installed:

```
participant@machine:~$ java -version
openjdk version "1.8.0_151"
OpenJDK Runtime Environment (build 1.8.0_151-8u151-b12-0ubuntu0.16.04.2-b12)
OpenJDK 64-Bit Server VM (build 25.151-b12, mixed mode)
```

2. Install **python**

```
sudo apt-get install python-dev
```

3. Get the library **libmysqlclient**

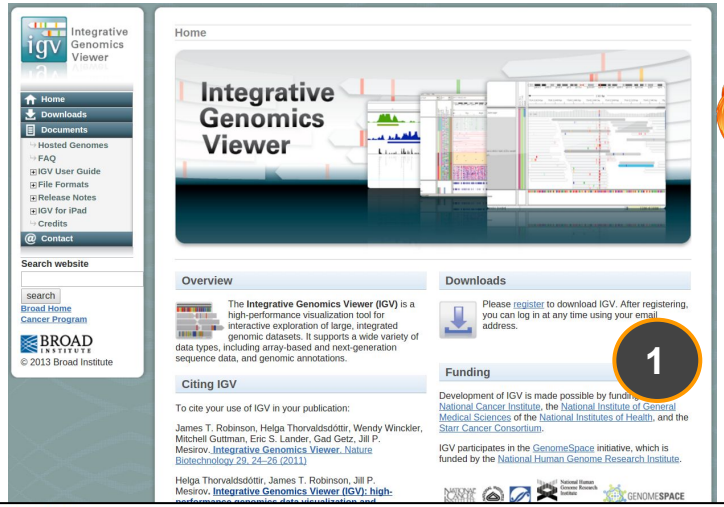
```
sudo apt-get install libmysqlclient-dev
```

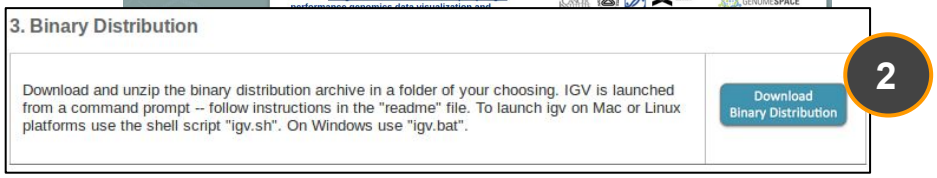
4. **Perl Modules ::**
Rubioseq built on perl modules which must be correctly installed. If one of these is missing, the pipeline will get a crash.

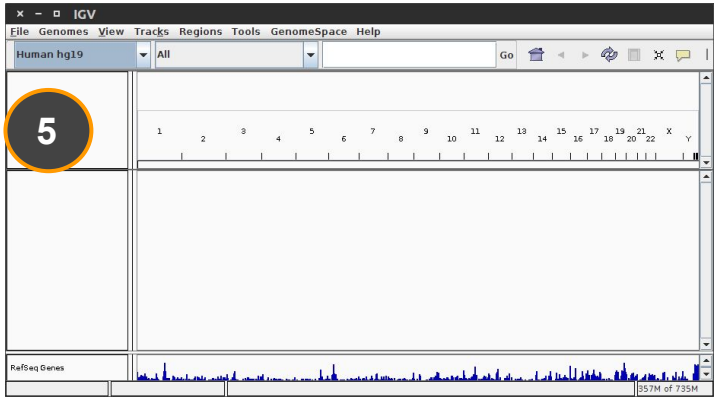
- DBI
- DBD::mysql
- XML::LibXML
- Carp
- FindBin
- File::Basename
- File::Spec
- File::Copy
- Getopt::Long
- Class::Inspector

```
sudo cpan DBI DBD::mys
```

Integrative Genomics Viewer (IGV)







Open Firefox.
https://www.broadinstitute.org/igv/

1. Download section: Register and Fill out the form.

2. Download the Binary distribution: file.zip
Alternatively, you can use this link:
http://data.broadinstitute.org/igv/projects/downloads/IGV_2.3.66.zip

3. Extract the Zip file.

4. Go to the new Directory. Click on igv.jar