FAANG Task Force Year1 Reports: metaFAIR

Authors: Peter Harrison (EMBL-EBI) and Annik Gmel (Agroscope)

20th November 2023

Membership list

1. Adetula, Adeyinka abiola
2. Brooks, Samantha
3. Chaudhari, Chirag
4. Giuffra, Elisabetta
5. Gmel, Annik †
6. Harrison, Peter †
7. Hunter, Christopher
8. Kalbfleisch, Ted
9. Sokolov, Alexey
10. Triant, Deb
11. Wang, Ying

Leadership list

1. Harrison Peter, EMBL-EBI, peter@ebi.ac.uk
2. Gmel, Annik, Agroscope, annik.gmel@agroscope.admin.ch

Task Force goals

Task Force priorities

The objective of the FAANG Metadata, Ontologies and FAIR Data Sharing Task Force is to ensure FAANG datasets are richly described, comply with FAIR (Findable, Accessible Interoperable and Reusable) data principles and are shared openly through the FAANG Data Policy. The task force recommends, develops and reviews new metadata rulesets, ontology improvements, data sharing methodologies, the FAANG Data Policy and data dissemination and presentation methods. It recommends best practice and promotes open data sharing amongst the FAANG community.

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.

Planned activities

- Organisation of periodic zoom calls to address Task Force key topics.
- Develop a whitepaper to state FAANGs position and promote to data providers the FAANG principles of sharing of FAANG data and phenotypes. Solutions for Nagoya or commercially sensitive datasets (preventing re-identification). Frame the problem to the community, even if we don't have the full solution at present.
- Prepare to lead roundtable/workshop discussion at PAG FAANG Friday event to help identify further gaps in requirements and to promote best practices from the task force.
- Promote FAANG principles of open data sharing to wider community.
- Identify training requirements for the FAANG consortia.
- Position task force against frame of AG2PI and AgBioData.

Summary of current activities

In this first year we established the metaFAIR task force regular meetings, that have a rotating responsibility between Peter and Annik to organise and chair. The group decided to initially aim to meet quarterly, which was felt the appropriate frequency to discuss arising larger metadata discussion points. It was decided that the group would switch to a more frequent meeting schedule as required, for example when the whitepaper writing is underway. The group met three times virtually and once in person at PAG 2023 (for those members that were present). This addresses our planned activity "Organisation of periodic zoom calls to address Task Force key topics."

The group has planned to deliver a talk as part of PAG 2024 by Peter Harrison from the EMBL-EBI Data Coordination Centre. A DCC training workshop has also been organised the day before the PAG FAANG workshop. This partly addresses our planned activity “Prepare to lead roundtable/workshop discussion at PAG FAANG Friday event to help identify further gaps in requirements and to promote best practices from the task force.” Further effort will be needed in 2024.

The EMBL-EBI DCC will deliver a workshop at PAG 2024 in San Diego, titled “FAANG: Hands on metadata validation and data submission training workshop“. The workshop will cover:

- How to prepare your FAANG sample, experimental and analysis metadata for submission.
- How to use the FAANG Validation tools to meet the FAANG standards.
- How to perform FAANG sample, experiment and analysis archive submissions.
- How to use the new FAANG ontology improvement tool.
- How to submit and visualise genomic track hubs.
- Using the FAANG Data Portal (https://data.faang.org/ <https://data.faang.org/>)
- Workshop will include Interactive hands on training.

The DCC also delivered two training courses on ontology improvement for agriculture and a summer school workshop on “Improving Cellular Model metadata descriptions for quality control, reproducibility and accelerating science”, as part of the EuroFAANG Initiative and projects. This partly addresses “Identify training requirements for the FAANG consortia”. Further effort will be needed in 2024.

Through Peter Harrison the group has started to interact with AgBioData to form a more coordinated view for metadata around both single cell and phenotyping. This partly addresses “Position task force against frame of AG2PI and AgBioData”, but more is needed in the next phase for example with AG2PI.

TF future activities planned for year 2.

In the second year the task force will deliver its planned training workshop and FAANG workshop talks at PAG 2024 in San Diego. EMBL is also running again its highly popular Livestock Genomics course https://www.ebi.ac.uk/training/events/livestock-genomics-0/

The group will continue to consider training requirements for FAANG, particularly those that can be delivered by the DCC, at other major conferences and also the possibility of virtual training videos and programmes.

The group plans to begin work on the whitepaper to state FAANGs position and promote to data providers the FAANG principles of sharing of FAANG data and phenotypes.

A particular focus is needed on phenotyping metadata, and there is huge potential to interact positively with crop communities through AgBioData, with a new working group established to tackle issues in this area, and also the emerging phenotyping Research Infrastructures through Elixir in Europe.

Single Cell Task Force 2022-2023 Annual report

Current Organizers:

Chris Tuggle	Iowa State University, Iowa cktuggle@iastate.edu
Christa Kuehn	Research Institute for Farm Animal Biology, Germany, kuehn@fbn-dummerstorf.de
Fernando Biase	Virginia Polytechnic Institute and State University, Virginia, fbiase@vt.edu
Yvonne Drechsler	Western University of Health Science, CA ydrechsler@westernu.edu
Christine Elsik	University of Missouri, Missouri, elsikc@missouri.edu
Peter Harrison	EMBL-EBI, United Kingdom, peter@ebi.ac.uk
James Koltes	Iowa State University, Ames, IA jekoltes@iastate.edu
Alex Clop	Centre for Research in Agricultural Genomics, Barcelona Spain, alex.clop@cragenomica.es
Richard Crooijmans	Wageningen University & Research, richard.crooijmans@wur.nl

Members (participated in initial organization, and are included in the current email list):

1. Acloque, Herve
2. Adetula, Adeyinka abiola
3. Archibald, Alan
4. Babatunde, Akeem
5. Becker, Doreen
6. Bhati, Meenu
7. Biase, Fernando
8. Chaudhari, Chirag
9. Cheng, Hans
10. Cheng, Hao
11. Clark, Emily
12. Clop, Alex
13. Crooijmans, Richard
14. Drechsler, Yvonne
15. Elsik, Christine
16. Fang, Lingzhao
17. Garrido, Juan j.
18. Giuffra, Elisabetta
19. Groenen, Martien
20. Guan, Dailu
21. Harrison, Peter
22. Hayes, Ben
23. He, Yanghua
24. Herrera uribe, Juber
25. Huang, Lusheng
26. Işık, Raziye
27. Kapoor, Muskan
28. Kerns, Karl
29. Kijas, James
30. Kim, Heebal
31. Koltes, James
32. Kuehn, Christa
33. Li, Congjun
34. Lien, Sigbjorn
35. Lindgren, Gabriella
36. Liu, George
37. Madsen, Ole
38. Menarim, Bruno
39. Murdoch, Brenda
40. Naboulsi, Rakan
41. Ng, Theros
42. Plastow, Graham
43. Psifidi, Androniki
44. Raine, Amanda
45. Reecy, James
46. Silverstein, Jeffrey
47. Tosser-klopp, Gwenola
48. Tuggle, Christopher
49. Vidal, Rodrigo
50. Wang, Ying
51. Watson, Mick
52. White, Stephen
53. Wiarda, Jayne
54. Zaldivar-lopez, Sara
55. Zhou, Huaijun

Activities

The FAANGsinglecell TF (SCTF) has met as a group via Zoom 4 times in 2022 and 2023:

March 30, 2022	February 6, 2023
May 16, 2022	April 3, 2023
Sept 15, 2022	June 5, 2023
Nov 17, 2022	November 28, 2023

Approximately 12-25 people participated in these calls. In early 2022, the calls focused on developing planned activities and providing the organizing text to the FAANG Steering Committee. In addition, 12-15 members of the SCTF group participated in the two face-to-face small group discussions at the PAG meeting in January 2023 and the ISAG meeting in July 2023. The priorities identified by Zoom discussions and the PAG/ISAG discussions is detailed below.

Speakers scFAANG seminar series (arranged by Gabriella Lindgren)

Fernando Biase, Virginia Tech "Single-cell transcriptome data produced with the SMART-seq framework" Monday, Sept 12, 2022

Amanda Raine, SciLifeLab, Uppsala University "scSPLAT, a method for single-cell DNA methylation profiling" Oct 10, 2022

Peter Harrison at EMBL-European Bioinformatics Institute (EMBL-EBI) "EMBL-EBI standardisation and archiving of single cell metadata and datasets. Current and Future". Feb 13, 2023

Alex Stretton, Parse Biosciences. "10x More Cost Effective than the Competition – How Parse Biosciences' Instrument-free single cell transcriptomics kits enable your projects to scale!" May 8, 2023

Ole Madsen, Wageningen University "Single cell sequencing at Wageningen University: Preliminary assessment of cell composition in intestinal organoids Jun 12, 2023

Hervé Acloque "Gene networks controlling functional cell interactions in the pig embryo revealed by scRNAseq and scOmics studies" Sep 11, 2023

Single cell TF priorities

The purpose of the groups will be identifying current projects and protocols and respective future needs, as well as further common interests and approaches with the overall goal of improving protocol sharing and standardization for both data creation, storage and analysis. Development of mechanisms and tools to exchange such information is needed to advance each component of single cell analysis.

1. **FAIR Data creation, storage and sharing**
    a. Sample acquisition and processing protocols
    b. Sample description/metadata schema
    c. Raw data (pre-analytical) data storage and sharing
    d. Cross-lab training in protocols
2. **Biological Analysis, Interpretation and Sharing**
    a. Raw data processing protocols

    b. Query-specific analysis tools- benchmarking/sharing
    c. Results visualization, storage and sharing
    d. Public data exploratory tools
    e. Experimental/biological validation of findings
    f. Cross-lab exchanges of methods and training in use of tools
    g. Cyber-infrastructure needs for extremely large/comparative projects

SCTF Priorities identified across the PAG/ISAG meetings.

PAG:

1. Infrastructure and funding to generate data and submit data/metadata to the FAANG data portal
    - especially to establish experimental data creation standards, as there are several competing technologies in sc analysis, in order to generate a cell atlas.
    - validate existing FAANG data portal infrastructure to submit and store sc data
2. Infrastructure to coordinate data visualization and meta-analysis across groups
3. Coordinate and organize opportunities for training in new wet-lab or computational training within the task force or within FAANG
    - priority high for workshops in data submission standards and protocols
    - EuroFAANG to do a training session. Use animal published data.
4. Identify methods and protocols members are working in

ISAG:

1. Collect community information
    - Tissues, species, objectives, protocols and technologies. These will need to be standardized.
2. Infrastructure for data sharing and data analysis
    - This will include established and newly developed tools as well as other related analytical information.
3. Training on bioinformatics and computational analysis
    - EuroFAANG to do a training session.
    - AgBioData Single Cell Working group is being formed to work on plant and animal data curation and standards.
4. Sharing data with EBI for single cell annotation in a standardized manner
5. Carry bi-monthly online talks
    - the contributors will be both, external experts on single cell protocols and data analysis as well FAANG’s single cell task force members to present their related research.

Future planned activities identified across all group members, both short and long term. Those to be addressed in 2024 priorities are labeled as such.

1. Organization of small groups focused on specific aspects of the two main topics (group on single cell data curation in line with AgBioData and the FAIR principles; group on cell annotation).

2. Identification of existing resources for data sharing and analysis from other single cell consortia. (ADDRESSED IN TOP PRIORITY LIST #4)
3. Establishment of spaces for collections of shared protocols and analytical pipelines.
4. Test of additional technologies in addition to 10x Genomics: PIPseq (Fluid BioSciences), Splitseq (Parse Biosciences), Singleron (All steps in one single plate. Richard C, WUR:), BD Biosciences (interrogates a custom set of genes. Chris T, ISU). Compare to 10x Genomics.
5. Identification of the most affordable technologies for single cell omics studies.
6. Publication of an article benchmarking existing technologies. (ADDRESSED IN TOP PRIORITY LIST #4)
7. Search for potential funding for a SC Atlas data portal. A component of this is expected to be collaboration with EBI in meta-analysis and data integration for automated annotation to integrate FAANG data into the EBI gene expression atlas. (ADDRESSED IN TOP PRIORITY LIST #3)
9. Share data with EBI for cell annotation.
10. Investigate approaches to translate, integrate, compare data from different platforms/protocols/technologies. (ADDRESSED IN TOP PRIORITY LIST #4)
11. Investigate how to deal with batch effects. (ADDRESSED IN TOP PRIORITY LIST #4)
12. Meet with the metaFAIR Task Force to discuss current data standards. (ADDRESSED IN TOP PRIORITY LIST #4)
13. Carry a survey to include all protocols used by the community (FAANG and others working on animals). (ADDRESSED IN TOP PRIORITY LIST #1)
14. Carry a similar survey to include the tools for data analysis and data structure used by the community. (ADDRESSED IN TOP PRIORITY LIST #4)
15. Identify possibilities and encourage additional data training:
    -EuroFAANG
    -FAANG meta-data workshop at PAG'24
16. TF communication.
17. Continue with the talks from external experts (e.g., Saket Choudhary) and FAANG members on single cell methodologies.

<u>Activities that have been identified as top priority (and will be undertaken in 2024):</u>

1. To conduct surveys of the community working on single-cell studies in farm animals, regarding the protocols and technologies being used.
    OUTCOME: a list of protocols and technologies that the community is using.
2.To carry a technology benchmarking to identify the most affordable – but still of high quality and satisfactory throughput – technology/assay for single cell studies:
    OUTCOME: a list of these technologies.

The results from these two activities will lead to the publication of an article benchmarking all these technologies in comparison to the 10X Genomics gold standard.

3.To develop a community resource for cell type annotation.
4. Identify the community members that are interested in sharing their single cell data and in collaborating with the aim to test existing pipelines and annotation tools to identify the optimal ones. This will be partly addressed in the surveys we are already conducting.