FAANG: Submission, retrieval and management of Next Generation Sequencing data

Peter Harrison

Data Coordination and Presentation Coordinator

EMBI-EBI

peter@ebi.ac.uk

EMBL

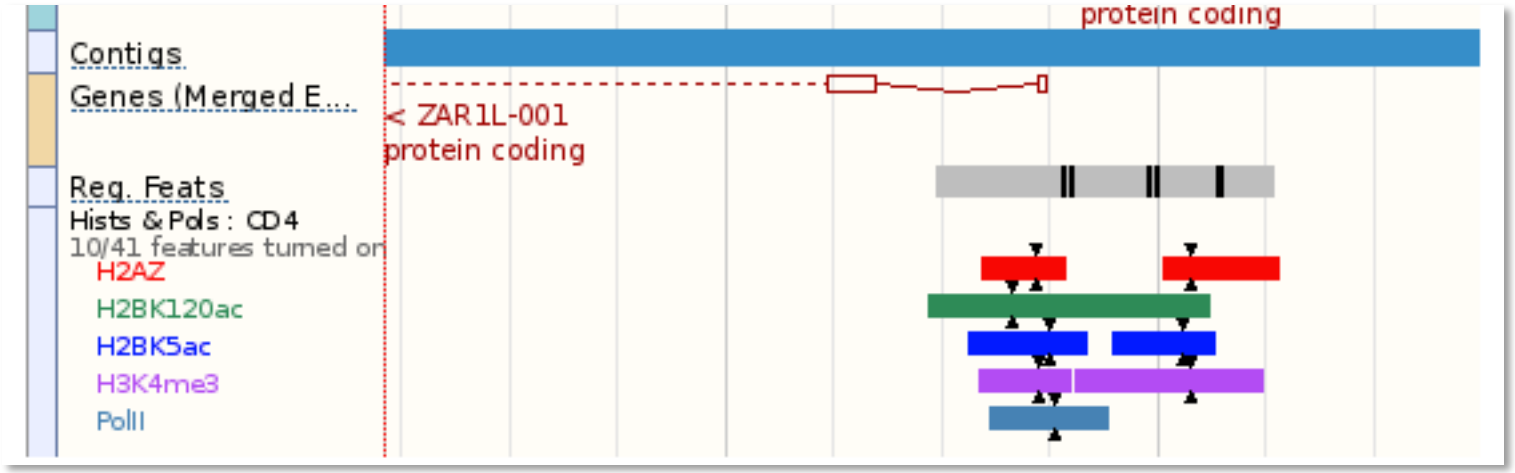
EMBL-European Bioinformatics Institute

Our mission is to help scientists to realise the potential of **‘big data’** by making the worlds biological data freely available through archives, bioinformatics services, training and tools.

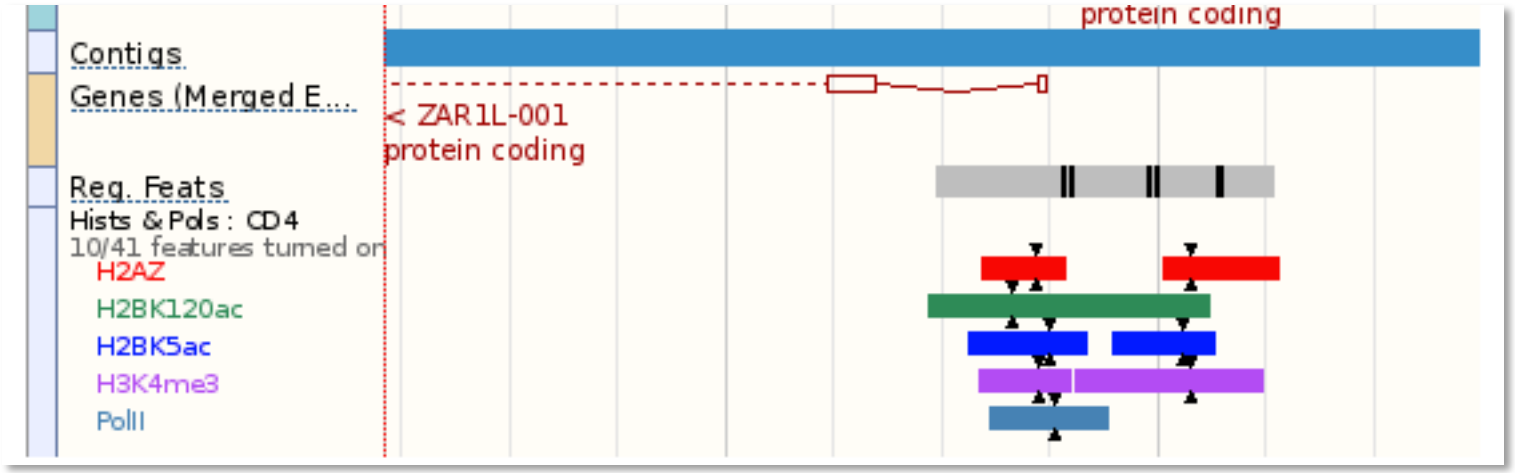
FAANG
Functional Annotation of Animal Genomes

EMBL-EBI

BBSRC
bioscience for the future

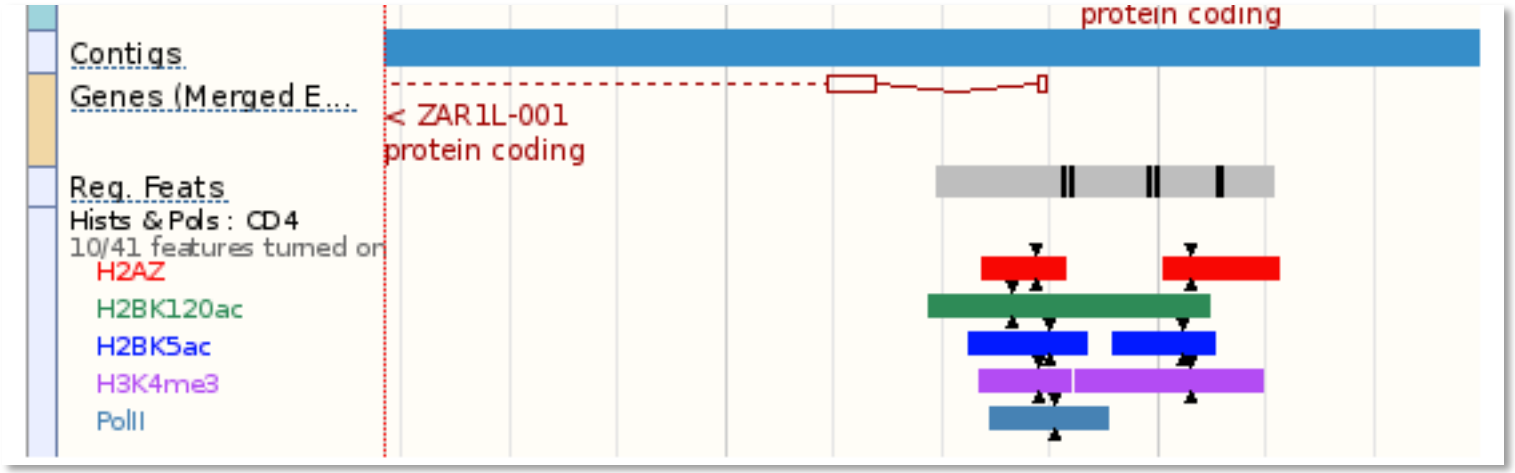


What is Chip-Seq data useful for?



What is Chip-Seq data useful for?

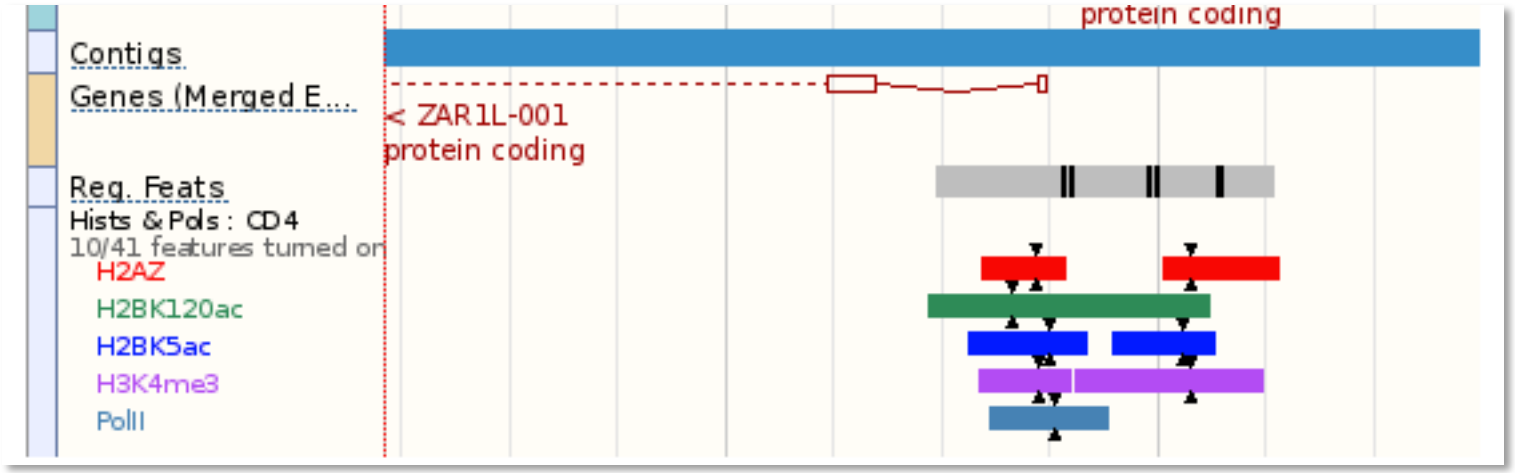
What species is this from?



What is Chip-Seq data useful for?

What species is this from?

What sex, was it healthy?

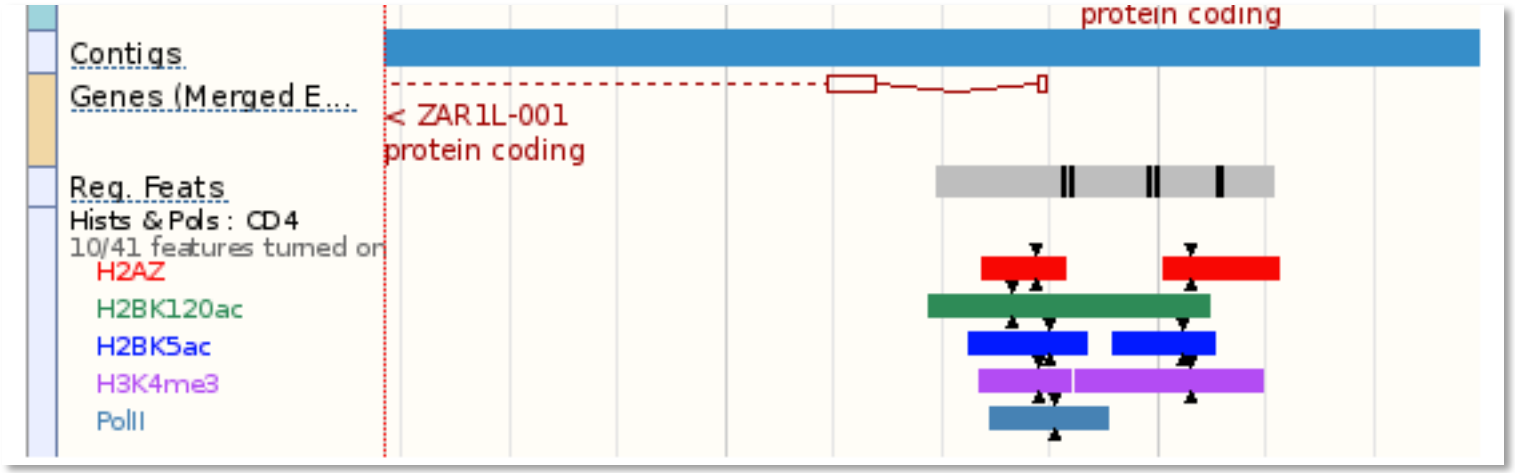


What is Chip-Seq data useful for?

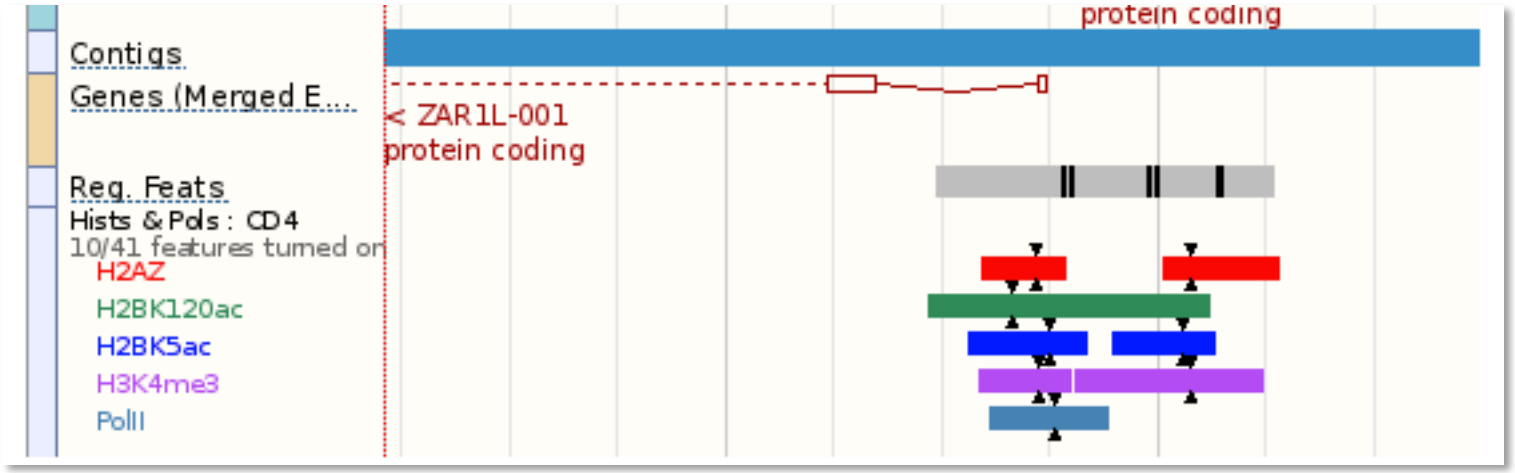
What species is this from?

What sex, was it healthy?

If you don't know is the data still useful?



The reality is that without context, data is largely meaningless.



The reality is that without context, data is largely meaningless.

For the submission, retrieval and management of data, how we document and describe the data is key.

Thinking about your analyses you have run this week, what do you need to know about the data to make it useful?

Thinking about your analyses you have run this week, what do you need to know about the data to make it useful?

Sample information

- Species
- Breed
- Location
- Age
- Tissue
- Sample collection protocol
- .....

Data information

- Storage and time stored
- Date and location
- Antibody provider
- Antibody catalog number
- Antibody lot
- Fragment size range
- .....

Analysis information

- Pipeline used
- Pipeline version
- Reference genome
- Assay type
- Protocol
- Platform
- .....

Thinking about your analyses you have run this week, what do you need to know about the data to make it useful?

Sample information

- Species
- Breed
- Location
- Age
- Tissue
- Sample collection protocol
- .....

Data information

- Storage and time stored
- Date and location
- Antibody provider
- Antibody catalog number
- Antibody lot
- Fragment size range
- .....

Analysis information

- Pipeline used
- Pipeline version
- Reference genome
- Assay type
- Protocol
- Platform
- .....

Good metadata is even more crucial for phenotypic studies.

A full metadata solution for FAANG

- The FAANG Metadata and Data Sharing Committee set out to identify all of the different sampling and experiment questions we would need for any study for FAANG:

Required 158 different metadata questions

- 67 sample attributes (Organisms, Specimens, pools of specimens, cell cultures, cell lines).
- 66 experiment attributes (Chip-Seq, RNA-Seq, ATAC-Seq, Hi-C, WGS, Dnase-Seq, BS-Seq).
- 25 analysis attributes (Analysis type, Pipeline, reference genome)

FAANGs Metadata ruleset goals

To accurately describe any livestock dataset to:

- be fully understood, repeatable and reusable.
- use consistent terminology.
- be well structured, rich and specific.

ChIP-seq DNA-binding proteins (7 rules)

Applied under these conditions:

assay type is "ChIP-seq"

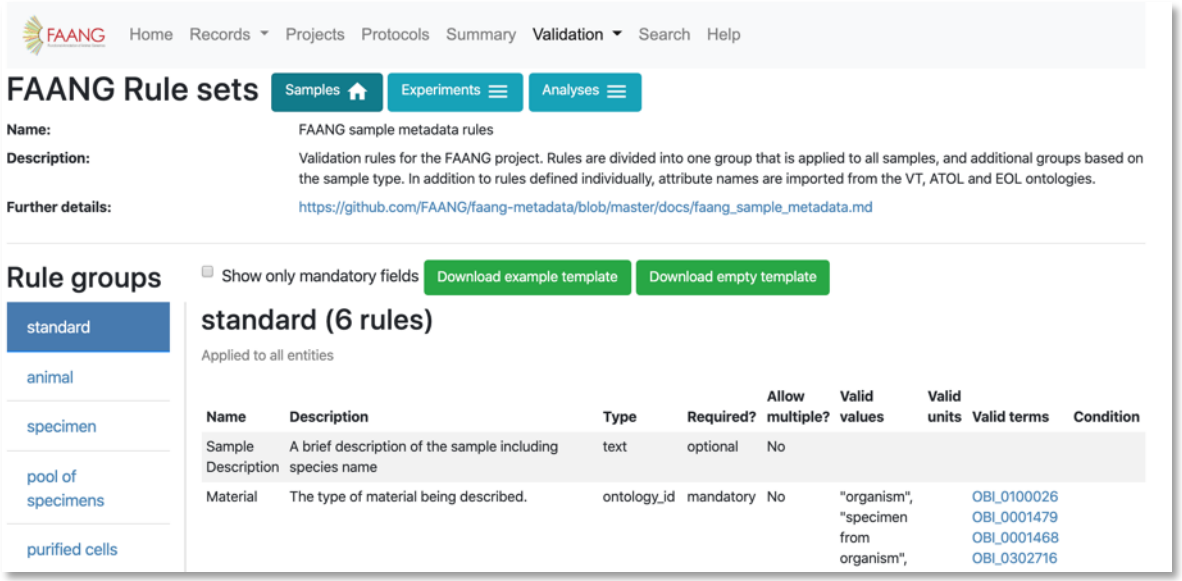
experiment target is lead node descendant of SO_0001700

Name	Description	Type	Required?	Allow multiple?	Valid values	Valid units	Valid terms	Condition
chip target	The target of the ChIP-seq experiment e.g. H3K4Me3, H3K4Me1, H3K27Me3, H3K27Ac, CTCF. If your target is not in the list, please contact the faang-dcc helpdesk.	ontology_text	mandatory	No			CHEBI_15358 OMIT_0038500 NCIT_C17804 NCIT_C34071	
control experiment	Experiment alias (in this submission) or ENA experiment accession (if submitted previously) of the ChIP-seq input DNA experiment which acts as the	text	recommended	No				

https://data.faang.org/ruleset/experiments

Biology evolves and so must the rules to describe it.

- FAANG rulesets now on version 3.8 (26 releases).
  - Describing new technology types and improvements to our understanding of what's important for livestock data.
- Rulesets are version controlled through Github, but are rendered to human readable tables in the FAANG data portal.



https://data.faang.org/ruleset/experiments

EMBL

Why do you need good metadata?

- Make your data usable
  - Reduce ambiguity
  - Facilitate reproduction of results
  - Improve integration across labs and projects
- Make your data discoverable
  - Other researchers
  - Informatics services (Ensembl, Gene Expression Atlas)
- Improve your own analysis
  - Easier to find batch effects and confounding factors

Photo: Peter Harrison

Rules are only useful if they are followed

We therefore need to validate and curate all FAANG datasets against the FAANG rulesets.

What is wrong with this data, how can you improve it? How many errors can you see?

animal ID	species	sex	date of birth	Provider	site	collection date	specimen type	assay type
IS-RJ-01	Goat	male	13-Mar-14	Ainsworth Husbandry	CB1 2PD UK	27-Oct-14	liver	RNA-seq
AU_RL200	Capra hircus	F	2011	CSRIO	144°15'0"E 32°53'0"S	Dec-11	bone	ChIP-Seq
UG-F50647	Bos taurus	male	unknown	University of Uganda	-27.516 152.566	02/01/2012	thalamus	RNA-seq
UG-F50545	Bos taurus	female	unknown	University of Uganda	KAGOOGE UGANDA	19/12/2011	brain	RNA-Seq

What is wrong with this data, how can you improve it? How many errors can you see?

animal ID	species	sex	date of birth	Provider	site	collection date	specimen type	assay type
IS-RJ-01	Goat	male	13-Mar-14	Ainsworth Husbandry	CB1 2PD UK	27-Oct-14	liver	RNA-seq
AU_RL200	Capra hircus	F	2011	CSRIO	144°15'0"E 32°53'0"S	Dec-11	bone	ChIP-Seq
UG-F50647	Bos taurus	male	unknown	University of Uganda	-27.516 152.566	02/01/2012	thalamus	RNA-seq
UG-F50545	Bos taurus	female	unknown	University of Uganda	KAGOOGE UGANDA	19/12/2011	brain	RNA-Seq

- Potentially 9 types of error that need to be checked or fixed.

Small variations are big metadata issues

- How many ways to say...
  - Dermal fibroblas
  - Dermal Fibroblast
  - Dermal fibroblast
  - dermal fibroblast
  - Dermal Fibroblasts
  - Dermal fibroblasts
  - Dermal Skin Fibroblast
  - Dermal skin fibroblasts
  - Fibroblasts
  - Human Dermal Fibroblast Cells
  - human dermal fibroblasts
  - Human dermal fibroblasts
  - NHDF-Ad-Der Fibroblasts

Small variations are big metadata issues

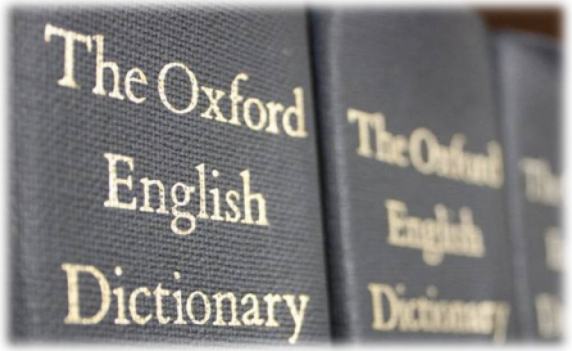
- How many ways to say…
  - Dermal fibroblas
  - Dermal Fibroblast
  - Dermal fibroblast
  - dermal fibroblast
  - Dermal Fibroblasts
  - Dermal fibroblasts
  - Dermal Skin Fibroblast
  - Dermal skin fibroblasts
  - Fibroblasts
  - Human Dermal Fibroblast Cells
  - human dermal fibroblasts
  - Human dermal fibroblasts
  - NHDF-Ad-Der Fibroblasts

The human brain is an amazing pattern-recognition machine, but a computer sees 13 different items.

This was real data from a project before we applied standards and validation.

How to consistently describe information around the world

- In language, we have agreed on how to describe things and how to specifically spell those words, we record this in dictionaries.
- In biology we can do the same thing, using ontologies.
- An ontology:
  - Unambiguously describes a biological entity
  - Has a short description of the entity.



There are many ways to describe the same entity.

- Issue in language and biology, solution is synonyms.
- In ontologies, all synonyms map to the same unique code.

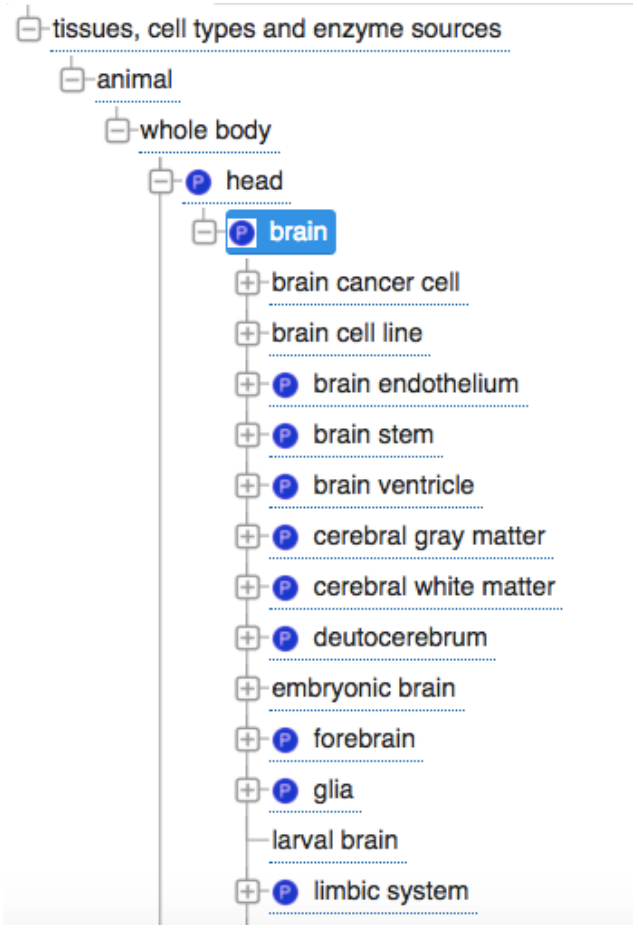
endocrine system disease

Synonyms: Endocrinopathy, ENDOCRINE DISORDER NOS, Diseases, Endocrine System, Hormone disturbance NOS, Endocrine gland disease NOS, Endocrine System Diseases, ENDOCRINE DIS, Disease, Endocrine System, Hormone disorders, Diseases of Endocrine System, Disease of endocrine gland, Endocrine disturbance, Disorder of endocrine system (disorder), Endocrine Diseases and Manifestations, System Disease, Endocrine, Endocrine disorder NOS (disorder), Disease, Endocrine, Disorder of endocrine gland, Hormone disturbance, Endocrine disorder, Endocrine disturbance NOS, Endocrine disease, Hormone abnormality, Disorder of endocrine system, Endocrine gland disease NOS (disorder), Hormone abnormality (finding), Endocrine System Disorder, ENDOCRINE DISORDERS, Endocrinopathy, NOS, DIS ENDOCRINE SYSTEM, Endocrine Diseases, Diseases, Endocrine, System Diseases, Endocrine, Endocrine disturbance NOS (disorder), ENDOCRINE SYSTEM DIS, Unspecified endocrine disorder

EFO_0001379

Ontologies have hierarchy

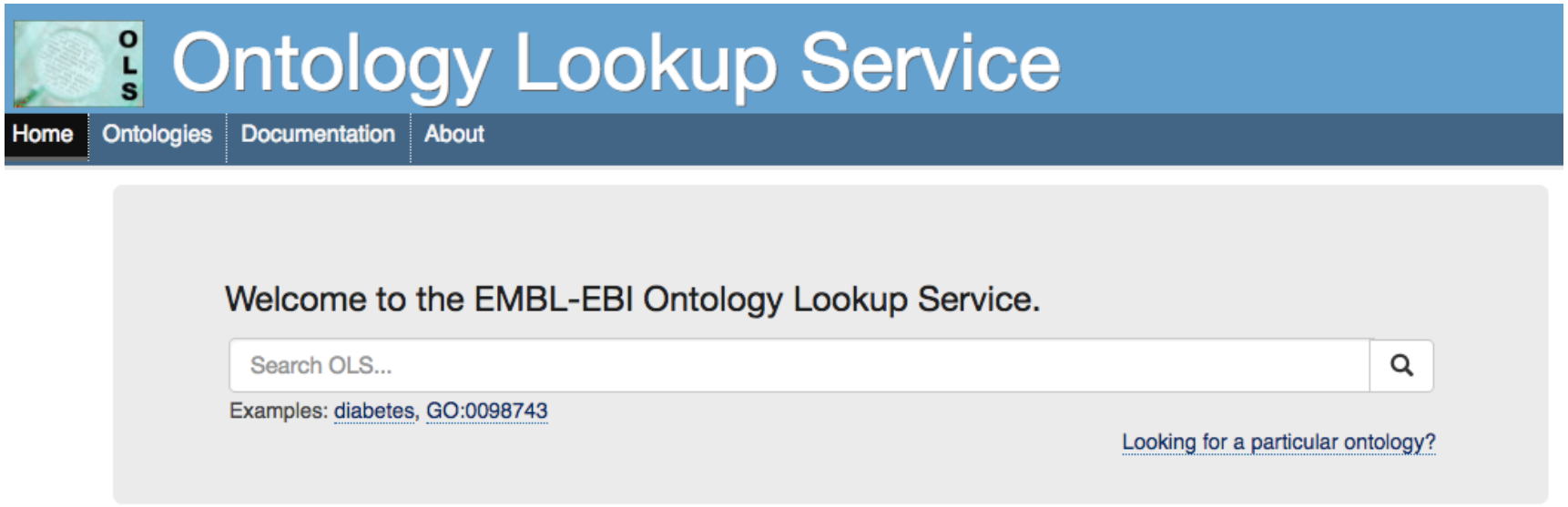
- Another key feature is that ontologies have a hierarchical structure.
- Know the position in the world view.
- Allows searches for brain and then return all results that are below brain in the hierarchy.
- When writing rulesets we can also restrict valid metadata values to particular parts of the ontology.

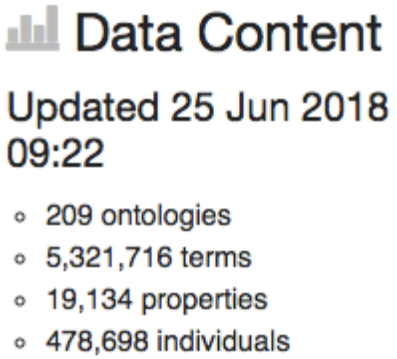


There are specific ontologies for every usage

- Species -> NCBI Taxonomy ontology
- Breed -> Livestock Breed ontology
- Organism part -> UBERON Anatomical ontology
- Sex -> Phenotype and Trait ontology
- Cell type -> Cell Line ontology
- Organisational role -> European Factor ontology
- ……..

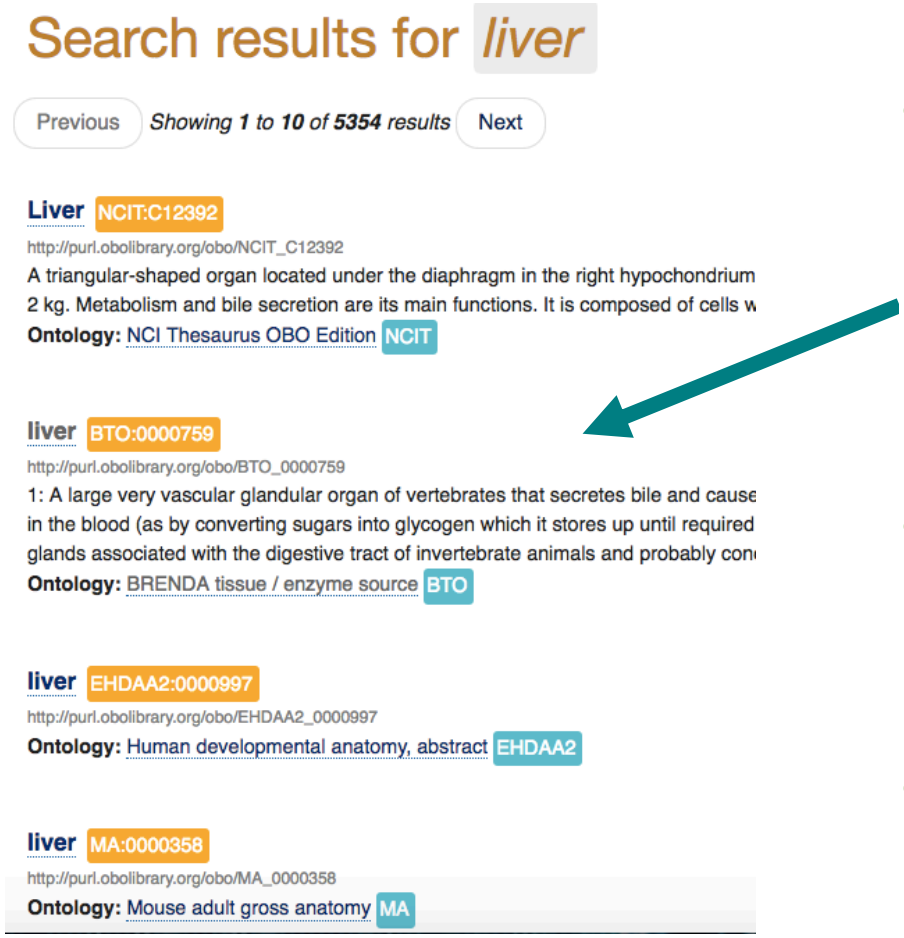
How do I know which ontology to use?





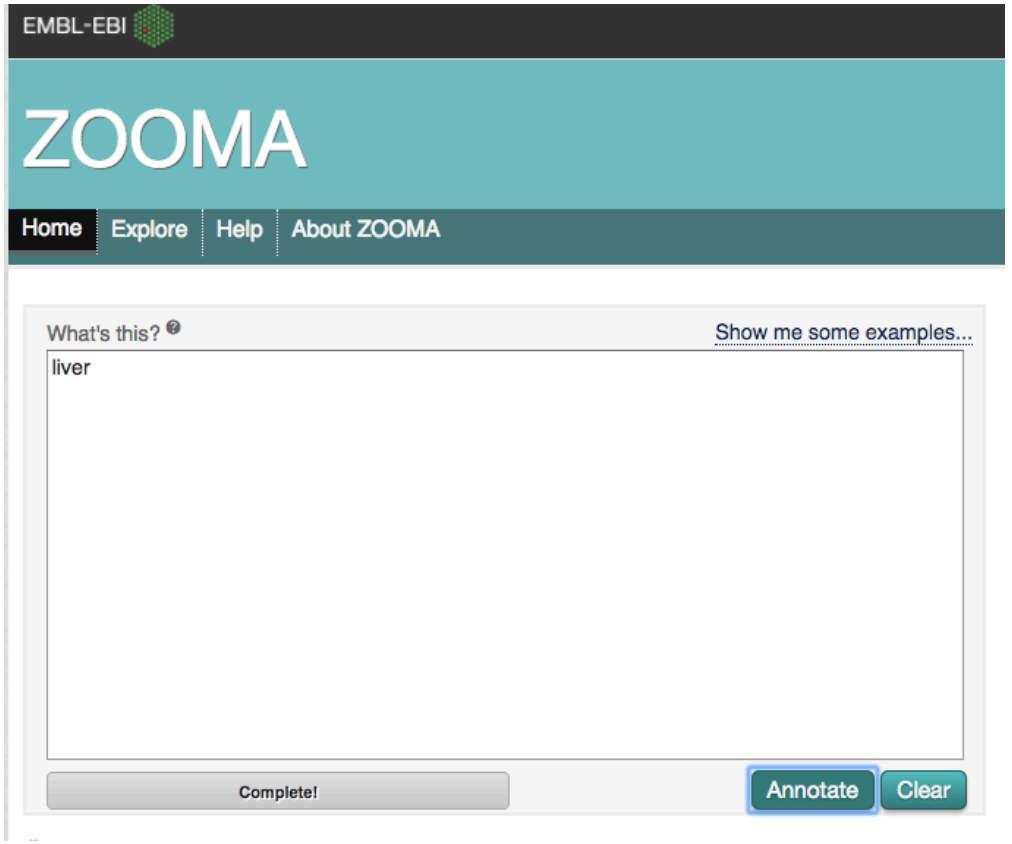
- Universal search over all ontologies.
- Shows the hierarchy, allows selection of a more specific term?
- Finds the ontology code regardless of synonym used.

Ontologies often overlap, FAANG has a set of preferred ontologies listed in its ruleset



- FAANG prefers this ontology for organism parts as its descriptions are more livestock friendly.
- Ontologies are typically human, mouse and disease focussed.
- Lots of movement to improve this.

Ontology assignment can be automated



- EMBL-EBI Zooma tool automates conversion from user provided free text to a controlled ontology term.
- It learns, can be customised for specific projects and provides a confidence level to allow automated assignment.

Results

The table below shows a report describing how ZOOMA annotates text terms supplied above.

Hide results that did not map?

Download my results

Term Type	Term Value	Ontology Class Label	Mapping Confidence	Ontology Class ID	Source
[NO TYPE]	liver	liver	High	UBERON_0002107	Expression Atlas

The European Gene bank problem

- The Innovative Management of Animal Genetic Resources (IMAGE) project is creating a central data portal for national genebanks across Europe.
- Problem: each Gene bank records data in its national language.
- Solution: we are teaching Zooma French, German, Italian, Hungarian.....
- It stores these as project specific synonyms and can then automatically assign the ontology code regardless of national language the next time it sees the same term.
- The unified database can then display English ontologies.

Not all metadata is created equal

- Mandatory: Fields are always required, validation will fail if not provided.

Not all metadata is created equal

- Mandatory: Fields are always required, validation will fail if not provided
- Recommended: Should always be provided, validation will fail unless a specific ‘missing’ term is supplied and warn even if it is:
  - 'not applicable'
  - 'not collected' (i.e. will always be missing)
  - 'not provided' (i.e. may be added later)
  - 'restricted access' (i.e. it isn't missing, we just can't include it in a public document)

Not all metadata is created equal

- Mandatory: Fields are always required, validation will fail if not provided
- Recommended: Should always be provided, validation will fail unless a specific ‘missing’ term is supplied and warn even if it is:
  - 'not applicable'
  - 'not collected' (i.e. will always be missing)
  - 'not provided' (i.e. may be added later)
  - 'restricted access' (i.e. it isn't missing, we just can't include it in a public document)
- Optional: Useful but not required to pass validation.

Summary: What makes a good metadata system

- Controlled vocabulary with consistent ontologies.
- Clear minimal data requirements.
- Encouraging engagement and completion of data.
- Ease of entering data and clear instructions.
- Developing batch upload to avoid form fatigue and inconsistency.
- Validation against known information, preventing errors.

"I liked it better before big data and metadata when we just had good old regular data."

Focus thus far on how to describe your data as this is a key aspect of data submission and management.

You will see later how this is also what powers data retrieval and effective downstream analysis.

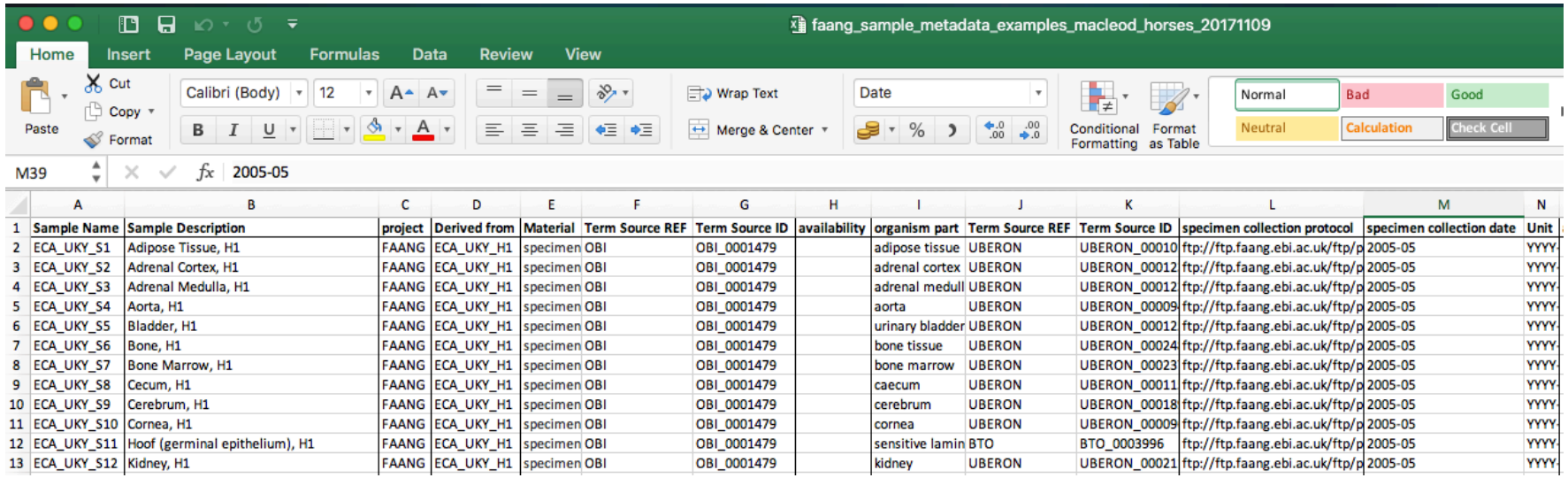
So for FAANG how and where do you submit your FAANG Next Generation Sequencing data.

Samples first

- Before you can submit your data you must have submitted your samples to the public archives and FAANG.
- You need the BioSample ID from each sample for your data submissions, they look like this ‘SAMEA3492638’

- Not going to cover sample submission today, we are going to focus on data and analysis submissions.
- Detailed instructions for sample submission is on the FAANG data portal.
- You can also email or find me on the EMBL-EBI stand from Sunday if you want to know more about sample submissions.

How best to collect data from the community.



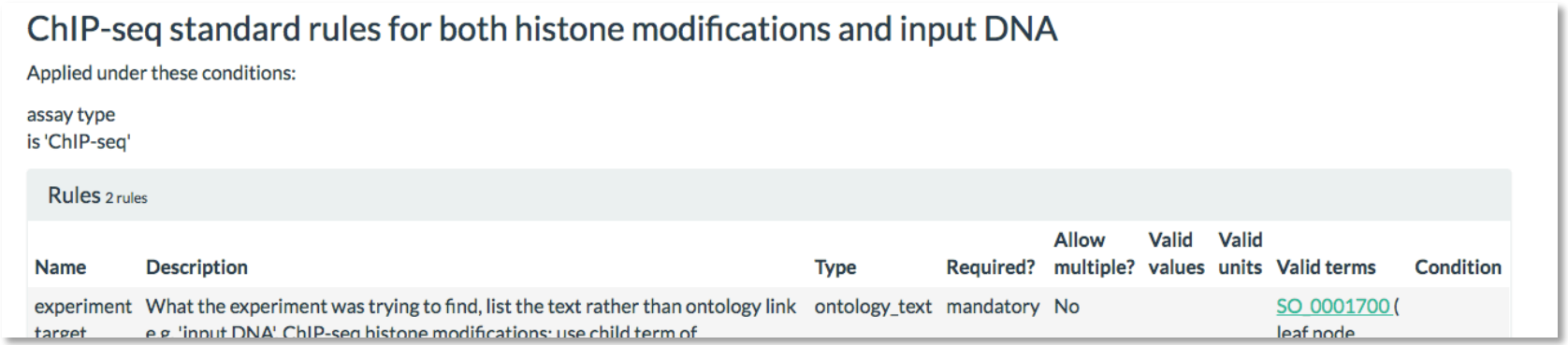
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Sample Name	Sample Description	project	Derived from	Material	Term Source REF	Term Source ID	availability	organism part	Term Source REF	Term Source ID	specimen collection protocol	specimen collection date	Unit
2	ECA_UKY_S1	Adipose Tissue, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		adipose tissue	UBERON	UBERON_00010	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
3	ECA_UKY_S2	Adrenal Cortex, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		adrenal cortex	UBERON	UBERON_00012	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
4	ECA_UKY_S3	Adrenal Medulla, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		adrenal medull	UBERON	UBERON_00012	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
5	ECA_UKY_S4	Aorta, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		aorta	UBERON	UBERON_00009	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
6	ECA_UKY_S5	Bladder, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		urinary bladder	UBERON	UBERON_00012	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
7	ECA_UKY_S6	Bone, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		bone tissue	UBERON	UBERON_00024	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
8	ECA_UKY_S7	Bone Marrow, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		bone marrow	UBERON	UBERON_00023	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
9	ECA_UKY_S8	Cecum, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		caecum	UBERON	UBERON_00011	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
10	ECA_UKY_S9	Cerebrum, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		cerebrum	UBERON	UBERON_00018	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
11	ECA_UKY_S10	Cornea, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		cornea	UBERON	UBERON_00009	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
12	ECA_UKY_S11	Hoof (germinal epithelium), H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		sensitive lamin	BTO	BTO_0003996	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY
13	ECA_UKY_S12	Kidney, H1	FAANG	ECA_UKY_H1	specimen	OBI	OBI_0001479		kidney	UBERON	UBERON_00021	ftp://ftp.faang.ebi.ac.uk/ftp/p	2005-05	YYYY

- Spreadsheets accessible by all and easily passed between lab members to complete.
- Can be automated through CSV conversion.
- Clear record and organisation of data.
- Cloud wrapper scripts to support automated processing coming soon.

For preparing your experiment and analysis submissions you will need

- The FAANG data standards
  - https://data.faang.org/ruleset/experiments
  - https://data.faang.org/ruleset/analyses

- The FAANG submission guide
  - https://dcc-documentation.readthedocs.io/en/latest/faq/



ChIP-seq standard rules for both histone modifications and input DNA

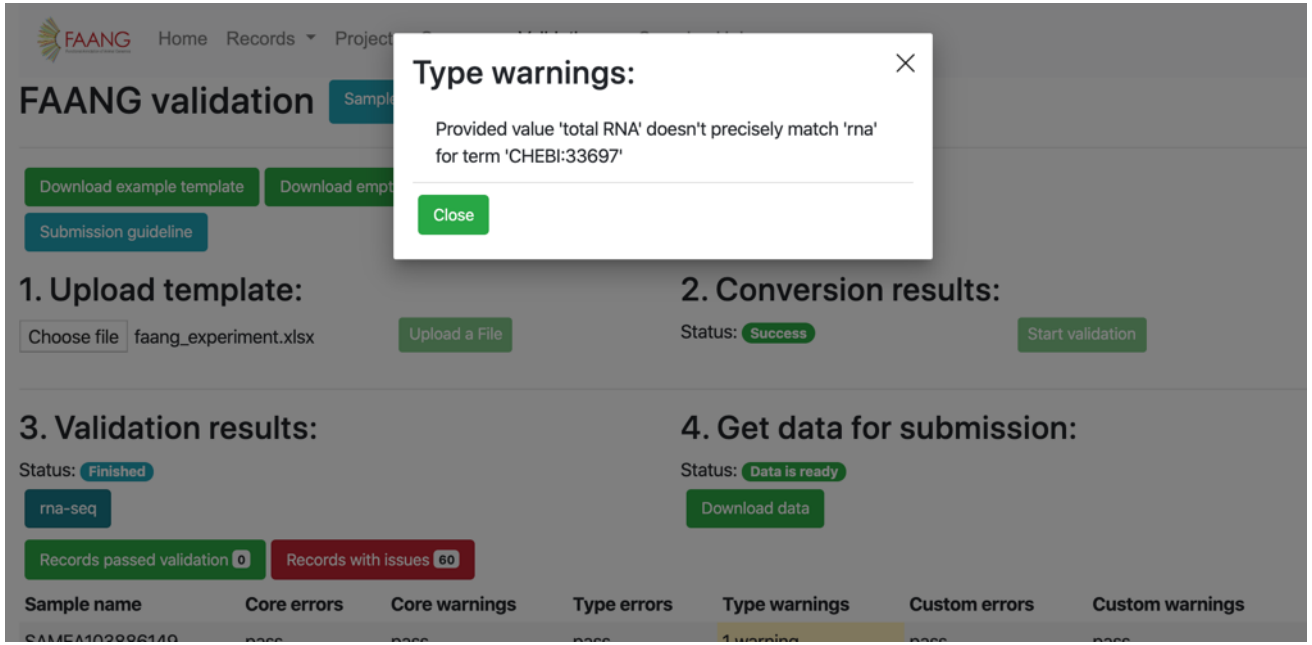
Applied under these conditions:

assay type
is 'ChIP-seq'

Rules 2 rules

Name	Description	Type	Required?	Allow multiple?	Valid values	Valid units	Valid terms	Condition
experiment target	What the experiment was trying to find, list the text rather than ontology link e.g. 'input DNA' ChIP-seq histone modifications: use child term of	ontology_text	mandatory	No			SO_0001700 (leaf node	

FAANG validation tools



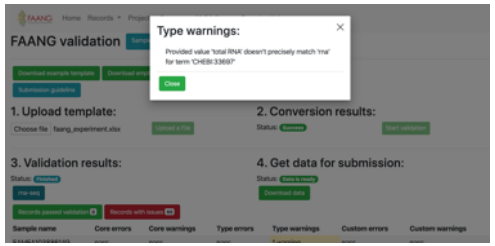
- Checks ontologies (scope, accuracy, terms).
- Relationships (familial, breeds).
- Minimum standards and validity.

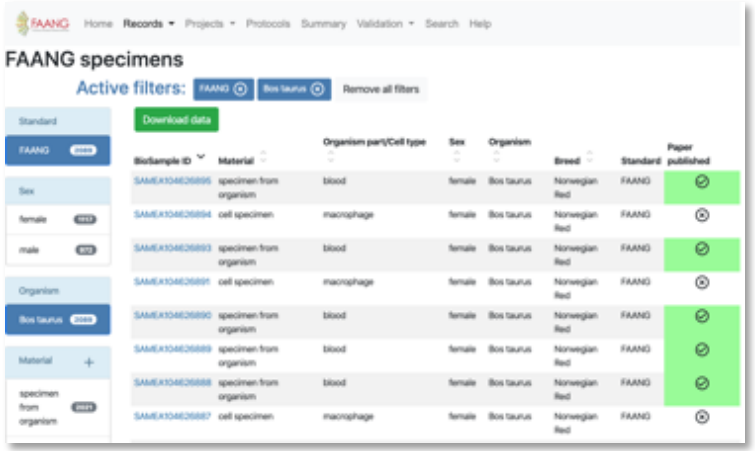
Validating you NGS data through the FAANG validation service

- Time for a demo.

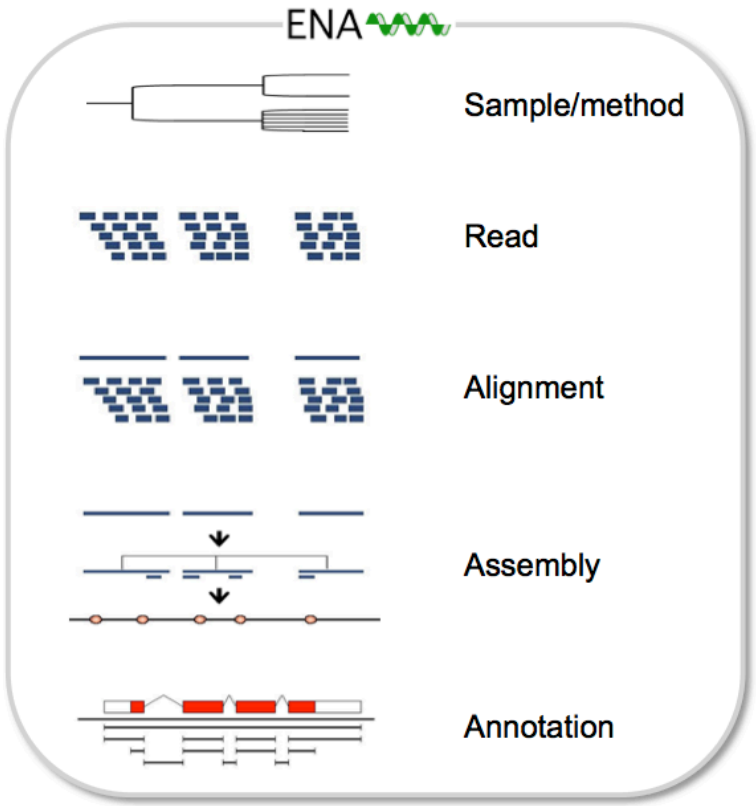
Submitting your data to the public archives

- You don’t submit your data direct to FAANG data portal.
- Instead we submit it to the public archives and then the FAANG data portal automatically indexes the data.
- Main archives for NGS data from EMBL-EBI for FAANG are the European Nucleotide Archive and European Variation Archive.



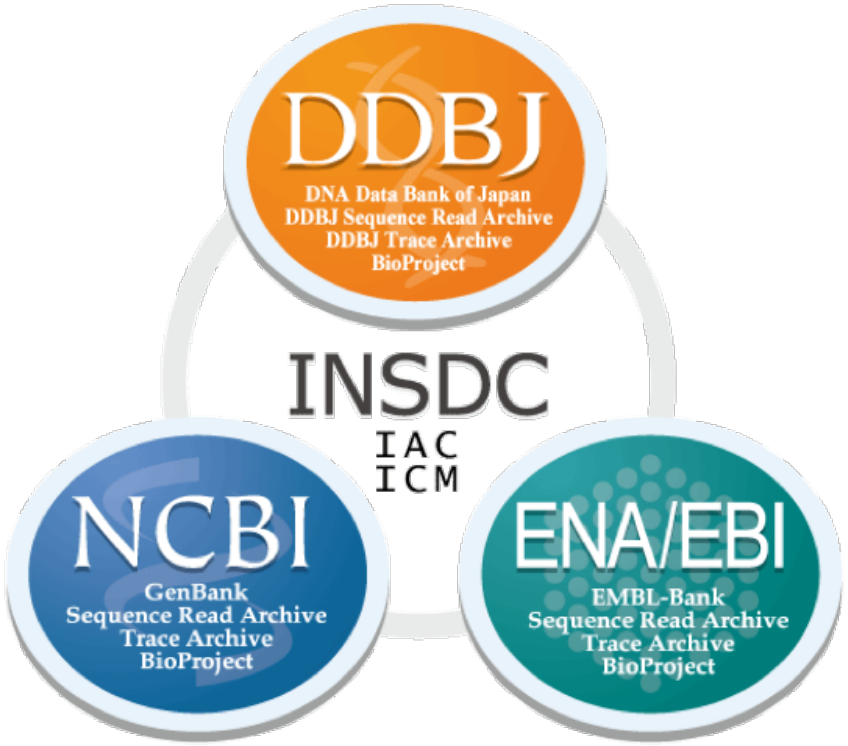


ENA: Comprehensive nucleotide resource



- Comprehensive database for archiving worlds nucleotide data from an array of sequencing technologies.
- Established in the early 1980s, continually extended for new technologies and applications.
- Globally comprehensive scientific record, recognised as a fundamental critical core data resource for the life sciences.
- Archive a new dataset every 6 minutes

International Nucleotide Sequence Database Collaboration



- Regular data exchange across the world.
- Recognise data accessioned by any partner and agree on data standards
- Open and unrestricted access.
- Hugely resilient.
- Globally comprehensive coverage, neutrality.
- Long term collaboration, each site independently funded and are experts in supporting their userbase and local context.

Submitting to the European nucleotide archive

Validated XML

NGS files

MD5s of NGS files

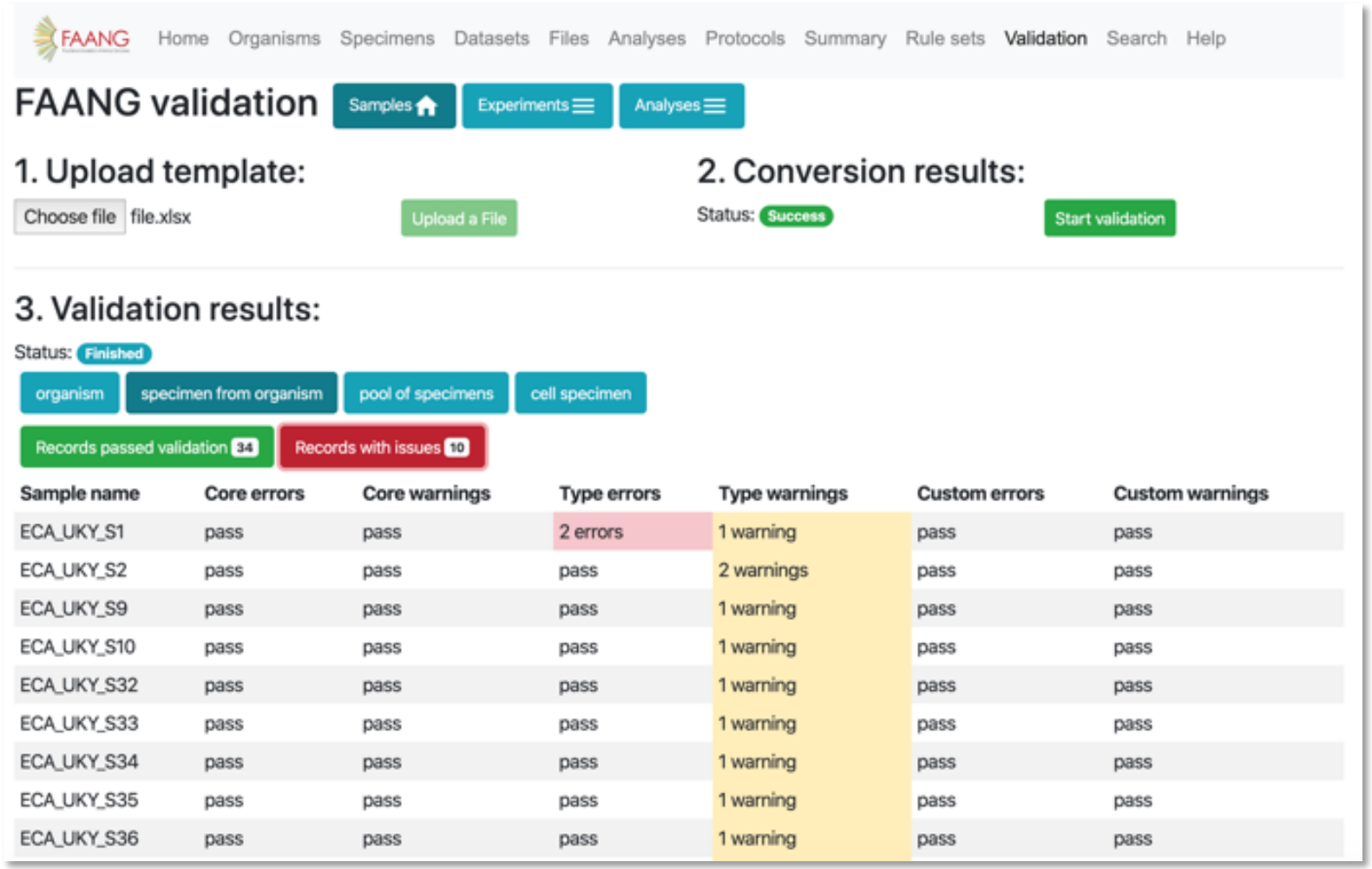
ENA
European Nucleotide Archive

Docs » Sequencing data submission » Submission to ENA

Submit the XML files to ENA

1. Register for an ENA submission account if you do not already have one https://www.ebi.ac.uk/ena/submit/sra/#registration
   - More detailed instructions on account registration can be found here http://ena-docs.readthedocs.io/en/latest/reg_01.html
2. Upload your sequencing data to your account area
   - Follow one of the submission methods detailed on the following page http://ena-docs.readthedocs.io/en/latest/upload_01.html
     - The main options are to use the webin Java client, FTP transfer or the Aspera client. These will require you to have registered for an ENA submission account first.
     - Please read this page and make sure to use one of the supported file formats http://ena-docs.readthedocs.io/en/latest/format_01.html

Currently developing automated brokering submission

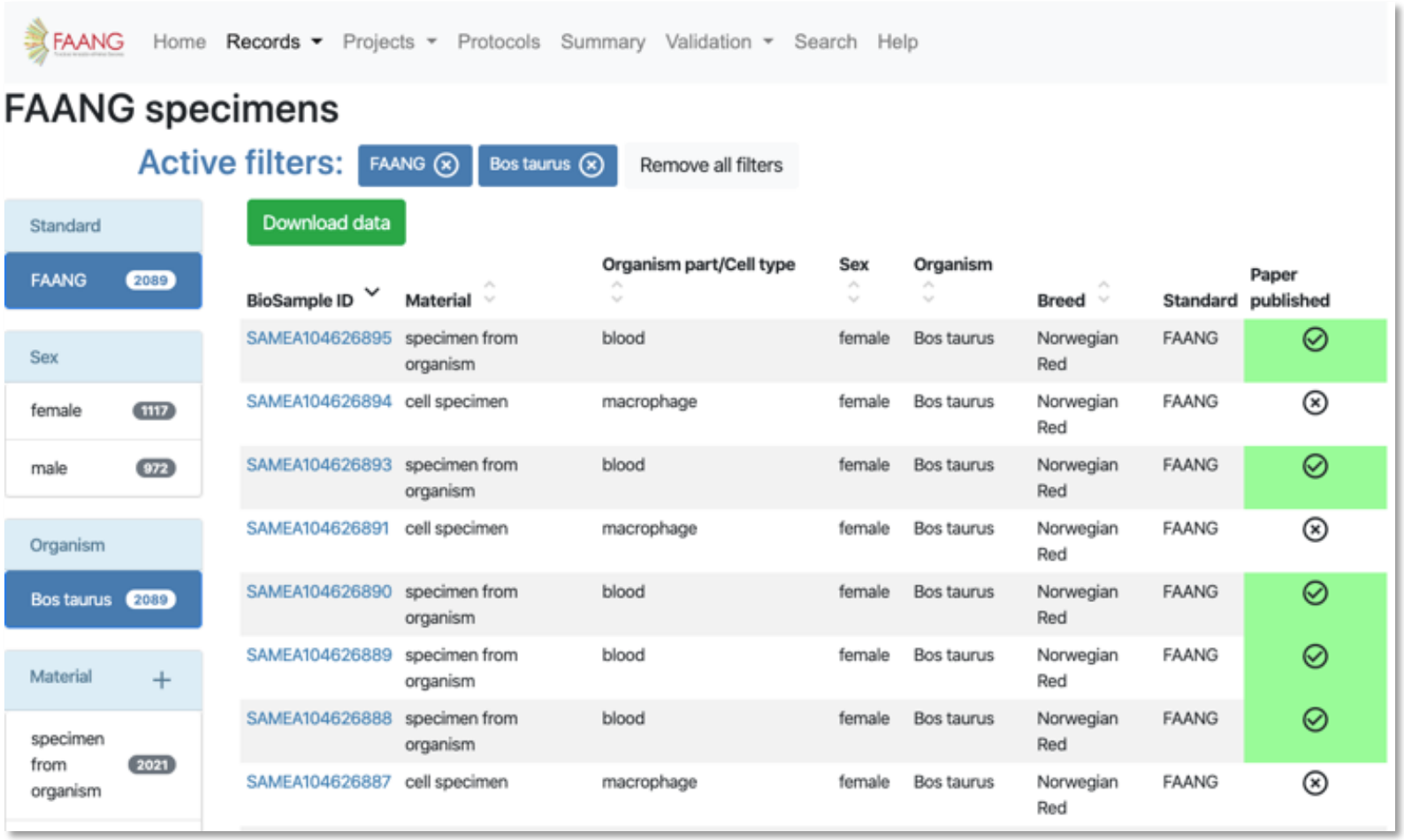


Sample name	Core errors	Core warnings	Type errors	Type warnings	Custom errors	Custom warnings
ECA_UKY_S1	pass	pass	2 errors	1 warning	pass	pass
ECA_UKY_S2	pass	pass	pass	2 warnings	pass	pass
ECA_UKY_S9	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S10	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S32	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S33	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S34	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S35	pass	pass	pass	1 warning	pass	pass
ECA_UKY_S36	pass	pass	pass	1 warning	pass	pass

- FAANG DCC currently investigating automated submission direct from FAANG data portal.
- Simplifies data submission to underlying public archives.
- Uses token based authentication.

NGS data automatically imported to FAANG data portal

- This is because all data run through validation service is tagged as FAANG.



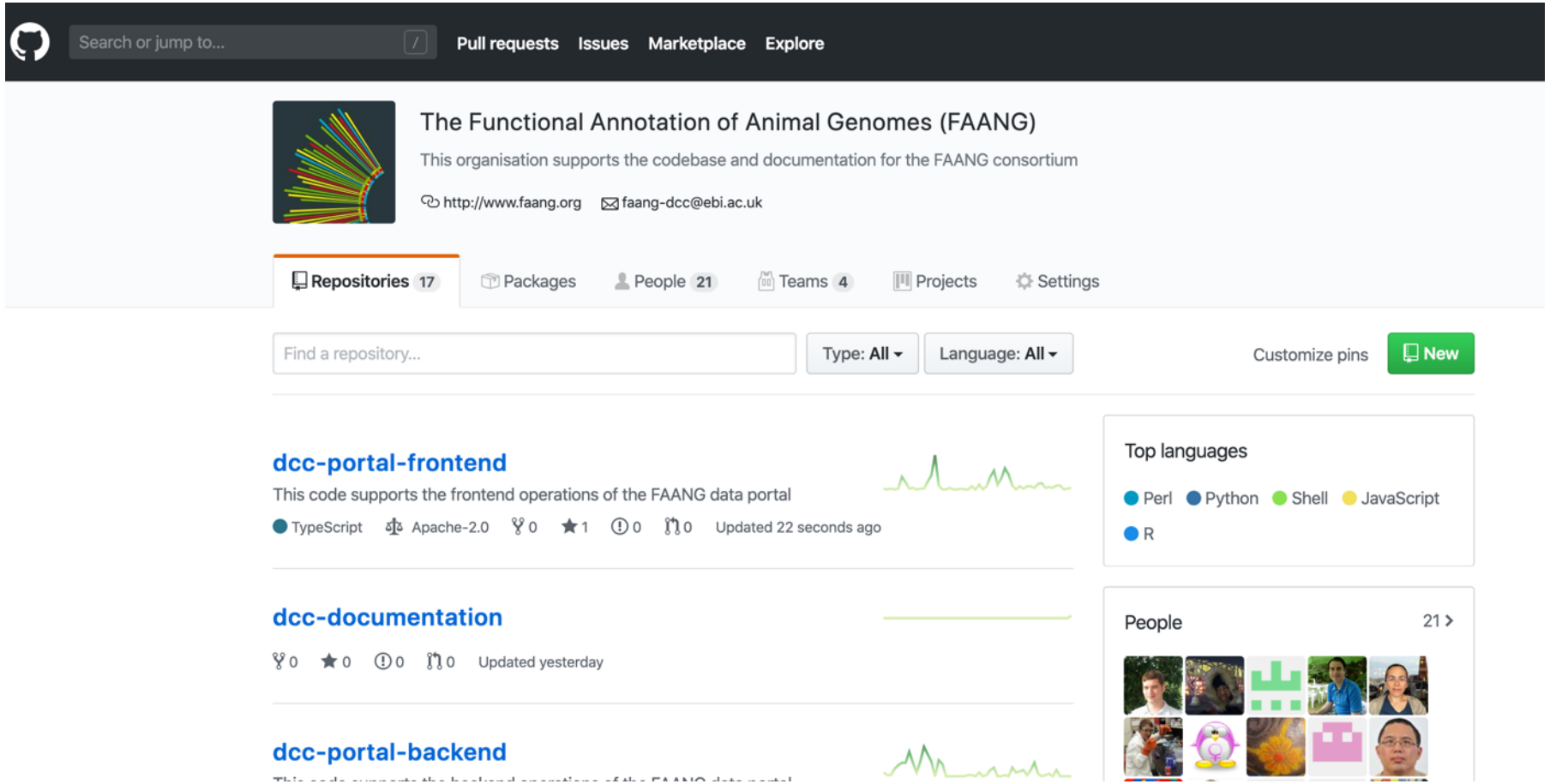
FAANG data search and retrieval

- Time for a demo.

Time for you to start preparing and utilising FAANG pipelines, other than these training courses, what resources are there to help.

FAANG GitHub repository is collating FAANG pipelines

https://github.com/FAANG/



Vision: FAANG standardized workflow distribution

FAANG Functional Annotation of Animal Genomes

FAANG Bioinformatics working groups

Community contributions

GitHub

FAANG Functional Annotation of Animal Genomes

docker

Master record

CYVERSE™

KBase

amazon web services™

EMBASSY cloud

EMBL

Next skills needed for developing your pipelines for FAANG

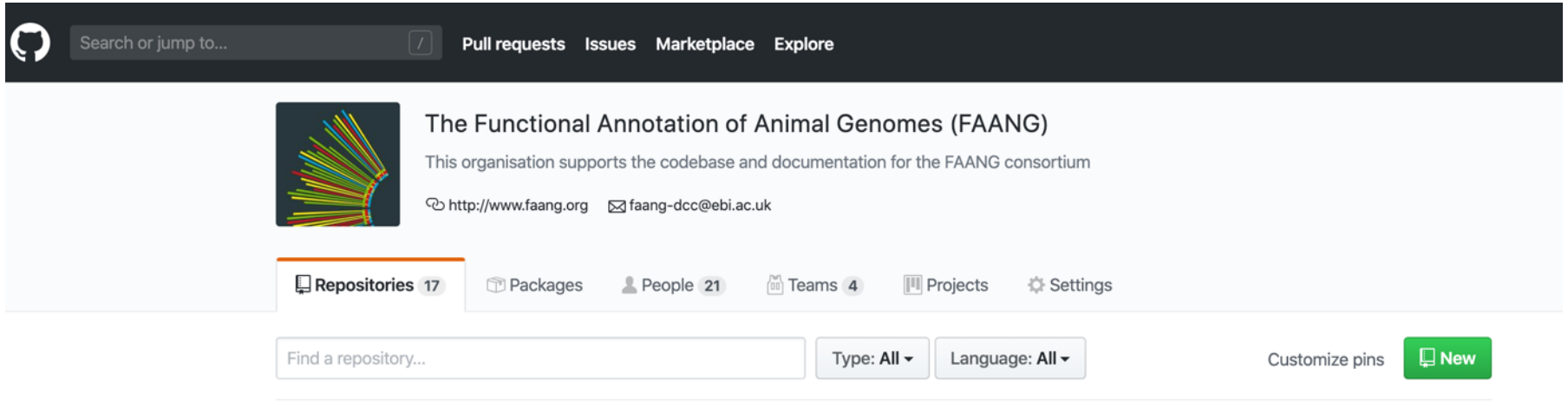
- Containerisation. e.g. Docker
  - Control software environment to ensure your pipeline is the same wherever its installed and allow use on desktop, HPC or cloud.
- Workflow managers e.g NextFlow.

  - enable easy pipeline construction from existing components and manage data flow.
- Effective utilisation of cloud resources. e.g. Embassy cloud, CyVerse or AWL
  - Rapid analysis through parralelisation.

- The above maximise reproducibility of pipelines and effective

FAANG coding standards and best practice

- Develop openly on GitHub with permissive licences e.g GNU or Apache.
- Write code to enable others to contribute and reuse with good documentation.
- Distribute your code in containers and use workflow managers.
- Don’t reinvent the wheel, reuse and adapt existing pipelines from others, look at FAANG GitHub repository for available pipelines.



If you want to know more, other FAANG talks at PAG

- I am giving the following other PAG talks:
  - The FAANG Data Coordination Centre: New European Perspectives for Our Continued Global Effort
    - Pacific salon 3, 16:15 today
  - A vision for Bioinformatics within the global FAANG project
    - Pacific salon 3, 17:15 today
  - The European Nucleotide Archive: Submission, Retrieval and Our Role As the FAANG Data Coordination Centre
    - California room, 09:30 Sunday
- Go chat to CyVerse stand 429 or attend their talks.
- Come find me on the EMBL-EBI stand 519 if you want to chat about anything

Why should I share my data?

Why should I share my data?

- Because my journal/funder makes me?

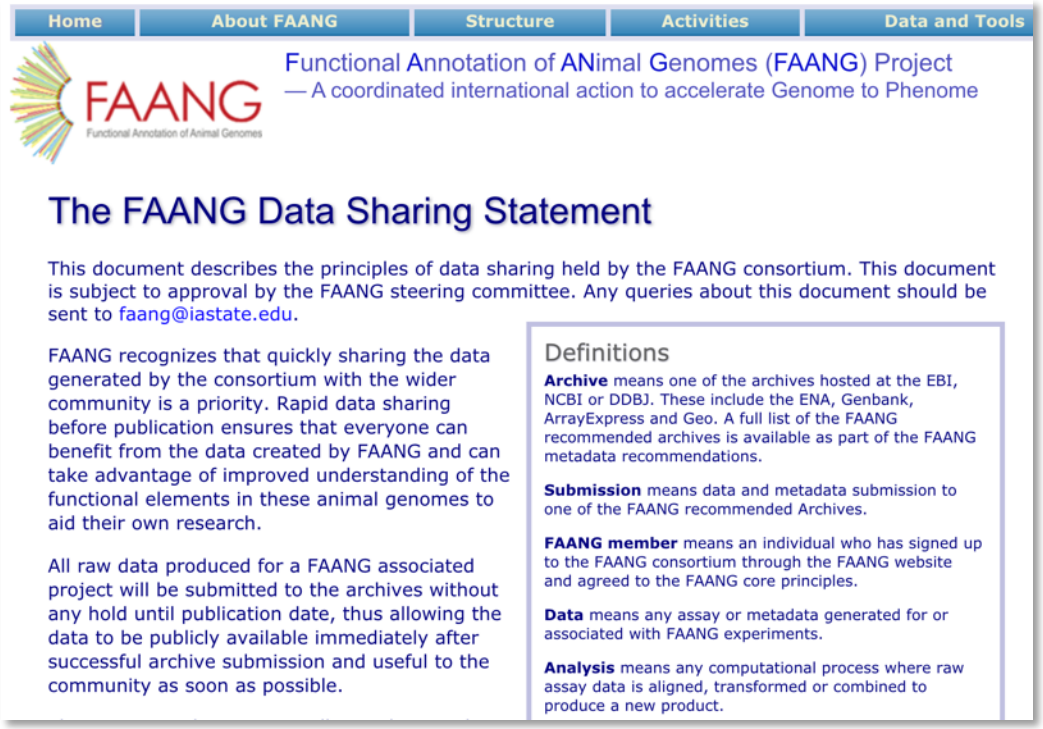
Why should I share my data?

- Because my journal/funder makes me?

- Better reasons:
  - It accelerates research.
  - Increases your citation rates and standing in the community.
  - Maximises value to public and charities (if applicable).
  - Increases impact of your research, fosters collaboration.
  - Stops wasteful duplication of data.
  - Promotes innovation through downstream and comparative analyses.

FAANG data sharing statement

- Promotes rapid pre-publication data release.
- As soon as samples taken, or data generated, they should be submitted to the public archives.
- Reserves right to data provider to first publication with data, but accelerates research by enabling others to begin working on future publications and using data for comparative analyses and benchmarking.


Home | About FAANG | Structure | Activities | Data and Tools

FAANG
Functional Annotation of Animal Genomes

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

The FAANG Data Sharing Statement

This document describes the principles of data sharing held by the FAANG consortium. This document is subject to approval by the FAANG steering committee. Any queries about this document should be sent to faang@iastate.edu.

FAANG recognizes that quickly sharing the data generated by the consortium with the wider community is a priority. Rapid data sharing before publication ensures that everyone can benefit from the data created by FAANG and can take advantage of improved understanding of the functional elements in these animal genomes to aid their own research.

All raw data produced for a FAANG associated project will be submitted to the archives without any hold until publication date, thus allowing the data to be publicly available immediately after successful archive submission and useful to the community as soon as possible.

Definitions

Archive means one of the archives hosted at the EBI, NCBI or DDBJ. These include the ENA, Genbank, ArrayExpress and Geo. A full list of the FAANG recommended archives is available as part of the FAANG metadata recommendations.

Submission means data and metadata submission to one of the FAANG recommended Archives.

FAANG member means an individual who has signed up to the FAANG consortium through the FAANG website and agreed to the FAANG core principles.

Data means any assay or metadata generated for or associated with FAANG experiments.

Analysis means any computational process where raw assay data is aligned, transformed or combined to produce a new product.

Without context, your data is meaningless.

Without sharing your data, your science is not believable.

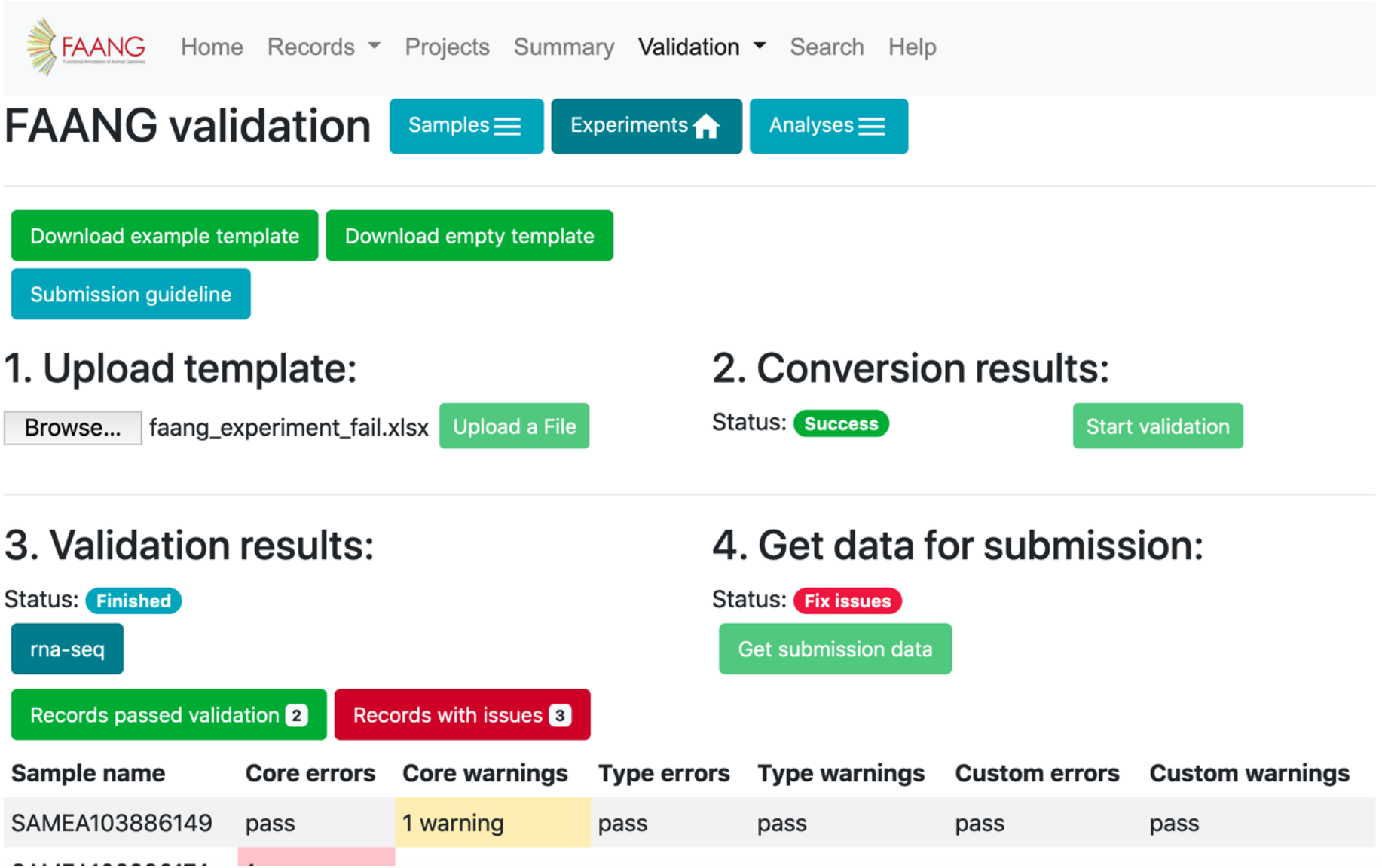
Without context, your data is meaningless.

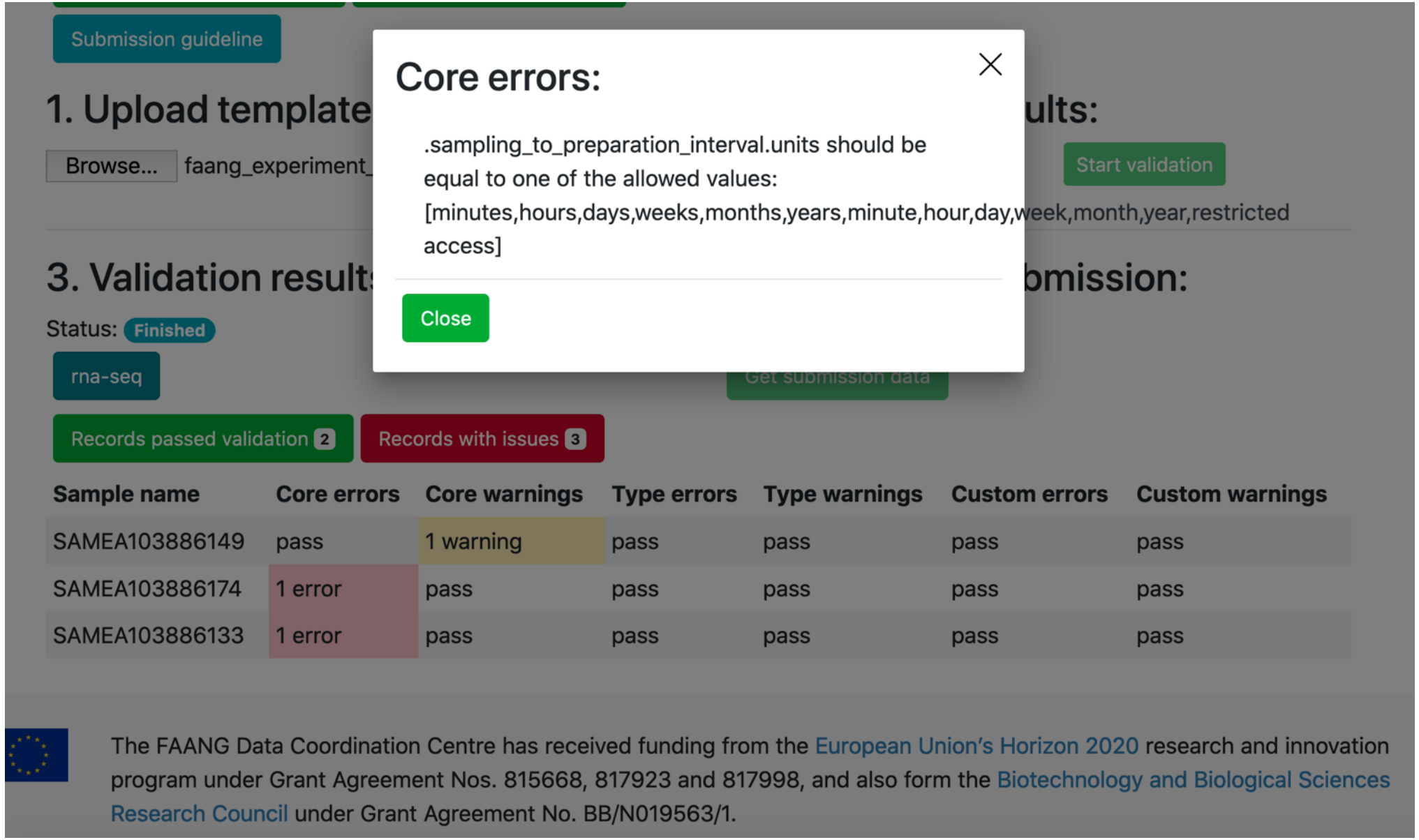
Without sharing your data, your science is not believable.

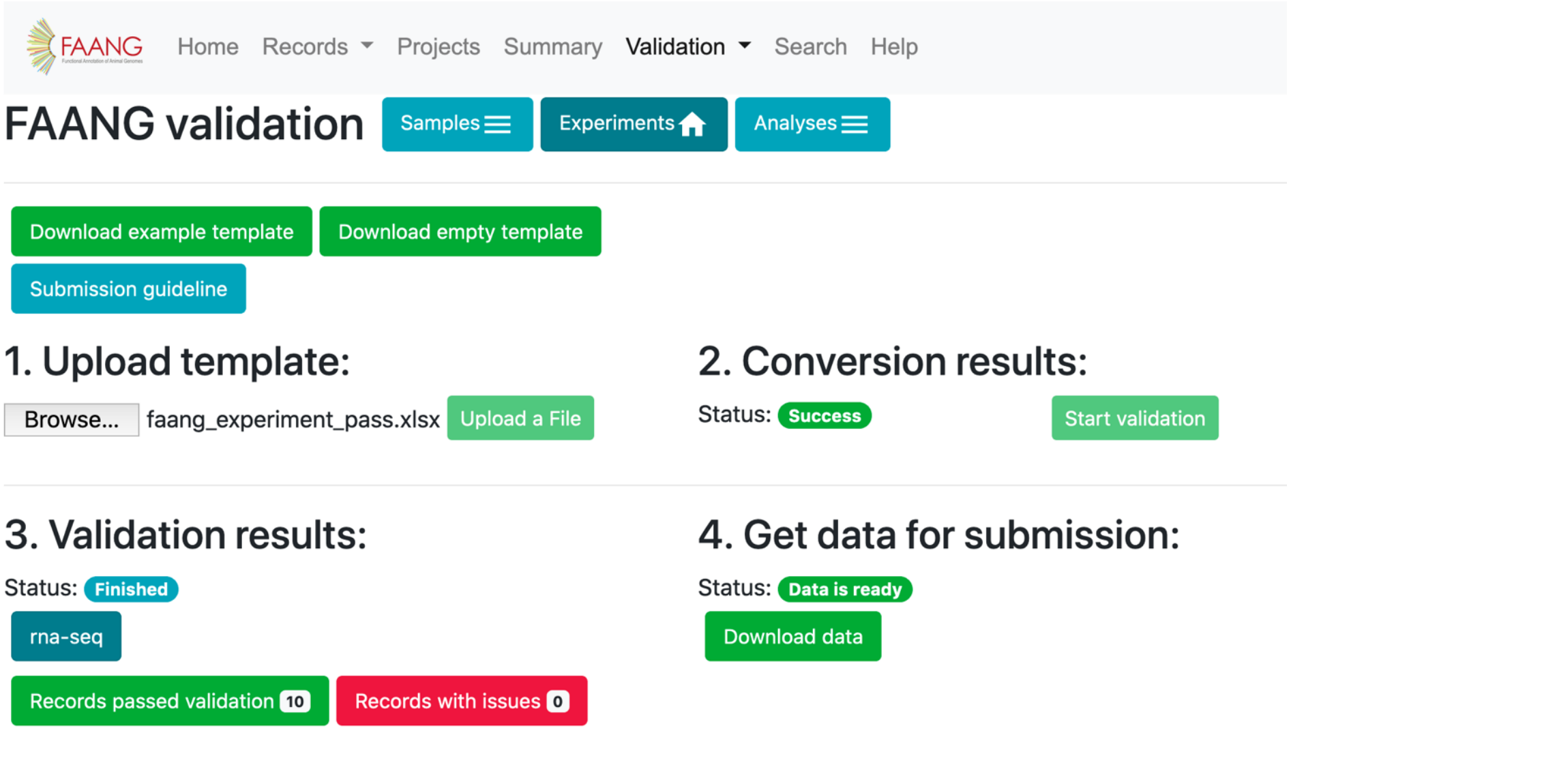
peter@ebi.ac.uk
@peterwharrison

- Alexey Sokolov
- Jun Fan
- Guy Cochrane
- Paul Flicek
- Daniel Zerbino

AQUA-FAANG
BovReg Understanding cattle genomes
GENE-SWitCH
FAANG Functional Annotation of Animal Genomes
Horizon 2020 European Union Funding for Research & Innovation
BBSRC bioscience for the future
cost European Cooperation in Science and Technology

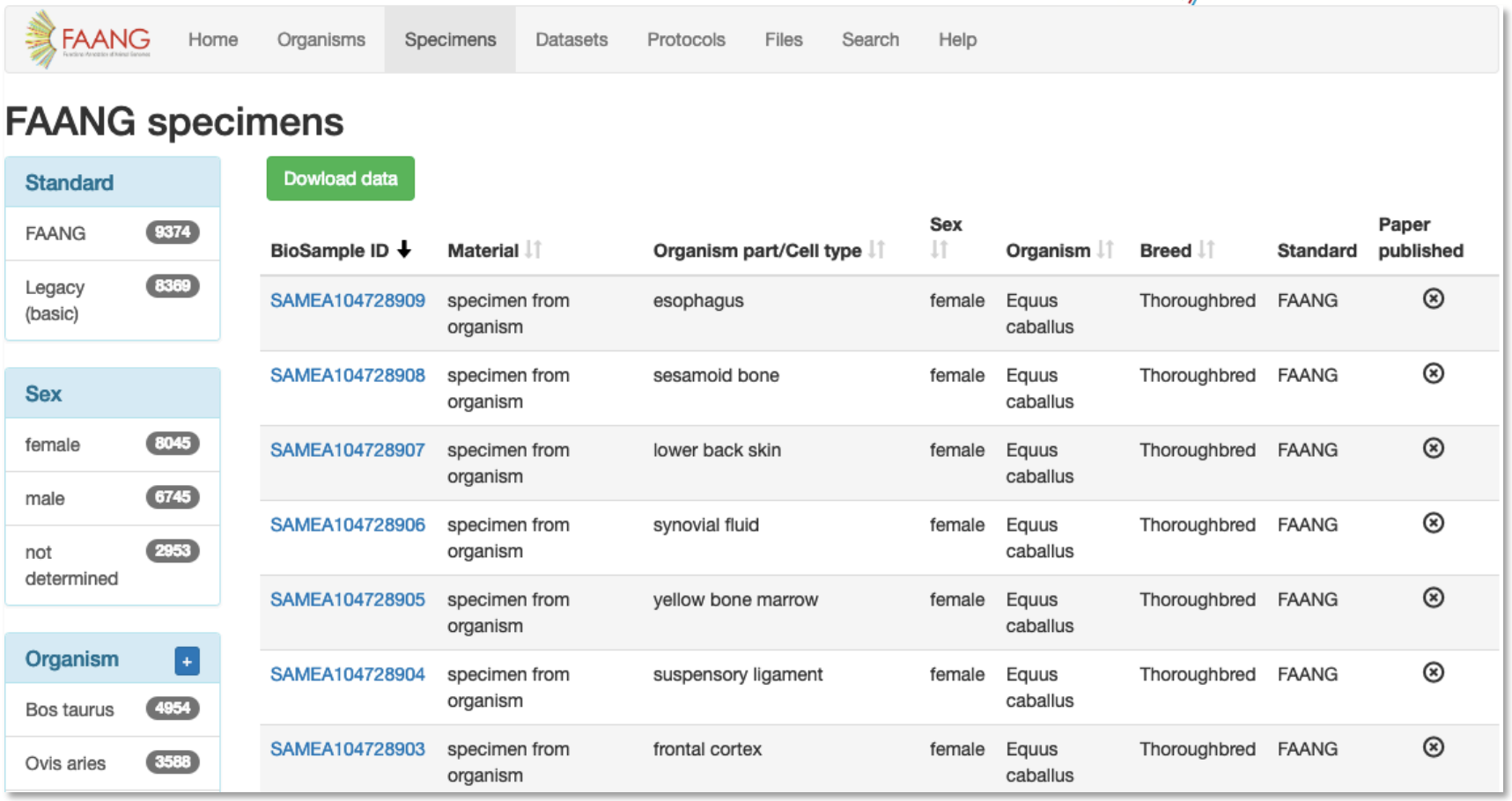

FAANG
Home
Records
Projects
Summary
Validation
Search
Help
FAANG validation
Samples
Experiments
Analyses
Download example template
Download empty template
Submission guideline
1. Upload template:
Browse...
faang_experiment_fail.xlsx
Upload a File
2. Conversion results:
Status: Success
Start validation
3. Validation results:
Status: Finished
rna-seq
Records passed validation 2
Records with issues 3
4. Get data for submission:
Status: Fix issues
Get submission data
Sample name
Core errors
Core warnings
Type errors
Type warnings
Custom errors
Custom warnings
SAMEA103886149
pass
1 warning
pass
pass
pass
pass


Submission guideline
1. Upload template
ults:
Browse...
faang_experiment_
Start validation
3. Validation result
bmission:
Status: Finished
rna-seq
Get submission data
Records passed validation 2
Records with issues 3
Sample name
Core errors
Core warnings
Type errors
Type warnings
Custom errors
Custom warnings
SAMEA103886149
pass
1 warning
pass
pass
pass
pass
SAMEA103886174
1 error
pass
pass
pass
pass
pass
SAMEA103886133
1 error
pass
pass
pass
pass
pass
Core errors:
.sampling_to_preparation_interval.units should be equal to one of the allowed values: [minutes,hours,days,weeks,months,years,minute,hour,day,week,month,year,restricted access]
Close
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 form the Biotechnology and Biological Sciences Research Council under Grant Agreement No. BB/N019563/1.


FAANG
Home
Records
Projects
Summary
Validation
Search
Help
FAANG validation
Samples
Experiments
Analyses
Download example template
Download empty template
Submission guideline
1. Upload template:
Browse...
faang_experiment_pass.xlsx
Upload a File
2. Conversion results:
Status: Success
Start validation
3. Validation results:
Status: Finished
rna-seq
Records passed validation 10
Records with issues 0
4. Get data for submission:
Status: Data is ready
Download data

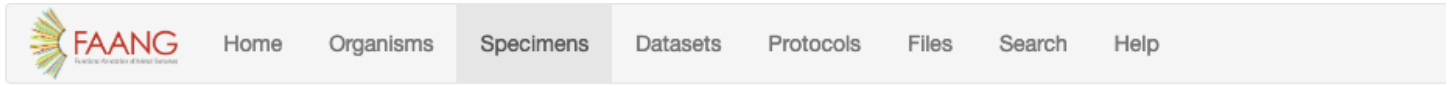
FAANG Data portal

- Single access point for all FAANG data.
- Indexes data held across different biological archives for direct download.
- Richly described FAANG data enables powerful filters to narrow down to records of interest.



http://data.faang.org/

FAANG Data portal



SAMEA5079418

Name:	OAR_RI_AF3_Oma
BioSample ID:	SAMEA5079418
Release date:	2016-11-17
Update date:	2018-09-06
Sex:	female
Organism BioSample Id:	SAMEA6265168
Organism:	Ovis aries
Breed:	Texel sire x Scottish Blackface dam
Health status:	normal
Description:	omasum from an adult female texel x blackface sheep, OAR_RI_AF3
Standard met:	FAANG
Project:	FAANG
Organisation:	BBSRC (funder) EMBL-EBI (curator) The Roslin Institute and Royal Dick School of Veterinary Studies (biomaterial provider) The Roslin Institute and Royal Dick School of Veterinary Studies (institution)
Material:	specimen from organism
Derived from:	SAMEA6265168
Specimen collection date:	2014-05-14
Animal age at collection:	2 year
Developmental stage:	adult
Organism part:	omasum
Specimen collection protocol:	ROSLIN_SOP_Harvest_of_Large_Animal_Tissues_20160516.pdf

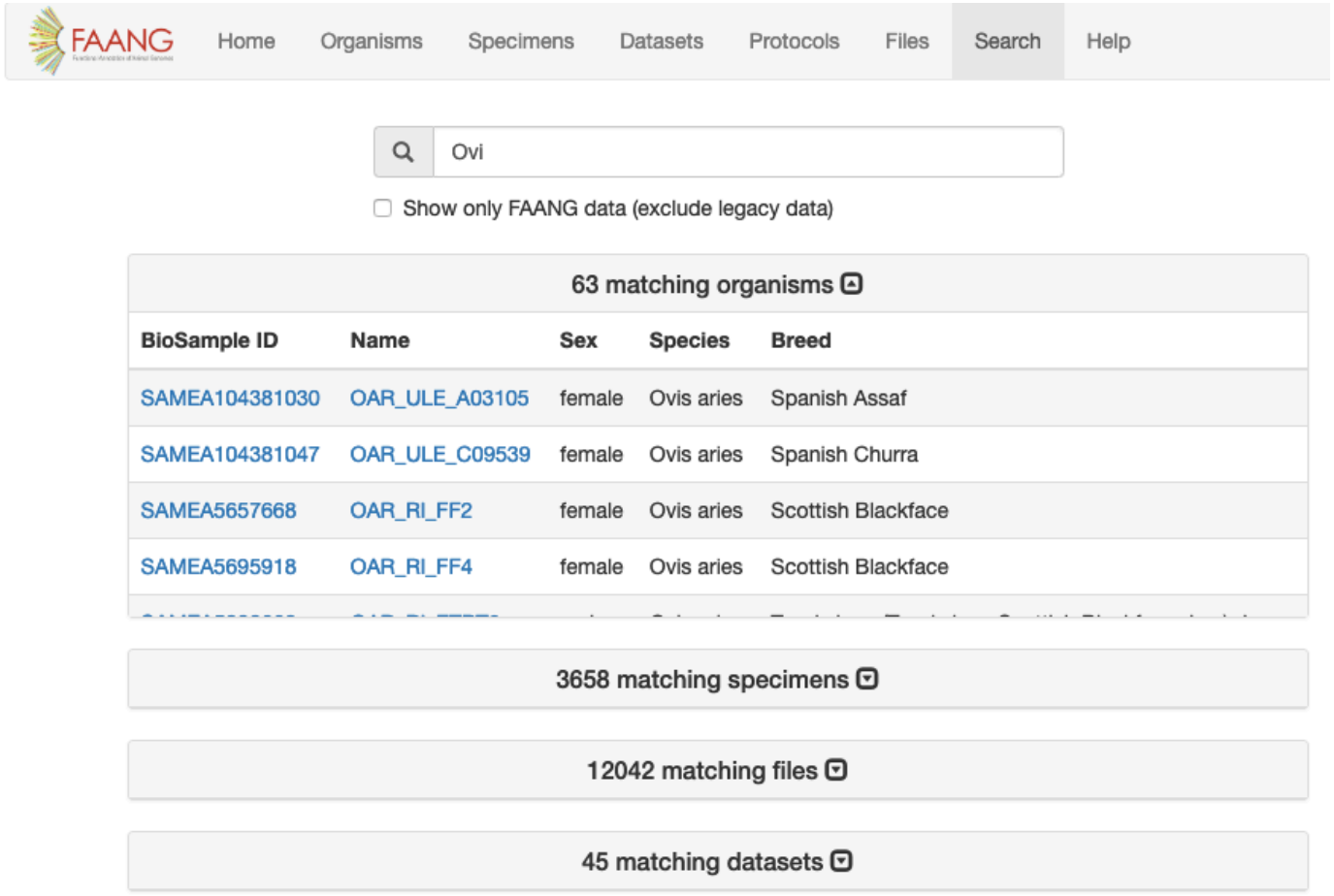
Files

Name	Archive	Experiment	Run	Download
ERR2073971_1	ENA	ERX2133147	ERR2073971	
ERR2073971_2	ENA	ERX2133147	ERR2073971	

- Individual detail pages provide full metadata, to ensure data matches your requirements and is comparable to other sets.
- Each record provides access to sample and experiment protocols.
- Download any data files associated with the sample with direct links to data held in the archives.

FAANG Data portal search

- Predictive text based search across all FAANG metadata fields.
- Returns organisms, specimens, datasets and individual analysis files.
- Search results link directly to individual detail pages.



http://data.faang.org/search