

GENE-SWitCH

The regulatory GENome of SWine and CHicken: functional annotation during development

Deliverable D3.1 GENE-SWitCH Data Management Plan

Deliverable leader: EMBL-EBI

Authors: Peter Harrison (EMBL-EBI) and Guy Cochrane (EMBL-EBI)

Version: 1.2

Due date of deliverable (as in DoA):	M6, updated on M18
Actual submission date:	18/12/2019, month M6, 15/12/2020 M18

Dissemination level:

PU Public	X
CO Confidential, only for members of the consortium (including the Commission Services)	

Research and Innovation Action, SFS-30-2018-2019-2020 Agri-Aqua Labs
Duration of the project: 01 July 2019 – 30 June 2023, 48 months



Table of contents

- 1 Summary 3**
- 2 Data Management Plan..... 3**
- 3 Conclusion 3**
- 4 Deviations or delays..... 3**
- 5 Annexes..... 3**
 - 5.1 GENE-SWitCH Data Management Plan..... 3
 - 5.2 FAANG Data Sharing Statement 13
- 6. Glossary..... 14**



1 Summary

The objective of this deliverable was to update the version 1.0 of the GENE-SWitCH Data Management plan (DMP) previously submitted. As a living document it is expected that the DMP will be updated periodically throughout the lifetime of the project. In order to develop the DMP, the H2020 template was used to ensure it covered the expectations of the commission and matched the standard structure that the community would expect from a DMP document. We utilised version 1.0 of the H2020 template, developed in 2016 by the commission, this template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. Version 1.2 of the GENE-SWitCH DMP will be distributed to the entire consortium and made available through the consortium collaborative platform, future updates to the DMP will also be distributed in this manner. The DMP was written by Peter Harrison (EMBL-EBI) utilising the input of the entire consortium into the GENE-SWitCH project proposal.

2 Data Management Plan

Version 1.2 of the GENE-SWitCH DMP has been prepared and distributed to the consortium. The plan was prepared in accordance to guidance and template set out by H2020. The DMP covers the projects commitment to FAIR open data access and the external requirements stipulated by its membership of the FAANG coordinated action, that is covered in the FAANG Data Sharing Statement (see Annex 5.2). The data management plan also complies with EU and H2020 requirements for open access, long term archival access to data, and strives for highest level of compliance with FAIR data principles.

For clarity version 1.2 of the DMP has been included in Annex 5.1 of this deliverable document and the table history of changes relates the revisions from previous versions.

The Data Management Plan ensures that all GENE-SWitCH data will be publicly available pre-publication and will conform to FAIR (findable, accessible, interoperable and reusable) standards. It documents the key strategies for ensuring long term data security through the use of the EMBL-EBI public archives.

3 Conclusion

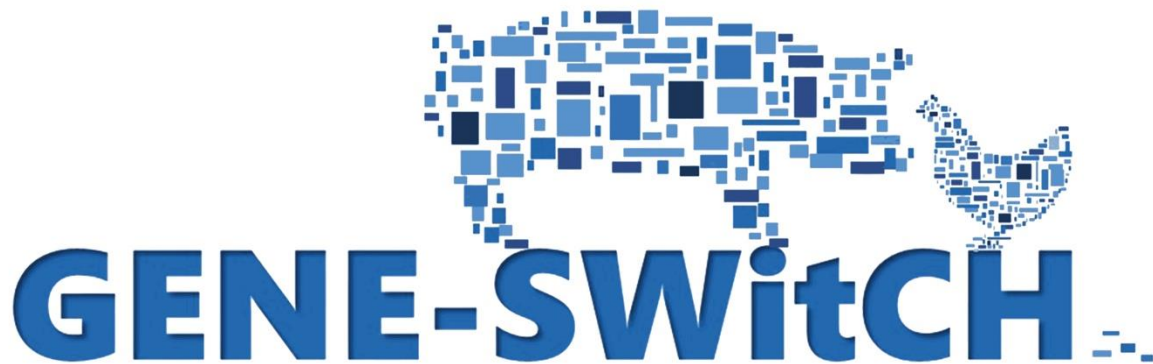
Version 1.2 of the GENE-SWitCH Data Management Plan document has been finalised and distributed to all members of the consortium. The Data Management Plan is intended to act as a living document, therefore updates to the document will be similarly distributed to all partners in the consortium. The data management plan and any updated versions will be available to the consortium from the GENE-SWitCH sharing platform <https://sites.inra.fr/site/gene-switch/SitePages/Home.aspx>.

4 Deviations or delays

None, Version 1.2 of the DMP updated at the end of the first reporting period as planned.

5 Annexes

5.1 GENE-SWitCH Data Management Plan



1 DATA MANAGEMENT PLAN

Project¹ Number: 817998

Project Acronym: GENE-SWitCH

Project title: The regulatory GENomE of SWine and CHicken: functional annotation during development

Author: Peter Harrison EMBL-EBI

Version: 1.2

Disclaimer

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

¹ The term 'project' used in this template equates to an 'action' in certain other Horizon 2020 documentation



History of changes

VERSION	PUBLICATION DATE	CHANGES
1.0	06.12.2019	Initial version from WP3
1.1	23.12.2019	Minor revisions from consortium comments
1.2	11.11.2020	Minor updates to Table 1, data security text and the FAANG Data Sharing Statement.



Table of Contents

1	Data Summary	6
2	FAIR data	8
2.1.	Making data findable, including provisions for metadata	8
2.2.	Making data openly accessible.....	8
2.3.	Making data interoperable	9
2.4.	Increase data re-use (through clarifying licences).....	10
3	Allocation of resources	11
4	Data security	11
5	Ethical aspects	12
6	Other issues	12
10.2	FAANG Data Sharing Statement	13
6.	Glossary	14

1 Data Summary

The GENE-SWitCH project aims to deliver to the livestock community two functionally mapped monogastric genomes, for chicken and pig. These functional maps and associated data collections will enable immediate translation into the pig and poultry sectors for developments in sustainable production. For example, within the project the datasets will be employed to evaluate the effect of maternal diet on the epigenome of pig foetuses. A key aspect of the generated datasets will be the extensive associated rich and controlled metadata, information that is key for the development of phenome to genome resources. The project will utilise existing pig and chicken datasets from the FAANG collection and wider community from the public archives as test datasets for the development of its openly developed bioinformatic pipelines and to enrich its new functional maps; it will also use the existing reference genomes from the community as a starting point for its own improved mapped genomes. The tissue samples for sequencing are being collected for pig from the INRA experimental unit GenESI in Lusignan, France and chicken samples from the UK National Avian Research Facility (www.narf.ac.uk). In total the project will generate data from 13 different assays, including both the core FAANG assays and additionally DNA methylation, Hi-C and Whole Genome Sequencing (Table 1). The Data Management Plan will be periodically updated throughout the project to reflect changes in the data produced by the project and any changes in storage and release. A future update will include the sizes of the datasets produced by the project.



Table 1. The number of GENE-SWitCH assays and samples across developmental stages.

Assays		Annotations	Number of assays	Number of samples				
				Early organogenesis 7 tissues * 2 species * 4 replicates (WP1)	Late organogenesis 7 tissues * 2 species * 4 replicates (WP1) 2 tissues * 84 replicates (WP5)	Newborn/hatched 7 tissues * 2 species * 4 replicates (WP1)	Weaned piglets 2 tissues * 84 replicates (WP5)	Fattening pigs 3 tissues * 300 replicates (WP4)
				Pillar 1	Pillar 1&2	Pillar 1	Pillar 1&2	Pillar2
Core FAANG assays	ATAC-seq	Accessible chromatin regions.	168 + 336	56	56 + 168	56	168	
	ChIP-seq	H3K4me3	Promoters, enhancers, repressors.	168	56	56	56	
		H3K4me1		168	56	56	56	
		H3K27me3		168	56	56	56	
		H3K27Ac	Chromatin domain boundaries.	168	56	56	56	
		CTCF	168	56	56	56		
	RNA-seq	mRNA-seq	Characterisation of transcripts structure (start and stop sites) and categories.	168	56	56	56	
		small RNA-seq		168	56	56	56	
		lrrNA-seq		42	14 (1 sample)	14 (1 sample)	14 (1 sample)	
		mRNA-seq*	Transcripts expression levels.	900 + 336		168		168
DNA methylation	RRBS	Active and inactive promoters	126	42 (3 replicates)	42 (3 replicates)	42 (3 replicates)		
	WGBS	Active and inactive chromatin	42	14 (1 sample)	14 (1 sample)	14 (1 sample)		
Hi-C	Capture Hi-C	Structural domain annotations Enhancer/promoter interactions	24	8 (2 tissues, 2 species, 2 replicates)	8 (2 tissues, 2 species, 2 replicates)	8 (2 tissues, 2 species, 2 replicates)		
WGS	Whole Genome Sequencing	Genetic variants	300				300	
TOTAL	13 different assays		3354 molecular assays	Pillar 1 : Deep molecular characterisation of 168 tissues Pillar 2 : Specific molecular characterisation of 288 and 900 samples representing respectively 2 tissues (skeletal muscle and liver) and 3 tissues (skeletal muscle, ileum, liver)				

Core FAANG assays; (*) at lower depth than above in order to process several samples.



All processed data generated in the project will be shared using standard bioinformatic data file formats (i.e. FASTQ, FASTA, SAM/BAM, GFF/GTF, BED, and VCF). The project will make extensive use of open access existing legacy datasets and of additional datasets generated during the lifetime of the project identified and accessed from EMBL-EBI public archives. Overall the project contributes to the global FAANG coordinated effort with i) the deliverance of high quality functional genome maps for the pig and the chicken, ii) demonstrable impact of these new data resources on developments in the breeding industry, and iii) the production of cutting edge bioinformatic pipelines and experimental techniques that will be of wide benefit to the scientific research community.

2 FAIR data

2.1. Making data findable, including provisions for metadata

The proposed data deposition of GENE-SWitCH data through FAANG to the EMBL-EBI public archives will ensure the generated data is highly discoverable. GENE-SWitCH utilises the FAANG metadata standards (<https://data.faaang.org/ruleset/samples#standard>). All data submissions will be validated through the FAANG validation and submission tools, that are in fact being updated as part of the GENE-SWitCH project and are accessible at (<https://data.faaang.org/validation/samples>). The deposition in the public archives gives every data file a unique accession. These accessions are globally recognised by the comparable archives at the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) and DNA Databank of Japan (DDBJ; <https://www.ddbj.nig.ac.jp/index-e.html/>). Different assay files are linked through the inclusion of the BioSamples identifier in all data submissions so that all of the datasets generated on each sample can be easily grouped and accessed from downstream presentation resources. GENE-SWitCH will conform with the FAANG record naming conventions. The FAANG data portal utilises ElasticSearch to ensure that all ontology validated metadata fields are keyword searchable using its predictive simultaneous search across samples, reads and analyses (<https://data.faaang.org/search>). It will be possible to search for GENE-SWitCH data as part of an all search or pre-limit the search specifically to only return GENE-SWitCH project data results. The data portal utilises the rich ontology supported metadata to provide filters that allow a user to explore the GENE-SWitCH data based on species, technology, breeds, sex, material, organism part, cell type, assay type, archive, and sequencing instrument. All software will be appropriately versioned using an agreed versioning structure from its coding standards document of work package 2.

2.2. Making data openly accessible

All samples and omics data will be deposited in the EMBL-EBI public archives that includes BioSamples, the European Nucleotide Archive, the European Variation Archive, PRIDE and BiImage archive. These are widely recognised and approved repositories for the long-term storage of biological data and the deposition routes are established with the FAANG Data Coordination Centre (DCC), that itself is based within the Molecular Archives cluster at EMBL-EBI. Apart from the reserved right of first publication stipulation set out in the FAANG Data Sharing statement (<https://www.faaang.org/data-share-principle>), there are no restrictions on use of the data, no data access committee is required and apart from anonymous usage analytics no tracking of individual data use will be made. The following data sharing statement is available both via the websites and Application Programmatic Interfaces (machine readable) of the public archives and FAANG data portal.

"This study is part of the FAANG project, promoting rapid prepublication of data to support the research community. These data are released under Fort Lauderdale principles, as confirmed in the Toronto Statement (Toronto International Data Release Workshop. Birney et al. 2009. Pre-publication data sharing. Nature 461:168-170). Any use of this dataset must abide by the FAANG data sharing principles. Data producers reserve the right to make the first publication of a global analysis of this data."



If you are unsure if you are allowed to publish on this dataset, please contact the FAANG Data Coordination Centre and FAANG Consortium (email faang-dcc@ebi.ac.uk and copy faang@iastate.edu) to enquire. The full guidelines can be found at <http://www.faang.org/data-share-principle>."

The EMBL-EBI public archives are fully aware and accepting of incoming FAANG data including the data of the GENE-SWitCH project. Whilst the FAANG metadata is fully machine readable and the license is available to both web and programmatic users, further improvements will be investigated to further improve the machine readability, in collaboration with the requirements of the other H2020 SFS30 projects. The FAANG DCC will investigate specific license API endpoints, html embedding of license links and license structure formatting to improve machine-based access, a key component of FAIR compliance.

The submission model for GENE-SWitCH will make the data available for direct download from both the FAANG data portal that utilises the underlying public archives infrastructure and from the public archives themselves. This provides by default a range of data access methods including web browser download, FTP, Aspera, Globus and API access to give flexibility to data consumers. All of these download options are open source and the archives have extensive documentation on the various data access options. The FAANG data portal collates the files from the various underlying archives to a single access point. The FAANG API provides programmatic users with the access FTP addresses to make a secondary call to download the data files themselves.

All GENE-SWitCH software will be publicly developed on the FAANG GitHub repository, so that the development process is open to community input and available pre-publication. All of the GENE-SWitCH repositories will be given the prefix 'proj-gs-' within the FAANG GitHub repository (<https://github.com/FAANG>). GENE-SWitCH will ensure that in the FAANG data portal the bioinformatic pipeline that was used to generate the analysis file is linked from the analysis file results page. This ensures that the analysis file, the raw data that generated it, the protocols and the bioinformatic pipelines are all downloadable from the same location. The software will include complete documentation, nextflow workflow management and be containerised in Docker. No specific tools are required to access the data from the data portals or the FAANG data portal, as they will use standard accepted file formats of the public archives. The FAANG data portal will provide a GENE-SWitCH project slice that will allow the data portal and programmatic access interface to provide a bulk download of all GENE-SWitCH data at once, this will be available at <http://data.faang.org/projects/gene-switch>.

2.3. Making data interoperable

GENE-SWitCH data will be submitted through the FAANG DCC that will ensure the data is interoperable with other FAANG datasets and highly reusable by the wider livestock community. To ensure interoperability with all other FAANG datasets, including the other three H2020 SFS30 projects, GENE-SWitCH will employ the latest version of the FAANG metadata standards (and utilise all future updates to these standards), currently version 3.8 (<https://github.com/FAANG/dcc-metadata/tree/master/rulesets>), and in a more readable form at <https://data.faang.org/ruleset/samples#standard>. It will ensure its compliance with these standards by running all data through the FAANG validation software prior to submission to the public archives. GENE-SWitCH will develop coding standards to ensure that all pipelines developed by the consortium are easily utilised, they will be containerised to ease installation and reuse. For its pipelines it will utilise open software applications, that will be implemented with a nextflow workflow manager and containerised using Docker to ensure consistent reuse across the project and by downstream users.

To ensure interdisciplinary interoperability GENE-SWitCH will utilise the recommended ontologies of the FAANG metadata standards as set by the FAANG Metadata and Data Sharing Committee. A specific



action of the project will be through the coordination of the FAANG DCC to improve the coverage and quality of ontologies for use in livestock metadata recording, and the consortium will publish a manuscript on the state of the art and usage of ontologies. Wherever an ontology is not possible we will employ controlled lists to prevent erroneous metadata recording. The ontologies that will be utilised in the project will be:

OBI	https://www.ebi.ac.uk/ols/ontologies/obi
NCBI Taxonomy	https://www.ebi.ac.uk/ols/ontologies/ncbitaxon
EFO	https://www.ebi.ac.uk/ols/ontologies/efo
LBO	https://www.ebi.ac.uk/ols/ontologies/lbo
PATO	https://www.ebi.ac.uk/ols/ontologies/pato
VT	https://www.ebi.ac.uk/ols/ontologies/vt
ATOL	https://www.ebi.ac.uk/ols/ontologies/atol
EOL	https://www.ebi.ac.uk/ols/ontologies/eol
UBERON	https://www.ebi.ac.uk/ols/ontologies/uberon
CL	https://www.ebi.ac.uk/ols/ontologies/cl
BTO	https://www.ebi.ac.uk/ols/ontologies/bto
CLO	https://www.ebi.ac.uk/ols/ontologies/clo
SO	https://www.ebi.ac.uk/ols/ontologies/so
GO	https://www.ebi.ac.uk/ols/ontologies/go
NCIT	https://www.ebi.ac.uk/ols/ontologies/ncit
CHEBI	https://www.ebi.ac.uk/ols/ontologies/chebi

2.4. Increase data re-use (through clarifying licences)

GENE-SWitCH data will be publicly released in the EMBL-EBI archives at the earliest opportunity and pre-publication. This will be submitted to the archives without embargo so that it is immediately released to the public. This is in accordance with the FAANG data sharing principles (<https://www.faang.org/data-share-principle>), that is based upon the principles of the Toronto (<https://www.nature.com/articles/461168a>) and Fort Lauderdale (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>) agreements. This reserves the right for GENE-SWitCH to make the first publication with the data, whether a dataset has an associated publication is tracked clearly in the FAANG data portal (<https://data.faang.org/home>). All datasets will be clearly labelled with these data sharing principles, with the following statement:

"This study is part of the FAANG project, promoting rapid prepublication of data to support the research community. These data are released under Fort Lauderdale principles, as confirmed in the Toronto Statement (Toronto International Data Release Workshop. Birney et al. 2009. Pre-publication data sharing. Nature 461:168-170). Any use of this dataset must abide by the FAANG data sharing principles. Data producers reserve the right to make the first publication of a global analysis of this data. If you are unsure if you are allowed to publish on this dataset, please contact the FAANG Data Coordination Centre and FAANG Consortium (email faang-dcc@ebi.ac.uk and copy faang@iastate.edu) to enquire. The full guidelines can be found at <http://www.faang.org/data-share-principle>."

This enables the wider community to immediately make use of the data that GENE-SWitCH produces to provide maximal value to researchers. All software developed by the consortium will be openly licensed for reuse, an example of this is in the GENE-SWitCH RNA-Seq pipeline (<https://github.com/FAANG/proj-gs-rna-seq/blob/master/LICENSE>). In accordance with GENE-SWitCH coding standards, this license file will be displayed in the root folder of all repositories.



Data quality assurance processes and metrics will be investigated and implemented by work package 2 as part of the pipeline development process.

It is intended that through the accurate recording of metadata, associated protocols and analysis software, and deposition in public archives that the data will remain available for long after the project grant ends, for the lifetime of the underlying public archives. The data will therefore be reusable by any party, at some point the datasets may be superseded by those produced on newer technologies. There will be no restriction on third party use of the data. The data generation work packages will apply the latest recommended community standards for data quality, comply with any standards set by FAANG working groups or the public archives, and for the generation and execution of bioinformatics analysis will utilise the latest open source, published and recognised analysis software for the construction of its pipelines.

3 Allocation of resources

GENE-SWitCH directly funds the activity of the FAANG Data Coordination centre (DCC) to conduct data management and coordination for the project. The proposal has specific tasks and deliverables that will ensure the data generated in the project will conform to FAIR data principles. This in particular enhances the existing FAANG metadata standards, archival support tools, data portal discovery and data visualisations to improve findability, accessibility, interoperability and reusability of GENE-SWitCH data. These enhancements will also benefit the entire FAANG community as improvements will apply to all FAANG data. Thus the costs associated with ensuring GENE-SWitCH data is FAIR have been fully factored into the costs provided to EMBL-EBI in work package 3. All work packages that generate and analyse data have appropriate funding for the accurate recording and provision of metadata, through the validation and submissions software provided by the DCC. Data management is the responsibility of the FAANG Data Coordination Centre at EMBL-EBI that is operated by Peter Harrison and Guy Cochrane.

GENE-SWitCH will use the EMBL-EBI public archives for the long-term preservation of its generated data, these resources have separate long term funding that will persist the data long after the grant ends. The inclusion of the data within the FAANG consortium data portal (<https://data.faang.org/home>) and Ensembl browser (<https://www.ensembl.org/index.html>) also ensures the functional annotation of genomes will remain accessible by the community in the long term, as these are likely to continue to receive separate funding.

4 Data security

GENE-SWitCH will at the earliest opportunity submit all data to the public archives at EMBL-EBI. Intermediate results and ongoing analyses will be conducted and stored on the EMBL-EBI embassy cloud platform that is located in the same data centre as the public archives. Access to the GENE-SWitCH embassy cloud analysis platform is controlled by user specific Kubernetes tokens only issued to consortium members. As soon as an analysis is finished it will be submitted to the relevant EMBL-EBI archive for immediate public release without embargo.

The EMBL-EBI archives are internationally recognised repositories for the long-term secure storage of scientific data. All data will be assigned a unique identifier for long term identification and preservation of the datasets. The EMBL-EBI data centres that host the public archives providing the long-term data storage, and the embassy cloud platform for the analysis and intermediate processing of GENE-SWitCH data are state of the art. EMBL-EBI uses three discrete Tier III plus data centres in different geographical locations to ensure long-term security. Research data is also replicated through the International Nucleotide Sequence Database Collaboration (INSDC; <http://www.insdc.org/>) agreements that sees the data replicated at the National Center for Biotechnology Information (NCBI;



<https://www.ncbi.nlm.nih.gov/>) and DNA Databank of Japan (DDBJ; <https://www.ddbj.nig.ac.jp/index-e.html>) that agree to recognise each other centres accessioned datasets.

EMBL-EBI commits to store the data for the lifetime that the archives remain active, this will be far beyond when the GENE-SWitCH grant ends, ensuring this data remains available to the scientific community for years to come.

5 Ethical aspects

The proposed GENE-SWitCH data management plan complies fully with all international, EU and national legal and ethical requirements. GENE-SWitCH data sharing and long-term preservation is not subject to informed consent. GENE-SWitCH will fully comply with General Data Protection Regulations for its activities and web services.

6 Other issues

As well as complying with H2020 procedures for data management, the GENE-SWitCH project will abide by the data sharing policy of the Functional Annotation of Animal Genomes (FAANG) coordinated action (<https://www.faaang.org/data-share-principle>). This statement outlines the expectations of all FAANG projects that contribute to the coordinated action in terms of data recording, archiving and sharing. The statement includes the principles of the Toronto (<https://www.nature.com/articles/461168a>) and Fort Lauderdale (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>) agreements. The requirements set out in the FAANG data sharing principles do not conflict with those imposed by the EU H2020 data management principles.



5.2 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.

The FAANG analysis group will turn the raw data into primary and integrated analysis results. Primary analysis results consist of sample level analysis such as alignment to a reference genome or quantification of signal in the assay. Integrated analysis results represent analyses which drawn together data from multiple samples and/or experiments such as genome segmentation or differential analysis results.

The majority of these analysis results will not be archived before publication but FAANG recognizes the need to share them both within the consortium and with the community. Initially all files that are not archived will be shared between FAANG members in private shared storage hosted at the EMBL-EBI. Any individual who signs up to FAANG and agrees to [the Toronto principles](#)¹ will be allowed access to this. There will be metadata files in the private data sharing area, which make credit for different datasets as clear 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.

Internal means data that is only accessible via the FAANG private shared storage.

Private shared storage means a storage space hosted at EMBL-EBI that has password access via FTP, aspera and Globus Grid FTP technologies.

Public means all data available through the FAANG public FTP site, which has no password and is accessible to everyone.



FAANG expects to make multiple releases each year. A data release will involve declaring a data freeze and copying all files associated with that data freeze from the private shared storage to the public FTP site. In the first instance these data freezes will contain the primary analysis results. As FAANG's analyses progress, the data freeze will be expanded to include integrative analysis too. The data freeze process will be coordinated by the FAANG Data Co-ordination Centre and will be based on consultation with FAANG members. FAANG will also aim to release all data associated with a paper before publication even if it lies outside this standard freeze cycle. The public data will be available to the whole community.

All FAANG public data is released under [Fort Lauderdale principles](#)². The FAANG website, data portal and FTP site will all have clear data reuse statements on them.

When considering internal FAANG data, if one FAANG member wishes to publish using data generated by another FAANG member they should first contact the data generator and clarify the member's publication strategy. Collaboration is for everyone's benefit and is strongly encouraged. The FAANG Steering Committee commits to report to journal editors and the laboratories involved any event that disregards the rights of data creators (including biological measurements as well as analysis of such measurements).

All members of FAANG can and will continue to do experimental and analysis work outside of FAANG and the other data generated is not required to meet the same data sharing expectations.

Only FAANG data can be distributed through the private storage and public FTP site.

REFERENCES:

1. [Toronto International Data Release Workshop](#): Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.
2. [Fort Lauderdale principles](#): Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects.

(Approved by the FAANG Steering Committee on May 26, 2015)

6. Glossary

DMP – Data Management Plan.