Université de Paris

UMR7216 Epigénétique & Destin Cellulaire

BIBS platform

Maturing In the Time of Coronavirus

Logo BIBS ?

14/09/2020

Magali Hennion

BIBS : BioInformatics and BioStatistics

Steering committee (CoPil)

Valérie Lallemand-Mezger

Claire Rougeulle

Olivier Kirsh

Jean-François Ouimette

Florent Hubé

Magali Hennion

Pierre Poulain (IJM)

UMR 7216 Epigénétique & Destin Cellulaire

Active members

Magali Hennion

Olivier Kirsh

Call for projects

"Projets Pilotes"

Call initiation : March 3rd

Goal for the platform : building an efficient working environment

- how to communicate with the project holder / between us ?
- how to do the analyses to make them reproducible, reusable, independent of the computer/cluster ?

Call for projects

"Projets Pilotes"

Call initiation : March 3rd

Goal for the platform : building an efficient working environment

- how to communicate with the project holder / between us ?
- how to do the analyses to make them reproducible, reusable, independent of the computer/cluster ?

March 30th → 4 projects (Nataliya, Madeleine, Agathe, Guillaume)

CoPil meeting -> decision made public on April 21st

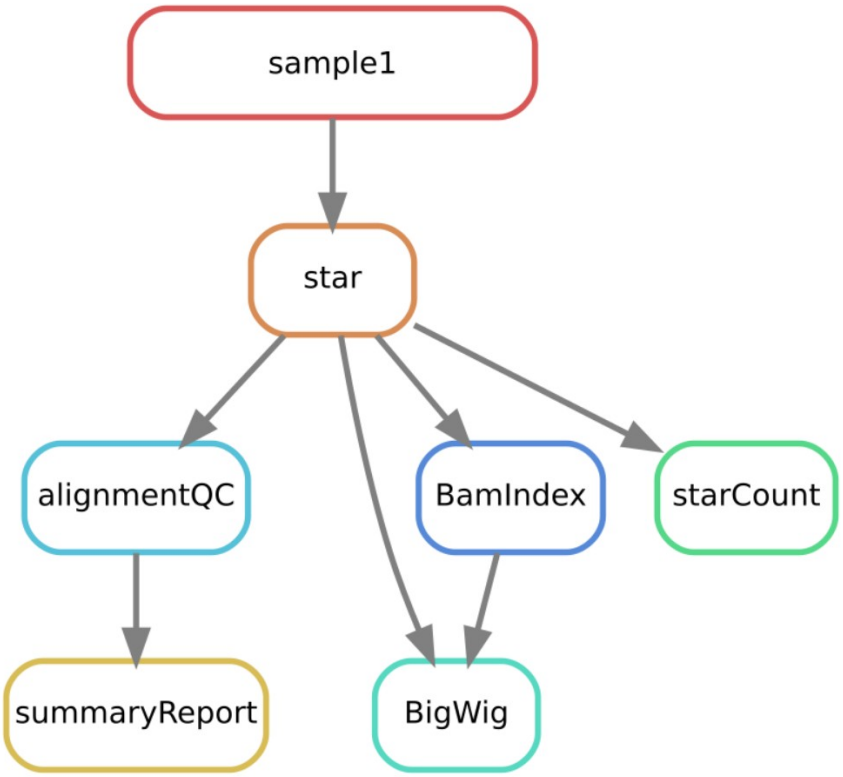
- Try to help everyone
- Start developing a RNA-seq analysis pipeline on Madeleine's data

→ Use it on other datasets, adapt it to other kind of analysis, ...

Why using a workflow manager?

One analysis : multiple steps, multiple samples

Simple example, one sample

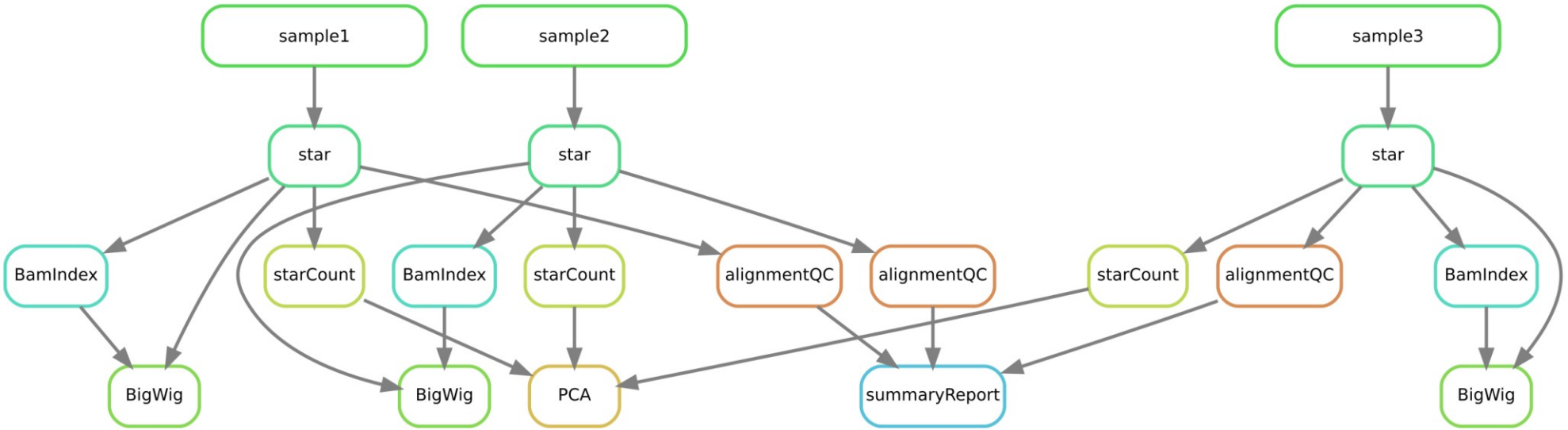


Some tasks can run **in parallel**

Some tasks have to **wait** for one or several others to complete

Why using a workflow manager?

Simple example, more samples (3)



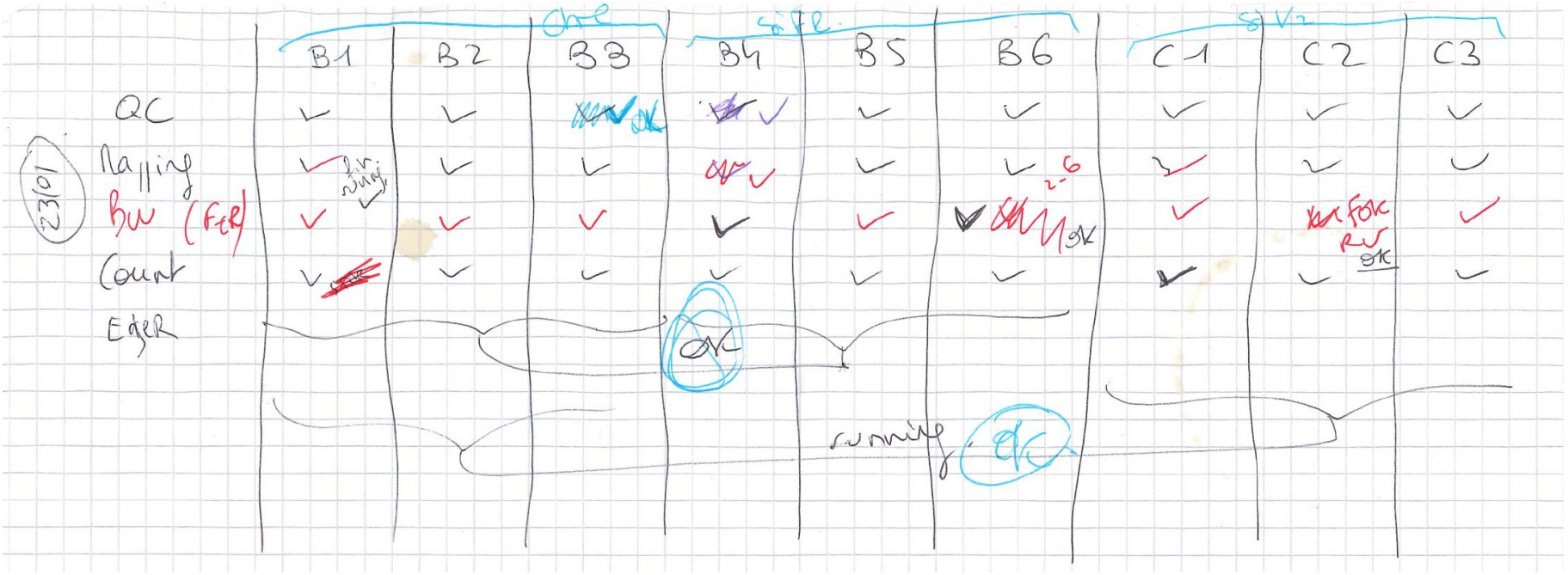
Some tasks are done **for each sample**

Some **combine** the information from several samples

→ Usually more samples, without workflow manager you may get lost...

Why using a workflow manager?

And/or end up with something like that...



Workflow manager

- more efficient, faster
- more reproducible
- modular → library of reusable blocks (any language)

RNA-seq analysis workflow

→ Facilitate primary and secondary analysis of RNA-seq data

Based on a published workflow

Zhang and Jonassen *BMC Bioinformatics* (2020) 21:110
https://doi.org/10.1186/s12859-020-3433-x

BMC Bioinformatics

SOFTWARE Open Access

RASflow: an RNA-Seq analysis workflow with Snakemake

Xiaokang Zhang and Inge Jonassen*

RNA-seq analysis workflow

→ Facilitate primary and secondary analysis of RNA-seq data

Based on a published workflow

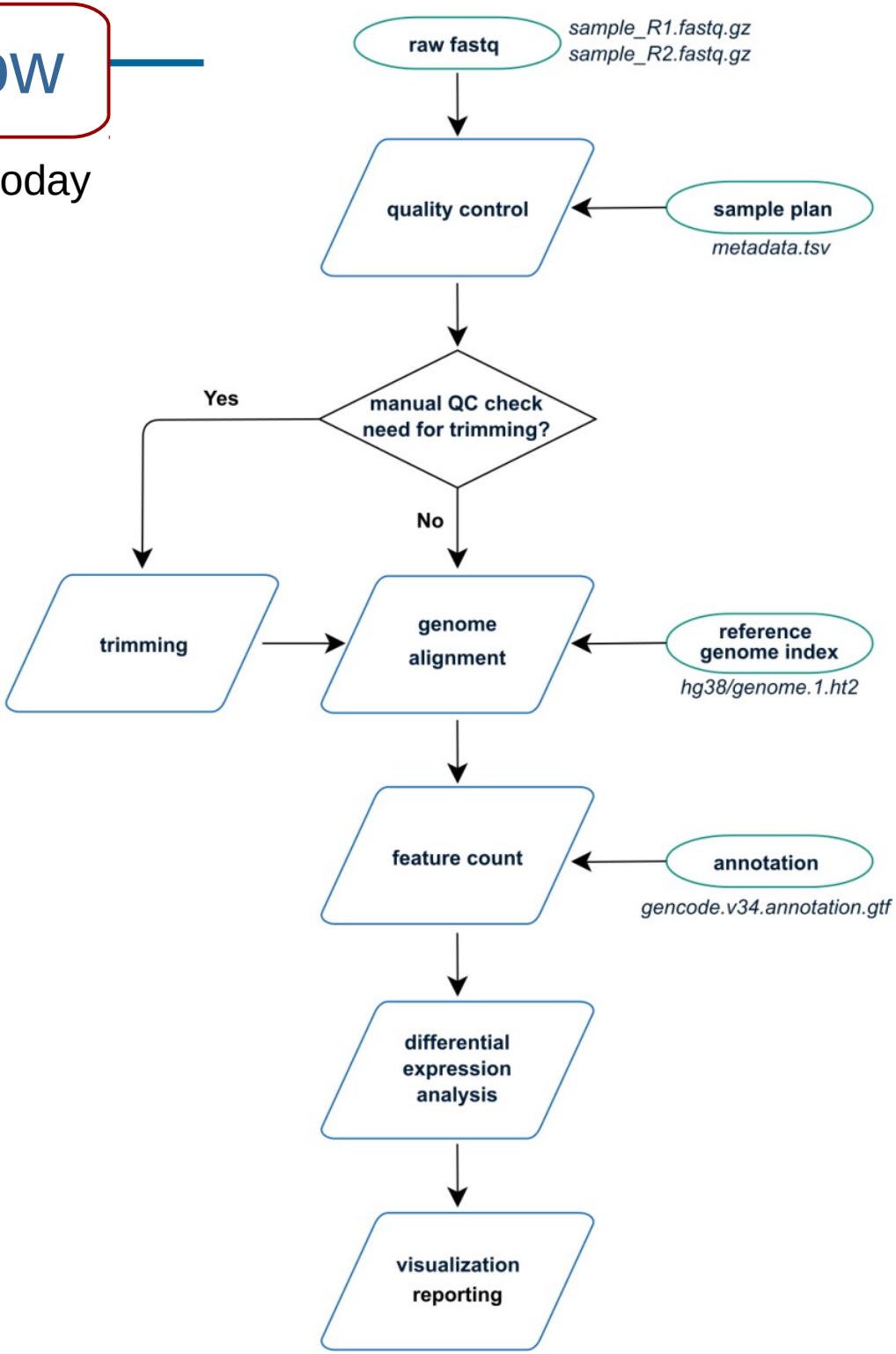
Modified

- to run on IFB and RPBS clusters
- to be computationally more efficient
- to include more tools and more parameters

→ Modular, highly flexible

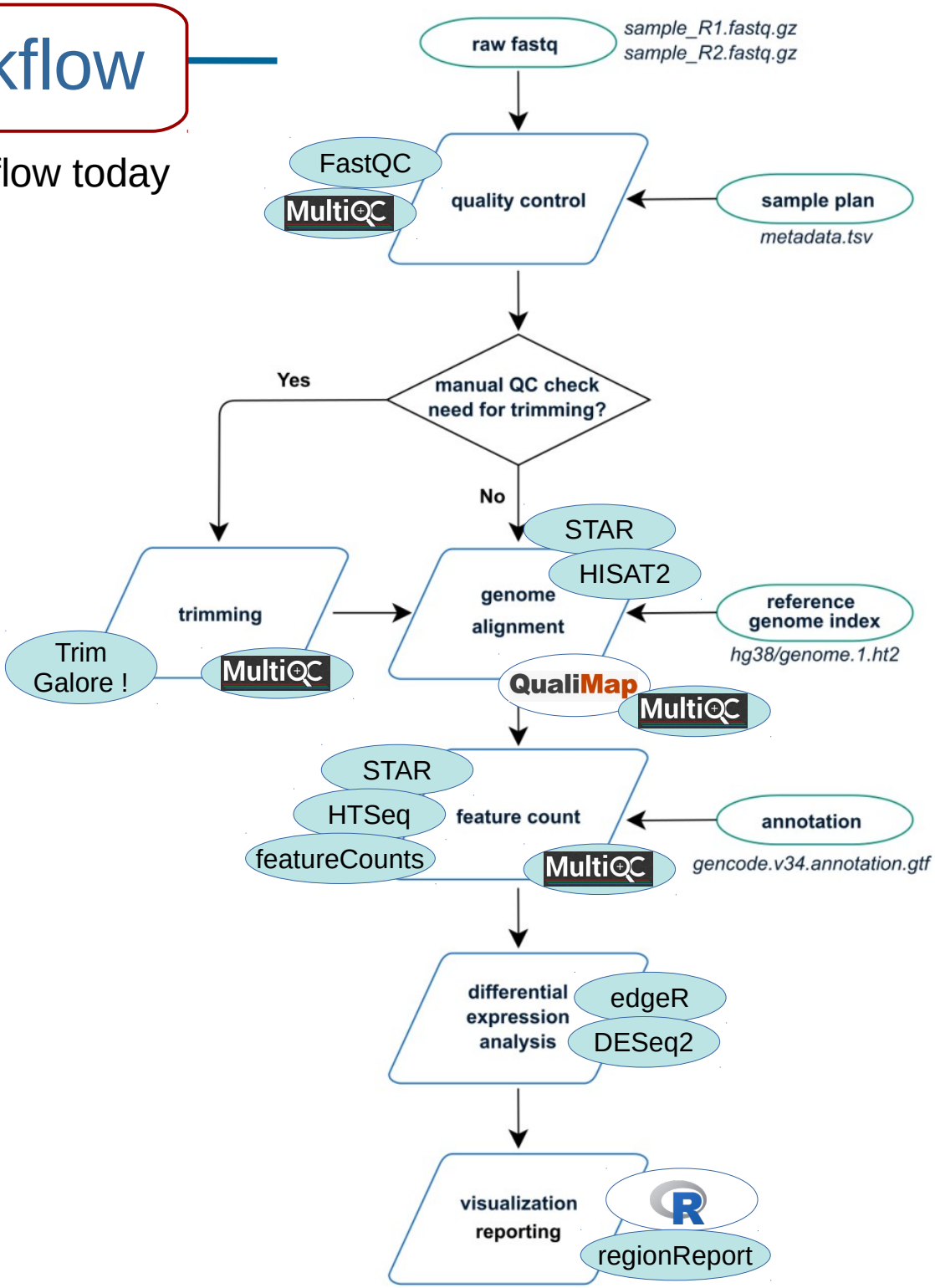
RNA-seq analysis workflow

The workflow today



RNA-seq analysis workflow

The workflow today



RNA-seq analysis workflow

Where to compute ?

- IFB core cluster (for now one installation / project)
- RPBS cluster (one installation for everyone)
- personal computer

RNA-seq analysis workflow

Where to compute ?

- IFB core cluster (for now one installation / project)
- RPBS cluster (one installation for everyone)
- personal computer

Prerequisites

- Minimal shell knowledge to copy your files at the right place, navigate between folders...
- Minimal knowledge about the cluster
- Knowledge about the tools (are they adapted to your question?)

→ The platform is here to help you !

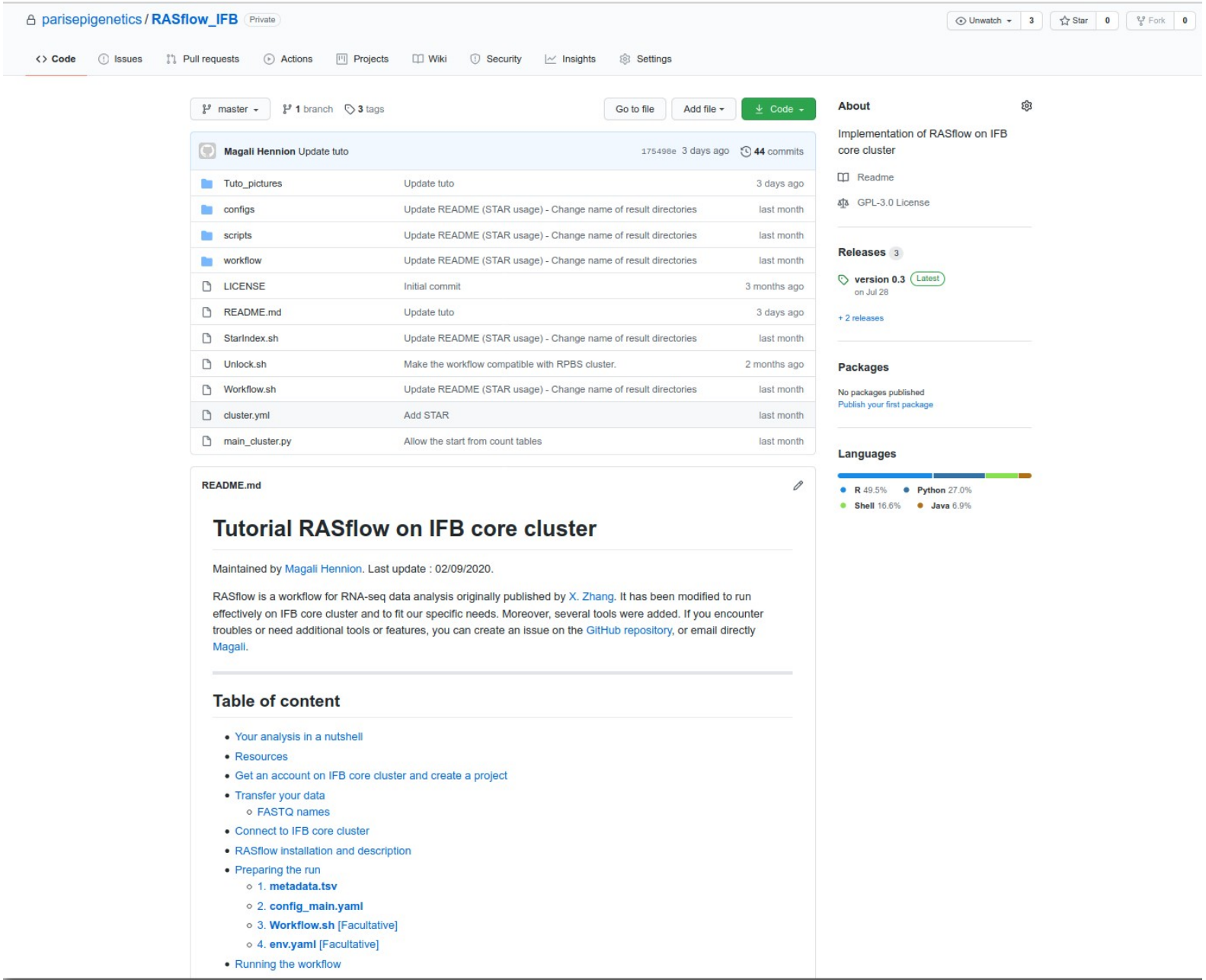
Not needed: SnakeMake, scripting

Advanced user: For now, tools with default parameters
→ add configuration options

RNA-seq analysis workflow

How to start your analysis ?

Tutorial and code available on GitHub



parisepigenetics / RASflow_IFB Private

Unwatch 3 | Star 0 | Fork 0

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

master | 1 branch | 3 tags | Go to file | Add file | Code

Magali Hennion Update tuto — 175498e 3 days ago — 44 commits

Tuto_pictures	Update tuto	3 days ago
configs	Update README (STAR usage) - Change name of result directories	last month
scripts	Update README (STAR usage) - Change name of result directories	last month
workflow	Update README (STAR usage) - Change name of result directories	last month
LICENSE	Initial commit	3 months ago
README.md	Update tuto	3 days ago
StarIndex.sh	Update README (STAR usage) - Change name of result directories	last month
Unlock.sh	Make the workflow compatible with RPBS cluster.	2 months ago
Workflow.sh	Update README (STAR usage) - Change name of result directories	last month
cluster.yml	Add STAR	last month
main_cluster.py	Allow the start from count tables	last month

About

Implementation of RASflow on IFB core cluster

Readme

GPL-3.0 License

Releases 3

version 0.3 Latest on Jul 28

+ 2 releases

Packages

No packages published

Publish your first package

Languages

R 49.5% Python 27.0% Shell 16.6% Java 6.9%

README.md

Tutorial RASflow on IFB core cluster

Maintained by Magali Hennion. Last update : 02/09/2020.

RASflow is a workflow for RNA-seq data analysis originally published by X. Zhang. It has been modified to run effectively on IFB core cluster and to fit our specific needs. Moreover, several tools were added. If you encounter troubles or need additional tools or features, you can create an issue on the GitHub repository, or email directly Magali.

Table of content

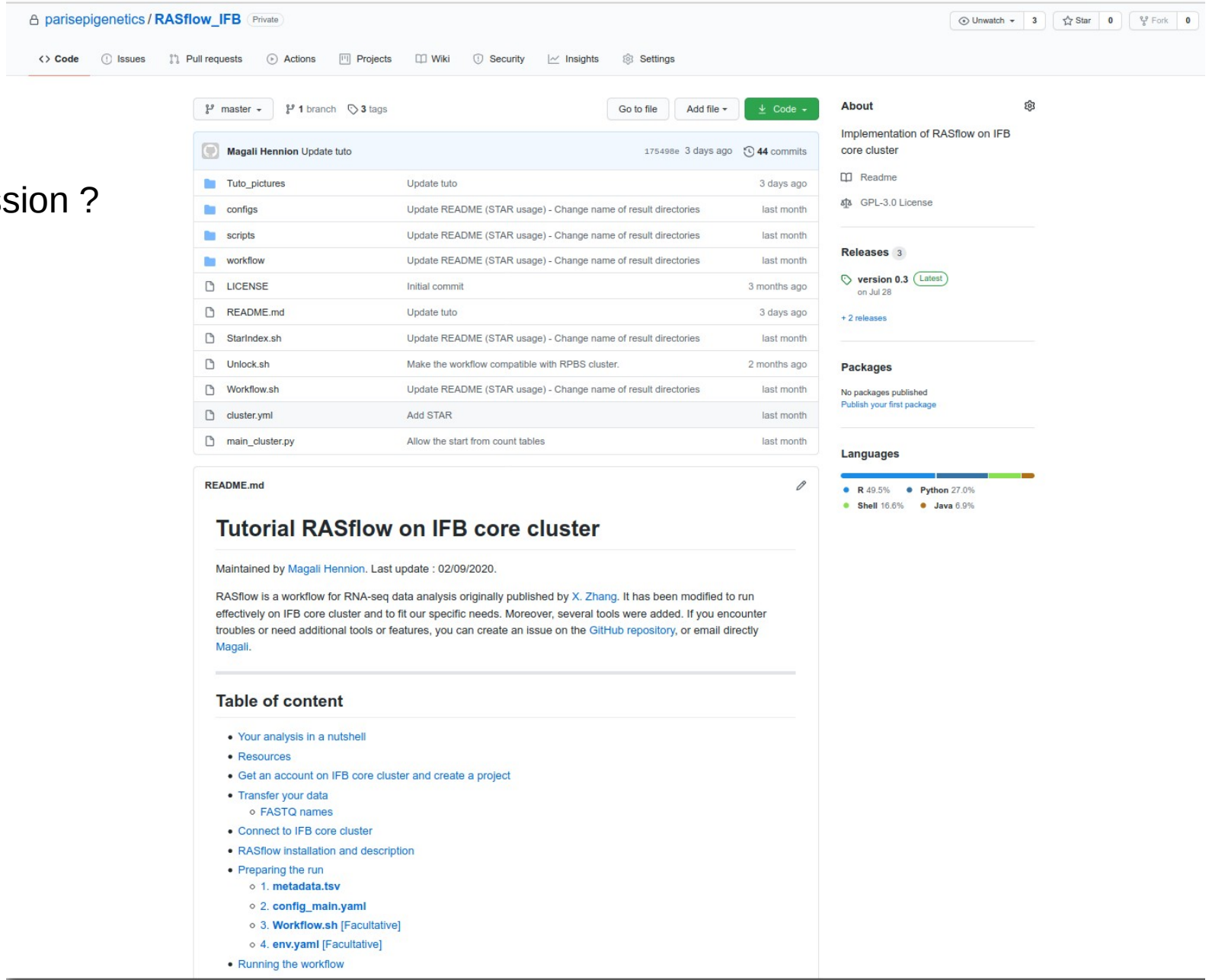
- Your analysis in a nutshell
- Resources
- Get an account on IFB core cluster and create a project
- Transfer your data
  - FASTQ names
- Connect to IFB core cluster
- RASflow installation and description
- Preparing the run
  - 1. **metadata.tsv**
  - 2. **config_main.yaml**
  - 3. **Workflow.sh** [Facultative]
  - 4. **env.yaml** [Facultative]
- Running the workflow

RNA-seq analysis workflow

How to start your analysis ?

Tutorial and code available on GitHub

Need for a training session ?



parisepigenetics / RASflow_IFB Private

Unwatch 3 | Star 0 | Fork 0

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

master | 1 branch | 3 tags | Go to file | Add file | Code

Magali Hennion Update tuto 175498e 3 days ago 44 commits

Tuto_pictures	Update tuto	3 days ago
configs	Update README (STAR usage) - Change name of result directories	last month
scripts	Update README (STAR usage) - Change name of result directories	last month
workflow	Update README (STAR usage) - Change name of result directories	last month
LICENSE	Initial commit	3 months ago
README.md	Update tuto	3 days ago
StarIndex.sh	Update README (STAR usage) - Change name of result directories	last month
Unlock.sh	Make the workflow compatible with RPBS cluster.	2 months ago
Workflow.sh	Update README (STAR usage) - Change name of result directories	last month
cluster.yml	Add STAR	last month
main_cluster.py	Allow the start from count tables	last month

About

Implementation of RASflow on IFB core cluster

Readme

GPL-3.0 License

Releases 3

version 0.3 Latest on Jul 28

+ 2 releases

Packages

No packages published

Publish your first package

Languages

R 49.5% Python 27.0% Shell 16.6% Java 6.9%

README.md

Tutorial RASflow on IFB core cluster

Maintained by Magali Hennion. Last update : 02/09/2020.

RASflow is a workflow for RNA-seq data analysis originally published by X. Zhang. It has been modified to run effectively on IFB core cluster and to fit our specific needs. Moreover, several tools were added. If you encounter troubles or need additional tools or features, you can create an issue on the GitHub repository, or email directly Magali.

Table of content

- Your analysis in a nutshell
- Resources
- Get an account on IFB core cluster and create a project
- Transfer your data
  - FASTQ names
- Connect to IFB core cluster
- RASflow installation and description
- Preparing the run
  - 1. **metadata.tsv**
  - 2. **config_main.yaml**
  - 3. **Workflow.sh** [Facultative]
  - 4. **env.yaml** [Facultative]
- Running the workflow

RNA-seq analysis workflow

Your analysis in a nutshell

- Get an account on IFB core cluster and create a project
- Transfer your data to the cluster
- Clone RASflow_IFB repository
- Modify `metadata.tsv` and `config_main.yaml`
- Run the workflow typing `sbatch Workflow.sh`
- Look at the results

Two files to modify

- sample plan (metadata.tsv)

sample	group	subject
T27	J0_WT	1
T28	J0_WT	2
T29	J0_WT	3
T30	J0_KO	1
T31	J0_KO	2
T32	J0_KO	3
T33	J10_WT	1
T34	J10_WT	2
T35	J10_WT	3
T36	J10_KO	1
T37	J10_KO	2
T38	J10_KO	3

RNA-seq analysis workflow

Two files to modify
- sample plan
- configuration file

```
# Please check the parameters, and adjust them according to your circumstance

# Project name
PROJECT: EXAMPLE

# ================= Control of the workflow ==================

## Do you need to do quality control?
QC: yes  # "yes" or "no". If set to "yes", the workflow will stop after the QC to let you decide whether you want to trim

## Do you need to do trimming?
TRIMMED: yes  # "yes" or "no"?

## Which mapping reference do you want to use? Genome or transcriptome?
REFERENCE: genome  # "genome" or "transcriptome", I haven't implemented transcriptome yet.

## Do you want to do Differential Expression Analysis (DEA)?
DEA: yes  # "yes" or "no"

## Do you want to visualize the results of DEA?
VISUALIZE: yes  # "yes" or "no"
```

RNA-seq analysis workflow

Two files to modify
- sample plan
- configuration file

```
# ================== Shared parameters for some or all of the sub-workflows ==================

## key file if the data is stored remotely, otherwise leave it empty
KEY:

## the path to fastq files
READSPATH: /shared/projects/YourProjectName/Raw_fastq

## the meta file describing the experiment settings
METAFILE: /shared/projects/YourProjectName/RASflow_IFB/configs/metadata.tsv

## paths for intermediate and final results
BIGDATAPATH: /shared/projects/YourProjectName/RASflow_IFB/data # for big files
RESULTPATH: /shared/projects/YourProjectName/RASflow_IFB/results

## is the sequencing paired-end or single-end?
END: pair  # "pair" or "single"

## number of cores you want to allocate to this workflow
NCORE: 30  # Use command "getconf _NPROCESSORS_ONLN" to check the number of cores/CPU on your machine
```

RNA-seq analysis workflow

Two files to modify
- sample plan
- configuration file

```
# ================== Configuration for alignment to genome and feature count ==================

## aligner
ALIGNER: hisat2 # "STAR" or "hisat2"

## genome and annotation files
INDEXPATH: /shared/bank/homo_sapiens/hg38/hisat2 # or index/STAR # folder containing index files
INDEXBASE: genome  # for hisat2, base of the name of the index files (ie genome.1.ht2)
ANNOTATION: /shared/projects/YourProjectName/RASflow_IFB/gtf/gencode.v34.annotation.gtf # GTF file

## bigwig option
BWSTRANDED: both # "no": bw merging forward and reverse reads, "yes": get 2 bw files, one forward a

## tool for feature count
COUNTER: featureCounts # "featureCounts" or "htseq-count" or "STARcount" (only with STAR aligner, -

## counting options
ATTRIBUTE: gene_id  # the attribute used in annotation file. It's usually "gene_id", but double che
STRAND: "reverse" # "no", "yes", "reverse". For ht-seq counts: For stranded=no, a read is considere
FEATURE: transcript # "exon" or "transcript"
```

[...]

→ **start the analysis**

RNA-seq analysis workflow

Running time

Data from Madeleine, 12 samples (pair-end), 140 GB in total, 4-8 GB / fastq files)

- Data transfer : ~ 1h (from the lab)

- Creating Conda environment : 15-30 min

- QC : 40-60 min

- trimming: 2-3 h

- mapping: 5-7 h

- DEA: 5-10 min

- visualization : 5-10 min

RNA-seq analysis workflow

Running time

Data from Madeleine, 12 samples (pair-end), 140 GB in total, 4-8 GB / fastq files)

- Data transfer : ~ 1h (from the lab)

- Creating Conda environment : 15-30 min

Need for a human brain

- QC : 40-60 min

- trimming: 2-3 h

- mapping: 5-7 h

- DEA: 5-10 min

- visualization : 5-10 min

→ TOTAL : ~ 12h computing + human time

RNA-seq analysis workflow

Quality control before and after trimming

file:///home/mag/Documents/Communication/20200914_LabSeminar/report_quality_control.html

file:///home/mag/Documents/Communication/20200914_LabSeminar/report_quality_control_after_trimming.html

RNA-seq analysis workflow

Results

Quality control before and after trimming

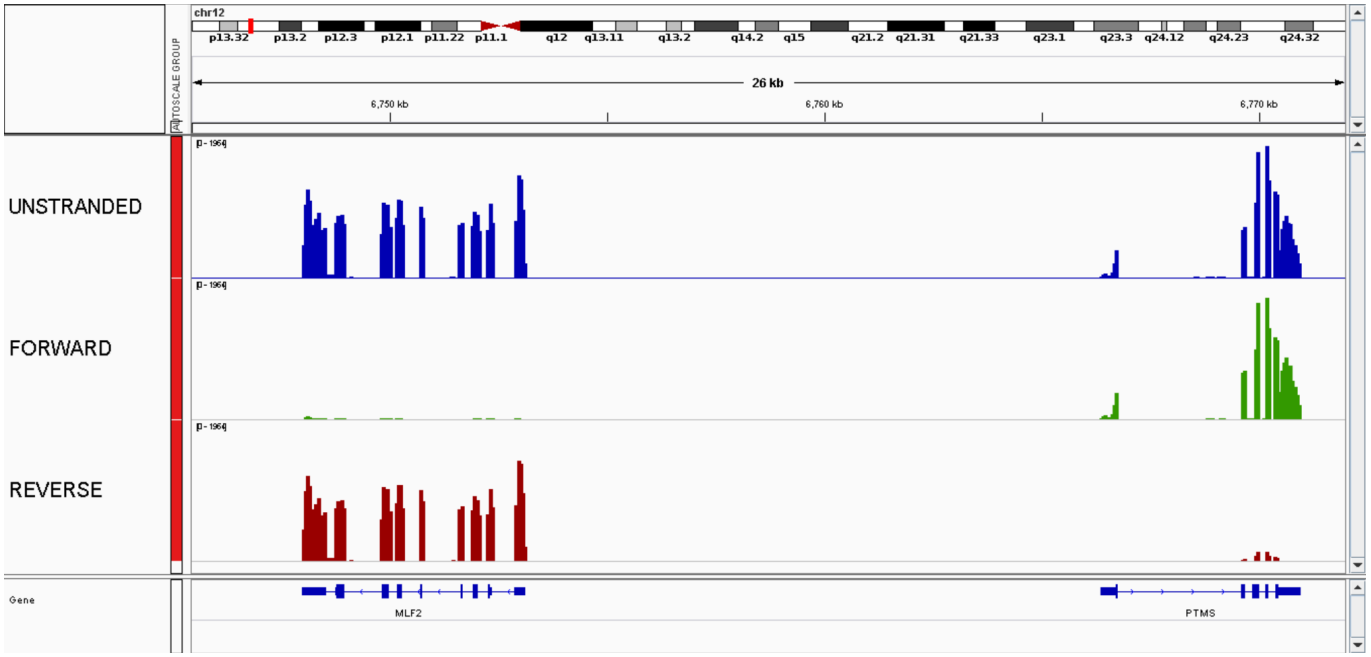
file:///home/mag/Documents/Communication/20200914_LabSeminar/report_quality_control.html

file:///home/mag/Documents/Communication/20200914_LabSeminar/report_quality_control_after_trimming.html

Quality control after mapping and counting

file:///home/mag/DataMadeleine/TIMING/hisat2/report_align_count_featureCounts.html

BigWig files to look at the data on genome browser



RNA-seq analysis workflow

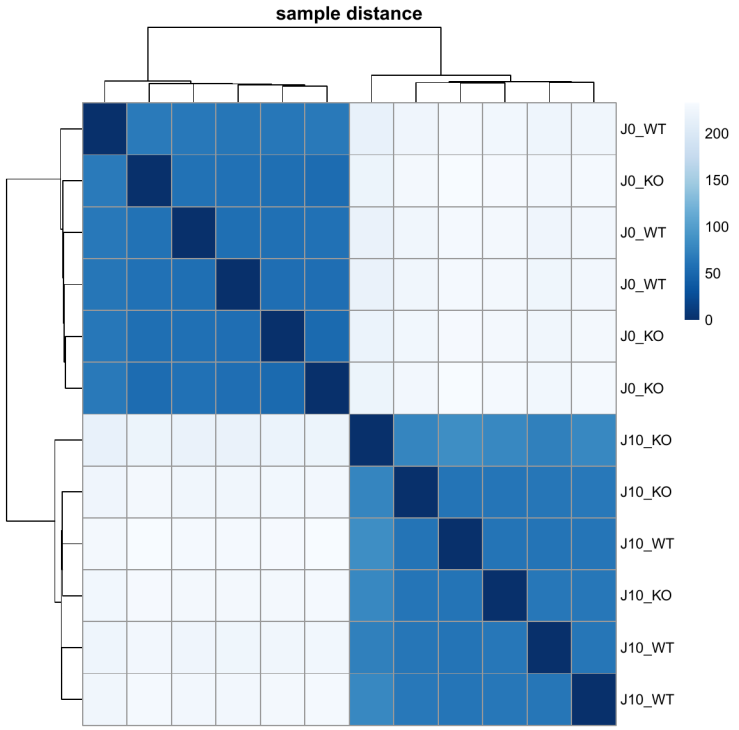
Results

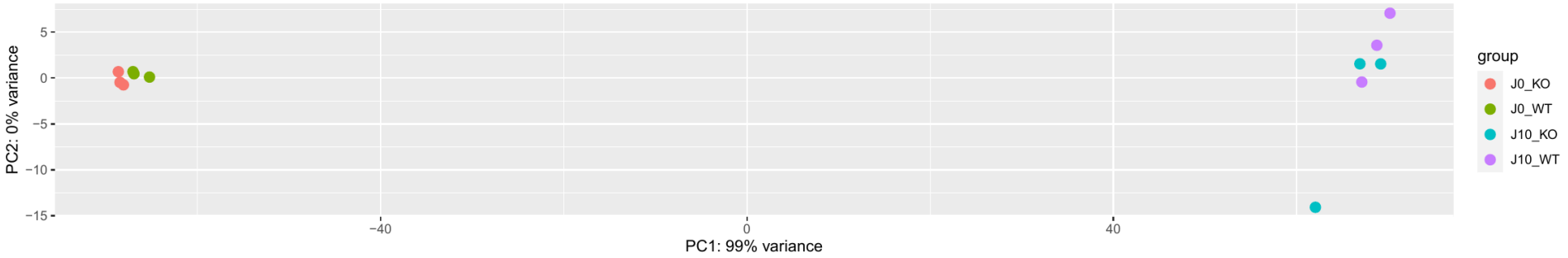
Raw and normalized count files

D197-D192T35_count.tsv

ENSG00000223972.5	1
ENSG00000227232.5	263
ENSG00000278267.1	0
ENSG00000243485.5	1
ENSG00000284332.1	0
ENSG00000237613.2	1
ENSG00000268020.3	0
ENSG00000240361.2	0
ENSG00000186092.6	0
ENSG00000238009.6	26
ENSG00000239945.1	0
ENSG00000233750.3	0
ENSG00000268903.1	19
ENSG00000269981.1	13
ENSG00000239906.1	3
ENSG00000241860.7	15
ENSG00000222623.1	0
ENSG00000241599.1	0
ENSG00000279928.2	4
ENSG00000279457.4	517
ENSG00000273874.1	0
ENSG00000228463.10	11
ENSG00000286448.1	0
ENSG00000236679.2	0
ENSG00000236601.2	0
ENSG00000237094.12	75

Heatmap and PCA of all the samples





RNA-seq analysis workflow

Results

Differential expression analysis (between pairs of conditions)

Tables (all genes / significantly different)

dea_J0_WT_J10_WT.tsv

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj		
ENSG00000138722.10	28643.5576775341	12.8878559048439	0.334254252259555	38.5570439799108	0	0		
ENSG00000150637.9	44495.3489697712	12.1138628938676	0.225254162861642	53.7786416018794	0	0		
ENSG00000248131.6	23035.2235024924	-11.5102220691977	0.214583147815272	-53.6399162114375	0	0		
ENSG00000236673.5	9410.62778439526	-11.1702993224953	0.284876900393565	-39.2109690433419	0	0		
ENSG00000010671.16	17434.4042351806	11.1368058134479	0.247095134871919	45.0709230645948	0	0		
ENSG00000274090.1	12673.5010973939	-11.0431466139289	0.243020448209864	-45.4412239598557	0	0		
ENSG00000128683.14	38392.4960153405	10.6771649471065	0.165881192381676	64.3663383039795	0	0		
ENSG00000261371.6	21011.2196060373	10.5882903827317	0.202428901612002	52.3062186200387	0	0		
ENSG00000185559.16	9291.39445851381	10.5672720975195	0.264977072858931	39.8799487952127	0	0		
ENSG00000151702.17	27888.8642370209	9.90260964734258	0.151987244928666	65.1542150921303	0	0		
ENSG00000111348.9	12188.4835524031	9.72369598445457	0.186453617244828	52.150750026408	0	0		
ENSG00000077420.16	6537.36530736273	9.72092375413297	0.241392363666131	40.2702206751576	0	0		
ENSG00000165702.14	9343.36263764007	9.59803131814038	0.227418626840593	42.2042444433016	0	0		
ENSG00000162511.8	5696.04146023551	9.51650005466677	0.249728319897167	38.1074123214599	0	0		
ENSG00000255929.5	7598.70343369437	-9.47019261095421	0.19136293672142	-49.4881233179476	0	0		
ENSG00000164512.18	6990.32718933163	9.40628158397668	0.212743801625526	44.2141275661404	0	0		
ENSG00000163737.4	41625.3614482218	9.3987254652196	0.16974012515012	55.3712650848306	0	0		
ENSG00000064201.16	10395.1831045548	9.38771319937487	0.191501351377414	49.0216551050514	0	0		
ENSG00000158578.21	21622.2181893878	9.30066251795876	0.150229090820593	61.9098635767279	0	0		
ENSG00000164035.10	6915.59463518514	9.20058963397617	0.205829692780933	44.7000114981879	0	0		
ENSG00000254277.1	12497.463812679	-9.02174609448843	0.140908284960578	-64.0256610674984	0	0		
ENSG00000182866.17	3254.84007527242	-8.82544806879715	0.222680250843156	-39.6328279467105	0	0		
ENSG00000104903.5	5556.34779269385	8.75726129253258	0.200868468802165	43.5969933198307	0	0		
ENSG00000152208.13	169397.89006299	-8.65256319895797	0.12072759446843	-71.6701366995317	0	0		
ENSG00000268555.2	3961.66735219303	8.31225086070514	0.196847711819128	42.2268096686986	0	0		
ENSG00000095303.17	19954.8652152089	8.281811632252	0.134548170511187	61.5527628565071	0	0		
ENSG00000134138.20	4943.42492987928	8.21668929248445	0.188177877289831	43.6644807074167	0	0		
ENSG00000226792.7	5929.13871166827	-8.19648712538188	0.14617436684561	-56.0733547355745	0	0		
ENSG00000236780.7	13522.677227173	-8.14693091270317	0.120096819744012	-67.8363584486953	0	0		
ENSG00000244342.5	4412.35248366687	-8.10918207356467	0.161825975718805	-50.1105093761678	0	0		
ENSG00000133083.14	5994.21884434801	-8.006668768897	0.141959393523884	-56.4011198565026	0	0		
ENSG00000149781.12	25395.5559785894	7.99738481605514	0.132915719822231	60.1688410276174	0	0		
ENSG00000118308.15	4085.4630243289	7.96054054065835	0.181538060028329	43.8505321661811	0	0		
ENSG00000164076.17	2044.74195604111	-7.9085982226867	0.209760276503151	-37.7030310720815	0	0		
ENSG00000166105.16	2718.07816642691	-7.90155905419819	0.190344071656288	-41.5119787311599	0	0		
ENSG00000172578.12	5552.74698183465	7.87262797383369	0.161041248685017	48.8857857109135	0	0		
ENSG00000165092.13	10490.0915575169	7.86110735462206	0.124118978589746	63.3352565734979	0	0		

Report

file:///home/mag/DataMadeleine/TIMING/hisat2/dea_featureCounts/Report_DESeq2/J0_WT_J0_KO/DESeq2Exploration.html

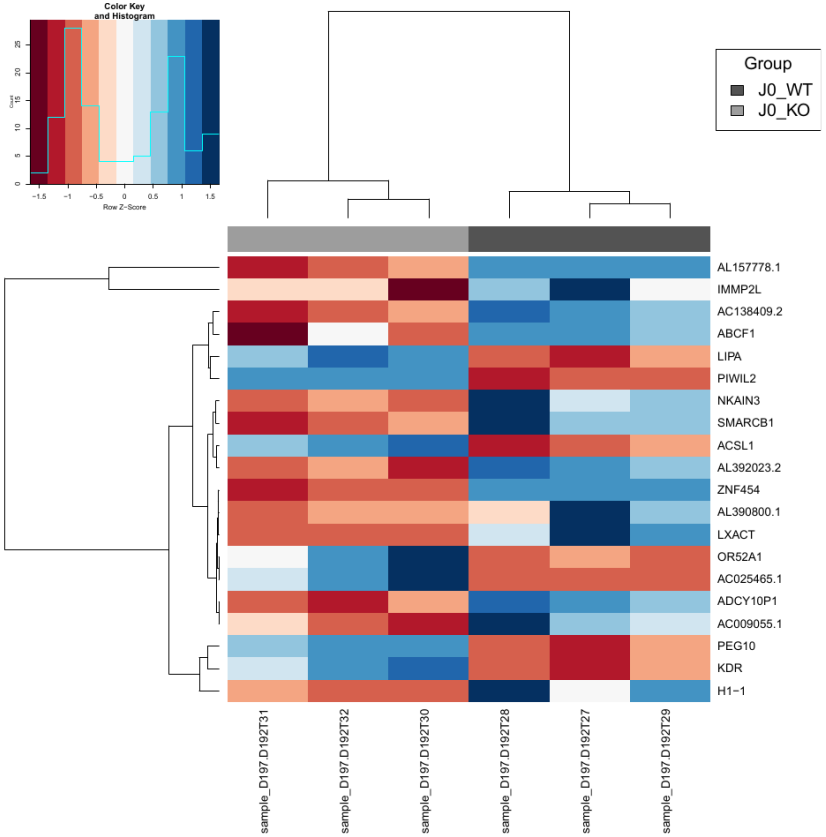
regionReport (with code)

RNA-seq analysis workflow

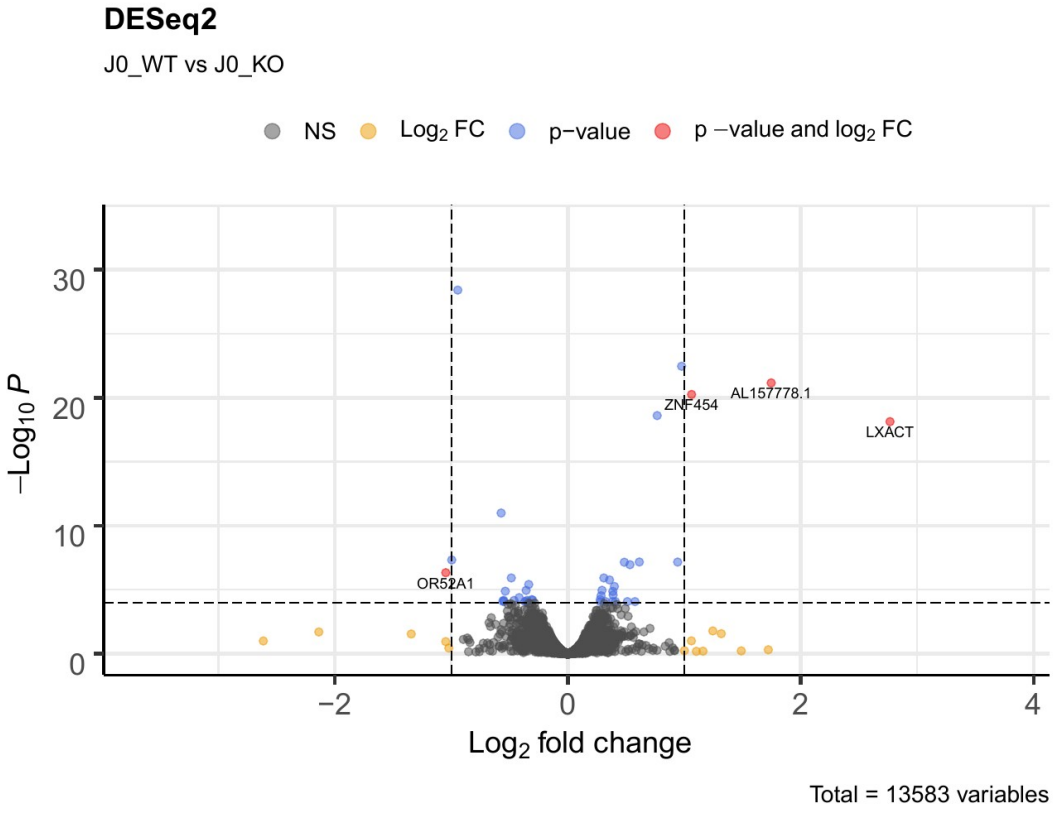
Results

Differential expression analysis (between pairs of conditions)

Heatmap Top 20 genes



Volcano plot

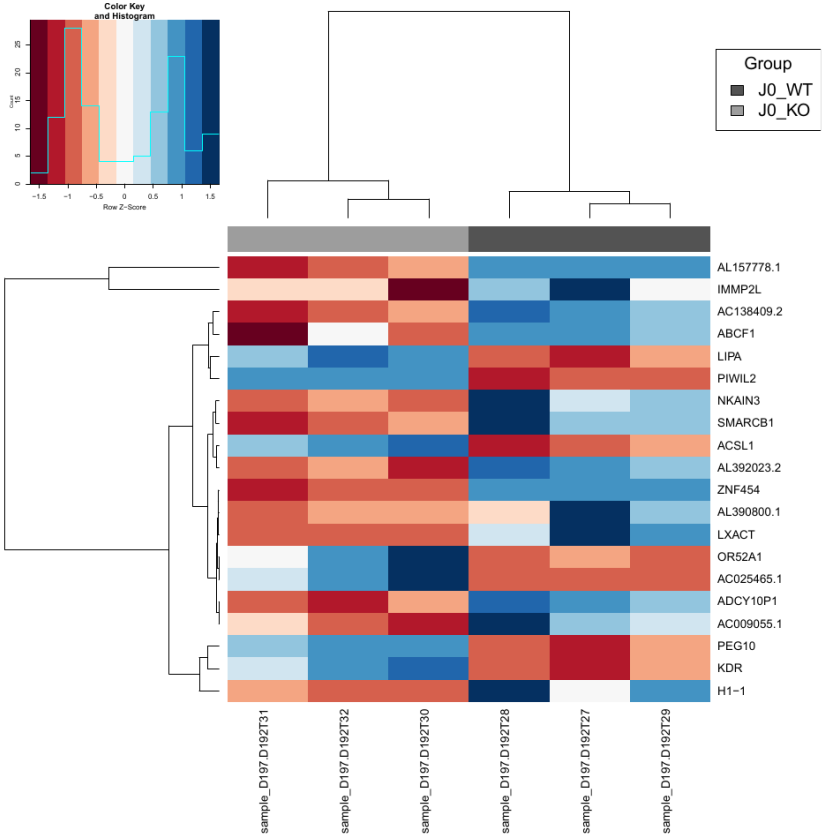


RNA-seq analysis workflow

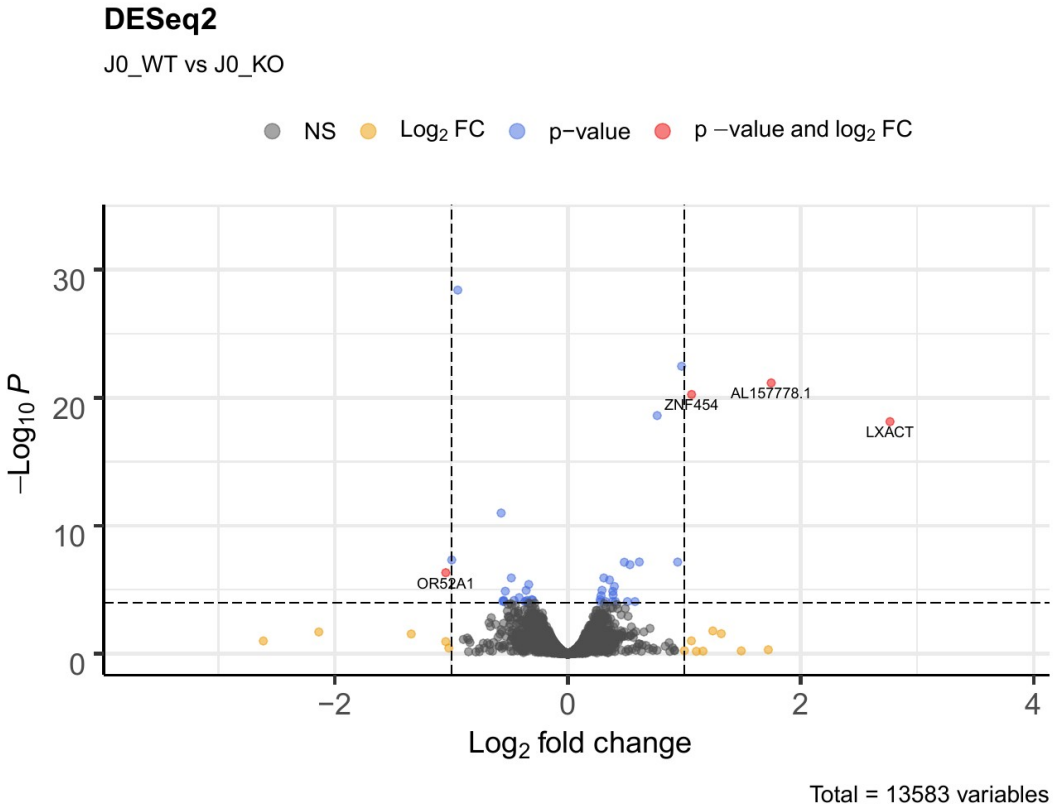
Results

Differential expression analysis (between pairs of conditions)

Heatmap Top 20 genes



Volcano plot



! Your analysis is not over ! This is just a starting point...

RNA-seq analysis workflow

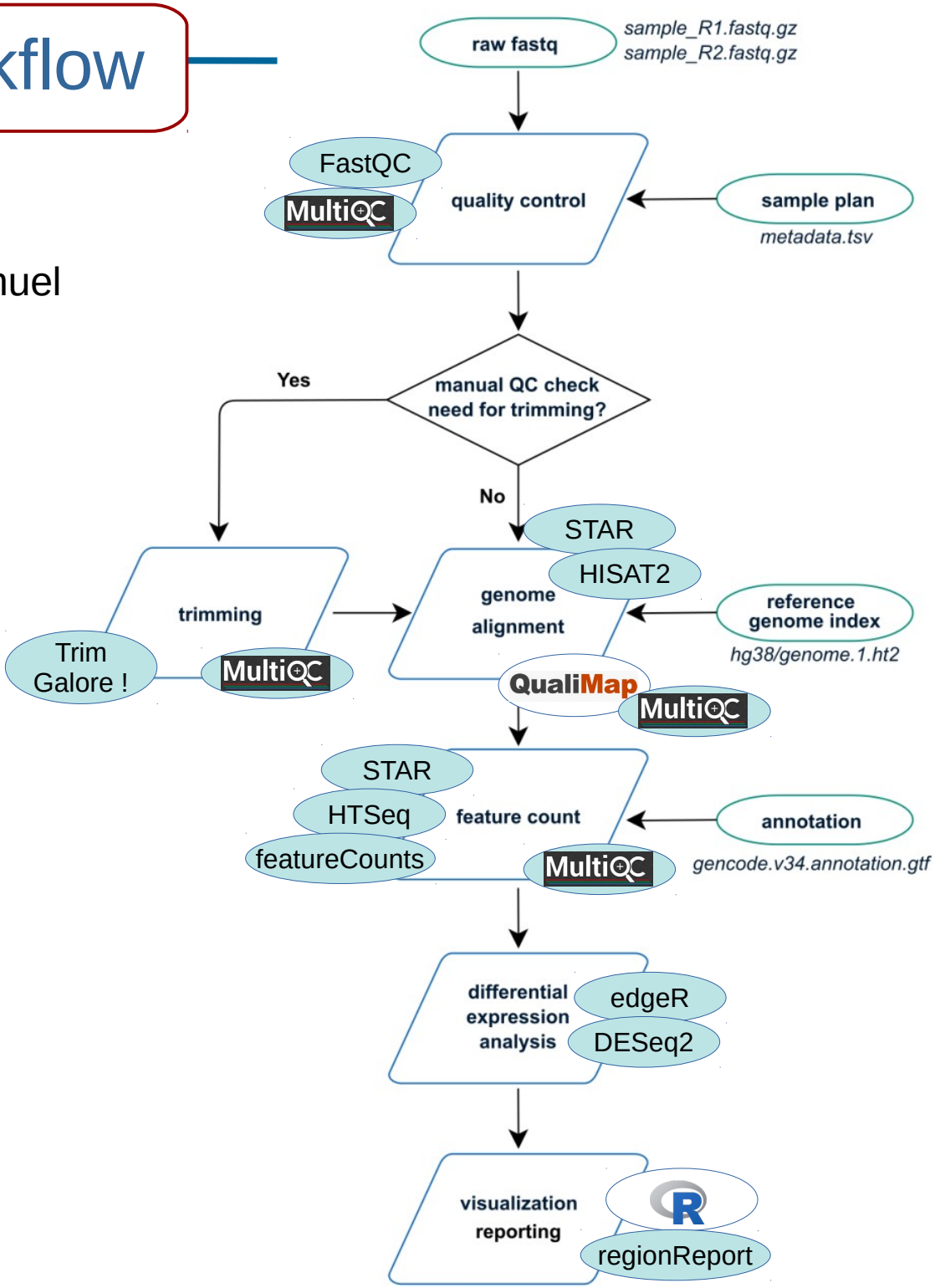
beta-testers : Olivier, Madeleine, Emmanuel

To be added

- Salmon (pseudo mapping)
- GSEA / GO / KEGG
- export .RData objects
- ask what you need !

To be adapted to

- genomic repeats
- transcript reconstruction
- ChIP-seq
- ATAC-seq
- WGBS
- ask what you need !



RNA-seq analysis workflow

Reproducibility

- Conda environment → define the tools and their version

→ results identical on IFB and RPBS clusters

```yaml
name: rasflow_IFB
channels:
  - conda-forge
  - bioconda
  - r
  - defaults
dependencies:
  - snakemake=5.14.0
  - graphviz=2.42.3
# command tool installs
  - R=4.0
  - python=3.7.6
# r channel installs
  - r-yaml=2.2.1
  - r-statmod=1.4.34
  - r-gplots=3.0.3
  - r-magick=2.3
  - r-dt=0.13
  - r-sessioninfo=1.1.1
# conda-forge channel installs
  - r-heatmap.plus=1.3
  - r-readr=1.3.1
  - r-hash=3.0.1
  - r-pheatmap=1.0.12
  - r-rcolorbrewer=1.1_2
  - imagemagick=7.0.10
# bioconda channel installs
  - fastqc=0.11.9
  - trim-galore=0.6.5
  - multiqc=1.9
  - salmon=1.2.1
  - hisat2=2.2.0
  - samtools=1.10
  - subread=2.0.1  # featureCounts included
  - htseq=0.12.4  # htseq-count included
  - bioconductor-edger=3.30.0
  - bioconductor-deseq2=1.28.0
  - qualimap=2.2.2a
  - bioconductor-mygene=1.24.0
  - bioconductor-tximport=1.16.0
  - bioconductor-enhancedvolcano=1.6.0
  - bioconductor-biomart=2.44.0
  - deeptools=3.4.3
  - bioconductor-regionreport=1.22.0
  - star=2.7.5a
```

RNA-seq analysis workflow

Reproducibility

- reporting

    * logs

    * R libraries (regionReport)

    * MultiQC

11 Reproducibility

The input for this report was generated with edgeR (Robinson, McCarthy, and Smyth, 2010; McCarthy, Chen, and Smyth, 2012; Chen, Lun, and Smyth, 2014) and the resulting features were called significantly differentially expressed if their BH adjusted p-values were less than `alpha` = 0.1. This report was generated in path /home/mag/DataMadeleine using the following call to edgeReport():

```
## edgeReport(dge = y, object = lrt, project = "edgeR-example",
##     intgroup = "group", outdir = "edgeReport-example")
```

Date the report was generated.

```
## [1] "2020-06-22 11:36:02 CEST"
```

Wallclock time spent generating the report.

```
## Time difference of 14.73 secs
```

`R` session information.

```
## - Session info -------------------------------------------------------------
##  setting  value
##  version  R version 3.6.3 (2020-02-29)
##  os       Linux Mint 19.1
##  system   x86_64, linux-gnu
##  ui       X11
##  language en_US
##  collate  en_US.UTF-8
##  ctype    en_US.UTF-8
##  tz       Europe/Paris
##  date     2020-06-22
##
## - Packages -----------------------------------------------------------------
##  package              * version  date       lib source
##  acepack                1.4.1    2016-10-29 [1] CRAN (R 3.6.3)
##  annotate               1.64.0   2019-10-29 [1] Bioconductor
##  AnnotationDbi          1.48.0   2019-10-29 [1] Bioconductor
##  askpass                1.1      2019-01-13 [1] CRAN (R 3.6.1)
##  assertthat             0.2.1    2019-03-21 [1] CRAN (R 3.6.1)
##  backports              1.1.5    2019-10-02 [1] CRAN (R 3.6.1)
##  base64enc              0.1-3    2015-07-28 [1] CRAN (R 3.6.1)
##  bibtex                 0.4.2.2  2020-01-02 [1] CRAN (R 3.6.3)
##  Biobase              * 2.46.0   2019-10-29 [1] Bioconductor
##  BiocFileCache          1.10.2   2019-11-08 [1] Bioconductor
##  BiocGenerics         * 0.32.0   2019-10-29 [1] Bioconductor
##  BiocManager            1.30.10  2019-11-16 [1] CRAN (R 3.6.3)
##  BiocParallel         * 1.20.1   2019-12-21 [1] Bioconductor
##  BiocStyle            * 2.14.4   2020-01-09 [1] Bioconductor
##  biomaRt                2.42.0   2019-10-29 [1] Bioconductor
##  Biostrings             2.54.0   2019-10-29 [1] Bioconductor
##  bit                    1.1-15.2 2020-02-10 [1] CRAN (R 3.6.3)
##  bit64                  0.9-7    2017-05-08 [1] CRAN (R 3.6.3)
##  bitops                 1.0-6    2013-08-17 [1] CRAN (R 3.6.3)
##  blob                   1.2.1    2020-01-20 [1] CRAN (R 3.6.3)
##  bookdown               0.19     2020-05-15 [1] CRAN (R 3.6.3)
##  BSgenome               1.54.0   2019-10-29 [1] Bioconductor
##  bumphunter             1.28.0   2019-10-29 [1] Bioconductor
##  checkmate              2.0.0    2020-02-06 [1] CRAN (R 3.6.3)
##  cli                    2.0.2    2020-02-28 [1] CRAN (R 3.6.1)
##  cluster                2.1.0    2019-06-19 [1] CRAN (R 3.6.3)
##  codetools              0.2-16   2018-12-24 [1] CRAN (R 3.6.3)
##  colorspace             1.4-1    2019-03-18 [1] CRAN (R 3.6.1)
##  crayon                 1.3.4    2017-09-16 [1] CRAN (R 3.6.1)
##  crosstalk              1.0.0    2016-12-21 [1] CRAN (R 3.6.1)
##  curl                   4.3      2019-12-02 [1] CRAN (R 3.6.1)
##  data.table             1.12.8   2019-12-09 [1] CRAN (R 3.6.3)
##  DBI                    1.1.0    2019-12-15 [1] CRAN (R 3.6.1)
##  dbplyr                 1.4.2    2019-06-17 [1] CRAN (R 3.6.1)
##  DEFormats              1.14.0   2019-10-29 [1] Bioconductor
##  DelayedArray         * 0.12.2   2020-01-06 [1] Bioconductor
##  derfinder              1.20.0   2019-10-29 [1] Bioconductor
##  derfinderHelper        1.20.0   2019-10-29 [1] Bioconductor
##  DESeq2               * 1.26.0   2019-10-29 [1] Bioconductor
##  digest                 0.6.25   2020-02-23 [1] CRAN (R 3.6.1)
##  doRNG                  1.8.2    2020-01-27 [1] CRAN (R 3.6.3)
##  dplyr                  0.8.5    2020-03-07 [1] CRAN (R 3.6.3)
##  DT                   * 0.12     2020-02-05 [1] CRAN (R 3.6.1)
##  edgeR                * 3.28.1   2020-02-26 [1] Bioconductor
##  evaluate               0.14     2019-05-28 [1] CRAN (R 3.6.1)
##  fansi                  0.4.1    2020-01-08 [1] CRAN (R 3.6.1)
##  farver                 2.0.3    2020-01-16 [1] CRAN (R 3.6.1)
##  fastmap                1.0.1    2019-10-08 [1] CRAN (R 3.6.1)
##  foreach                1.5.0    2020-03-30 [1] CRAN (R 3.6.3)
##  foreign                0.8-76   2020-03-03 [1] CRAN (R 3.6.3)
##  Formula                1.2-3    2018-05-03 [1] CRAN (R 3.6.3)
##  genefilter             1.68.0   2019-10-29 [1] Bioconductor
##  geneplotter            1.64.0   2019-10-29 [1] Bioconductor
##  GenomeInfoDb         * 1.22.0   2019-10-29 [1] Bioconductor
##  GenomeInfoDbData       1.2.2    2020-03-10 [1] Bioconductor
##  GenomicAlignments      1.22.1   2019-11-12 [1] Bioconductor
##  GenomicFeatures        1.38.2   2020-02-15 [1] Bioconductor
##  GenomicFiles           1.22.0   2019-10-29 [1] Bioconductor
##  GenomicRanges        * 1.38.0   2019-10-29 [1] Bioconductor
##  ggplot2              * 3.3.0    2020-03-05 [1] CRAN (R 3.6.3)
##  glue                   1.3.1    2019-03-12 [1] CRAN (R 3.6.1)
##  gridExtra              2.3      2017-09-09 [1] CRAN (R 3.6.3)
##  gtable                 0.3.0    2019-03-25 [1] CRAN (R 3.6.1)
##  highr                  0.8      2019-03-20 [1] CRAN (R 3.6.1)
##  Hmisc                  4.3-1    2020-02-07 [1] CRAN (R 3.6.3)
##  hms                    0.5.3    2020-01-08 [1] CRAN (R 3.6.1)
##  htmlTable              1.13.3   2019-12-04 [1] CRAN (R 3.6.3)
##  htmltools              0.4.0    2019-10-04 [1] CRAN (R 3.6.1)
##  htmlwidgets            1.5.1    2019-10-08 [1] CRAN (R 3.6.1)
##  httpuv                 1.5.2    2019-09-11 [1] CRAN (R 3.6.1)
##  httr                   1.4.1    2019-08-05 [1] CRAN (R 3.6.1)
##  IRanges              * 2.20.2   2020-01-13 [1] Bioconductor
##  IRdisplay              0.7.0    2018-11-29 [1] CRAN (R 3.6.1)
##  IRkernel               1.1      2020-03-03 [1] Github (IRkernel/IRker
nel@1a45f74)
##  iterators              1.0.12   2019-07-26 [1] CRAN (R 3.6.3)
##  jpeg                   0.1-8.1  2019-10-24 [1] CRAN (R 3.6.3)
##  jsonlite               1.6.1    2020-02-02 [1] CRAN (R 3.6.1)
##  knitcitations          1.0.10   2019-09-15 [1] CRAN (R 3.6.3)
##  knitr                * 1.28     2020-02-06 [1] CRAN (R 3.6.3)
##  knitrBootstrap         1.0.2    2018-05-24 [1] CRAN (R 3.6.3)
##  labeling               0.3      2014-08-23 [1] CRAN (R 3.6.1)
##  later                  1.0.0    2019-10-04 [1] CRAN (R 3.6.1)
##  lattice                0.20-40  2020-02-19 [1] CRAN (R 3.6.1)
##  latticeExtra           0.6-29   2019-12-19 [1] CRAN (R 3.6.3)
##  lifecycle              0.2.0    2020-03-06 [1] CRAN (R 3.6.3)
##  limma                * 3.42.2   2020-02-03 [1] Bioconductor
##  locfit                 1.5-9.1  2013-04-20 [1] CRAN (R 3.6.3)
##  lubridate              1.7.4    2018-04-11 [1] CRAN (R 3.6.1)
##  magick                 2.3      2020-01-24 [1] CRAN (R 3.6.3)
##  magrittr               1.5      2014-11-22 [1] CRAN (R 3.6.1)
##  markdown               1.1      2019-08-07 [1] CRAN (R 3.6.1)
##  Matrix                 1.2-18   2019-11-27 [1] CRAN (R 3.6.1)
##  matrixStats          * 0.55.0   2019-09-07 [1] CRAN (R 3.6.3)
##  memoise                1.1.0    2017-04-21 [1] CRAN (R 3.6.1)
##  mime                   0.9      2020-02-04 [1] CRAN (R 3.6.1)
##  munsell                0.5.0    2018-06-12 [1] CRAN (R 3.6.1)
##  nnet                   7.3-13   2020-02-25 [1] CRAN (R 3.6.3)
##  openssl                1.4.1    2019-07-18 [1] CRAN (R 3.6.1)
##  pbdZMQ                 0.3-3    2018-05-05 [1] CRAN (R 3.6.1)
##  pheatmap             * 1.0.12   2019-01-04 [1] CRAN (R 3.6.3)
##  pillar                 1.4.3    2019-12-20 [1] CRAN (R 3.6.1)
##  pkgconfig              2.0.3    2019-09-22 [1] CRAN (R 3.6.1)
```

RNA-seq analysis workflow

Reproducibility

- reporting

    * logs

    * R libraries (regionReport)

    * MultiQC

Any volunteer to try it?

11 Reproducibility

The input for this report was generated with edgeR (Robinson, McCarthy, and Smyth, 2010; McCarthy, Chen, and Smyth, 2012; Chen, Lun, and Smyth, 2014) and the resulting features were called significantly differentially expressed if their BH adjusted p-values were less than `alpha` = 0.1. This report was generated in path /home/mag/DataMadeleine using the following call to edgeReport():

```
## edgeReport(dge = y, object = lrt, project = "edgeR-example",
##     intgroup = "group", outdir = "edgeReport-example")
```

Date the report was generated.

```
## [1] "2020-06-22 11:36:02 CEST"
```

Wallclock time spent generating the report.

```
## Time difference of 14.73 secs
```

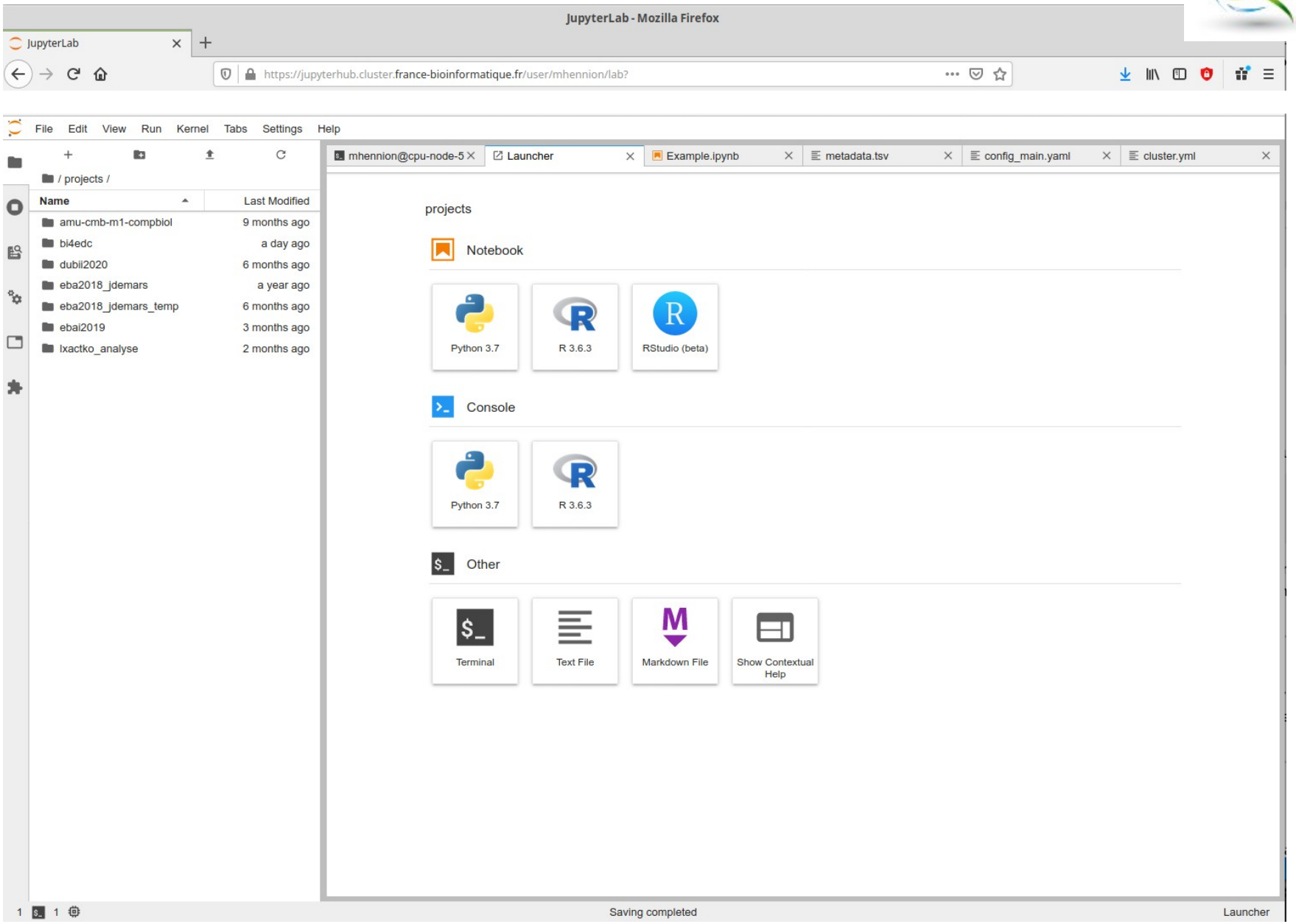
`R` session information.

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 3.6.3 (2020-02-29)
##  os       Linux Mint 19.1
##  system   x86_64, linux-gnu
##  ui       X11
##  language en_US
##  collate  en_US.UTF-8
##  ctype    en_US.UTF-8
##  tz       Europe/Paris
##  date     2020-06-22
##
## - Packages -------------------------------------------------------------------
##  package              * version  date       lib source
##  acepack                1.4.1    2016-10-29 [1] CRAN (R 3.6.3)
##  annotate               1.64.0   2019-10-29 [1] Bioconductor
##  AnnotationDbi          1.48.0   2019-10-29 [1] Bioconductor
##  askpass                1.1      2019-01-13 [1] CRAN (R 3.6.1)
##  assertthat             0.2.1    2019-03-21 [1] CRAN (R 3.6.1)
##  backports              1.1.5    2019-10-02 [1] CRAN (R 3.6.1)
##  base64enc              0.1-3    2015-07-28 [1] CRAN (R 3.6.1)
##  bibtex                 0.4.2.2  2020-01-02 [1] CRAN (R 3.6.3)
##  Biobase              * 2.46.0   2019-10-29 [1] Bioconductor
##  BiocFileCache          1.10.2   2019-11-08 [1] Bioconductor
##  BiocGenerics         * 0.32.0   2019-10-29 [1] Bioconductor
##  BiocManager            1.30.10  2019-11-16 [1] CRAN (R 3.6.3)
##  BiocParallel         * 1.20.1   2019-12-21 [1] Bioconductor
##  BiocStyle            * 2.14.4   2020-01-09 [1] Bioconductor
##  biomaRt                2.42.0   2019-10-29 [1] Bioconductor
##  Biostrings             2.54.0   2019-10-29 [1] Bioconductor
##  bit                    1.1-15.2 2020-02-10 [1] CRAN (R 3.6.3)
##  bit64                  0.9-7    2017-05-08 [1] CRAN (R 3.6.3)
##  bitops                 1.0-6    2013-08-17 [1] CRAN (R 3.6.3)
##  blob                   1.2.1    2020-01-20 [1] CRAN (R 3.6.3)
##  bookdown               0.19     2020-05-15 [1] CRAN (R 3.6.3)
##  BSgenome               1.54.0   2019-10-29 [1] Bioconductor
##  bumphunter             1.28.0   2019-10-29 [1] Bioconductor
##  checkmate              2.0.0    2020-02-06 [1] CRAN (R 3.6.3)
##  cli                    2.0.2    2020-02-28 [1] CRAN (R 3.6.1)
##  cluster                2.1.0    2019-06-19 [1] CRAN (R 3.6.3)
##  codetools              0.2-16   2018-12-24 [1] CRAN (R 3.6.3)
##  colorspace             1.4-1    2019-03-18 [1] CRAN (R 3.6.1)
##  crayon                 1.3.4    2017-09-16 [1] CRAN (R 3.6.1)
##  crosstalk              1.0.0    2016-12-21 [1] CRAN (R 3.6.1)
```

```
##  crosstalk              1.0.0    2016-12-21 [1] CRAN (R 3.6.1)
##  curl                   4.3      2019-12-02 [1] CRAN (R 3.6.1)
##  data.table             1.12.8   2019-12-09 [1] CRAN (R 3.6.3)
##  DBI                    1.1.0    2019-12-15 [1] CRAN (R 3.6.1)
##  dbplyr                 1.4.2    2019-06-17 [1] CRAN (R 3.6.1)
##  DEFormats              1.14.0   2019-10-29 [1] Bioconductor
##  DelayedArray         * 0.12.2   2020-01-06 [1] Bioconductor
##  derfinder              1.20.0   2019-10-29 [1] Bioconductor
##  derfinderHelper        1.20.0   2019-10-29 [1] Bioconductor
##  DESeq2               * 1.26.0   2019-10-29 [1] Bioconductor
##  digest                 0.6.25   2020-02-23 [1] CRAN (R 3.6.1)
##  doRNG                  1.8.2    2020-01-27 [1] CRAN (R 3.6.3)
##  dplyr                  0.8.5    2020-03-07 [1] CRAN (R 3.6.3)
##  DT                   * 0.12     2020-02-05 [1] CRAN (R 3.6.1)
##  edgeR                * 3.28.1   2020-02-26 [1] Bioconductor
##  evaluate               0.14     2019-05-28 [1] CRAN (R 3.6.1)
##  fansi                  0.4.1    2020-01-08 [1] CRAN (R 3.6.1)
##  farver                 2.0.3    2020-01-16 [1] CRAN (R 3.6.1)
##  fastmap                1.0.1    2019-10-08 [1] CRAN (R 3.6.1)
##  foreach                1.5.0    2020-03-30 [1] CRAN (R 3.6.3)
##  foreign                0.8-76   2020-03-03 [1] CRAN (R 3.6.3)
##  Formula                1.2-3    2018-05-03 [1] CRAN (R 3.6.3)
##  genefilter             1.68.0   2019-10-29 [1] Bioconductor
##  geneplotter            1.64.0   2019-10-29 [1] Bioconductor
##  GenomeInfoDb         * 1.22.0   2019-10-29 [1] Bioconductor
##  GenomeInfoDbData       1.2.2    2020-03-10 [1] Bioconductor
##  GenomicAlignments      1.22.1   2019-11-12 [1] Bioconductor
##  GenomicFeatures        1.38.2   2020-02-15 [1] Bioconductor
##  GenomicFiles           1.22.0   2019-10-29 [1] Bioconductor
##  GenomicRanges        * 1.38.0   2019-10-29 [1] Bioconductor
##  ggplot2              * 3.3.0    2020-03-05 [1] CRAN (R 3.6.3)
##  glue                   1.3.1    2019-03-12 [1] CRAN (R 3.6.1)
##  gridExtra              2.3      2017-09-09 [1] CRAN (R 3.6.3)
##  gtable                 0.3.0    2019-03-25 [1] CRAN (R 3.6.1)
##  highr                  0.8      2019-03-20 [1] CRAN (R 3.6.1)
##  Hmisc                  4.3-1    2020-02-07 [1] CRAN (R 3.6.3)
##  hms                    0.5.3    2020-01-08 [1] CRAN (R 3.6.1)
##  htmlTable              1.13.3   2019-12-04 [1] CRAN (R 3.6.3)
##  htmltools              0.4.0    2019-10-04 [1] CRAN (R 3.6.1)
##  htmlwidgets            1.5.1    2019-10-08 [1] CRAN (R 3.6.1)
##  httpuv                 1.5.2    2019-09-11 [1] CRAN (R 3.6.1)
##  httr                   1.4.1    2019-08-05 [1] CRAN (R 3.6.1)
##  IRanges              * 2.20.2   2020-01-13 [1] Bioconductor
##  IRdisplay              0.7.0    2018-11-29 [1] CRAN (R 3.6.1)
##  IRkernel               1.1      2020-03-03 [1] Github (IRkernel/IRker
nel@1a45f74)
##  iterators              1.0.12   2019-07-26 [1] CRAN (R 3.6.3)
##  jpeg                   0.1-8.1  2019-10-24 [1] CRAN (R 3.6.3)
##  jsonlite               1.6.1    2020-02-02 [1] CRAN (R 3.6.1)
##  knitcitations          1.0.10   2019-09-15 [1] CRAN (R 3.6.3)
##  knitr                * 1.28     2020-02-06 [1] CRAN (R 3.6.3)
##  knitrBootstrap         1.0.2    2018-05-24 [1] CRAN (R 3.6.3)
##  labeling               0.3      2014-08-23 [1] CRAN (R 3.6.1)
##  later                  1.0.0    2019-10-04 [1] CRAN (R 3.6.1)
##  lattice                0.20-40  2020-02-19 [1] CRAN (R 3.6.1)
##  latticeExtra           0.6-29   2019-12-19 [1] CRAN (R 3.6.3)
##  lifecycle              0.2.0    2020-03-06 [1] CRAN (R 3.6.3)
##  limma                * 3.42.2   2020-02-03 [1] Bioconductor
##  locfit                 1.5-9.1  2013-04-20 [1] CRAN (R 3.6.3)
##  lubridate              1.7.4    2018-04-11 [1] CRAN (R 3.6.1)
##  magick                 2.3      2020-01-24 [1] CRAN (R 3.6.3)
##  magrittr               1.5      2014-11-22 [1] CRAN (R 3.6.1)
##  markdown               1.1      2019-08-07 [1] CRAN (R 3.6.1)
##  Matrix                 1.2-18   2019-11-27 [1] CRAN (R 3.6.1)
##  matrixStats          * 0.55.0   2019-09-07 [1] CRAN (R 3.6.3)
##  memoise                1.1.0    2017-04-21 [1] CRAN (R 3.6.1)
##  mime                   0.9      2020-02-04 [1] CRAN (R 3.6.1)
##  munsell                0.5.0    2018-06-12 [1] CRAN (R 3.6.1)
##  nnet                   7.3-13   2020-02-25 [1] CRAN (R 3.6.3)
##  openssl                1.4.1    2019-07-18 [1] CRAN (R 3.6.1)
##  pbdZMQ                 0.3-3    2018-05-05 [1] CRAN (R 3.6.1)
##  pheatmap             * 1.0.12   2019-01-04 [1] CRAN (R 3.6.3)
##  pillar                 1.4.3    2019-12-20 [1] CRAN (R 3.6.1)
##  pkgconfig              2.0.3    2019-09-22 [1] CRAN (R 3.6.1)
```

JupyterLab

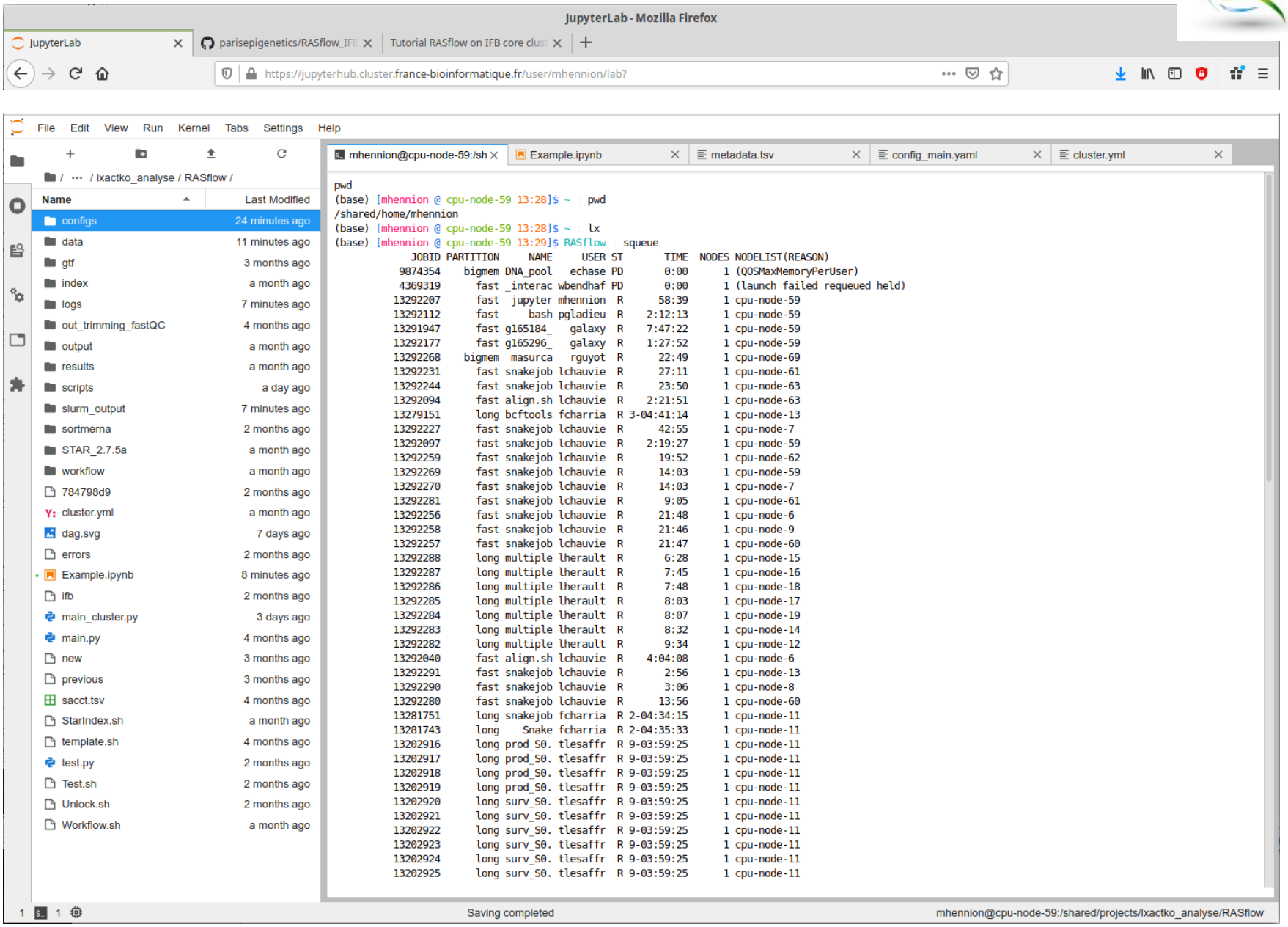
Make it easier



All your analysis from online hub

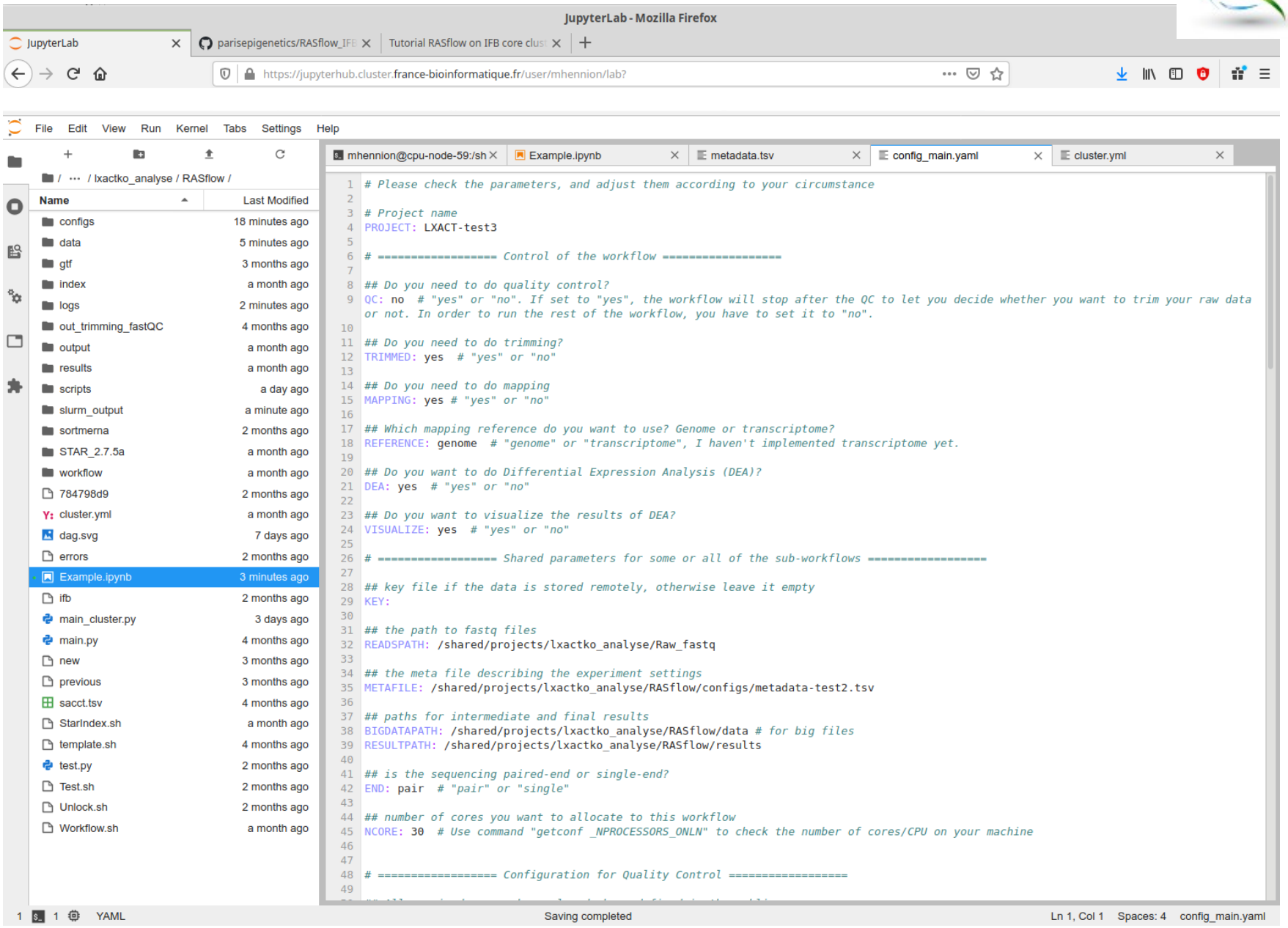
JupyterLab

Make it easier



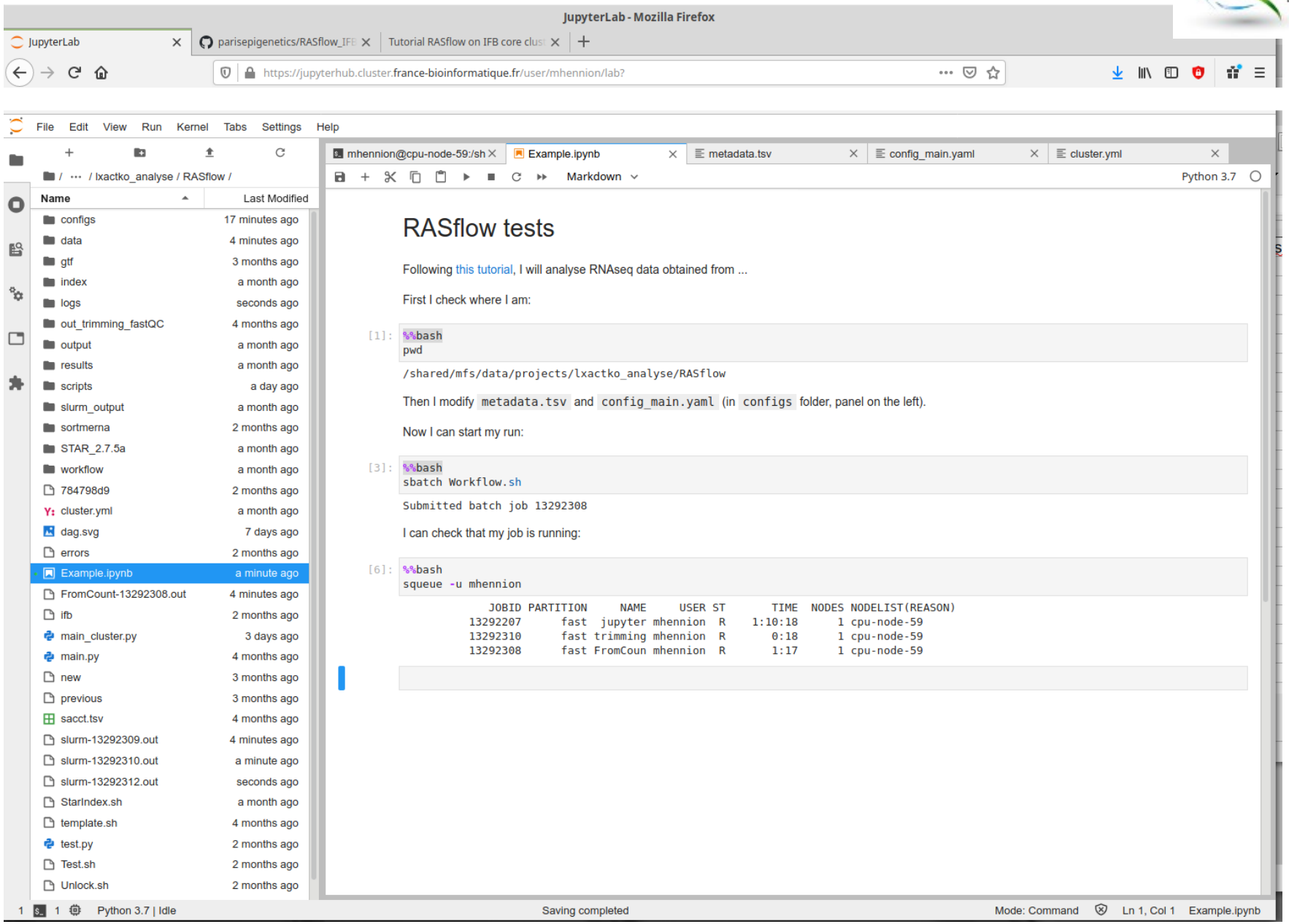
JupyterLab

Make it easier



JupyterLab

Make it easier

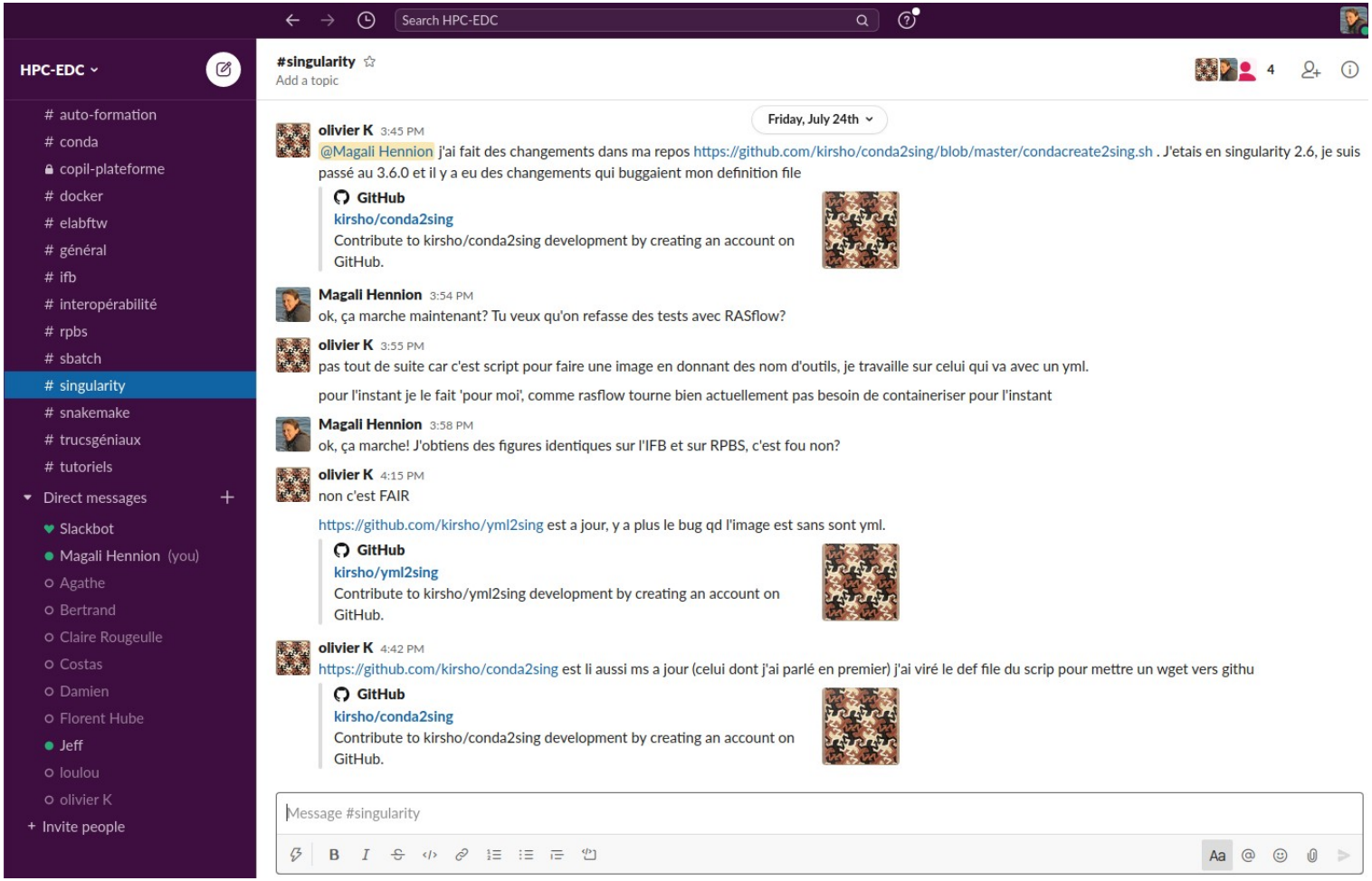


On going ...

Communication

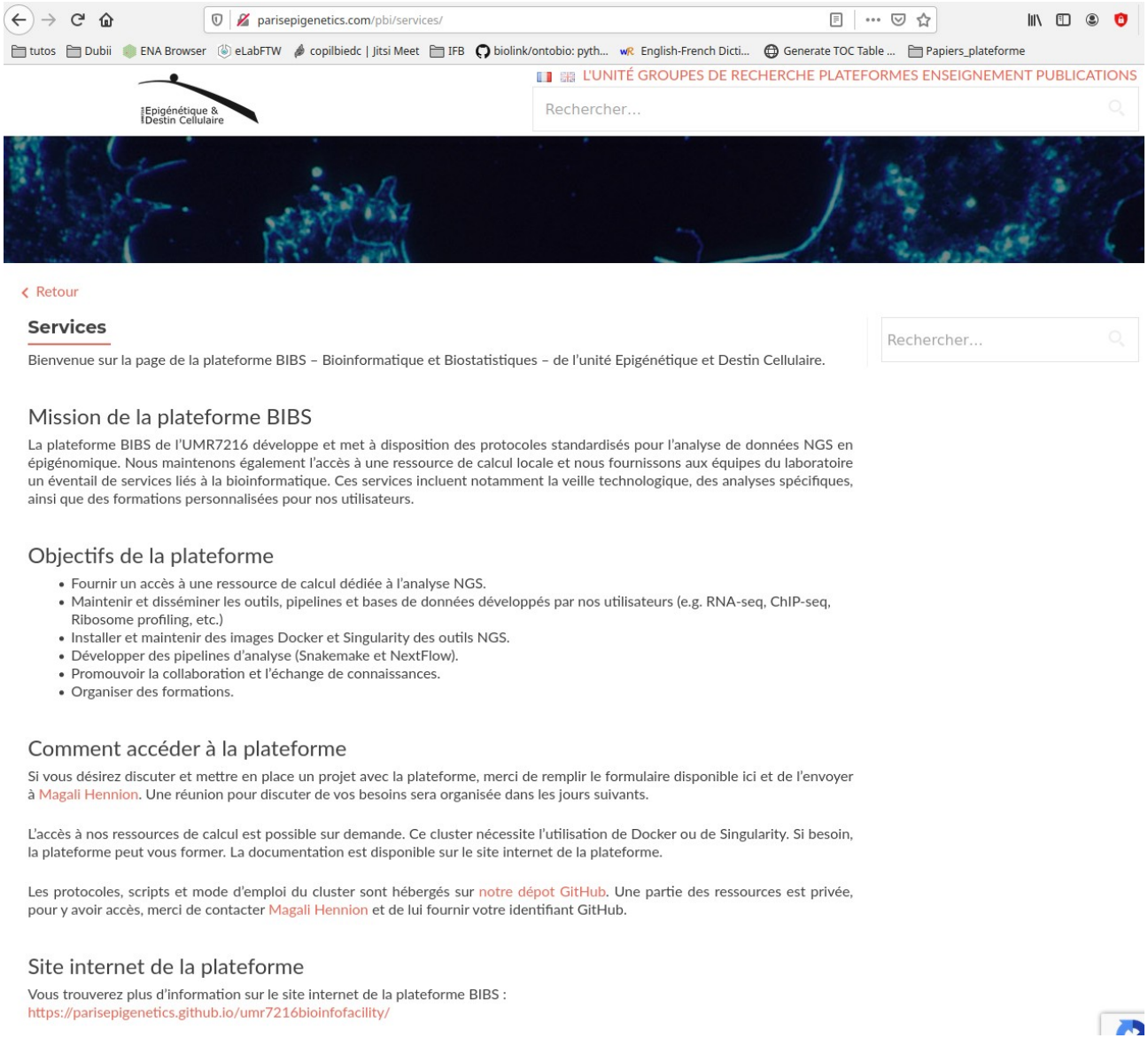
Slack : hpc-edc.slack.com

→ Ask your questions, help the others, complain, ...



Communication

Webpage on EDC website : http://parisepigenetics.com/pbi/services/



parisepigenetics.com/pbi/services/

tutos | Dubii | ENA Browser | eLabFTW | copilbiedc | Jitsi Meet | IFB | biolink/ontobio: pyth... | English-French Dicti... | Generate TOC Table ... | Papiers_plateforme

Epigénétique & Destin Cellulaire

L'UNITÉ GROUPES DE RECHERCHE PLATEFORMES ENSEIGNEMENT PUBLICATIONS

Rechercher...

Retour

Services

Rechercher...

Bienvenue sur la page de la plateforme BIBS – Bioinformatique et Biostatistiques – de l'unité Epigénétique et Destin Cellulaire.

Mission de la plateforme BIBS

La plateforme BIBS de l'UMR7216 développe et met à disposition des protocoles standardisés pour l'analyse de données NGS en épigénomique. Nous maintenons également l'accès à une ressource de calcul locale et nous fournissons aux équipes du laboratoire un éventail de services liés à la bioinformatique. Ces services incluent notamment la veille technologique, des analyses spécifiques, ainsi que des formations personnalisées pour nos utilisateurs.

Objectifs de la plateforme

- Fournir un accès à une ressource de calcul dédiée à l'analyse NGS.
- Maintenir et disséminer les outils, pipelines et bases de données développés par nos utilisateurs (e.g. RNA-seq, ChIP-seq, Ribosome profiling, etc.)
- Installer et maintenir des images Docker et Singularity des outils NGS.
- Développer des pipelines d'analyse (Snakemake et NextFlow).
- Promouvoir la collaboration et l'échange de connaissances.
- Organiser des formations.

Comment accéder à la plateforme

Si vous désirez discuter et mettre en place un projet avec la plateforme, merci de remplir le formulaire disponible ici et de l'envoyer à Magali Hennion. Une réunion pour discuter de vos besoins sera organisée dans les jours suivants.

L'accès à nos ressources de calcul est possible sur demande. Ce cluster nécessite l'utilisation de Docker ou de Singularity. Si besoin, la plateforme peut vous former. La documentation est disponible sur le site internet de la plateforme.

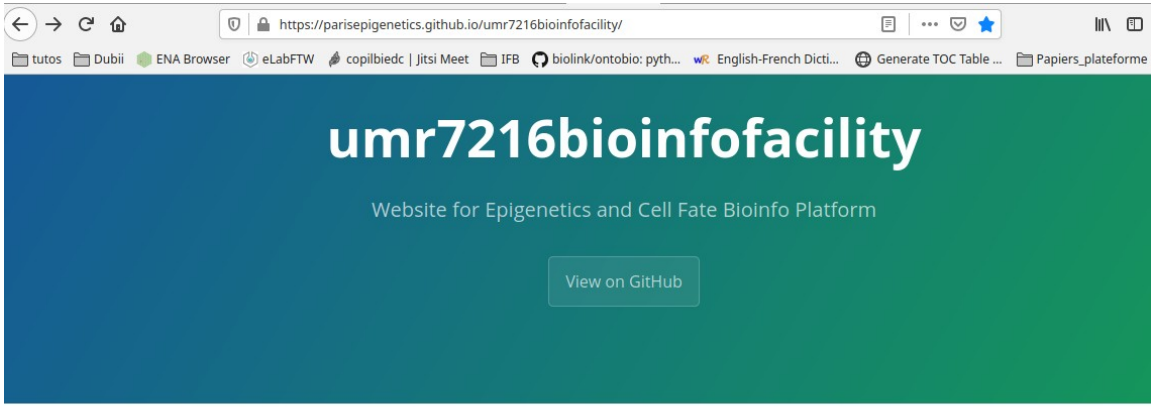
Les protocoles, scripts et mode d'emploi du cluster sont hébergés sur notre dépot GitHub. Une partie des ressources est privée, pour y avoir accès, merci de contacter Magali Hennion et de lui fournir votre identifiant GitHub.

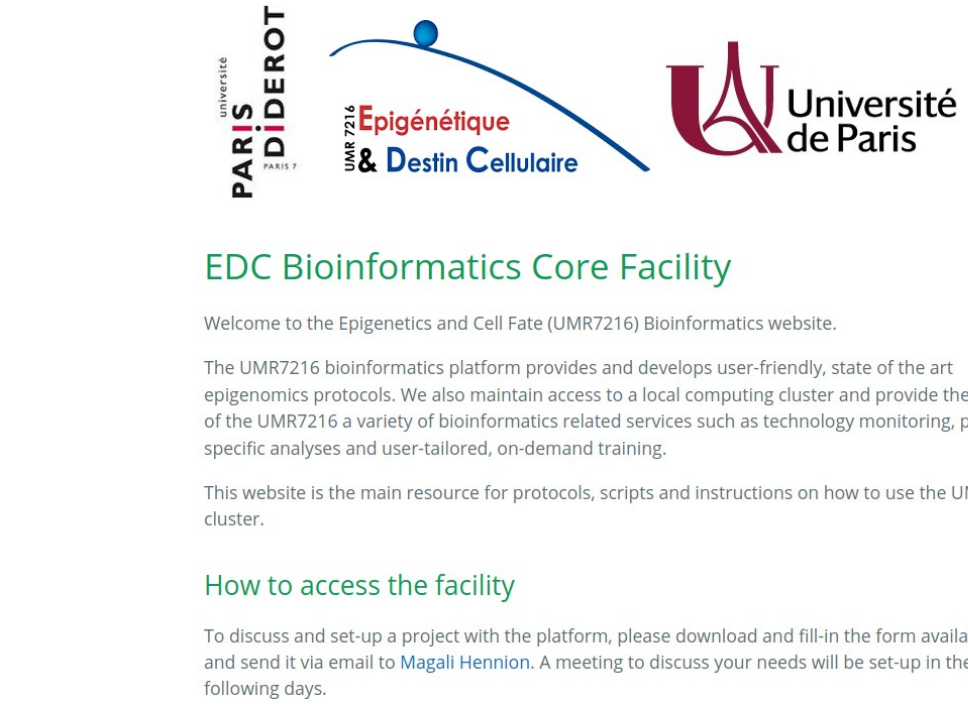
Site internet de la plateforme

Vous trouverez plus d'information sur le site internet de la plateforme BIBS :
https://parisepigenetics.github.io/umr7216bioinfofacility/

Communication

Website : https://parisepigenetics.github.io/umr7216bioinfofacility/





Update on going...

Feedback welcome !

A word about iPOP-UP

Mission statement

- Organize a resource for bioinformatics in the U-Paris landscape
- Assist/advise users on specific projects
- Train users to the use of the resource

RPBS cluster will be upgraded and **open** to the community

→ goal : starting production in January 2021

Technical committee

Pierre Tuffery (BFA, RPBS)
Julien Rey (BFA, RPBS)
Jean-Philippe Jais (Necker)
Pierre Poulain (IJM)
Christophe Cerin (Paris 13 Nord)
Magali Hennion (EDC)
Sjoerd DeVries (BFA, RPBS)
Niclas Setterblad (Saint-Louis)
Yves Clement (IJM)
Guillaume Seith (IGBMC, IFB)
Benjamin Saintpierre (Cochin)

A word about iPOP-UP

Where we are

- a system administrator (Abdeslam Tahari) has been recruited
- new equipment will be ordered and installed soon
- IR open position (EDC)

A word about iPOP-UP

Where we are

- a system administrator (Abdeslam Tahari) has been recruited

- new equipment will be ordered and installed soon

- IR open position (EDC)

→ If you have specific needs or great ideas, it's time to let me know!

→ RPBS people want us (and other partners) to be more involved

If you want to be part of the task force that manage the resource and organize the services,... or to do beta-testing, please let me know.

UMR7216 Epigénétique & Destin Cellulaire

Acknowledgments

Olivier

Valérie and Claire

The CoPil

The beta-testers


Community Support

Julien Rey

Thank you for your attention!