FAANGPrediction Task Force 2022-2023 Annual report

Current Organizers:
Hao Cheng (University of California Davis, USA), qtlcheng@ucdavis.edu
Daniela Lourenco (University of Georgia, USA), danilino@uga.edu
Mario Calus (Wageningen University & Research, Netherlands), mario.calus@wur.nl

Members:

1. Adebiyi, Marion
2. Adetula, Adeyinka abiola
3. Brito, Luiz
4. Bermann, Matias
5. Calus, Mario
6. Chaudhari, Chirag
7. Cheng, Hao
8. Duenk, Pascal
9. Estellé, Jordi
10. Guan, Dailu
11. Kadarmideen, Haja
12. Kizilaslan, Mehmet
13. Liu, George
14. Lourenco, Daniela
15. Qu, Jiayi
16. Ramon, Manuel
17. Regitano, Luciana
18. Vidal, Rodrigo
19. Wang, Ying
20. Zhao, Tianjing
21. Zhou, Huaijun

FAANGPrediction TF priorities

The main aim of the Task Force FAANGPrediction is to contribute to unlocking the potential of the generated FAANG data in animal breeding programs in practice. To achieve this, FAANGPrediction will facilitate exchange and collaboration between quantitative geneticists working on routine genomic evaluations and implementation of genomic prediction models that use functional genomics information such as SNP-level functional annotation and individual-level intermediate omics data (e.g., transcriptomics, metabolomics), and molecular geneticists and bioinformaticians involved in producing and interpreting the FAANG data.

- Bridging the gap between molecular and quantitative geneticists by introducing details of FAANG data and genomic evaluation to each other.

- Creating a data repository to convert FAANG data into formats friendly to quantitative geneticists.
- Developing new statistical and machine learning methods for FANNG-enabled genomic prediction.
- Evaluating its potential in routine genomic evaluation.

Current Activities

In the inaugural year of our initiative, we implemented regular meetings among the task force organizers to deliberate on forthcoming activities. TF priorities were discussed in small group in-person discussions at PAG and ISAG. A monthly online seminar series has been planned, designed to showcase ongoing progress and brainstorm the next steps. We have identified several key data resources and practical questions/problems, which are detailed in the table below. This initiative aligns with our planned objective of 'organizing workshops to bridge the gap between molecular and quantitative geneticists.' These workshops aim to foster mutual understanding by introducing the intricacies of FAANG data and genomic evaluation to both disciplines.

Priorities identified across the PAG/ISAG meetings.

- Data generation, infrastructure, funding to generate data, broadening scope by integrating expertise and training. => Often the phenotypic and genomic data that we work with is owned by private parties, and this data typically cannot be shared. A common (freely available) dataset could be very useful to compare and benchmark models.
- A major goal is to incorporate farm data as routine practice and to share data across industry, which will be imperative to data access.
- The question ‘Can a national effort such as phenotyping be set up?’
- The need to incorporate other areas of expertise complementing the know-how that is already present was identified. This new expertise included molecular geneticists and computer scientists. => Our experience is that to make optimal use of any functional annotation data in genomic or phenotypic prediction, we need to exchange with experts on interpreting this kind of data. In the context of FAANG, it means that we need active exchange with member of other Task Forces.

a. What collaborative projects are being done in your TF that interest **you** the most? Please provide a short title, and your name if you wish.	GENE-SwitCH (Mario Calus; co-leader of WP4) GEroNIMO (Mario Calus; leader of WP4) Luciana (Brazil): Collaborative Nellore (bos indicus); 400 animals with multiple omics layers (RNA-seq different thesis), WGS imputed, proteomics. Developed a database that will be released next year. Clara Diaz (Spain): beef & dairy cattle, several omics (e.g.+ behavioural traits, also physiological data related to heat stress (400). Smaller: Sequence + transcriptomics diff. tissues. Maria Saura: tarbot (n=800) genotypes + phenotypes + microbiome data (e.g. the skin). Sheep (n=400), genotypes + microbiome, phenotypes. Sebastian (Spain): large database of phenotypes (100k), genotypes (5k). Start to integrate omics to improve predictions. (PhD student): Transcriptome of native breeds of cattle, WGS, pathogen presence Nick Kertz (US): predicting heifer fertility in beef cattle, using transcriptomic data. Luis Varona (Zaragoza): Legarra & Christensen; developing methods to integrate omics data. 300-400 animals with microbiota available. Jordi Estelle (INRAE): host-genetics – microbiote inter; national project funded on modelling this better. Interested to exchange, on sampling and method development. Isidore Houga (Roslin): developing GP for crossbred dairy cattle in Africa (n=2000). No functional genomics data yet. Norman Marwashe: focussing using genomic data in national (beef) cattle evaluations. Large scale national genotyping (& phenotyping for a long time). Just a few small-scale experiments in gene expression. Some collaboration with Brazil & South America on tolerance to ticks, wants to use expression data here. Siriluk Ponsuksili: colleagues already involved in FAANG. Working with omic from microbiome, some data in pigs (embryos?) + some metabolome data. Also some data in chicken. How to use these omics data to integrate in breeding (values)? <u>Summary:</u> • EuroFAANG projects that have an explicit WP on genomic prediction (https://www.geronimo-h2020.eu/ & https://www.gene-switch.eu/) • Several additional national projects were mentioned, with individual omics data of up to 400 individual animals.

b. What initiatives should our task force undertake to advance that project? What type of effort will this initiative require (grant writing, exchange of materials, organize Zoom calls, etc)?	(Brazil) Main restriction is people available; who is going to do the modelling? Both more people and more skills are needed (those skilled are already occupied). Workshops to train scientists & post-graduate students => building scientists. Developing proposals for funding: networking & discussion to identify possibilities. E.g. Cost networks in EU; could we have one on prediction? OECDE (Spanish funding instrument) also provides funds for networking. Combine datasets across country. General challenge: inclusivity. Commercial breeds are prioritized at the expense of local breeds. Clara: are doing simulation to decide on sampling, in the most informative and useful way, with implications for a breeding program (and to inform experimental set-ups). Local beef cattle breed, huge variability in response to stress (dairy as well). Standardization & harmonization? How to combine expression measured at different time-points? May not always be possible? Example: metabolomics between plasma & serum are very different. Are eQTL in one population good predictors in another breed? Brazil has datasets that may be useful for this. How to combine different datasets? Functional annotation approach could solve this, t Sharing experience of sampling, data generation. Plant people may have shared questions (@ INRAE); something to consider. What kind of omics (and what approach) could be most useful and most feasible? Perhaps microbiome is most robust? Do we reach a plateau across omics layers? Can we accumulate this across different experiments? Get expert opinions on what would be a good omics to start with. Generate meta-data from available data, and share across groups, to identify options to combine/evaluate (this could even include e.g. levels of LD). Do an inventory of available data? Across-species inventory could be very informative. <u>Summary:</u> 1) There are several relatively small datasets on individual omics being identified. <u>Suggested initiative:</u> Inventory of available data (Check with FarmGTEX to avoid & see how to collaborate.)

	2) Lack of (availability of) skilled people to undertake the required analyses. Suggested initiative: Organize a Summer School. 3) Build more of a formal network to collaborate and exchange.
c. Who is willing to lead or organize this initiative? Please indicate names and activities volunteered.	1) Inventory of available data (Check with FarmGTEX to avoid & see how to collaborate.) => To be led by Task force leaders (Hao, Daniela & Mario). 2) Summer school? => Will be organized by the GEroNIMO project. 3) Cost Action => Should be a number of (young) people from Europe? UK can participate. Probably needs to be led by an established group in this area.

Future planned activities

- Initiate collaborative research projects between molecular and quantitative geneticists. These projects could utilize FAANG data in innovative ways, fostering inter-disciplinary collaboration.

- Dedicate a portion of the regular meeting to exploring and evaluating potential new data sources. Conduct specific workshops or online sessions focused on the evaluation of new data sources. These can cover aspects such as data reliability, applicability to current research needs. Through a funded project, organizer Hao Cheng is developing an open database for publicly available AG2P data sets, analysis, and educational resources (e.g., a software package, a GitHub repository). The platform will host metadata linked to each data set and analysis resource for at least 100 public AG2P data sets and a similar number of data-analysis resources (currently focusing on genotypes and phenotypes).

- Organize a series of advanced workshops that delve deeper into specific aspects of FAANG data and genomic evaluation. These workshops can be more technical and focused.