COMPUTER RESOURCES

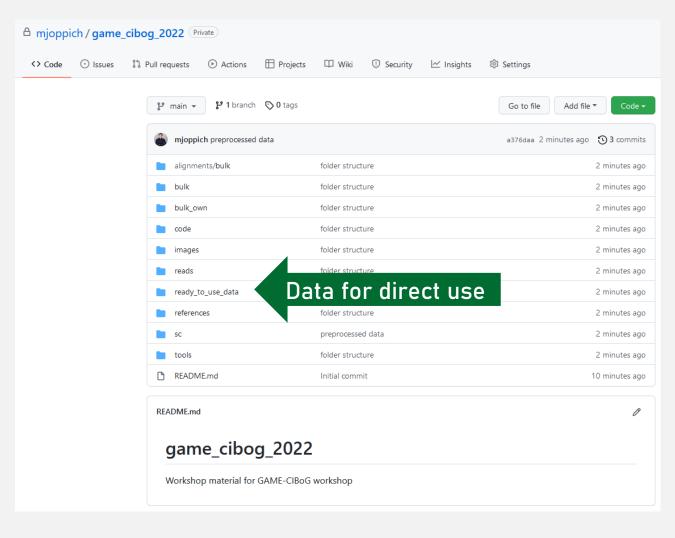
- Computer
- Some kind of Linux
 - You use Linux? → you're fine (but need Win/Mac to run Loupe browser)
 - You use Windows? → use Windows Subsystem for Linux → you're fine





- You use Mac OS? → you're mostly fine, but I can't help you
 - My Mac experience dates back to Mac OS X Leopard!

GIT REPOSITORY



- git clone <u>https://github.com/mjoppich/game_cibog_2022.git</u>
- Folder structure useful (not required though!) for workshop
- Send eMail to
 - joppich@bio.ifi.lmu.de
 - with Github username to get access

SOFTWARE RESOURCES

You need some R version

```
mjoppich@spectre3:/mnt/c/Users/mjopp$ R --version
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
```

- You may want to use jupyter+R kernel
 - install.packages("devtools")
 - devtools::install_github("IRkernel/IRkernel")
 - IRkernel::installspec()
- Some kind of IDE or jupyter-lab in the browser
 - Big fan of Visual Code



- View PDF, PNG, JPEG
- Supports jupyter notebooks and plots

```
options(jupyter.plot_mimetypes = 'image/png')
options(repr.plot.width = 1, repr.plot.height = 0.75, repr.plot.res = 100)
```

REQUIRED TOOLS FOR DOWNLOADING DATA

Getting the read data

- sratoolkit for downloading files
 - wget --output-document tools/sratoolkit.tar.gz http://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
 - cd tools && tar -vxzf sratoolkit.tar.gz
 - vdb-config -interactive
 - Make path to sratoolkit visible:
 - Add

```
export PATH=$PATH:/mnt/t/workshop/tools/sratoolkit.2.11.2-ubuntu64/bin to ~/.bashrc and source ~/.bashrc
```

 Then we can call fastq-dump as fastq-dump --gzip SRR9048093 for downloading specific SRA-files

REQUIRED FILES AND REFERENCES

Doing the alignment

- Hisat2
 - sudo apt install hisat2
 - http://daehwankimlab.github.io/hisat2/
 - Index/Reference genome: H. sapiens, GRCh38, https://genome-idx.s3.amazonaws.com/hisat/grch38_genome.tar.gz
 - tar xfz grch38_genome.tar.gz
- featureCounts / subread
 - sudo apt install subread
 - https://sourceforge.net/projects/subread/files/subread-2.0.3/
 - Genome Annotation File:
 - http://ftp.ensembl.org/pub/release- 105/gtf/homo_sapiens/Homo_sapiens.GRCh38.105.gtf.gz
 - gunzip Homo sapiens.GRCh38.105.gtf.gz

Binaries

Version: HISAT2 2.2.1

Release Date: 7/24/2020

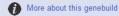
Source https://cloud.biohpc.swmed.edu/index.php/s/fE9QCsX3NH4QwBi/downlo

OSX_x86_64 https://cloud.biohpc.swmed.edu/index.php/s/zMgEtnF6LjnjFrr/download

Linux_x86_64 https://cloud.biohpc.swmed.edu/index.php/s/oTtGWbWjaxsQ2Ho/downlo

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.



Download FASTA files for genes, cDNAs, ncRNA, proteins

Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins

Update your old Ensembl IDs



R PACKAGES

```
# ensures we can install everything
library(devtools) => install.packages("devtools")
# accessing files, plotting
library(readr) => install.packages("readr")
library(ggplot2) => install.packages("ggplot2")
library(cowplot) => install.packages("cowplot")
library(ggrepel) => install.packages("ggrepel")
library(pheatmap) => install.packages("pheatmap")
library(RColorBrewer) => install.packages(" RColorBrewer ")
# Bioconductor package manager => required for basically all bioinformatics tasks
if (!requireNamespace('BiocManager', quietly = TRUE))
   install.packages('BiocManager')
library(EnhancedVolcano) => BiocManager::install('EnhancedVolcano') or devtools::install github('kevinblighe/EnhancedVolcano')
library(DESeq2) => BiocManager::install('DESeq2')
library(clusterProfiler) # => BiocManager::install('clusterProfiler')
library(DOSE) # => BiocManager::install('DOSE')
library(enrichplot) # => BiocManager::install('enrichplot')
library(ReactomePA) # => BiocManager::install('ReactomePA')
library(org.Hs.eg.db) # => BiocManager::install('org.Hs.eg.db')
library(disgenet2r) => devtools::install bitbucket("ibi group/disgenet2r")
```

DOWNLOADING CELLRANGER+LOUPE

Cell Ranger - 6.1.2 (October 25, 2021)

- Self-contained, relocatable tar file. Does not require centralized installation.
- Contains binaries pre-compiled for CentOS/RedHat 6.0+ and Ubuntu 12.04+.
- Download Linux 64-bit 768 MB md5sum: 310d4453acacf0eec52e76aded14024c



wget -0 cellranger-6.1.2.tar.gz "https://cf.10xgenomics.com/releases/cell-exp/cellranger
6.1.2.tar.gz?Expires=1642730466&

Policy=eyJTdGF0ZWllbnQiOlt7IIJlc291cmNlIjoiaHR0cHM6Ly9jZi4xMHhnZW5vbWljcy5jb20vcmVsZWFzZXMvY2VsbC11eH
AvY2VsbHJhbmdlci0ZLjEuM550YXIuZ3oiLCJDb25kaXRpbZ4iOnsiRGF0ZUxlc3NUaGFuIjp7IkFXUzpFcG9jaFRpbWUiOjE2NDI
3MzA0NjZ9fXldfQ_8Signature=USFWTdaZIygJwvTtlCSo7xZdWZmDq2CH~w~0tlWglCUdTVNLKbGcjqYXLN8FIAEWeaCckQu4AJrW-

bbNRGwF70MS9oB5WheoRAopZyKXuNaCt97tqKs9NN1UpwAbRdgD0JFz1QdKMCT3cg3~AXRLOxFnxWnNVvfjcpcjrN1dXv7qcS862t VezERtC0OKrwCR1syjFnCKKxmHhiO4xPRAdoOs8UhoFloXLqB-WsLM7f~CSNJ4-dV-RfRGS-

@ePVuU~s2mOGdDtb0F66je612kZo433Z7SleX11GWK6ktwpJEBSqkzNoJfYCDjTw0HCXU~F4X8nTh6rpGNEv5a8QEhA__&Key-Pair-Id=APKAT7S6ASRYOXRWRPDA"

- Human reference (GRCh38) dataset required for Cell Ranger.
- Download 11 GB md5sum: dfd654de39bff23917471e7fcc7a00cd
- Build steps

curl wget

wget https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-GRCh38-2020-A.tar.gz





Loupe Browser 6.0.0 (October 7, 2021)

Please follow the install instructions after downloading one of the installers below.

Read about what's new in Loupe Browser 6.0



filesize: 694 MiB

md5sum: 8fce42ab47c3d040d7a1538be8d33900



filesize: 784 MiB

md5sum: aee779daa98bc101092169666fd1c2aa

- https://support.10xgenomics.com/single-cell-geneexpression/software/downloads/latest
- Might require brief registration