



# The Metadata and Data Sharing Committee

Laura Clarke

10<sup>th</sup> January

# The Metadata and Data Sharing Group

- **To define metadata standards** for samples, experiments and analyses, ensuring FAANG data are maximally useful.
- **To establish best practice** within FAANG for **data archiving** at the NCBI and EMBL-EBI
- **To enable data sharing** both within FAANG and with the wider community

**SHARING DATA MEANS SHARING DATA DESCRIPTIONS**

Chris Tuggle

# Communication

- Email [faang-metadata@animalgenome.org](mailto:faang-metadata@animalgenome.org)
- Monthly phone calls – 2<sup>nd</sup> Thursday Month  
4pm UK time
- Next call Thursday February 11<sup>th</sup> 4pm
- Confluence -  
<http://www.ebi.ac.uk/seqdb/confluence/display/FAANG/Metadata+and+Data+Sharing>
- Github - <https://github.com/FAANG/faang-metadata>



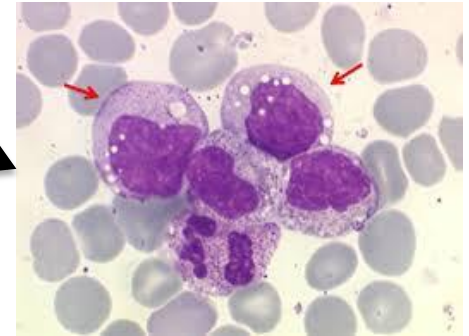
# BioSamples



Derived from



Derived from



**Alan's Pig**  
**Duroc**  
**LBO:0000358**

**Blood**  
**UBERON\_0000178**

**CD14-positive,  
CD16-negative  
classical monocyte**  
**CL\_0002057**

Related to



**Alan's other pig**  
**Duroc**  
**LBO:0000358**



### Animal

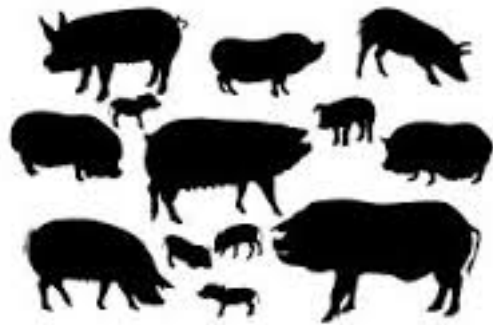
- Species - NCBI taxon ID.
- Sex (any child term of PATO\_0000047)
- Birth date
- Strain / Breed (ontology)

Links between entities are required

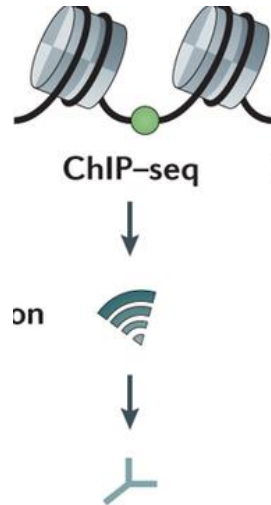
### Specimen

- Date at which specimen collection occurred
- Animal age at point of specimen collection
- Animal Disease / health status at point of collection
- Tissue (UBERON term preferred)
- Method of collection (protocol)
- Fasted status - either 'fed', 'fasted' or 'unknown'. Criteria must be specified in the protocol.

# Assay Archives – Metadata Structure



1 to  
many



1 to  
≥1



**STUDY**

**A collection of  
experiments**

**EXPERIMENT**

**e.g. Sequencing  
library**

**RUN**

**Sequencing  
lane or plex**

# Experiment Metadata

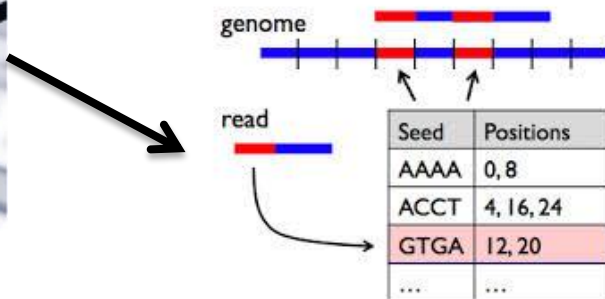
- Assay Type
- Experiment target
- Sample
- Sample storage
- Sample storage processing
- Sample to preparation interval / preparation date
- Experimental Protocol

# Analysis metadata



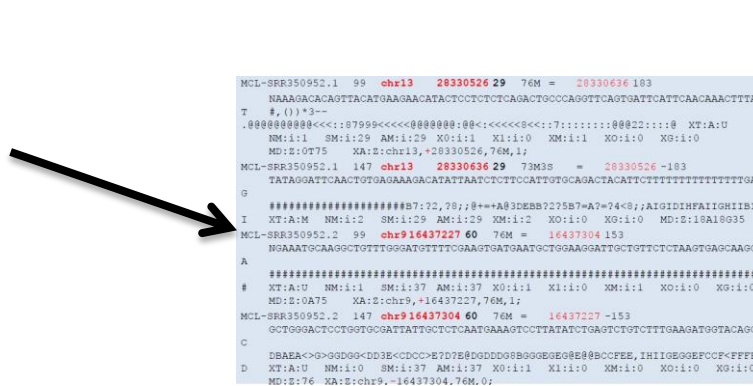
**RUN**

**Sequencing  
lane or plex**



**Analysis**

**Algorithm  
Parameters**



**Output**

**File Format  
Reference data**



# Analysis metadata

- Input data
- Reference data
- Analysis protocol
- URLs and versions for all software used (including in-house scripts)
- Full command line used to run the analysis
- Link to any VM or containers used, if applicable
- QC Attributes

# Filenames

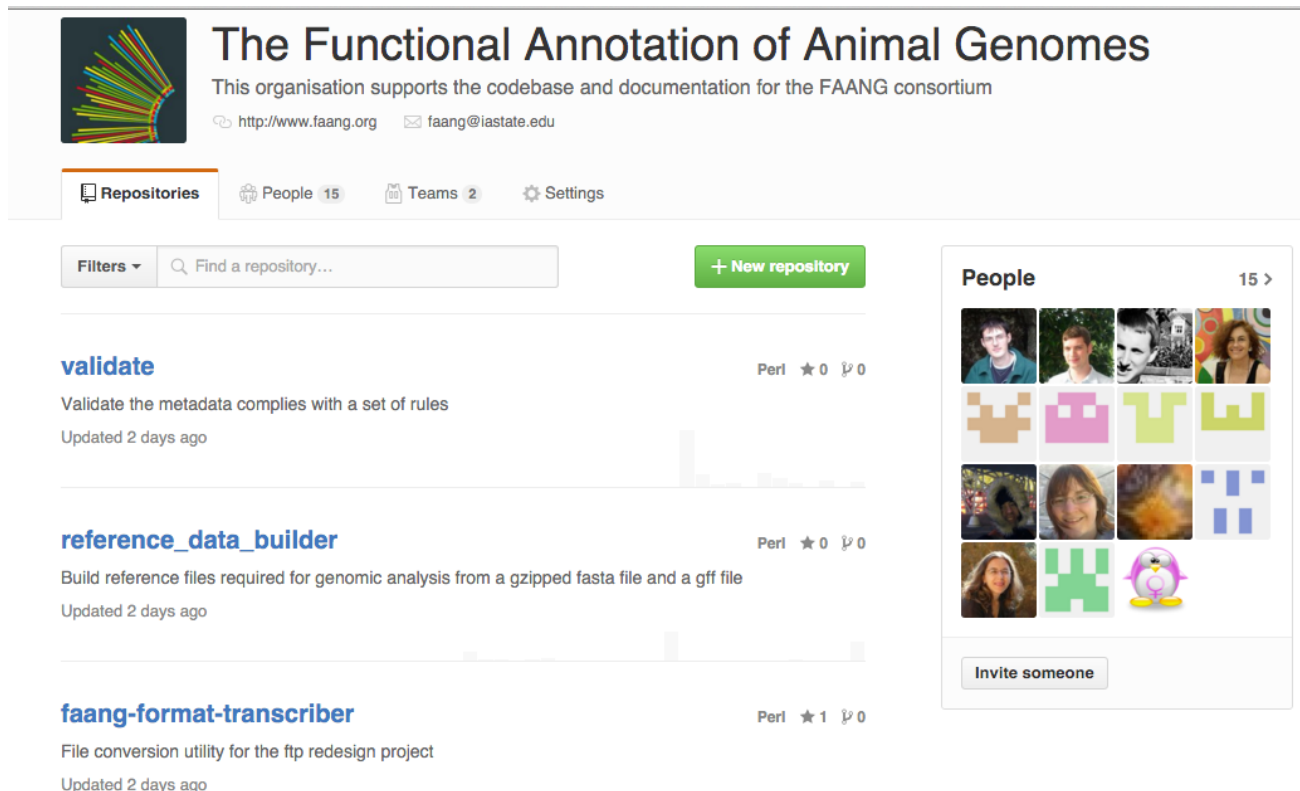
- species / assembly version, sample name sample description (tissue or cell type) assay type experiment target experiment ID analysis protocol name results type data freeze date file format
- OAR3\_1.OA\_Roslin001.liver.H3K27ac.ERX053278.FAANGUK\_chipv3.peaks.20150617.bb
- Peak calls from Roslin's first sheep liver, H3K27ac chipseq analysed with the chipv3 pipeline aligned to OAR3\_1

# Data Sharing

- Public FTP site <ftp://ftp.faang.ebi.ac.uk/ftp/>
- Internal dropbox
  - [drop-faa@ftp.faang.ebi.ac.uk](mailto:drop-faa@ftp.faang.ebi.ac.uk)
- Usernames, passwords and guidelines for use are on confluence
- <http://www.ebi.ac.uk/seqdb/confluence/display/FAANG/FAANG+internal+data+sharing>

# Sharing Standards and Code

- <https://github.com/FAANG> - 7 repositories
- Everyone is welcome to be a member



The screenshot shows the GitHub organization page for "The Functional Annotation of Animal Genomes". The organization's profile includes a logo of colorful lines, the name "The Functional Annotation of Animal Genomes", and a description: "This organisation supports the codebase and documentation for the FAANG consortium". Contact information includes the website <http://www.faaang.org> and email [faang@iastate.edu](mailto:faang@iastate.edu). Navigation tabs for "Repositories", "People 15", "Teams 2", and "Settings" are visible. A search bar for repositories and a "+ New repository" button are present. Three repositories are listed: "validate" (Perl, 0 stars, 0 forks), "reference\_data\_builder" (Perl, 0 stars, 0 forks), and "faang-format-transcriber" (Perl, 1 star, 0 forks). A "People" sidebar on the right shows 15 members with their profile pictures and avatars, and an "Invite someone" button.

# Declaring FAANG standards v1

- We aim to do this on February 11<sup>th</sup>
- Please give us your comments

# Validation against the standard

- EMBL-EBI writing a metadata validation tool
- <https://github.com/FAANG/validate>
- Configurable rules
- Run by submitters and the data coordination centre

# Supporting Archive submission

- EMBL-EBI wants to support data submission
- Who from your group/institute is submitting the data?
- Are they are member of the M&DS group?
- Please get them to join

# What next?

Submit your data to the archives