The FAANG Data Coordination Centre:

Driving metaFAIR, Global Coordination and EuroFAANG Objectives

Peter Harrison

Genome Analysis Team Leader

EMBL-EBI

peter@ebi.ac.uk

FAANG Functional Annotation of Animal Genomes 10 Years

EMBL 50 YEARS | 1974–2024

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**.

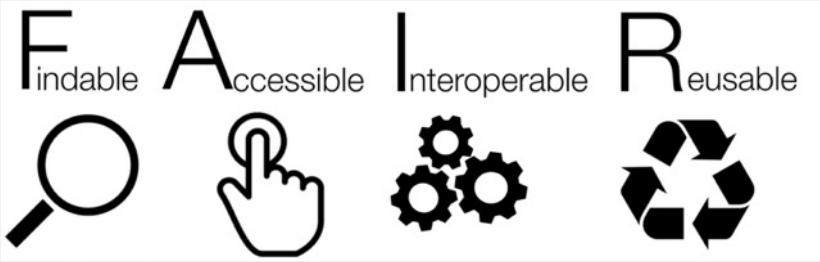
- Ensuring data is richly described, available and searchable.
- Ensuring data is consistently reported and presented.
- Facilitating standardisation of formats, protocols and analysis methods.
- Facilitating data openness, reusability and cross-project analysis.

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

EMBL-EBI

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.
- Links to standardised analysis workflows with evert analysis file.
- A data platform and community drive that ensures data is open and FAIR.
- Innovative data presentation and contextual genome visualisation.



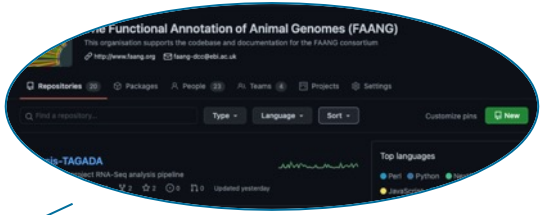
To accelerate genome to phenome research

Photo: CODATA

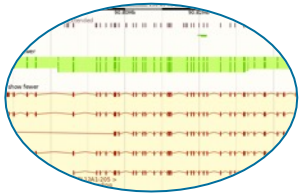
FAANGs standardised datasets

Rich, validated metadata

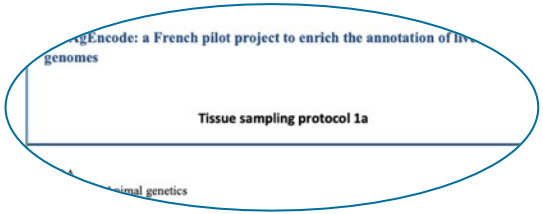
Publications



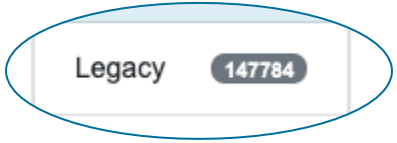
Analysis workflows and parameters



FAANG dataset



Detailed sampling and analysis protocols

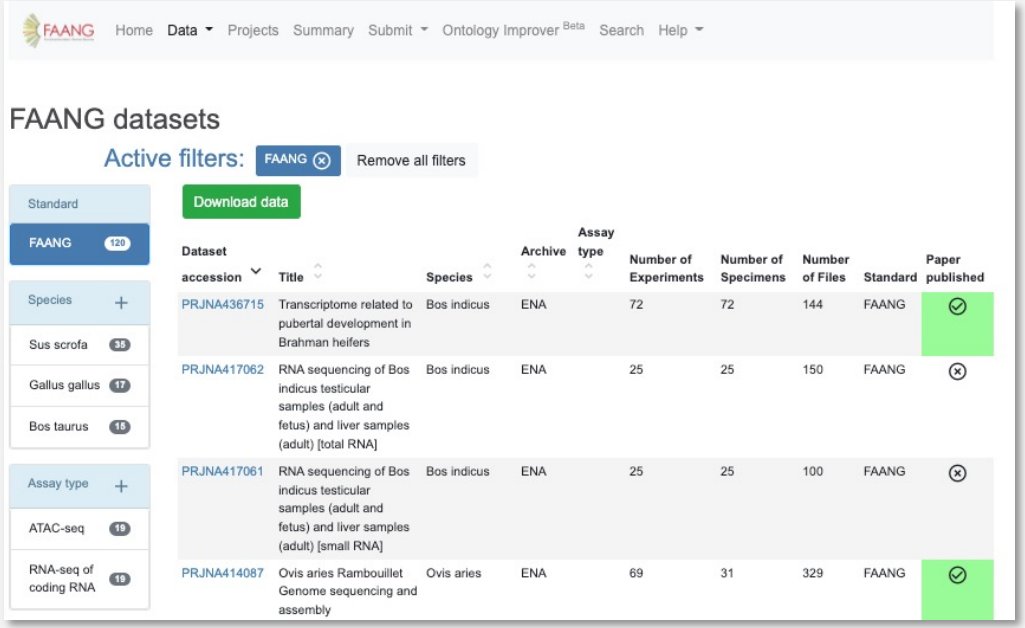


Legacy contextual datasets

Secondary analyses, track hubs and annotations

The FAANG Data Portal

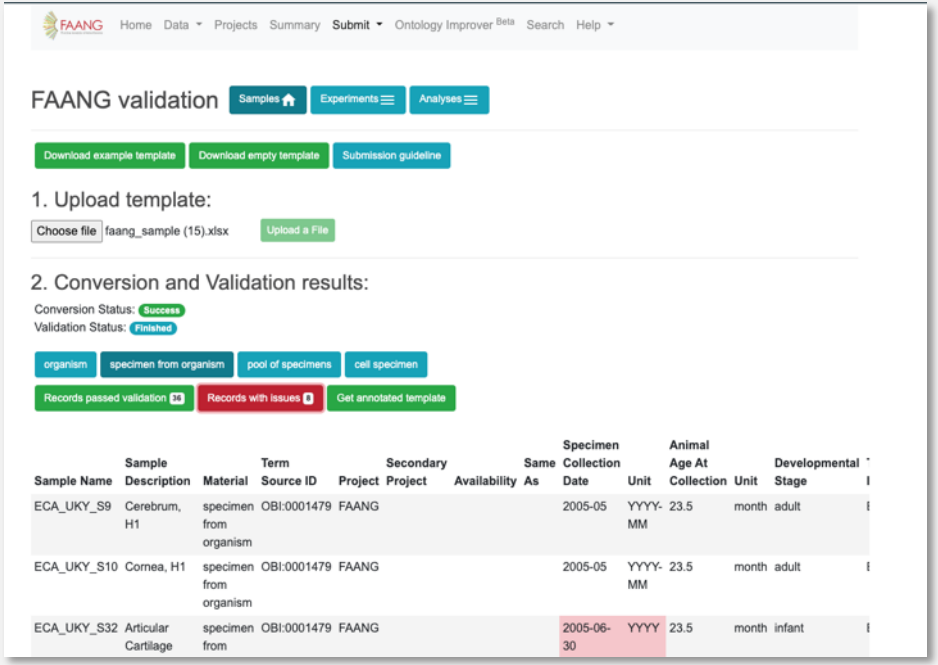
- A single access point to all FAANG metadata, data, pipelines and publications.
- Direct access to download data from underlying public archives.
- Automatically identifies dataset (re)use in publications, and links publications to datasets.
- Intuitive search and filtering.
- Data visualisation and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



https://data.faang.org/

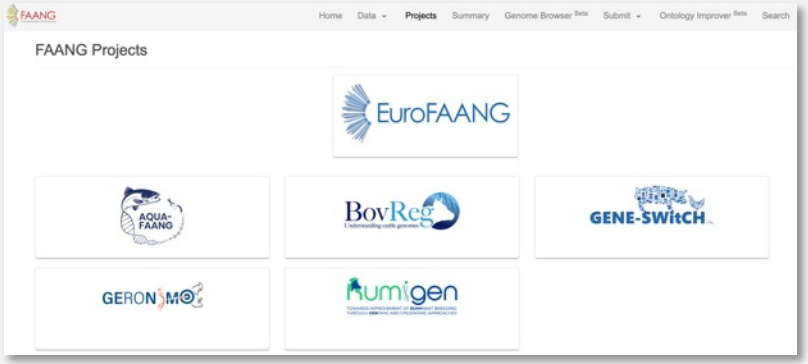
Validation and brokered submission

- All FAANG data has pre-submission validation, ensuring rich metadata.
- Not only highlights errors, suggests improvements such as more specific ontologies.
- Brokered submission to underlying archives simplifies process for users.
- Also submit protocols, track hubs and workflows with configuration files.



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

Project Specific Pages



- 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

metaFAIR task force

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

Sign up at www.faang.org

Task Force: metaFAIR
— FAANG Metadata, Ontologies and FAIR Data Sharing Task Force

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-EBI

Ensembl Genome Browser

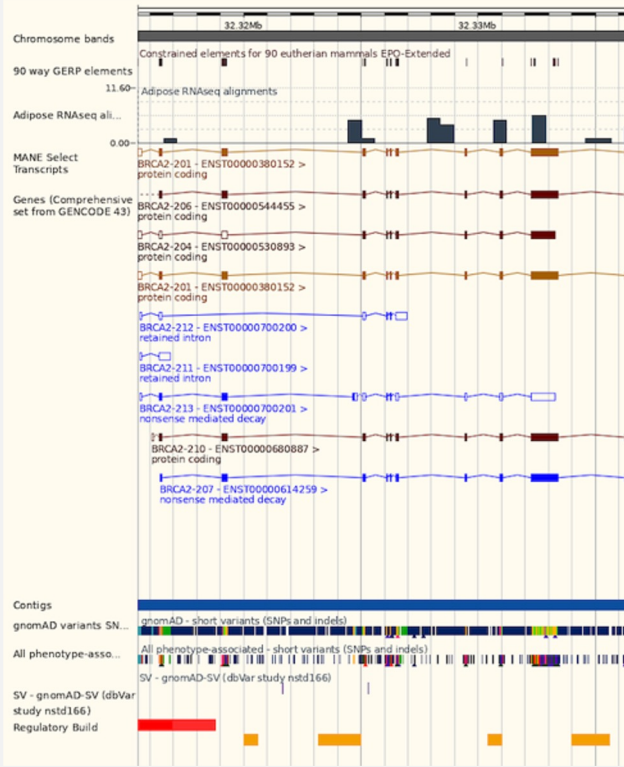
- A genomics platform and browser for enabling and accelerating downstream science.
- Standardized, high quality and richly described FAANG data is driving new annotations in Ensembl.
- Range of tools such as Variant Effect Predictor, EMBL industry promotion and ongoing training.

Comparative analysis

Gene and transcript annotation

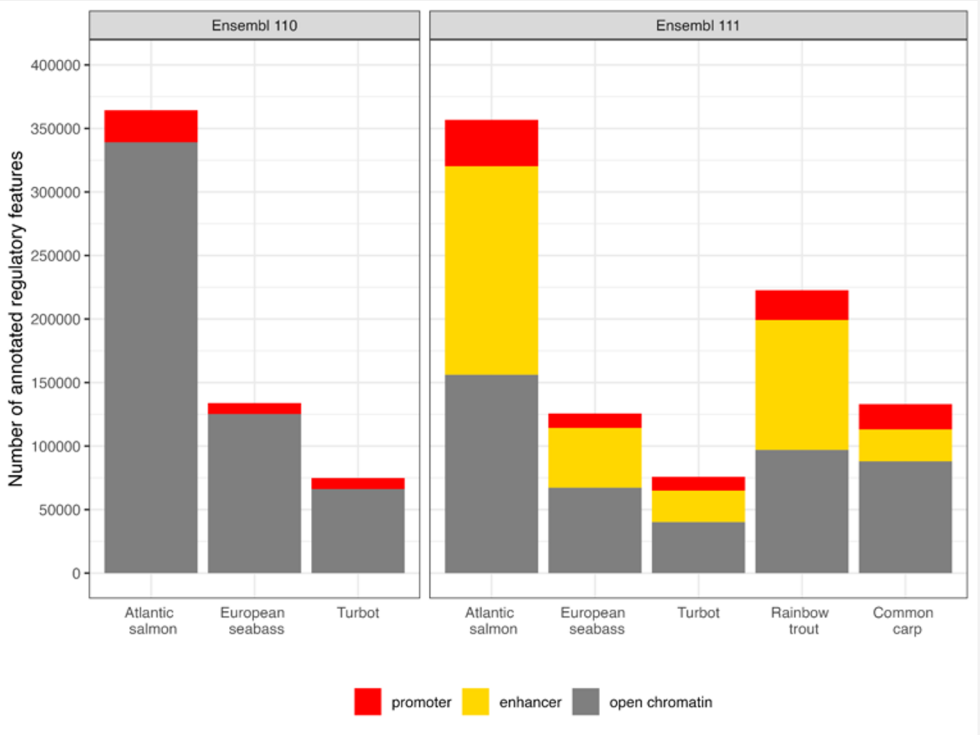
Variation

Regulatory elements



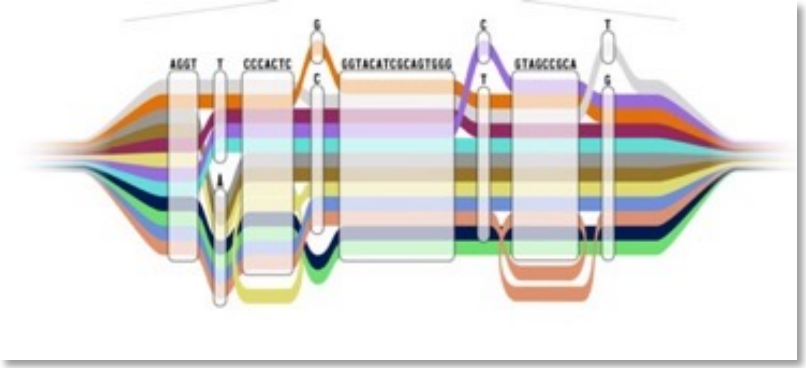
Ensembl First Regulatory builds

- In collaboration with EuroFAANG projects released Ensembl regulatory builds for chicken, pig and 5 fish species.
- Driven new Ensembl annotation methods and visualisation.
- Viewable in the Ensembl browser.



Future directions

- Ensembl will continue to update and improve reference annotations and looking forward to adding T2T references.
- Always looking for further species and breeds/strains of economic and research interest.
- This variation builds towards a pangenomic view for each species, following work being done in human and crops.
- Through Earth BioGenome Projects Ensembl is looking to annotate all eukaryotic species.



Research Infrastructure and DCC

- The FAANG DCC and Data Portal have been funded by EuroFAANG projects and is now a key component of the future EuroFAANG Research Infrastructure.
- Expanding DCC and portal to support new breeding, phenotyping, biobanking, genome editing and genomic technologies.
- Will retain a global focus and connection, guidance by FAANG steering committee and metaFAIR task force.

Horizon2020
European Union Funding
for Research & Innovation

Elixir Domestic Animal Genomes and Phenomes

- Launching new task force this year.
- Being driven by EuroFAANG and emerging Research Infrastructures on Phenotyping.
- Elixir brings together scientists from 23 countries and over 250 research institutes to:
  - Coordinate data and technology discussions
  - Develop standards
  - Standardise analysis and workflows
  - Organise training, workshops and hackathons
  - Connect to global efforts

EMBL-EBI

Summary

- Continuing to strive for high quality, richly described and standardised data.
- Request a FAANG Data Portal specific customised page for your project.
- Utilise the new track hub submission and visualisation tools to increase the usability and impact of your FAANG results/papers.
- Your data is driving improved Ensembl annotations.
- Sign up to the metaFAIR and Elixir Domestic Animal Genomes and Phenomes task forces.

Photo: Peter Harrison

Acknowledgments

Alexey Sokolov:
Agriculture and Biodiversity Coordination
Project Lead

Koosum Roochun:
Bioinformatician

faang-dcc@ebi.ac.uk

EuroFAANG
RESEARCH INFRASTRUCTURE

AQUA-FAANG

BovReg
Understanding cattle genomes

GENE-SWitCH

peter@ebi.ac.uk
@peterwharrison

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

Ensembl

wellcome

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