



VizFaDa: Towards interactive visualisations of FAANG data

Guillaume Devailly

► To cite this version:

Guillaume Devailly. VizFaDa: Towards interactive visualisations of FAANG data. 2. FAANG-Europe Workshop on Functional Annotation of Animal Genomes, FAANG-Europe, Feb 2020, Prague, Czech Republic. hal-03181770

HAL Id: hal-03181770

<https://hal.inrae.fr/hal-03181770>

Submitted on 25 Mar 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.

➤ **VizFaDa: Towards interactive visualisations of FAANG data**

Guillaume Devailly
[@G_Devailly](#)
Toulouse, France

The VizFaDa project

National funding call on 'Open Sciences' funding short projects.

Proposal: Interactive visualisations to foster FAANG data re-use

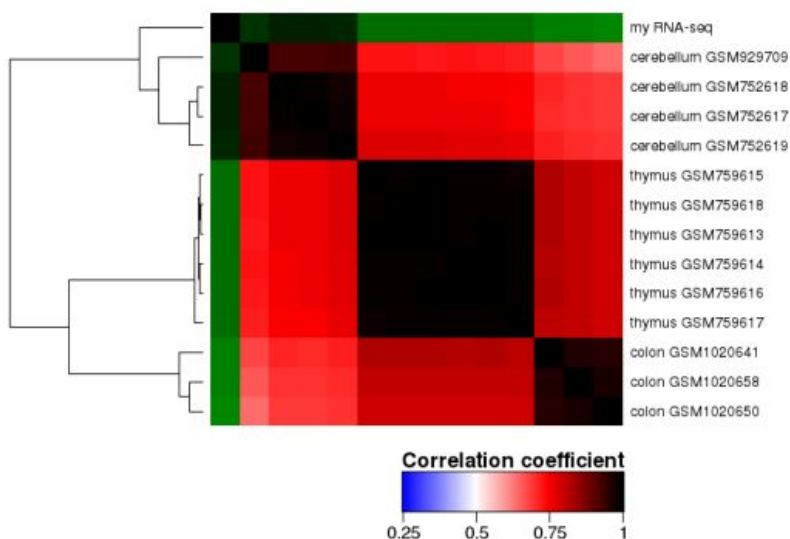
https://gdevailly.netlify.com/files/vizfada_proposal.pdf

76k€ for 2 years, started 2020-01-01

Proposed visualisations

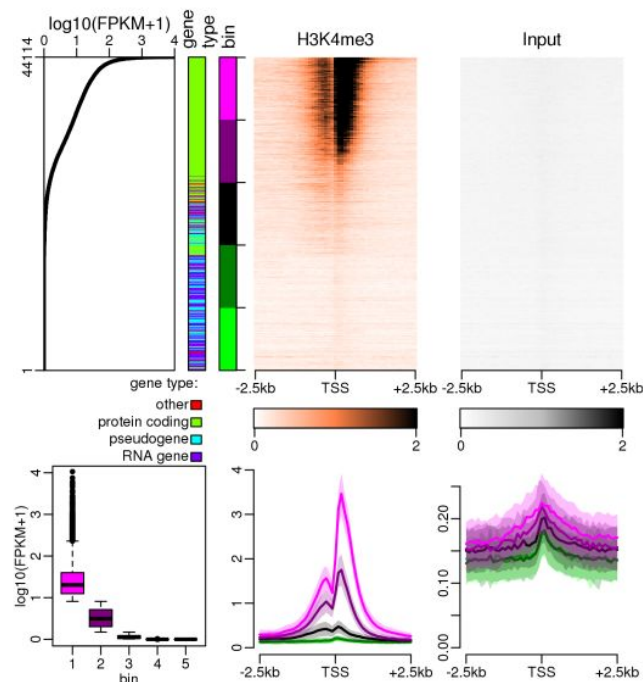
Correlation heatmaps

- ❖ quick overview of the datasets
- ❖ comparison with user-provided data
- ❖ identification of outlier experiments



Stack profiles of epigenomics tracks

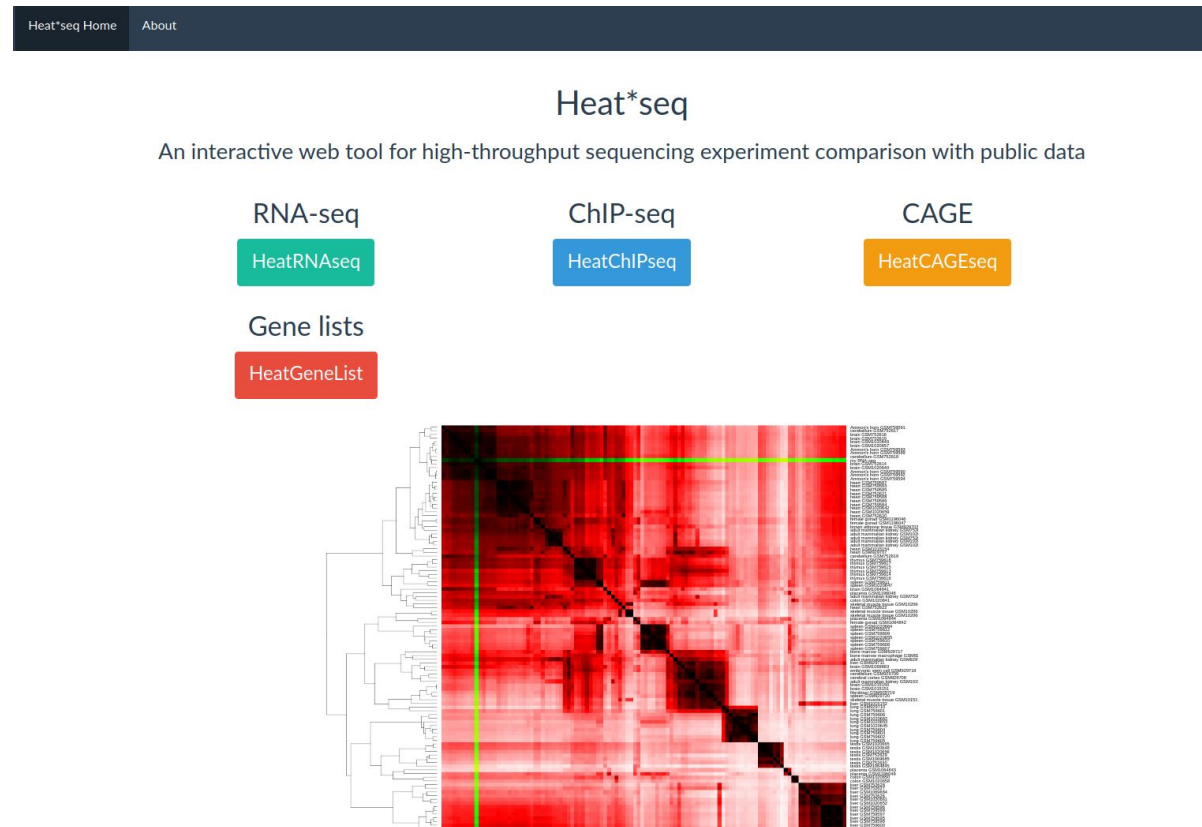
- ❖ complex but informative visualisations
- ❖ identifications of biological associations and of experimental biases



Correlation Heatmaps

Based of Heat*seq:

- Web app: www.heatstarseq.roslin.ed.ac.uk
- Source code: github.com/gdevailly/HeatStarSeq_gh
- Paper: doi.org/10.1093/bioinformatics/btw407



HeatRNAseq workflow

Expression matrix

Gene name (~40.000)	Exp 1	Exp 2	Exp 3	...	Exp 77
ENSG000000000003	3.983	2.361	10.216	...	80.583
ENSG000000000005	0.071	0.260	0.000	...	3.329
ENSG000000000419	10.277	2.893	14.153	...	42.639
...	...	...	...	...	...
ENSG00000273493	0.000	0.000	0.000	...	0.000

Correlation matrix

Pearson's, after log10 scaling

\	Exp 1	Exp 2	Exp 3	...	Exp 77
Exp 1	1	0.942	0.938	...	0.663
Exp 2	0.942	1	0.917	...	0.680
Exp 3	0.938	0.917	1	...	0.706
...	...	...	...	1	...
Exp 77	0.663	0.680	0.706	...	1

Clustered heatmap



HeatChIPseq workflow

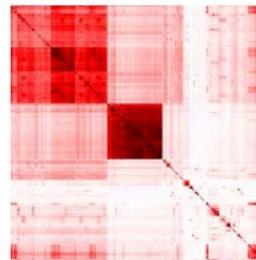
Binary peak matrix

Coordinates (~700.000 non overlapping regions)	Exp 1	Exp 2	Exp 3	...	Exp 690
chr1:10073-10413	F	T	F	...	F
chr1:16110-16390	F	F	T	...	T
chr1:29198-29688	F	F	F	...	F
...	...	...	...	...	...
chrY:28709160- 28709494	T	T	F	...	F

Correlation matrix

\	Exp 1	Exp 2	Exp 3	...	Exp 690
Exp 1	1	0.059	0.786	...	0.035
Exp 2	0.059	1	0.058	...	0.118
Exp 3	0.786	0.058	1	...	0.047
...	...	...	...	1	...
Exp 690	0.035	0.118	0.047	...	1

Clustered heatmap



Live Demo!



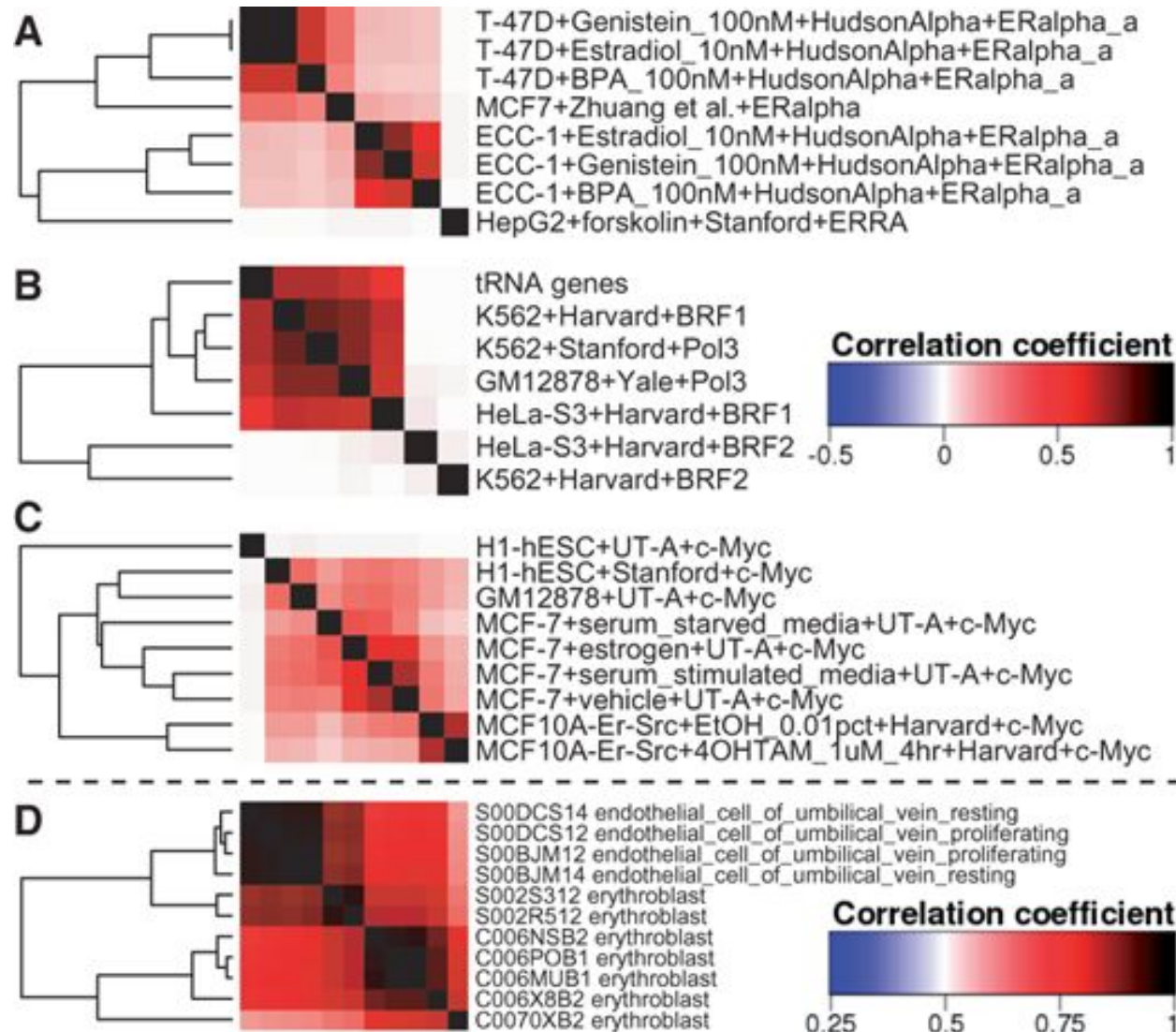
www.heatstarseq.roslin.ed.ac.uk



INRAE

VizFaDa: Towards interactive visualisations of FAANG data
2020/02/12 / FAANG workshop / Guillaume Devailly

If live demo fails...

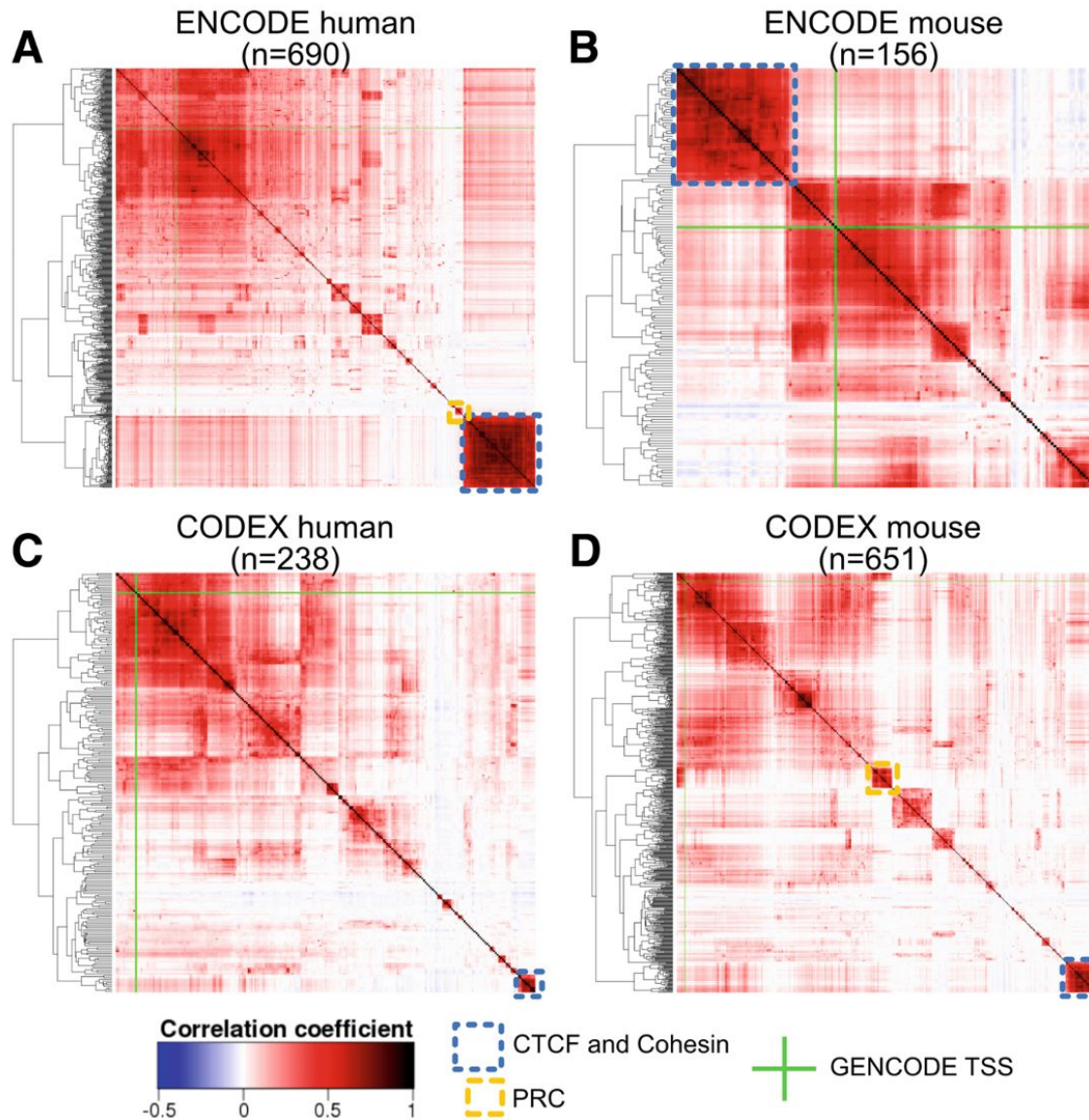


INRAE

VizFaDa: Towards interactive visualisations of FAANG data

2020/02/12 / FAANG workshop / Guillaume Devailly

If live demo fails...



Perspectives

Heatmaps by species and by data-type:

- ❖ Multi-species comparison?
- ❖ Histone marks and TF ChIP-seq on the same heatmaps?
- ❖ DNA methylation data?
- ❖ HiC data?

Methods:

- ❖ Best distance metrics: correlations coefficient, jaccard index, entropy, etc. ?
- ❖ Scaling beyond 1000 experiments per heatmaps?
- ❖ Technological stack?

Data portal integration:

- ❖ Semi-independent web app?
- ❖ 'Add experiment to cart' UI à la ENCODE?

Stack profiles of epigenetic marks

Based of PEREpigenomics

- Web app: joshiapps.cbu.uib.no/perepigenomics_app
- Source code: github.com/gdevailly/perepigenomicsAnalysis & forgemia.inra.fr/guillaume.devailly/perepigenomics_app
- Preprint: soon



Live Demo!

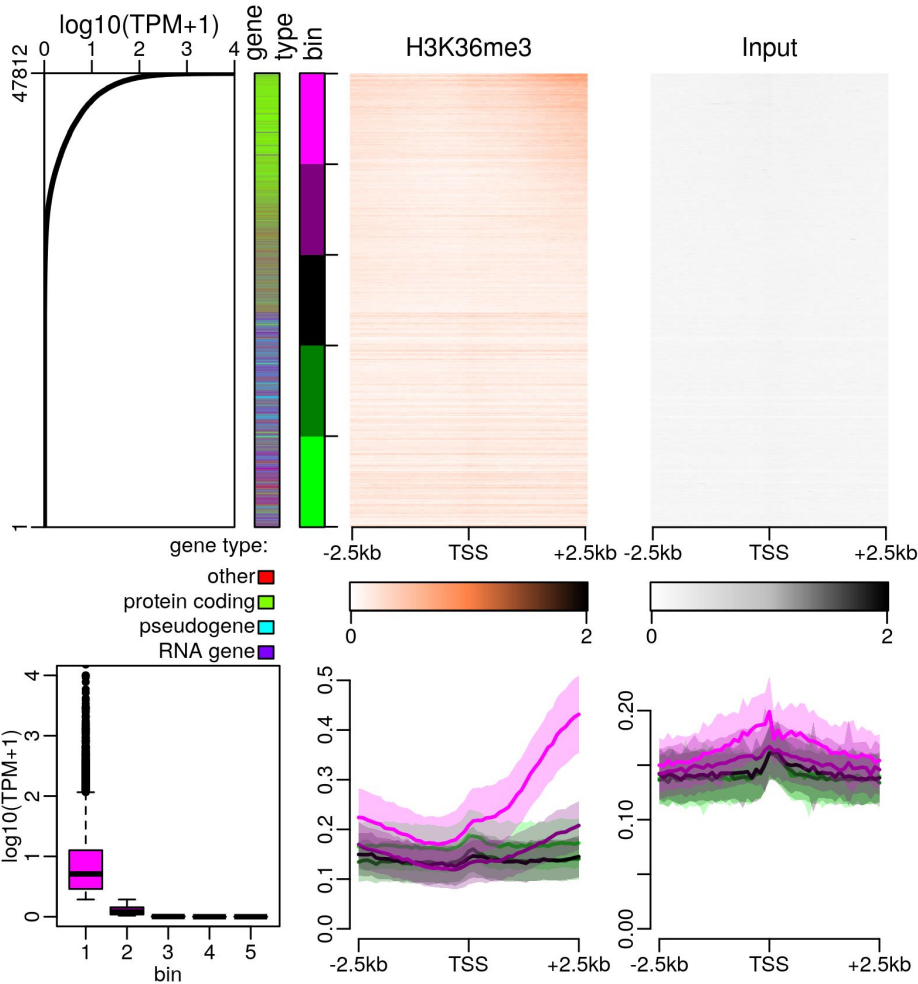


joshiapps.cbu.uib.no/perepigenomics_app/

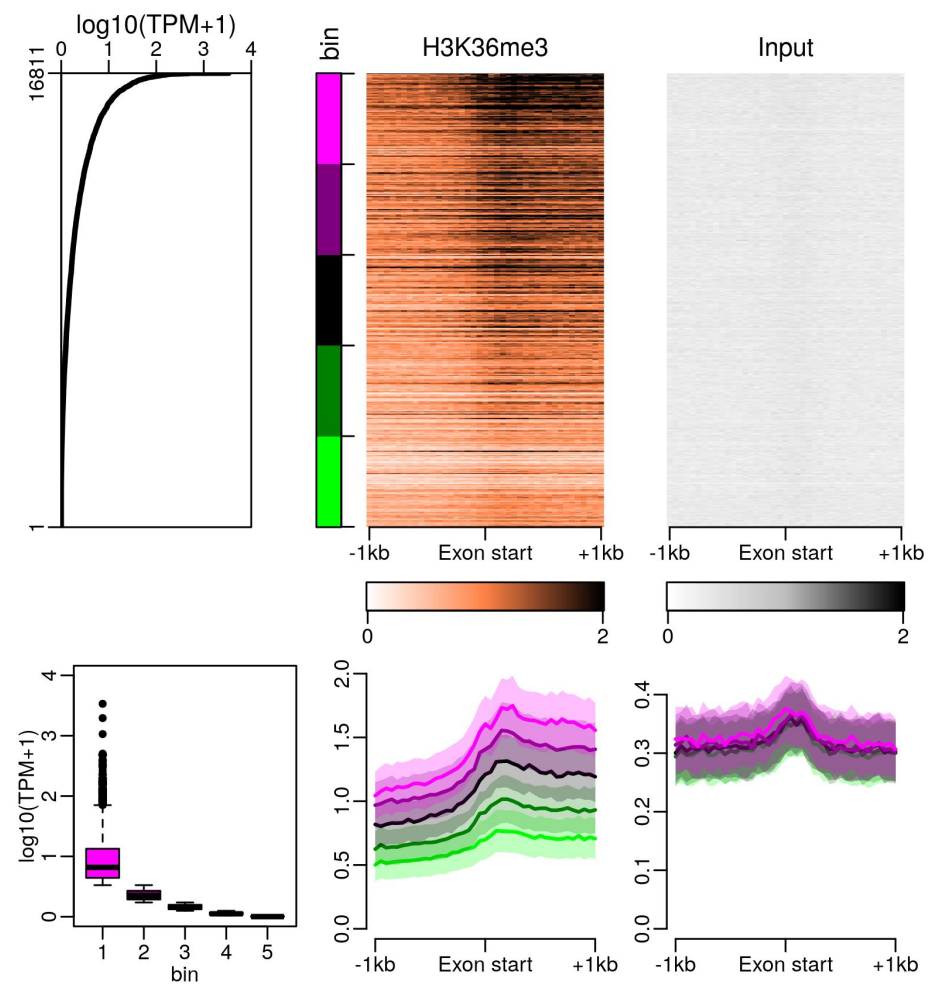


If live demo fails...

Adult liver, TSS

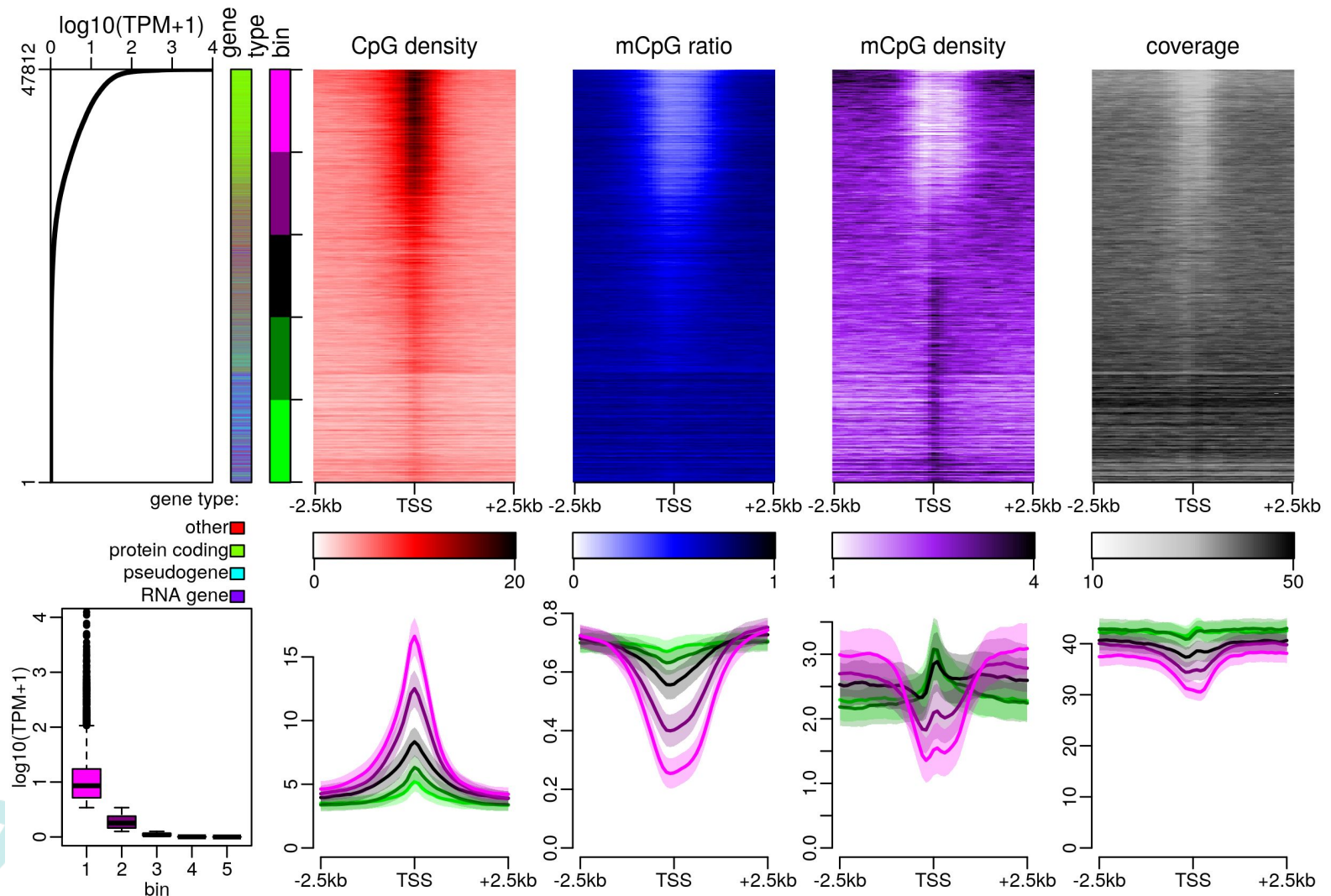


Adult liver, middle exons



If live demo fails...

DNA methylation (WGBS), Pancreas, TSS



Perspectives

- ❖ RNA-seq + epigenetic data in the same tissue

Data portal integration:

- ❖ Picture gallery in experiment page
- ❖ + independent global gallery?

Common processing of FAANG data:

- ❖ will it be on time?
- ❖ NextFlow nfc core pipeline?

Keeping up with new FAANG data?

Keeping up with new genome versions & annotations?

- ❖ Automated workflows?
- ❖ Re-running everything? Cache invalidation?



Consortia:

We are recruiting!

Bioinformatics Engineer, fixed term, 18 months

<https://genphyse.toulouse.inra.fr/content/bioinformatics-engineer-ie-fixed-term-18-months-inrae-toulouse-france>

guillaume.devailly@inrae.fr

Toulouse, France



- ❖ Guillaume Devailly (data processing and visualisations)
- ❖ Sylvain Foissac (scientific advisor)

Toulouse, France



- ❖ Philippe Bardou (web application development)
- ❖ Patrice Dehais (system administrator)



Hinxton, Royaume-Uni



- ❖ Peter Harrison (FAANG Data Coordination Center)
- ❖ Guy Cochrane (FAANG Data Coordination Center)

INRAE

VizFaDa: Towards interactive visualisations of FAANG data
2020/02/12 / FAANG workshop / Guillaume Devailly

Heat*seq & PEREpigenomics acknowledgments



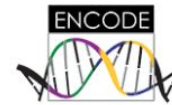
Anna Mantsoki



Anagha Joshi



Barry Horne
(IT)



Guillaume DEVAILLY received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement No PCOFUND-GA-2012-600181.

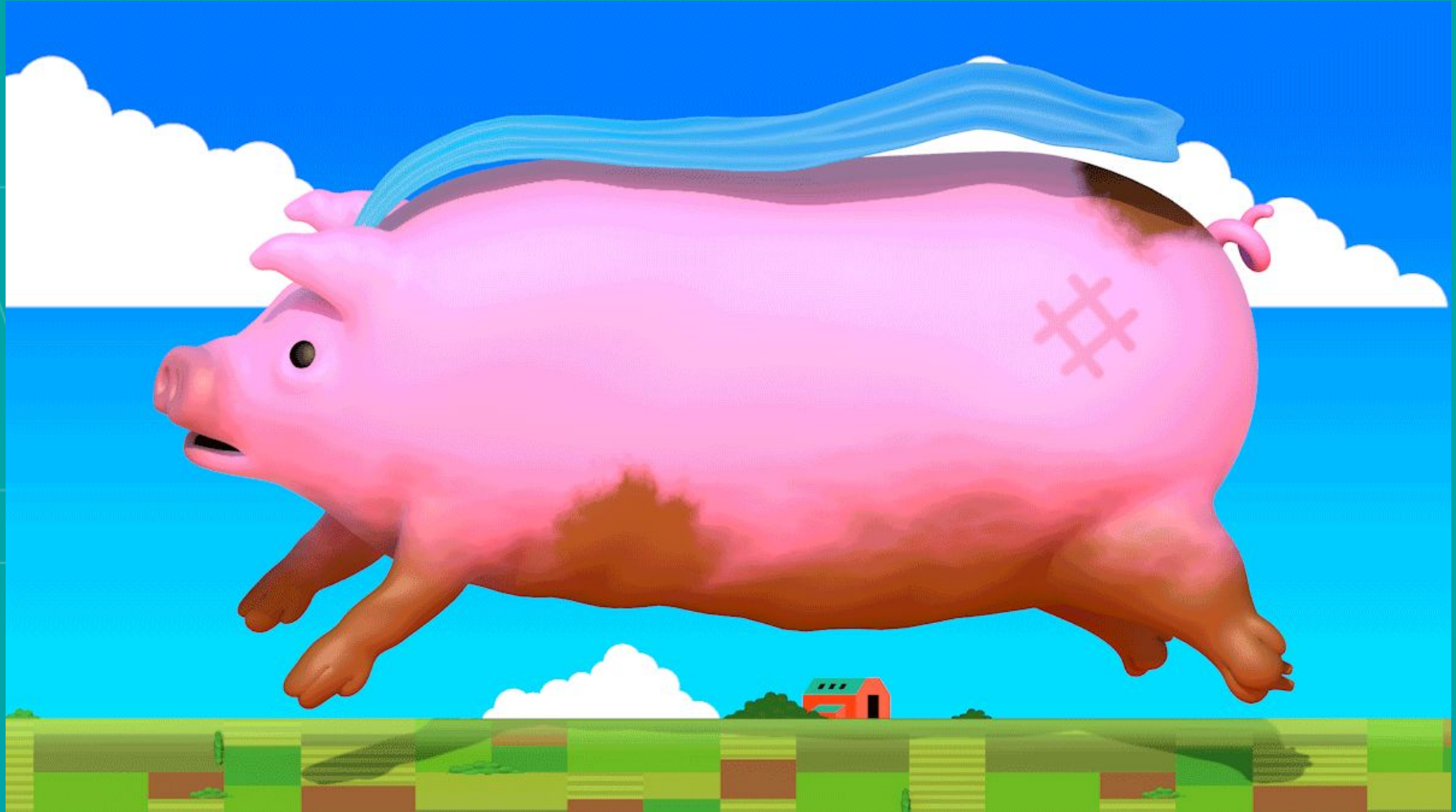


VizFaDa: Towards interactive visualisations of FAANG data
2020/02/12 / FAANG workshop / Guillaume Devailly

Thanks for your attention!

Slides available online: <https://frama.link/this-presentation>

INRAE



INRAE

➤ Bonus presentation:

https://gdevailly.github.io/2020-02-03_MIAT/VizualisationsEpigenomicsData.html