

The Global FAANG Data Portal and its role in the future EuroFAANG infrastructure

Peter W. Harrison

Genome Analysis Team Leader

peter@ebi.ac.uk



EuroFAANG

EMBL-EBI





Data Coordination Centre at EMBL-EBI

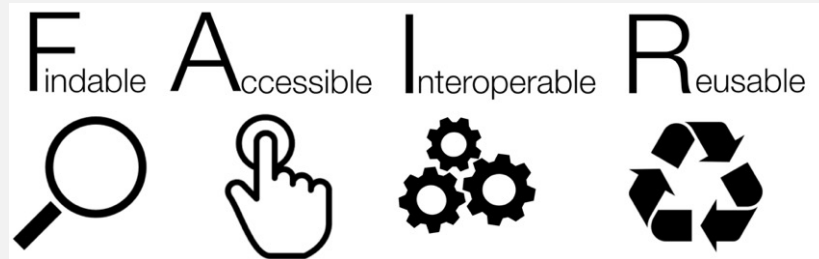
Supports core FAANG aims of data **openness, reusability, and standardisation** to create a harmonised rich **genome to phenome resource**.

- Ensuring data is richly described, available and searchable.
- Ensuring data is consistently reported and presented.
- Facilitating standardisation of formats, protocols and analysis methods.
- Facilitating data openness, reusability and cross-project analysis.

Funded by EU EuroFAANG projects, but serving the global FAANG community.

What makes FAANG datasets special

- Rich, consistent and validated metadata descriptions.
- Standardised set of core assays from each contributing project.
- Mandatory sampling, experiment and analysis protocols connected to each dataset and available with the datasets for download.
- Projects using standardised analysis pipelines, often now attached to each dataset.
- A data platform and community drive that ensures data is open and FAIR.
- Innovative data presentation and contextual genome visualisation.



To accelerate genome to phenome research

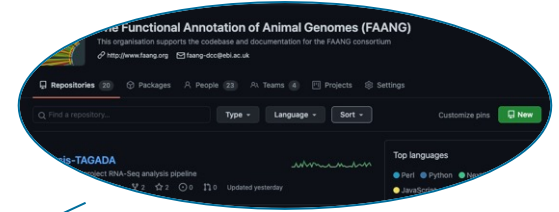
Photo: CODATA

FAANGs standardised datasets

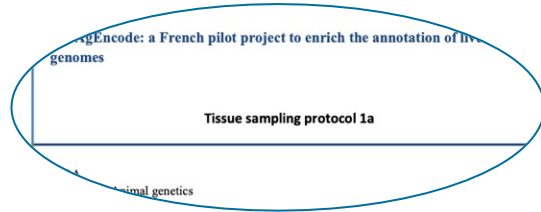


Rich, validated
metadata

Publications

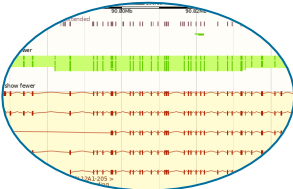


Analysis pipeline
and parameters



Detailed sampling and
analysis protocols

FAANG dataset



Secondary analyses,
track hubs and
annotations

Legacy 147784

Legacy contextual datasets

The FAANG Data Portal



- A single access point to all FAANG metadata, data, pipelines and publications.
- Direct access to download data from underlying public archives.
- Automatically identifies dataset (re)use in publications, and links publications to datasets.
- Intuitive search and filtering.
- Data visualisation and cross referencing.

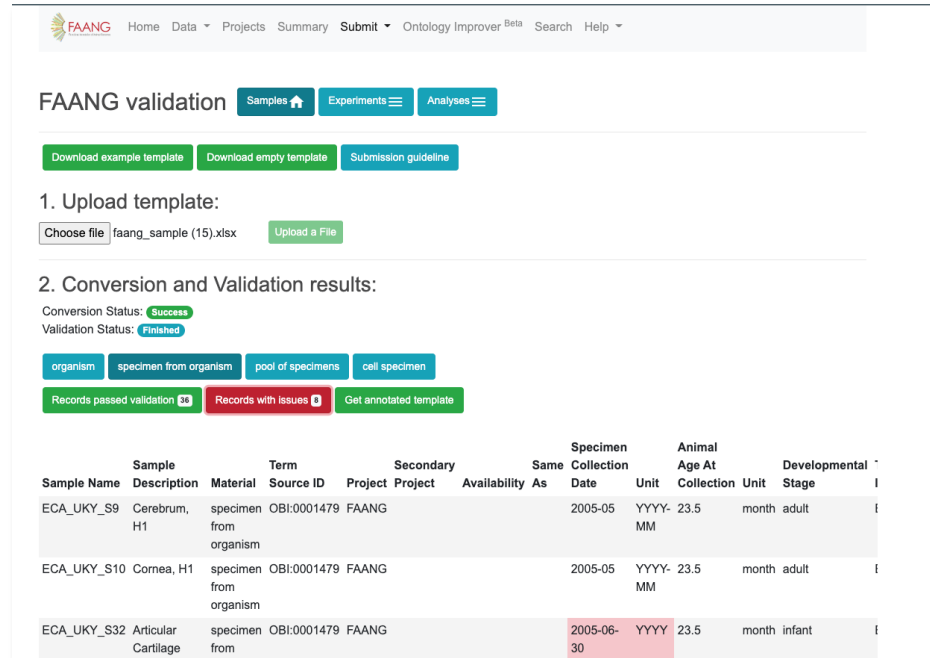
The screenshot shows the FAANG Data Portal interface. At the top, there is a navigation bar with the FAANG logo and links for Home, Data, Projects, Summary, Submit, Ontology Improver (Beta), Search, and Help. Below the navigation bar, the page title is "FAANG datasets". There are active filters for "FAANG" and a "Remove all filters" button. A "Download data" button is also visible. On the left side, there are filter panels for "Species" (with counts for Sus scrofa: 35, Gallus gallus: 17, Bos taurus: 15) and "Assay type" (with counts for ATAC-seq: 15, RNA-seq of coding RNA: 15). The main content is a table of datasets with columns for Dataset, Title, Species, Archive, Assay type, Number of Experiments, Number of Specimens, Number of Files, Standard, and Paper published.

Dataset	Title	Species	Archive	Assay type	Number of Experiments	Number of Specimens	Number of Files	Standard	Paper published
PRJNA436715	Transcriptome related to pubertal development in Brahman heifers	Bos indicus	ENA		72	72	144	FAANG	✓
PRJNA417062	RNA sequencing of Bos indicus testicular samples (adult and fetus) and liver samples (adult) [total RNA]	Bos indicus	ENA		25	25	150	FAANG	✗
PRJNA417061	RNA sequencing of Bos indicus testicular samples (adult and fetus) and liver samples (adult) [small RNA]	Bos indicus	ENA		25	25	100	FAANG	✗
PRJNA414087	Ovis aries Rambouillet Genome sequencing and assembly	Ovis aries	ENA		69	31	329	FAANG	✓

<https://data.faang.org/>

Validation and brokered submission

- Rich and detailed metadata rulesets.
- All FAANG data goes through pre-submission validation, that ensures rich data descriptions.
- Validation service not only highlights errors, it suggests improvements such as more specific ontologies.
- Brokered submission to underlying archives simplifies multi-archive submission for users.
- Can now also submit track hubs and Nextflow pipeline configuration files.





The screenshot displays the FAANG validation interface. At the top, there is a navigation bar with the FAANG logo and links for Home, Data, Projects, Summary, Submit, Ontology Improver Beta, Search, and Help. Below this, the 'FAANG validation' section is active, with buttons for 'Samples', 'Experiments', and 'Analyses'. Three main action buttons are visible: 'Download example template', 'Download empty template', and 'Submission guideline'. The interface shows two steps: '1. Upload template:' with a file upload area containing 'faang_sample (15).xlsx' and an 'Upload a File' button; and '2. Conversion and Validation results:' showing a 'Success' conversion status and a 'Finished' validation status. Below this, there are filter buttons for 'organism', 'specimen from organism', 'pool of specimens', and 'cell specimen'. Summary statistics show 'Records passed validation' (24), 'Records with issues' (1), and a 'Get annotated template' button. A table lists the validation results for three samples.

Sample Name	Sample Description	Material	Term Source ID	Secondary Project	Availability	Specimen Collection Date	Animal Age At Collection	Developmental Stage
ECA_UKY_S9	Cerebrum, H1	specimen from organism	OBI:0001479	FAANG	Same As	2005-05	YYYY-MM	23.5 month adult
ECA_UKY_S10	Cornea, H1	specimen from organism	OBI:0001479	FAANG	Same As	2005-05	YYYY-MM	23.5 month adult
ECA_UKY_S32	Articular Cartilage	specimen from	OBI:0001479	FAANG	Same As	2005-06-30	YYYY	23.5 month infant

<https://data.faang.org/validation/samples>

Project Specific Pages





[Login to BovReg Portal](#)


The BovReg consortium will provide a comprehensive map of functionally active genomic features in cattle and how their (epi)genetic variation in beef and dairy breeds translates into phenotypes.

Tweets from @BovReg

[Follow on Twitter](#)

BovReg @BovReg · Jan 9
Profiling of open chromatin in developing pig 🐷
BovReg's laboratory course on ATAC-Seq: data generation and bioinformatic analysis held on November 2022 here we present you the video presentation by Mazdak Salavati @roslininstitute

👍 Enjoy the video 🎥

 youtube.com
ATAC-Seq Profiling of open chromatin in developing pig -

For more information on the BovReg project: [BovReg](#)

Related Publications


[Column selection](#)

Title	Year	Journal
Comparative Analysis of the Circular Transcriptome in Muscle, Liver, and Testis in Three Livestock Species.	2021	Front Genet
Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle.	2019	Front Genet
Single-cell RNA sequencing of freshly isolated bovine milk cells and cultured primary mammary epithelial cells.	2021	Sci Data
Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in <i>Bos taurus</i> Bulls.	2020	Int J Mol Sci

« Previous **1** Next »

Related Pipelines

[Column selection](#)

Pipeline name	Assay type	Pipeline link	Pipeline documentation	Platform
BovReg/nf-core-maseq	RNA-seq	https://github.com/BovReg/nf-core-maseq	https://github.com/BovReg/nf-core-maseq#readme	


- Specific customised sub project views.
- Automatically displays just that projects data, pipelines, publications, metadata and track hubs.
- Can be further customised with project descriptions, figures, specific table views etc.
- Request one for your FAANG project.







<https://data.faang.org/projects>

Community sourced Ontology improvement tool

- Ontology descriptions frequently not appropriate for animal agriculture.
- New data portal tool allows users to flag ontologies that need improvement and collectively work and vote on changes.
- Authenticated and tracks provenance.
- Updates are forwarded to underlying ontologies.
- Look out for virtual workshops later in 2023 to train on using tool and kickstart collaborative improvement

Livestock Ontologies

Search livestock ontologies 

Term	Ontology Type	Ontology ID	Status ↑
Sus scrofa	organism	NCBITaxon_9823	 Verified
Female	sex	PATO_0000383	 Awaiting assessment
Capra hircus	organism	NCBITaxon_9925	 Awaiting assessment
Gallus gallus	organism	NCBITaxon_9031	 Verified
specimen from organism	material	OBI_0001479	 Needs Improvement
Sample			 Not supported

<https://data.faang.org/ontology>

Track Hub submission



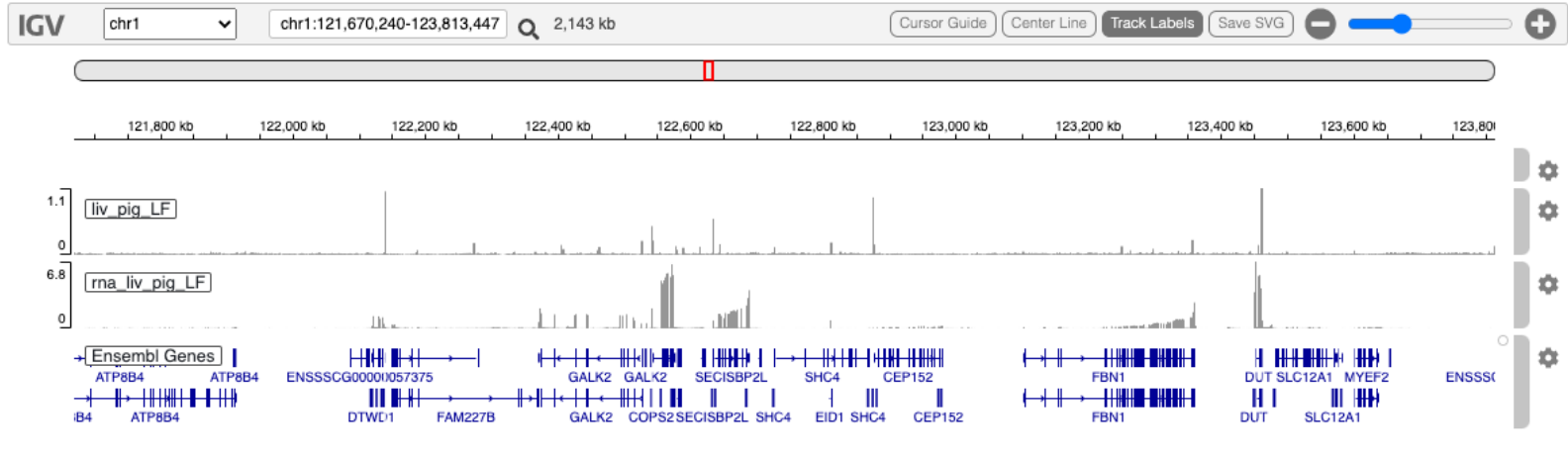
- New FAANG brokered submission, validation, storage and presentation of community genomic track hubs.
- Tracks are automatically linked to relevant data and connections shown in the FAANG Data portal to view in Genome browsers.

The screenshot shows the 'Track Hubs Submission' page on the FAANG website. At the top, there are navigation links: Home, Data, Projects, Summary, Genome Browser Beta, Submit, Ontology Improver Beta, and Help. Below the navigation, there are three buttons: 'Download example template', 'Download empty template', and 'Submission guidelines', along with a 'Logout' button. The 'Upload template' section shows a file upload area with 'Choose file' and 'trackhubs (2).xlsx' selected, and an 'Upload' button. A red error message states 'Error: Template validation failed'. Below this, there are radio buttons for 'Submit new trackhub' (selected) and 'Update existing trackhub'. The 'Validation Results' section shows a red error message: 'Fix errors in "Tracks Data" sheet'. Below this, there are tabs for 'Hub Data', 'Genome Data', and 'Tracks Data'. A table with the following data is displayed:

Name	Short Label	Long Label	Email	Description File Path
Test Hub	test	Hub with test data	test@someplace.ac.uk	test_hub.html

At the bottom, there is a 'Submission' section with a 'Submit Track Hub' button. The footer contains the European Union flag and text: 'The FAANG Data Coordination Centre has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement Nos. 815668, 817923 and 817998, and also from the Biotechnology and Biological Sciences Research Council under Grant Agreement No. BB/N019563/1.'

Track hub viewer



- New track hub genome browser embedded in FAANG Data Portal.
- Based on IGV browser, providing powerful visual interaction.
- Highly customisable viewer to present your project track hubs.
- Select your reference of choice, setup multi track views, and incorporate Ensembl annotations.

Track hub viewer



FAANG Track Hubs

GENE-Switch_WP5_ATAC-seq

gs-wp5

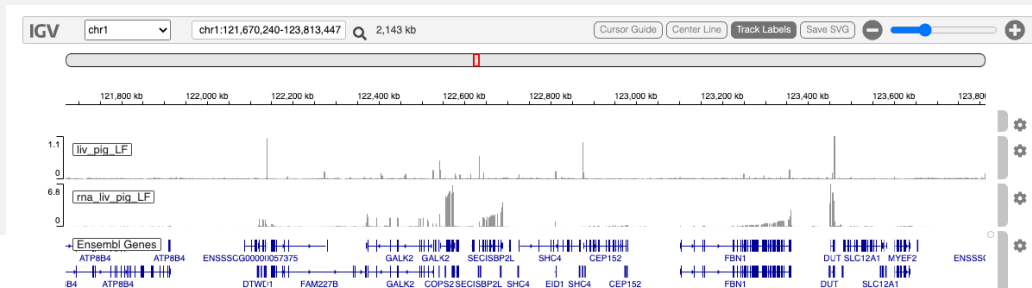
consensus consensus_peak

liv_fet_LF liver_fetus_low.fibre

liv_fet_HF-B liver_fetus_high.fibre.Beetpulp

liv_fet_HF-P liver_fetus_high.fibre.Pea.plus.acetate

- Will soon allow launch from FAANG project pages to track hub views you preselect.
- Solution for track hubs that were not working as desired in current genome browser technology (Ensembl and UCSC).
- May be replaced by Ensembl widget browser in the future.
- In beta, further improvements planned.



Future EuroFAANG infrastructure

- The FAANG DCC and Data Portal is a key component of the future EuroFAANG infrastructure.
- Will provide the common data structure and data access for the future infrastructure.
- Expanding DCC and portal to support new breeding, phenotyping, biobanking, genome editing and genomic technologies.
- Supporting transnational access to infrastructure and expertise through integrated web portal .
- Retain a global focus and connection.



metaFAIR task force



Functional Annotation of ANimal Genomes (FAANG) Project
— A coordinated international action to accelerate Genome to Phenome

Task Force: metaFAIR

— FAANG Metadata, Ontologies and FAIR Data Sharing Task Force

Sign up at
www.faang.org

The Task Force will review and recommend development to the FAANG Data Coordination Centre on the following key areas:

1. The FAANG metadata standards (<https://data.faang.org/ruleset/samples>).
2. The process and guidance documentation for FAANG validation and submissions.
3. The improvement of ontologies through the FAANG Ontology Improvement Tool (<https://data.faang.org/ontology>), including organising ontology focus workshops.
4. Track hub standardisation, submission and presentation for FAANG analysis results.
5. How FAIR is FAANG data and what could be improved?
6. Prepare FAANG for developments in single cell atlases, pangenomes and emerging technologies.
7. Phenotyping recording improvement across FAANG.
8. Management of Information Sharing. Recording of Nagoya and clearer recording of third party limitations on data/phenotyping data use and IP.*
9. Identify training requirements for the FAANG consortia.
10. Contribute to organised scientific dissemination to the whole community.

EMBL Virtual Ensembl training courses



COURSE AT EMBL-EBI

**EuroFAANG GENE-SWitCH workshop:
utilisation of GENE-SWitCH data in
Ensembl and beyond**



31st Jan – 1st Feb

COURSE AT EMBL-EBI

**EuroFAANG AQUA-FAANG workshop:
methods to use and reuse the AQUA-
FAANG data and Ensembl resources
to advance science**



17th Apr – 18th Apr



faang-dcc@ebi.ac.uk

Summary



- Request a FAANG Data Portal specific customised page for your project.
- Assist the community in improving animal agriculture ontology descriptions with the new improvement tools (attend the workshops).
- Utilise the new track hub submission and visualisation tools to increase the usability and impact of your FAANG results/papers.
- Sign up to the metaFAIR task force and/or share your DCC needs with us at: faang-dcc@ebi.ac.uk.



Photo: Peter Harrison

Acknowledgments



Alexey Sokolov:
Agriculture and Biodiversity Coordination
Project Lead



Akshatha Nayak:
Bioinformatician



Koosum Roochun:
Bioinformatician

 peter@ebi.ac.uk
 [@peterwharrison](https://twitter.com/peterwharrison)

<https://data.faang.org/>
<https://www.ensembl.org/>



faang-dcc@ebi.ac.uk



EuroFAANG



AQUA-FAANG



BovReg
Understanding cattle genomes



GENE-SWITCH



*This project has received funding from the European Union's
Horizon Programme under the grant agreement n°s **817923**,
815668, **817998**, and **101094718**.*