

The FAANG Data Sharing Statement

Version 2.0. (Update approved by the FAANG steering committee on 1st December 2021, original approved on 26th May 2015)

This document describes the principles of data sharing for the FAANG consortium. Any queries about this document should be sent to faang@iastate.edu and faang-dcc@ebi.ac.uk.

FAANG believes that pre-publication data-sharing, collaboration and data reuse is for everyone's benefit and is strongly encouraged.

For FAANG data consumers:

FAANG data are released under the **Fort Lauderdale and Toronto principles**^{1,2}. FAANG data creators **reserve the right to first publication of the results obtained from using a dataset in genome wide analysis** (see box 1 for clarifying examples). The publications made on any dataset can be checked on the FAANG Data Portal (<https://data.faang.org/>). If you are unsure if you are allowed to publish on a dataset, please contact the FAANG Data Coordination Centre and FAANG consortium (email faang-dcc@ebi.ac.uk and cc_faang@iastate.edu) to enquire.

When using FAANG data you should **cite relevant publications and preprints from the data creators as well as all of the data accession numbers (e.g. PRJEB19199) in the main body of the publication** (not in the supplementary materials).

The FAANG consortium is producing high quality and well-annotated datasets to support the community in generating a powerful genome to phenome resource and promotes rapid dissemination of data to accelerate research. FAANG datasets are high quality, focus on a standardised set of multi-omic assays, are accompanied by rich validated metadata, phenotypic information and detailed protocols. FAANG participants provide these data pre-publication to encourage data reuse for maximal benefit to the community.

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).

Fostering collaboration through joint data analyses is also highly encouraged so you are invited to contact data creators directly (or via faang-dcc@ebi.ac.uk), or seek collaborative partners amongst the FAANG working groups and membership.

Box 1.

Examples of **permitted use**, that must include **citation of relevant publications or preprints from the**

data creators and the dataset accession numbers in the resulting manuscript:

Any researcher may download sequence data and/or derived bed files from the data portal, map these data to a genome and may derive results from these mapped data to address limited questions in their own research projects such as:

Is a specific set of genes expressed in a distinct tissue or set of tissues?

Is a locus, or pathway impacted by a particular histone mark?

Are particular SNV allele(s) present in the FAANG dataset?

What functional elements are present in a genomic region of interest for a particular trait?

Examples of prohibited use without prior publication from the data creators or permission from the author:

What is prohibited is the publication either on-line, or in the peer reviewed literature, of the results of a genome wide analysis of these data. Examples include but are not limited to:

Publishing on-line or in the peer reviewed literature a genome wide gene annotation file (gtf or bed) detailing transcription and isoform variation for the species' genome.

Publishing on-line or in the peer reviewed literature a genome wide survey of allele specific expression of transcripts and isoforms.

Publishing on-line or in the peer reviewed literature results derived from an integrated analysis of these data with other datasets for a genome wide study.

The above examples are not an exhaustive list, if in doubt, please contact the FAANG Data Coordination Centre and FAANG consortium (email faang-dcc@ebi.ac.uk and cc faang@iastate.edu).

For FAANG data producers:

FAANG recognizes that rapid sharing of the sample metadata and raw data generated by the consortium with the wider community is a priority. FAANG aims to ensure that everyone can benefit from the data created by FAANG to aid their own research as rapidly as is possible.

- All sample metadata and raw data produced for a FAANG associated project will be submitted to the public archives, without any hold until publication date, as soon as possible after sampling or data generation and initial quality control checks.
- All primary and integrated analysis results produced for a FAANG associated project are also encouraged to be made public prior to publication without embargo. However, it is acceptable that primary and integrated analysis results are kept private until publication, as long as the sample metadata and raw data have been made public.
- All FAANG public data are released under **Fort Lauderdale and Toronto principles**^{1,2}. The FAANG website, dataset descriptions and Data Portal have clear data reuse statements. The FAANG submission guidelines describe the suggested statement to

include with your dataset submissions (https://dcc-documentation.readthedocs.io/en/latest/experiment/ena_template/).

- The Data Portal has developed mechanisms to clearly identify which datasets are unpublished and which have at least one publication.

For FAANG primary and integrated analyses not made available in archives pre-publication, FAANG recognizes the need to enable and promote collaboration amongst consortium and community members. FAANG therefore provides functionality for primary and integrated analyses to be privately shared between FAANG members in private shared storage hosted at the EMBL-EBI. This requires an agreement between the two parties and that all have agreed to the **Fort Lauderdale and Toronto principles**^{1,2}.

Only FAANG data can be submitted to the FAANG Data Portal.

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.

Software and analysis pipelines developed by FAANG consortium members are strongly encouraged to be released under permissive open source software licenses wherever possible, such as Apache 2.0.

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).

REFERENCES:

1. **Fort Lauderdale principles:** Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects.
2. **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.

Version 2.0. (Update approved by the FAANG steering committee on 1st December 2021, original approved on 26th May 2015)

Definitions

Archive means one of the archives hosted at the EMBL-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.

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 draw together data from multiple samples and/or experiments such as genome segmentation or differential analysis results.

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 access limited to agreed persons by the data provider

Public means all data is available through the FAANG public data portal and underlying public archives, without embargo and is accessible to everyone.