



VizFaDa: Visualisations of FAANG data

Guillaume Devailly, Laura Morel, Peter W Harrison

► To cite this version:

Guillaume Devailly, Laura Morel, Peter W Harrison. VizFaDa: Visualisations of FAANG data. JO-BIM 2021, Oct 2021, Paris, France. . hal-03401235

HAL Id: hal-03401235

<https://hal.inrae.fr/hal-03401235>

Submitted on 25 Oct 2021

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.



Distributed under a Creative Commons Attribution 4.0 International License

VizFaDa: Visualisations of FAANG data

Laura Morel¹, Peter W. Harrison², Guillaume Devailly¹

1: GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan, France
2: EMBL-EBI, Hinxton, United Kindgom



Abstract

- Functional Annotation of Animal Genomes (FAANG) is an open consortia gathering genome regulation data in farmed animals: 🐑 🐷 🐮 🐔 🐟
- Raw genomic data and rich sample metadata are freely available through a data portal hosted at EMBL-EBI: data.faang.org
- We aim to provide data visualisations of FAANG datasets (RNA-seq, ChIP-seq, DNA methylation), embedded to the FAANG data portal.
- Visualisations will provide an overview of the available data to foster data re-use.

➤ Data processing **nextflow** & **nf-core**

Netxflow pipelines based on nf-core:

- A Nextflow module to get metadata & download data
- Fast RNA-seq quantification using Salmon pseudo-aligner
 - Pearson correlation on $\log_{10}(\text{TPM} + 1)$
- ChIP-seq processing using bwa and macs2
 - Jaccard index from bedtools

➤ Notable by-products

Uniformly processed outputs on FAANG data:

- FastQC reports
- Gene and transcript expression matrices (count & TPM)
- ChIP-seq peaks and bigwig tracks

➤ Epigenetic profiles at TSS

Generated using the R package Epistack:

- github.com/GenEpi-GenPhySE/epistack

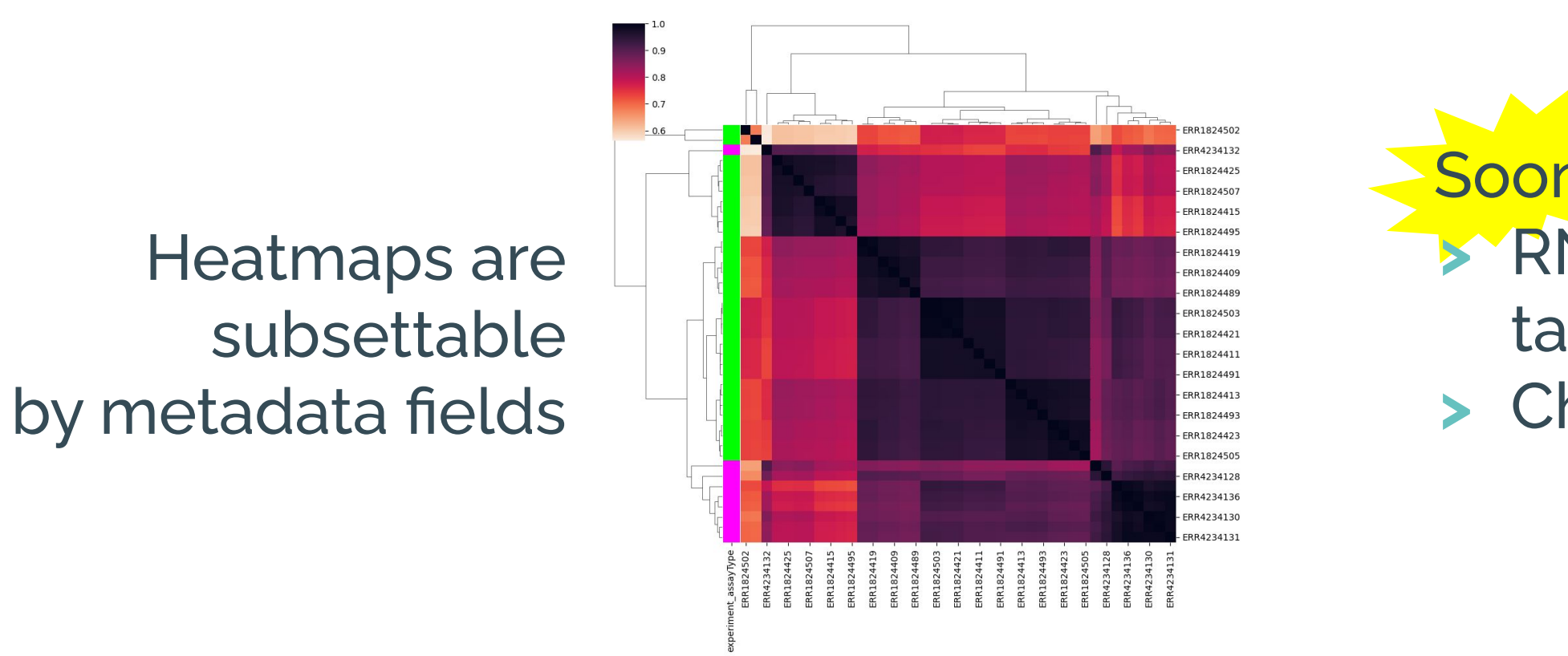
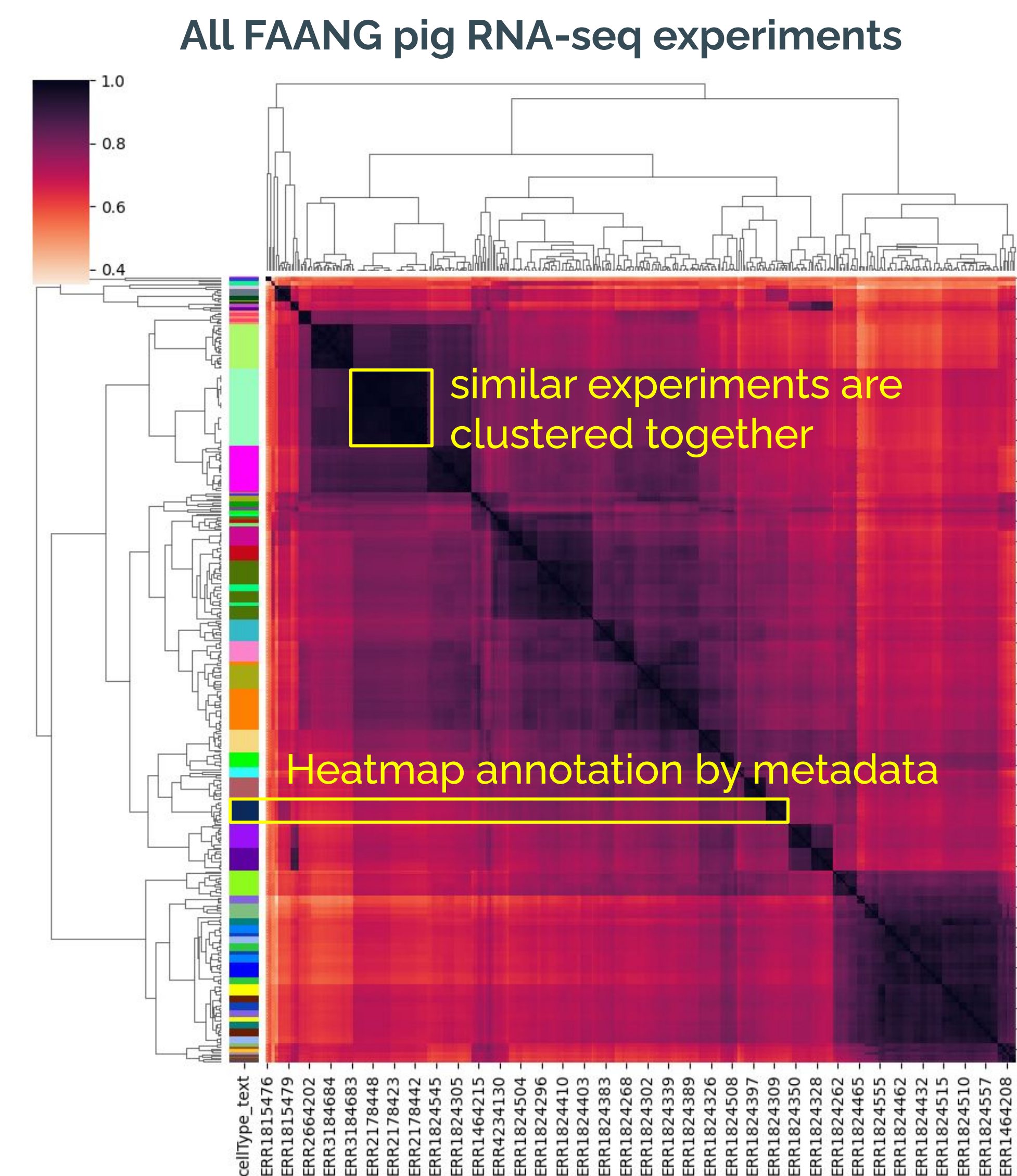
Integrative analysis of RNA-seq and ChIP-seq / DNA methylation data:

- data quality
- exploratory analysis

➤ Sample-to-sample correlation heatmaps

By species and by experiment types (RNA-seq, ChIP-seq, DNA methylation)

- overview of available data
- visualising the relationships between samples
- outlier identification



Heatmaps are subsettable by metadata fields

- Soon: upload your own data
- RNA-seq: gene expression table (count, TPM)
- ChIP-seq: peak bed file

➤ VizFaDa timeline

2020-07	2021-01	2021-07	2022-01	2022-07
VizFaDa starts	All RNA-seq processed	Web-app (alpha)	All ChIP-seq processed	Web-app (beta)
			All DNA methylation processed	Web app (RC)

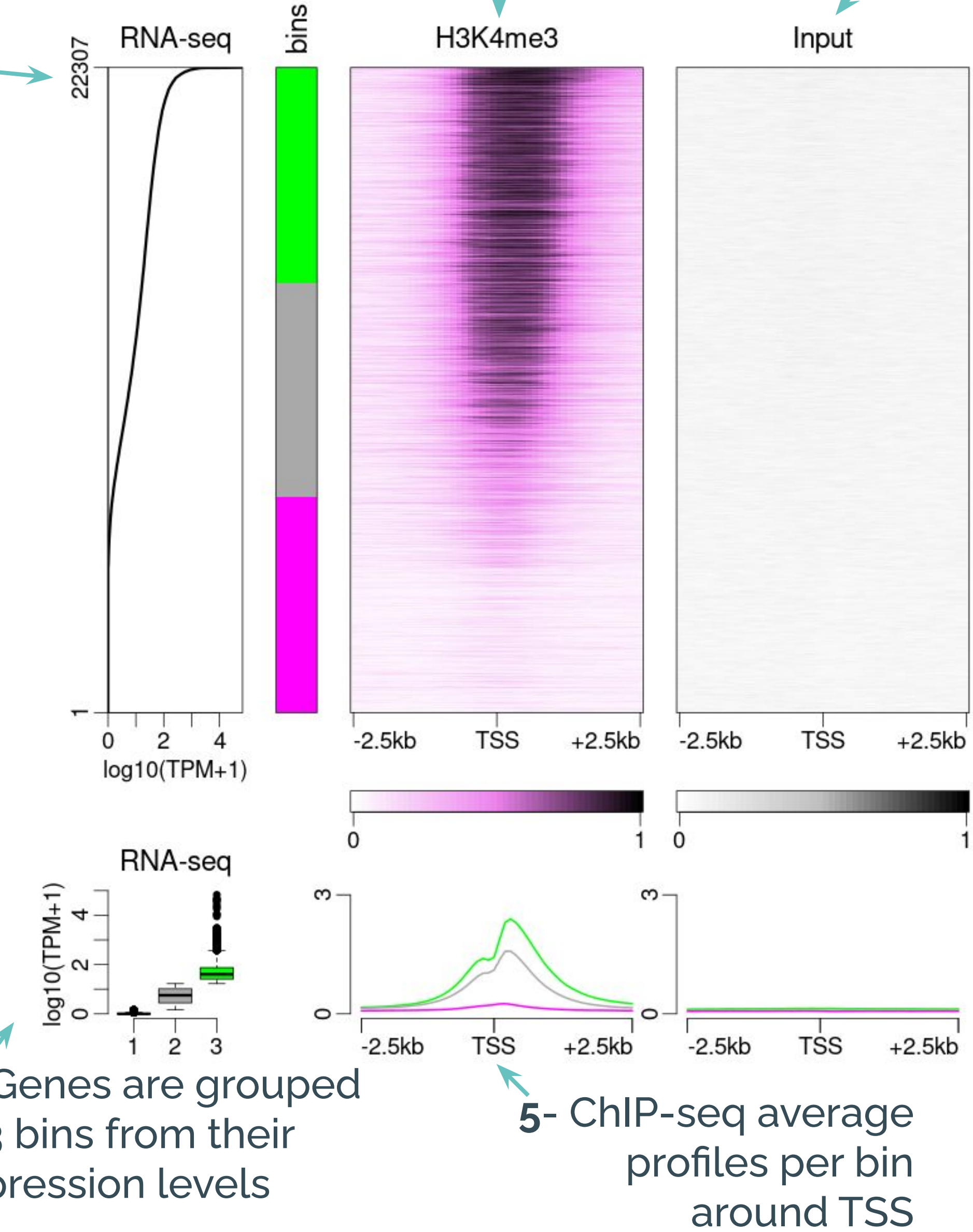
Maintenance:

- new FAANG data
- using new reference genomes and annotations

H3K4me3 ChIP-seq from a cow cerebellum

2- ChIP-seq profiles around Transcription Start Sites
3- ChIP-seq control profiles

1- The 22,307 annotated *Bos taurus* genes are sorted according to their expression level from a matching FAANG RNA-seq experiment



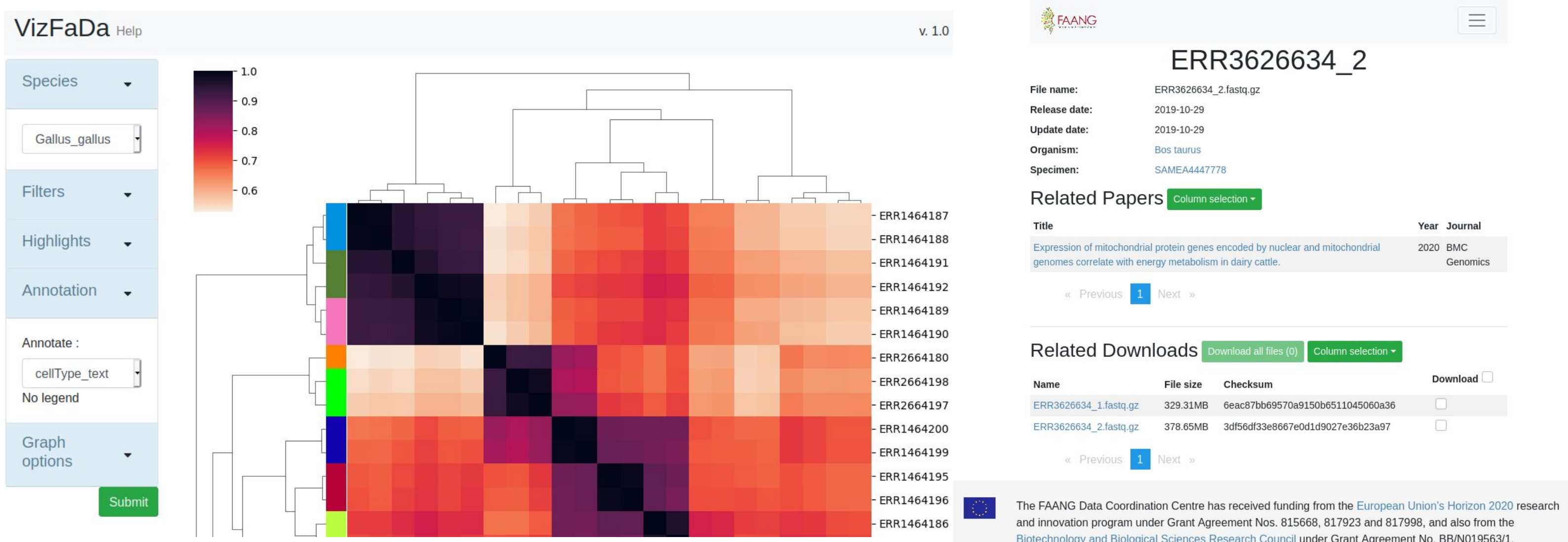
4- Genes are grouped in 3 bins from their expression levels

5- ChIP-seq average profiles per bin around TSS

➤ Web applications

➤ Soon: a dedicated web-app for visualisations

➤ Soon: integrations to the data.faang.org portal



➤ Fundings

VizFaDa is funded by an Open Science Grant from ANR. The FAANG data portal is supported by the H2020 awards to AquaFAANG (grant no. 817923), GENE-SWitCH (grant no. 817998) and BovReg (grant no. 815668) and the EMBL