



Visualisation of FAANG data with VizFaDa

Laura Morel, Peter Harrison, Guillaume Devailly

► To cite this version:

Laura Morel, Peter Harrison, Guillaume Devailly. Visualisation of FAANG data with VizFaDa. JO-BIM 2022, Jul 2022, Rennes, France. hal-03740369

HAL Id: hal-03740369

<https://hal.inrae.fr/hal-03740369>

Submitted on 29 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



> Visualization of FAANG Data with VizFaDa

Laura Morel – 07/07/2022



• What is FAANG ?

Functional Annotation of ANimal Genome

www.faang.org



• Aims

- Standardize core assays and experimental protocols
- Coordinate and facilitate data sharing
- Improve functional annotation of animal genomes



● FAANG Data Portal

<https://data.faang.org>



Home Records ▾ Projects Summary Submit ▾ Search Help

ERR3626635_2

File name: ERR3626635_2.fastq.gz
Release date: 2019-10-29
Update date: 2019-10-29
Organism: Bos taurus
Specimen: SAMEA4447778
BioSample ID: SAMEA4447799
File size: 371829560 (354.6MB)
Archive: ENA
Read count: 4361251
Base count: 1317097802
Submission: ERA2191575
Experiment: ERX3620479 [Experiment details](#)
Study: PRJEB35127
Study title: Bovine gene expression atlas

Related Downloads

[Download all files \(0\)](#)

[Column selection ▾](#)

Name	File size	Checksum
ERR3626635_1.fastq.gz	323.25MB	965be71a8bbca81c1d5827b9ef145a13
ERR3626635_2.fastq.gz	354.6MB	f154e18c8e391725cbfa456a3fc7aa57

● **VizFaDa**

- Funded by the ANR in a call for open science projects
- Aims
 - Interactive visualization
 - RNA-Seq
 - ChIP-Seq, ATAC-Seq
 - DNA methylation
 - Automatically process new data
 - Encourage FAANG data usage



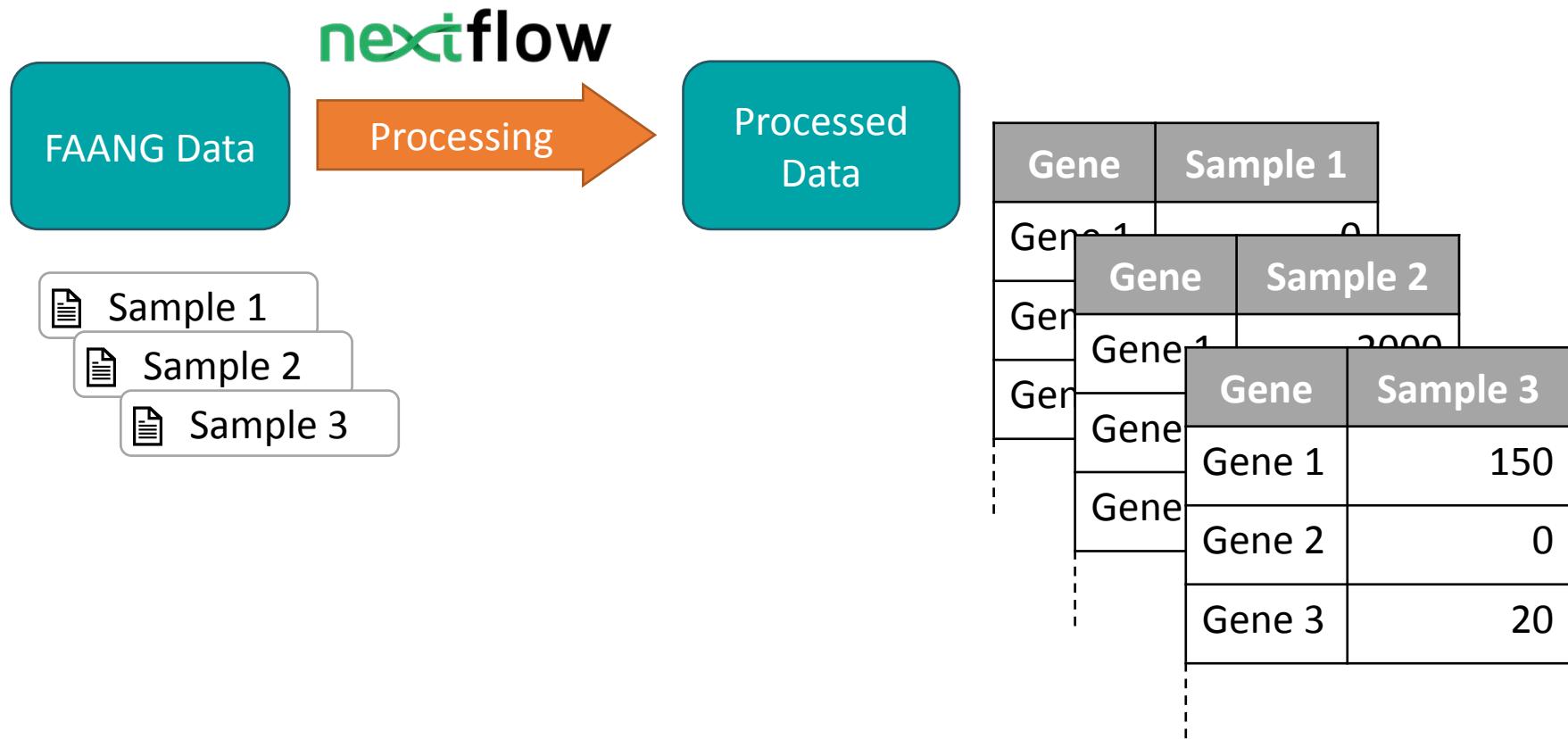
INRAE

Visualization of FAANG data with VizFaDa

07/07/2022 – Laura Morel

● Data Processing Overview

From FAANG data to Clustered Heatmaps



● Data Processing Overview

From FAANG data to Clustered Heatmaps

FAANG Data



Processed Data



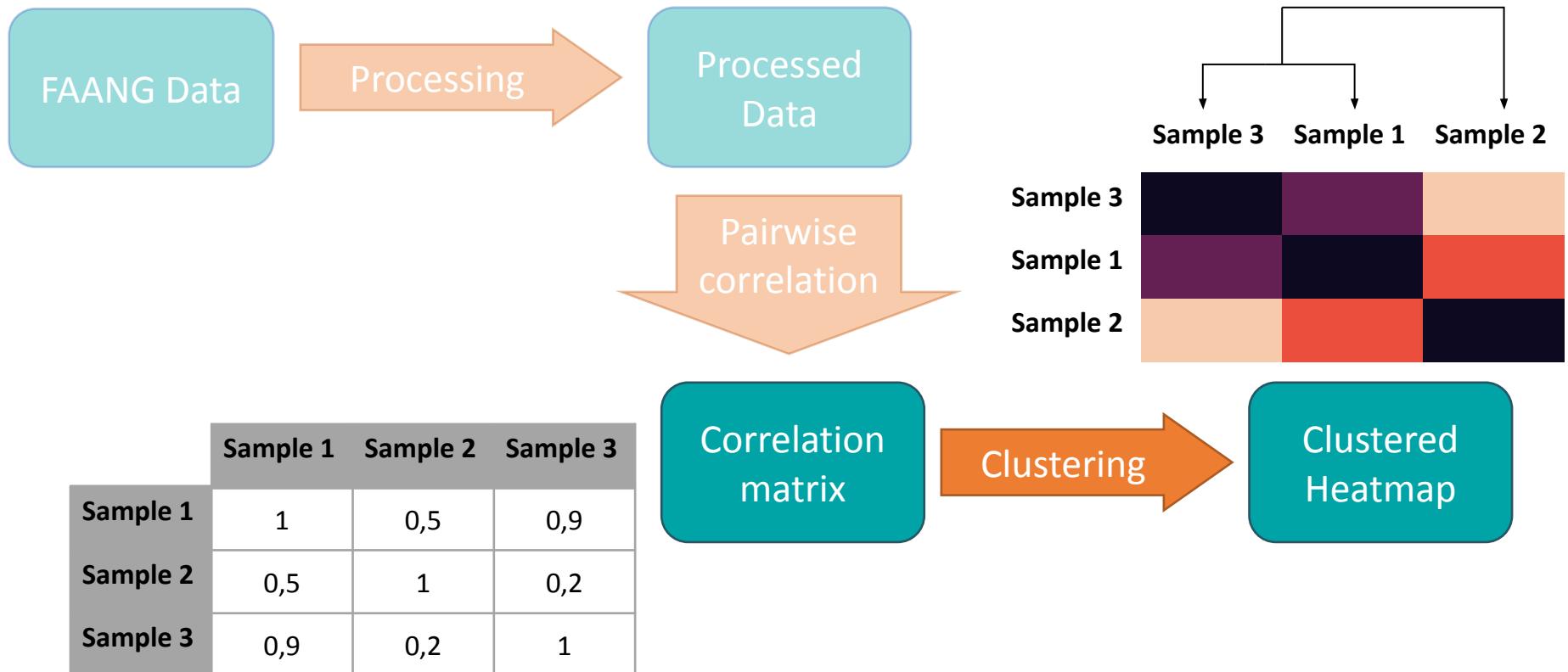
Correlation matrix

Gene	Sample 1
Gene	Sample 2
Gene	Sample 3
Gene 1	150
Gene 2	0
Gene 3	20

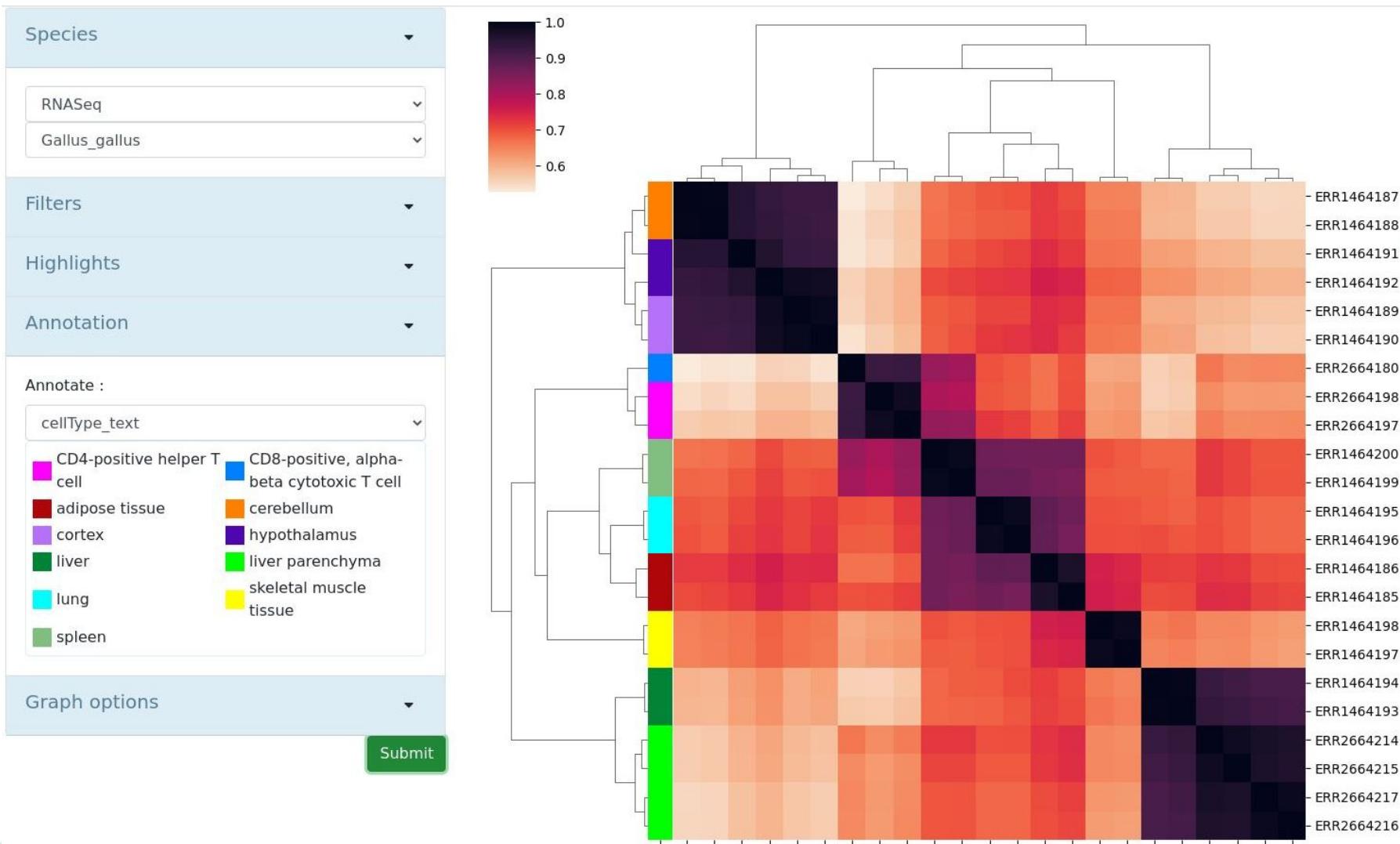
	Sample 1	Sample 2	Sample 3
Sample 1	1	0,5	0,9
Sample 2	0,5	1	0,2
Sample 3	0,9	0,2	1

● Data Processing Overview

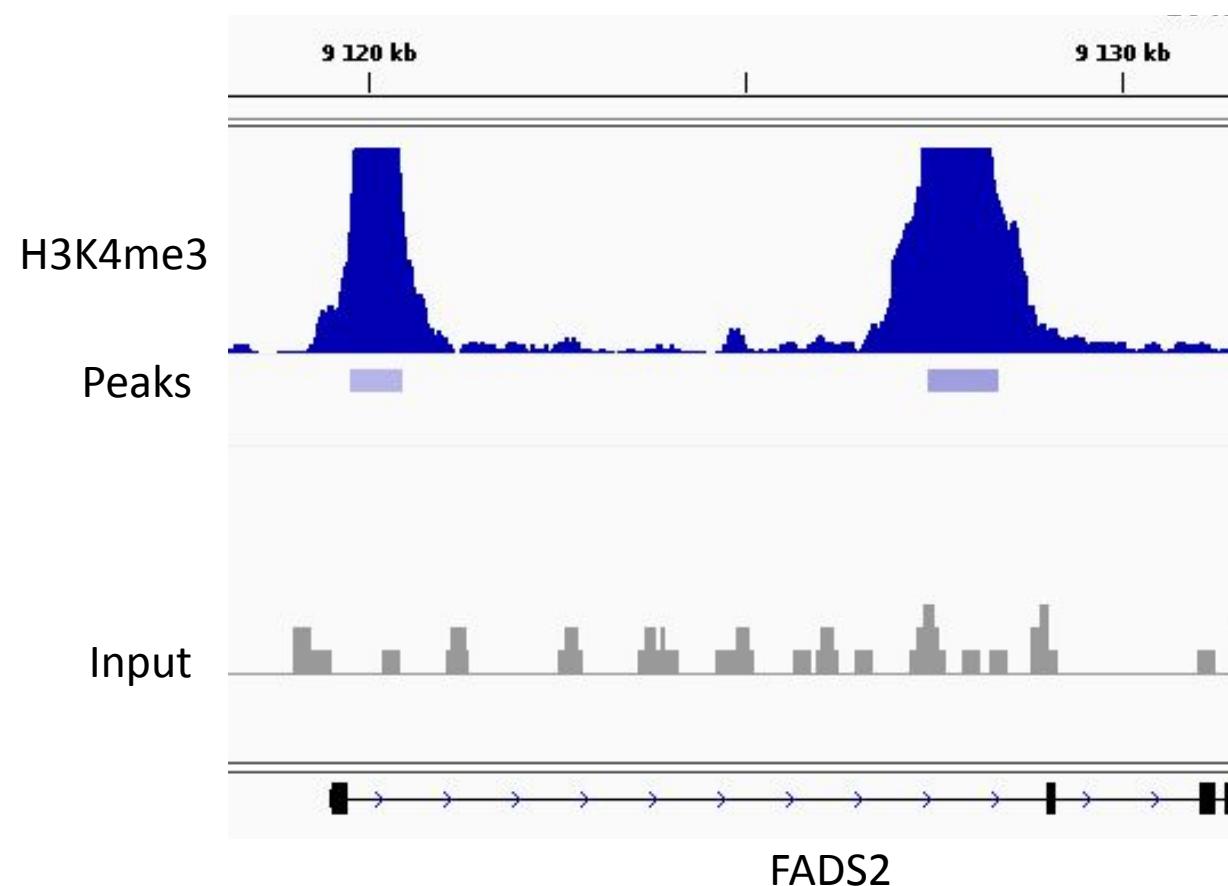
From FAANG data to Clustered Heatmaps



● Example of visualization



• ChIP-Seq processing



Jaccard index on peaks

$$J(A,B) = \frac{\text{The intersect of } A \text{ & } B}{\text{division}}$$

$$J(A,B) = \frac{\text{The union of } A \text{ & } B}{A \cup B}$$

• ChIP-Seq annotated by target

Species

ChIPSeq

Gallus_gallus

Filters

Highlights

Annotation

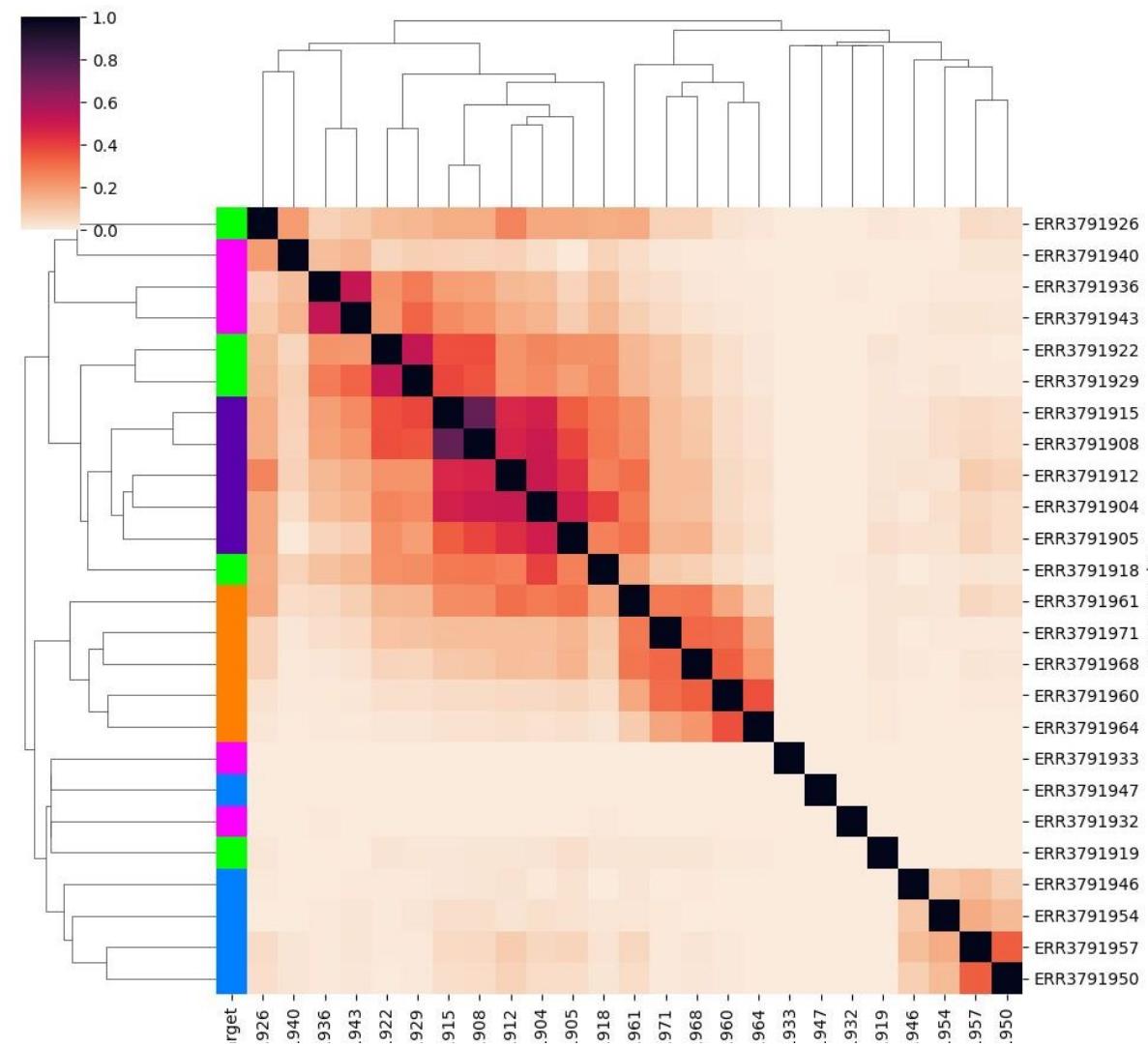
Annotate :

experiment_target

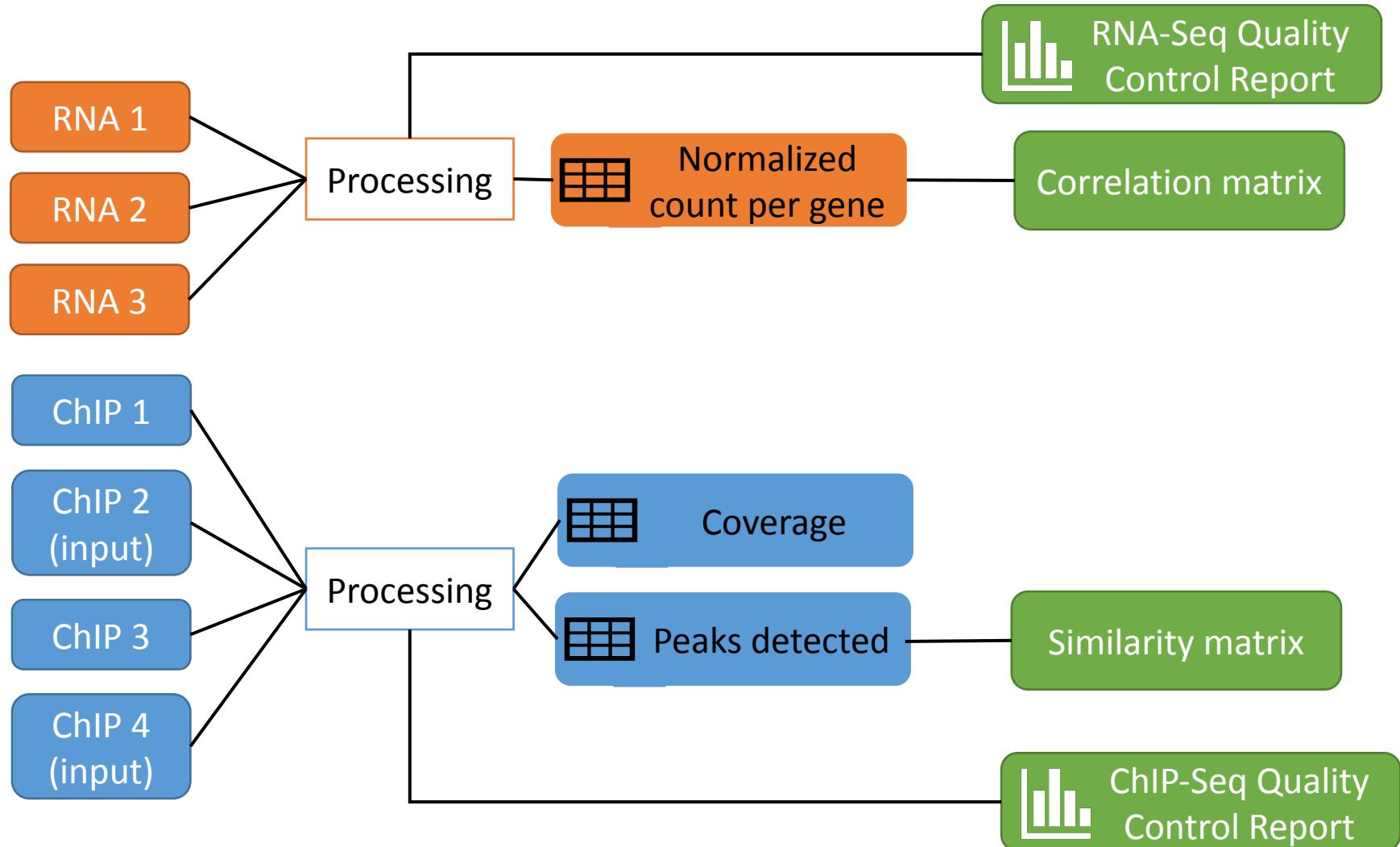
- █ CTCF
- █ H3K27Ac
- █ H3K27Me3
- █ H3K4Me1
- █ H3K4Me3
- █ input DNA

Graph options

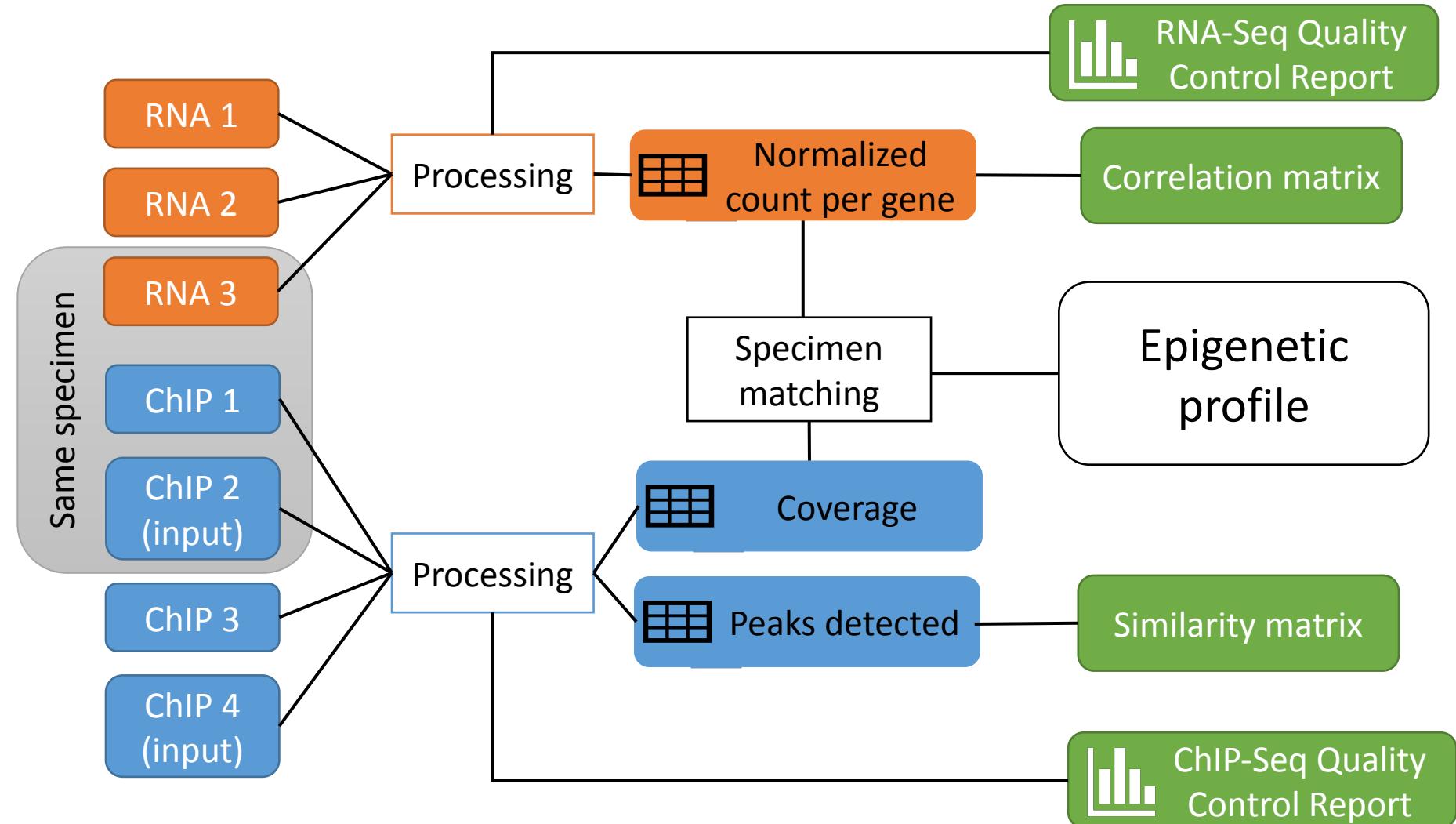
Submit



• Epigenetic profiles

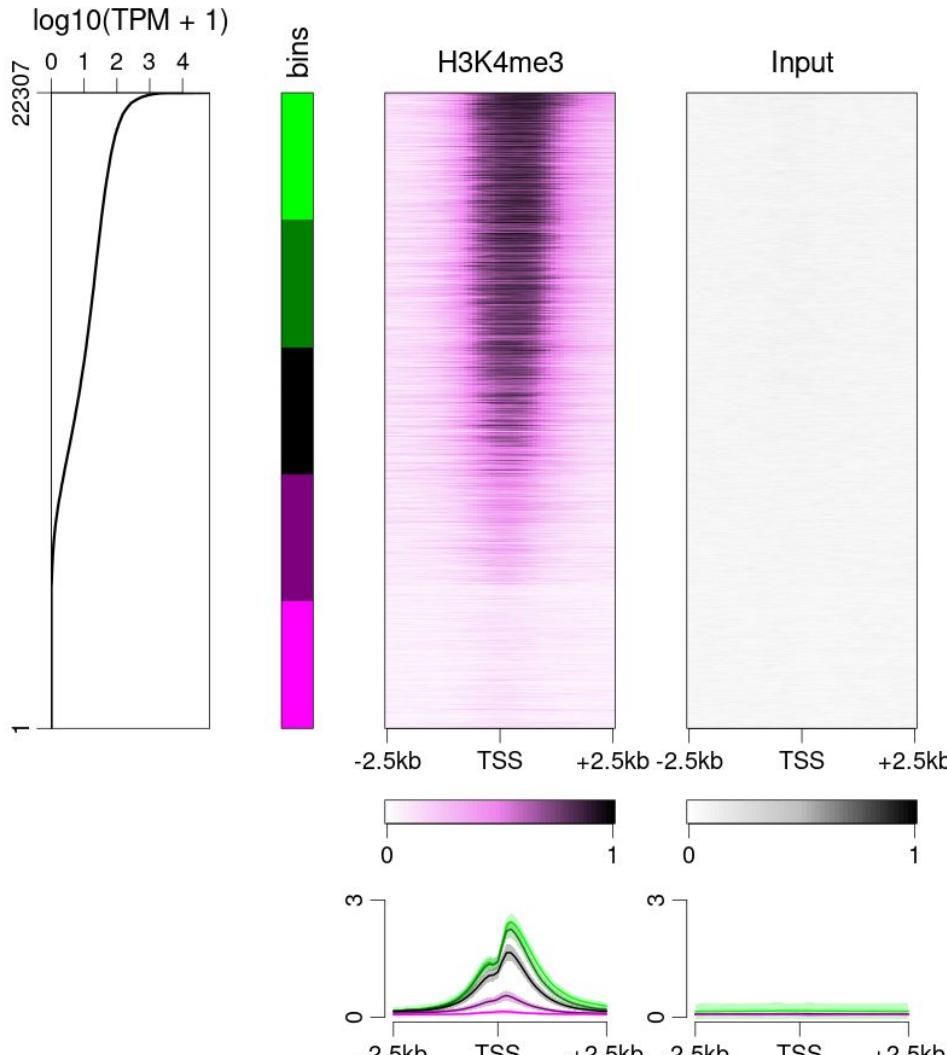


• Epigenetic profiles



● Epigenetic profiles

H3K4me3 ChIP-Seq in *Bos taurus* cerebellum



• Epistack

- R package created by Safia Saci and Guillaume Devailly

- **Work in progress**

- Web app improvement
 - UX/UI improvement
 - Allow user to add their own data
- Processing
 - Add support for DNA methylation data
 - Automatically integrate new data
- Add VizFaDa to FAANG data portal

- ## Acknowledgements



- Guillaume Devailly
- Sylvain Foissac
- Peter Harrison
- Alexey Sokolov
- Safia Saci

INRAe

Ae

INRAe

Visualization of FAANG data with VizFaDa

07/07/2022 – Laura Morel

Thank you for your attention



INRAE

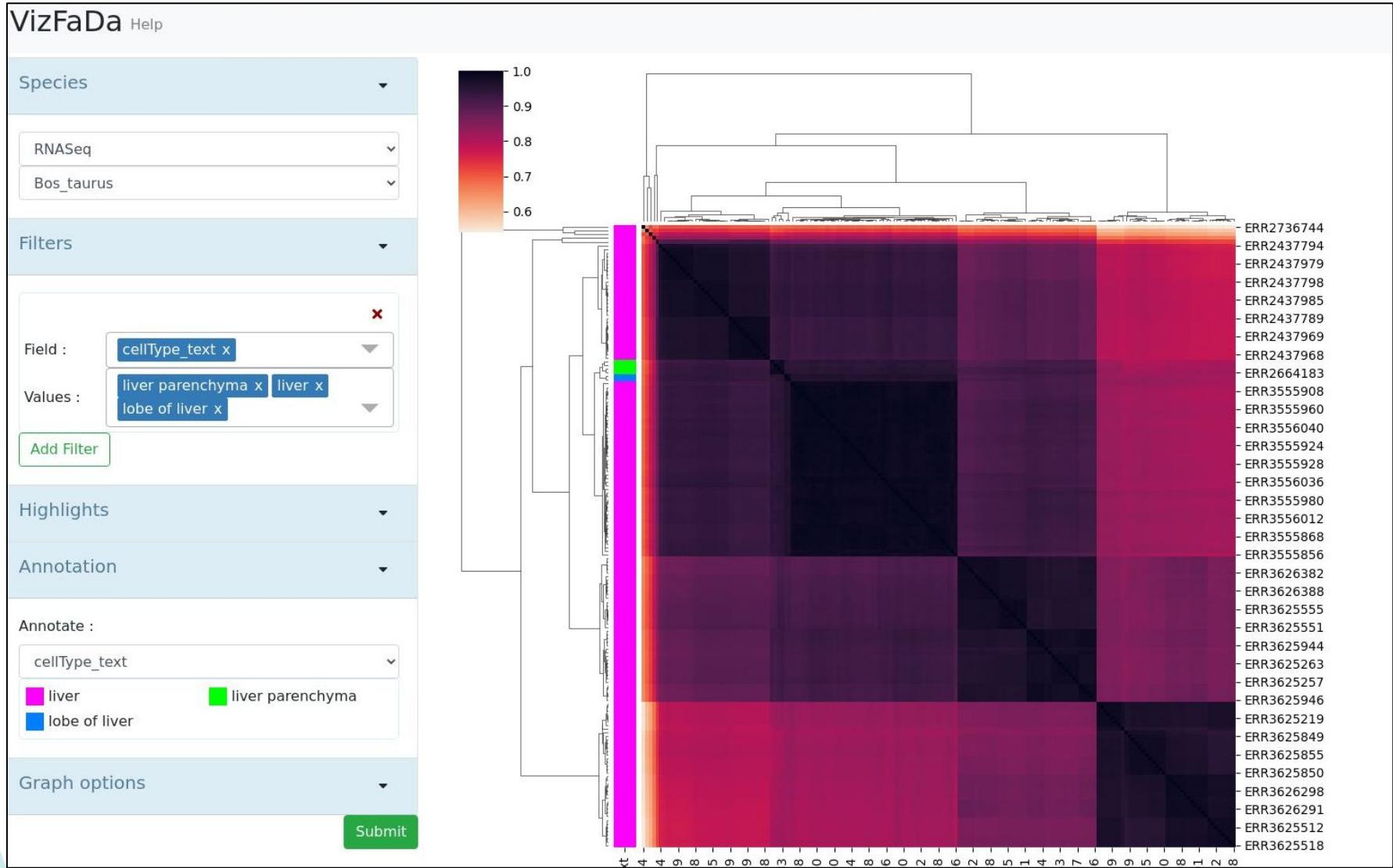
Visualization of FAANG data with VizFaDa

07/07/2022 – Laura Morel

p. 20

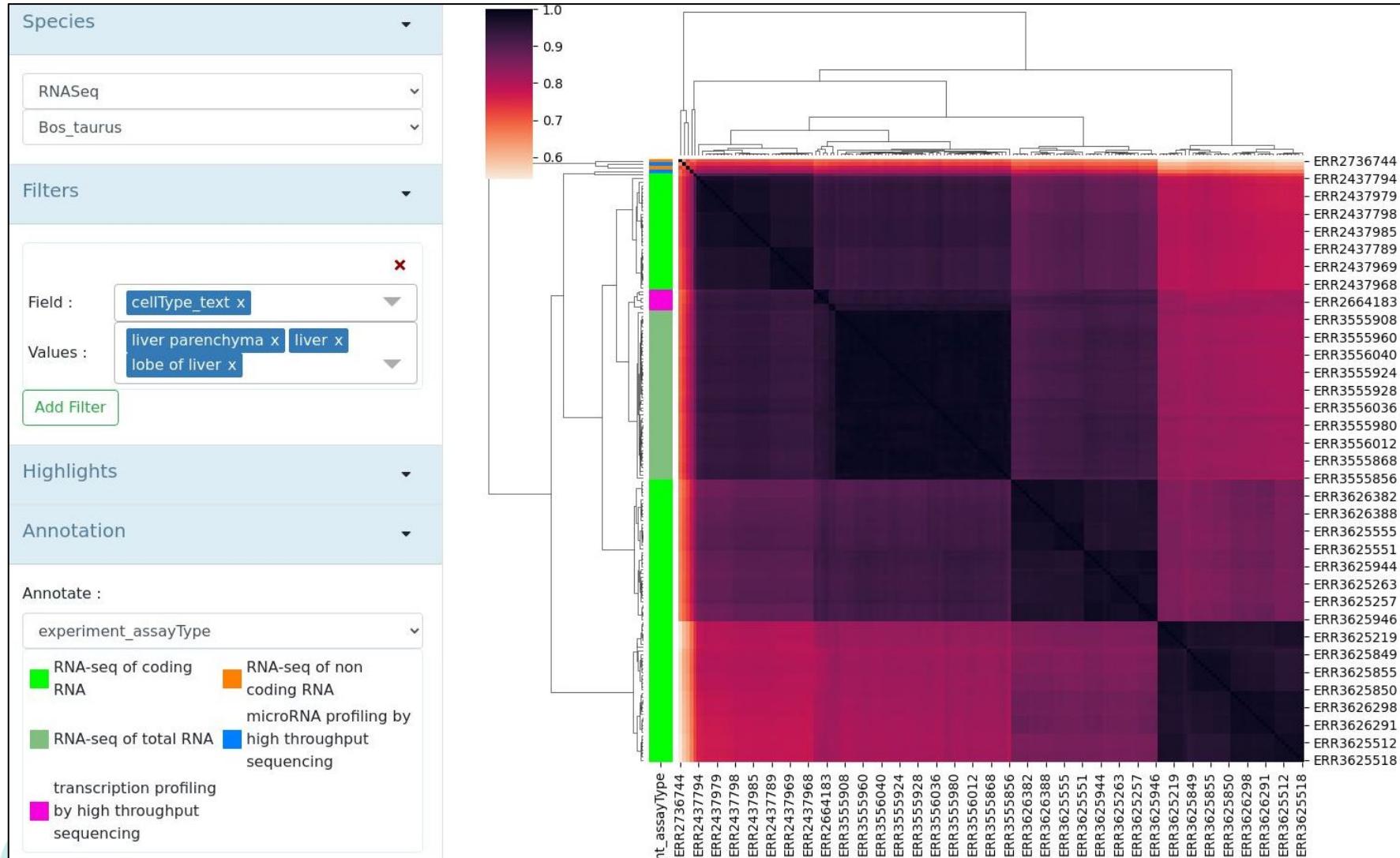
● Web Application

Example: RNA-Seq data of liver tissue in cows



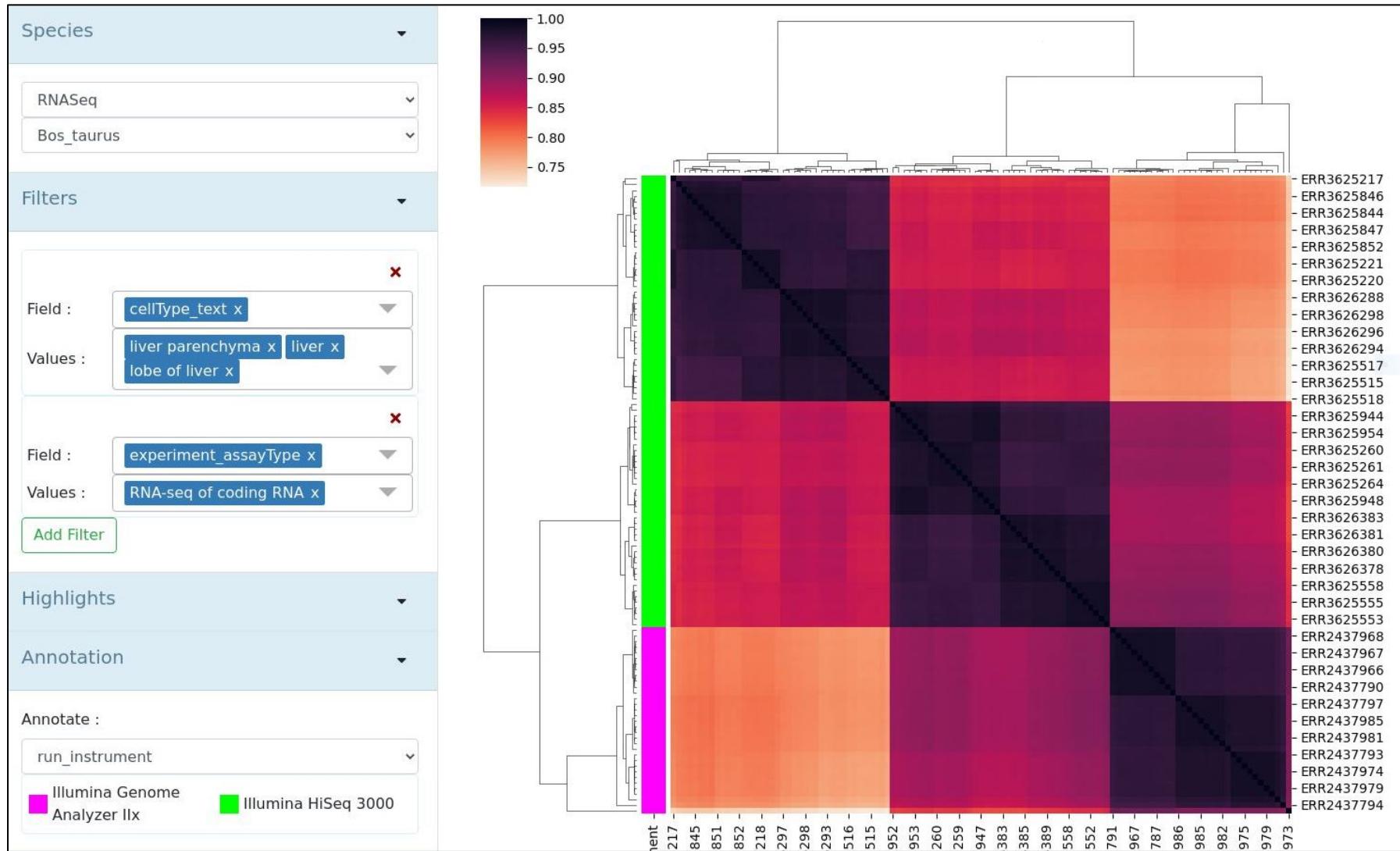
● Web Application

Example: RNA-Seq data of liver tissue in cows



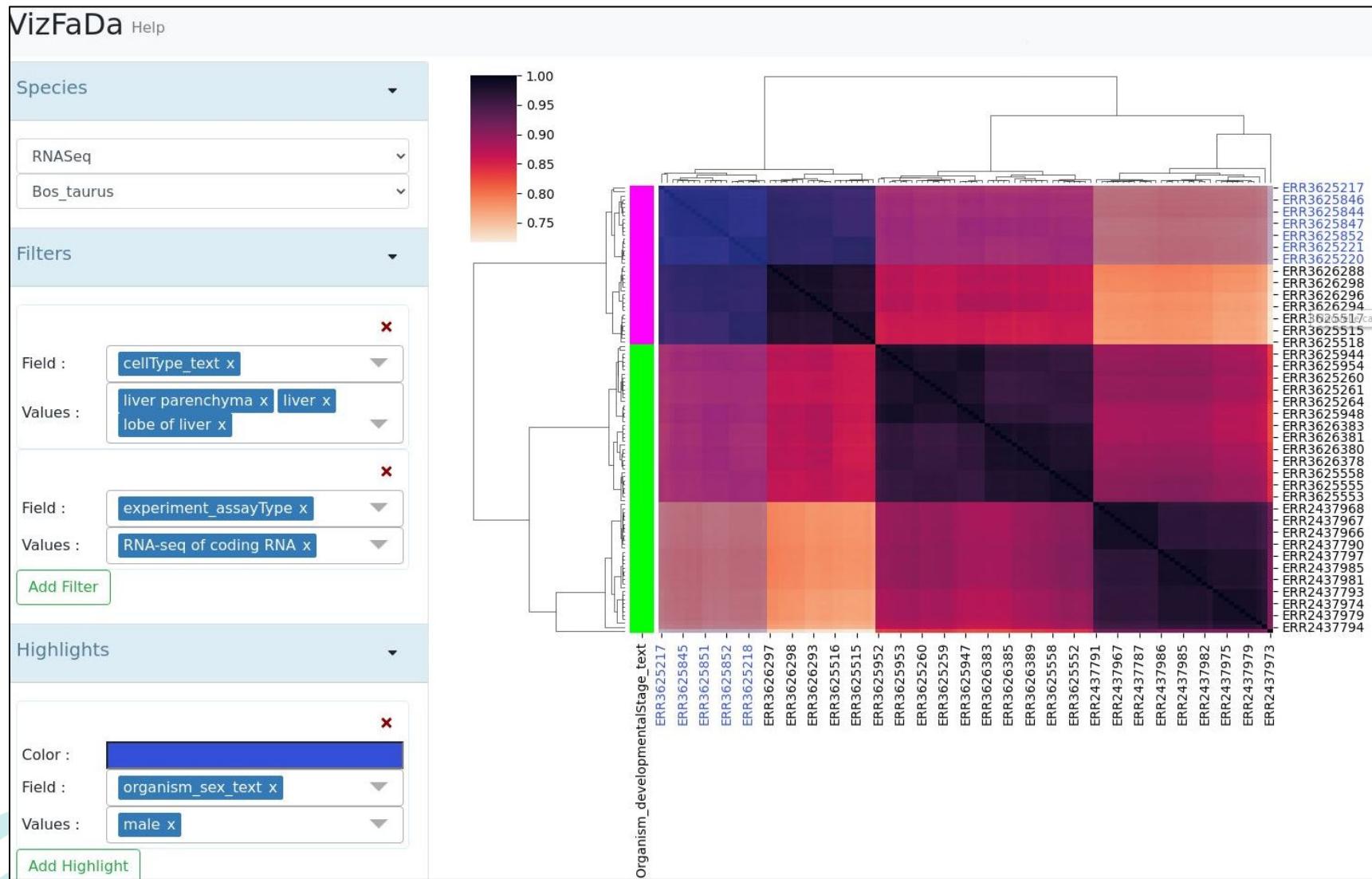
● Web Application

Example: RNA-Seq data of liver tissue in cows



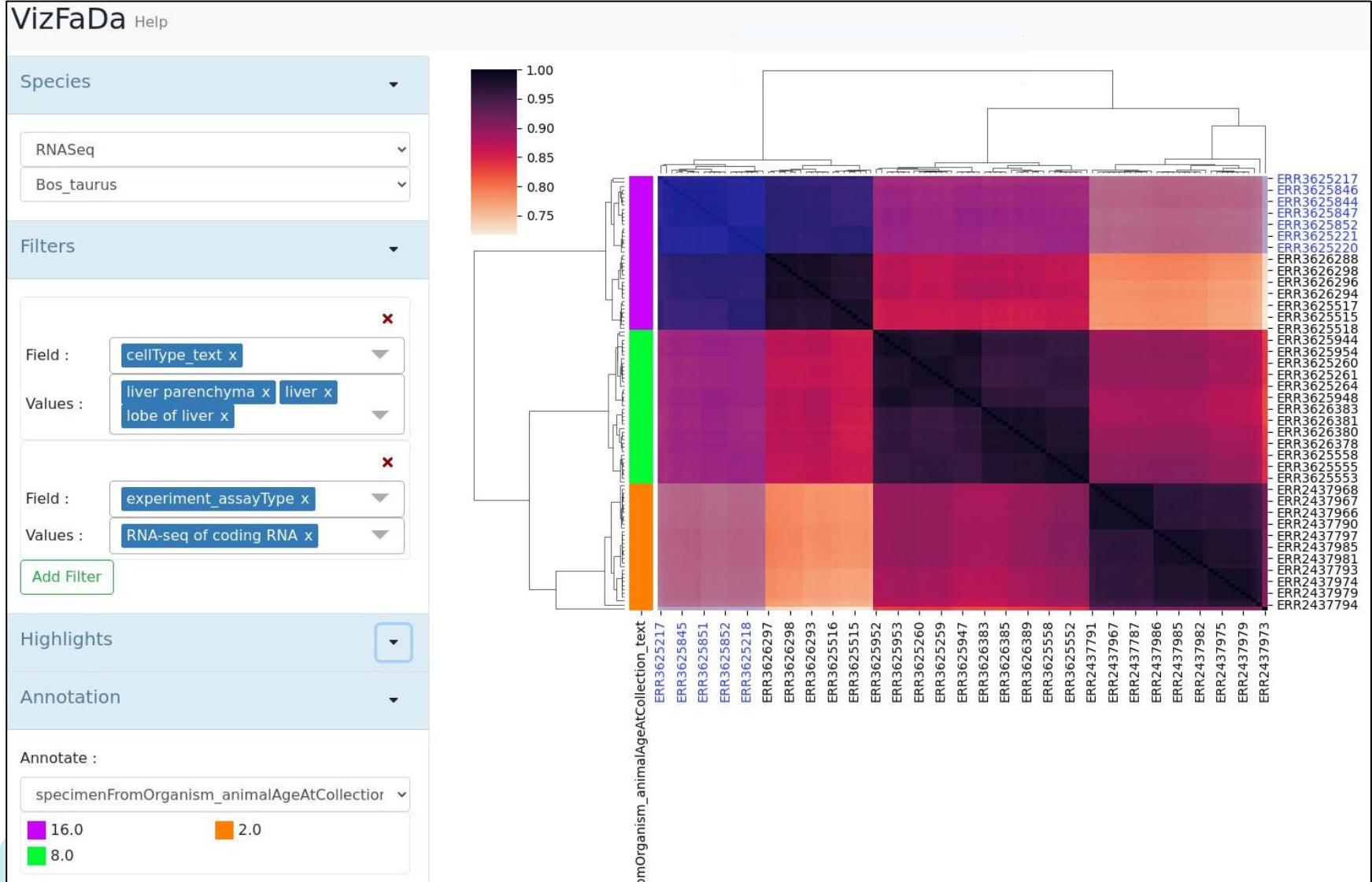
● Web Application

Example: RNA-Seq data of liver tissue in cows



● Web Application

Example: RNA-Seq data of liver tissue in cows



● FAANG Data Portal

<https://data.faang.org>

The screenshot shows the FAANG Data Portal interface. At the top, there's a navigation bar with links for Home, Records, Projects, Summary, Submit, Search, and Help. Below the navigation is a section titled "FAANG files". On the left, there's a sidebar with three filter sections: "Standard" (Legacy 145407, FAANG 20528), "Species" (Bos taurus 56139, Gallus gallus 34578, Sus scrofa 28920), and "Assay type" (RNA-Seq 91177). In the center, there's a green button labeled "Download data". To its right is a table with columns: File name, Study, Experiment, Species, Assay type, Target, Instrument, Specimen, Standard, and Paper published. The table contains six rows of data, each with a delete icon (a circled X) in the last column.

File name	Study	Experiment	Species	Assay type	Target	Instrument	Specimen	Standard	Paper published
SRR999281_2	SRP030134	SRX357350	Bos taurus	RNA-Seq	Unknown	Illumina Genome Analyzer IIx	SAMN02361031	Legacy	(X)
SRR999281_1	SRP030134	SRX357350	Bos taurus	RNA-Seq	Unknown	Illumina Genome Analyzer IIx	SAMN02361031	Legacy	(X)
SRR999280_2	SRP030134	SRX357349	Bos taurus	RNA-Seq	Unknown	Illumina Genome Analyzer IIx	SAMN02361031	Legacy	(X)
SRR999280_1	SRP030134	SRX357349	Bos taurus	RNA-Seq	Unknown	Illumina Genome Analyzer IIx	SAMN02361031	Legacy	(X)
SRR999279_2	SRP030134	SRX357348	Bos taurus	RNA-Seq	Unknown	Illumina Genome Analyzer IIx	SAMN02361030	Legacy	(X)
SRR999279_1	SRP030134	SRX357348	Bos	RNA-Seq	Unknown	Illumina	SAMN02361030	Legacy	(X)