From FAANG to Fork: Application of Highly Annotated Genomes to Improve Farmed Animal Production

Emily L. Clark[1*], Alan L. Archibald[1], Hans D. Daetwyler[2,3], Martien A. M. Groenen[4], Peter W. Harrison[5], Ross D. Houston[1], Christa Kühn[6,7], Sigbjørn Lien[8], Daniel J. Macqueen[1], James M. Reecy[9], Diego Robledo[1], Mick Watson[1], Christopher K. Tuggle[9] and Elisabetta Giuffra[10]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG, UK
[2]Agriculture Victoria, AgriBio Centre for AgriBioscience, Bundoora 3083, Victoria, Australia
[3]School of Applied Systems Biology, La Trobe University, Bundoora 3083, Victoria, Australia
[4]Animal Breeding and Genomics Centre, Wageningen University and Research, 6708 PB, Wageningen, The Netherlands
[5]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD UK
[6]Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, Genome Physiology Unit, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany
[7]University Rostock, Faculty of Agricultural and Environmental Sciences, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
[8]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, NO-1432, Norway
[9]Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA
[10]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France

Corresponding author for preprint email address: emily.clark@roslin.ed.ac.uk

Abstract

Here we describe a set of research priorities to meet present and future challenges posed to farmed animal production that build on progress, successes and resources from the Functional Annotation of ANimal Genomes (FAANG) project.

Keywords: FAANG, genomics, farmed animal, genotype-to-phenotype

Introduction

The Food and Agriculture Organisation of the United Nations (FAO) reports that by the year 2050 the global human population is likely to reach 9.1 billion, rising to 11.2 billion by the year 2100. This population growth poses several challenges to the global food system, which will need to produce more healthy food using fewer natural resources, reducing the environmental impact, conserving biodiversity and flexibly adjusting to changing societal expectations. Meeting this demand requires environmentally sustainable improvements to farmed animal health and welfare, and of efficiency and diversification (e.g. to include a broader range of locally adapted species) [1]. The changes in breeding strategies and management practises required to meet these goals will need to build on an improved ability to accurately use genotype to predict phenotype in the world's farmed animal species, both terrestrial and aquatic (Box 1).

Box One: Addressing the Challenges of Global Food Production in the 21st Century

Addressing the Challenges of Global Food Production in the 21st Century

- Sustainable improvements in efficient production of farmed animals, to provide healthy food, are needed to address a growing world population confronted with environmental and social challenges.
- Chief among the improvements required is the ability to more accurately use genotype to predict phenotype (G2P) in specific environments.
- Advances in the analysis of genome function promises to provide tools and knowledge for the genetic improvement of farmed animals and conservation of genomic biodiversity.
- The FAANG consortium is working to discover basic functional genomic knowledge underpinning the G2P question across major terrestrial and aquatic farmed animal species.
- In the coming decade, FAANG will associate large-scale genotypes and phenotypes in pedigreed animal populations to predict G2P relationships.
- The FAANG$_{\text{SingleCell}}$ and FAANG$_{\text{GTEx}}$ projects will apply exciting new technologies to model cells and tissues and verify which sequence variants affect phenotype.
- Public investment is needed to expand G2P research and training of new scientists to accelerate genetic improvement of diverse farmed animal populations to feed the global consumer.

Here we describe a set of research priorities to meet such present and future challenges that build on progress, successes and resources from the Functional Annotation of ANimal Genomes (FAANG) project [2]. The first stages of FAANG focused on foundational data generation to characterise expressed and regulatory genomic regions, curation and provision of annotated farmed animal genomes [2,3,4]. These were largely based on individual level, high depth approaches [4]. The primary challenge facing this community now is harnessing these resources to link genotype, phenotype and genetic merit in order to translate this research out of the laboratory and into industry application in the field. To achieve this effectively, we will need to generate functional genomic information for large populations of animals, rather than relying on a small number of deeply annotated individuals. Furthermore, to date, most of the datasets are from tissues consisting of heterogeneous cell populations, hindering the resolution of functional information and limiting our ability to understand the fundamental cellular and subcellular processes underlying phenotypes. Since the original FAANG white paper was published in 2015 [2], exciting new opportunities have arisen to tackle these challenges. We describe a set of research action priorities for FAANG for the next decade (Figure 1), in each of the sections below.

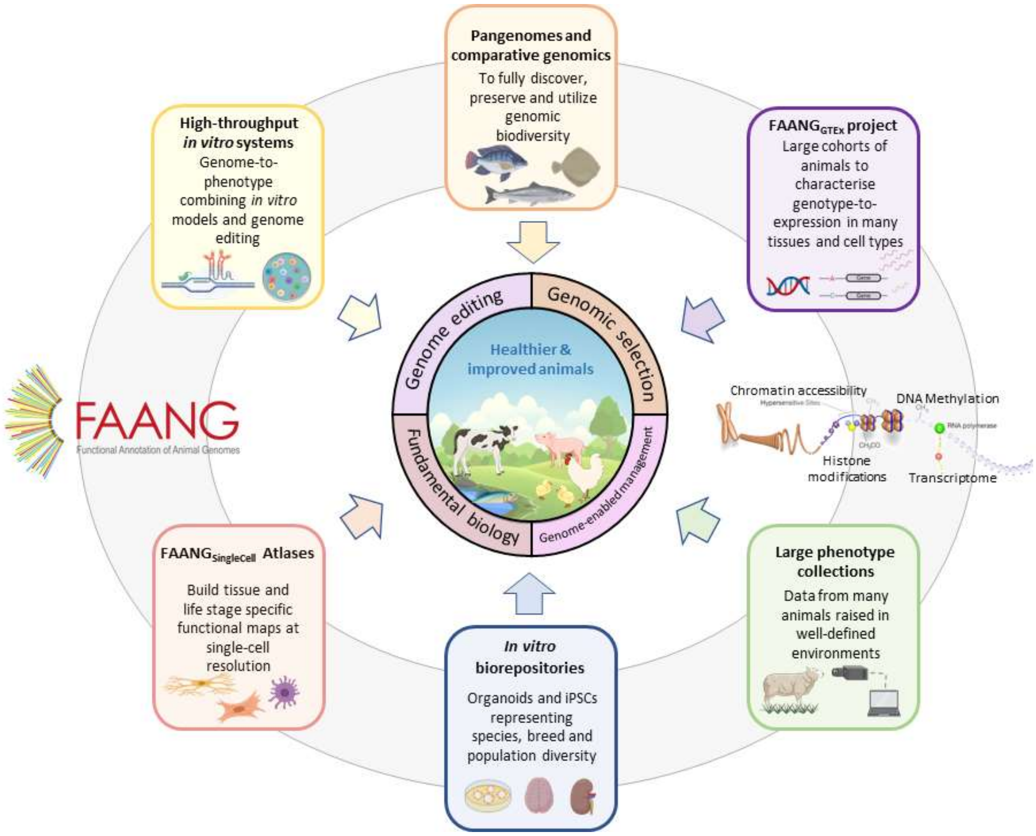

Figure 1: Research priorities for the next decade of FAANG research.

Omics empowered genomic selection

In the past twenty years genomic selection has substantially increased genetic gain in some farmed animal species through the use of large training populations [5]. However, prediction accuracy in genetically distant populations (i.e. across populations, breeds, and generations) remains limited due in part to the current reliance on neutral markers in incomplete linkage disequilibrium with causative genetic variants in the breeding population of interest [5]. Using variants more tightly linked to causative variants and supported by genomic information in a multi-breed training population can partially alleviate these limitations [6]. Large-scale whole-genome resequencing has produced inventories of many millions of variants for thousands of animals [7]. In such sequence datasets, the causative variants are directly genotyped among millions of neutral markers. This reduces the signal-to-noise ratio when all the data are used for genomic prediction without prior biological information. Efforts to detect causative variants have been successful for variants with large phenotypic effects, often deleterious, using a combination of quantitative, population and molecular genetics [5]. However, economically important traits have a polygenic architecture and causative variants are expected to have small effects, which makes their detection and quantification difficult. Most of these causal variants, with small effects, are likely to be located in regulatory sequences and impact complex traits through changes in gene expression [5]. Thus, it is expected that improvements in prediction accuracy can be achieved by filtering the genetic marker information based upon whether the genetic variants reside in functional sequences and developing robust prediction models that can accommodate the biological priors. As functional (expressed and regulatory) genomic elements are not easy to predict from sequence alone, FAANG will enhance current genome annotation with functional information from a range of relevant tissues, cell types and developmental stages. Recently, novel methods for the integration of biological information (e.g. methylation of regions of predicted functionality) into genomic prediction have been proposed e.g. [6]. These models, which are based on the combination and ranking of many diverse datasets from multiple animals, could facilitate further improvements in predicting genetic merit and consequently on genomic selection, as has been demonstrated in cattle [6]. As many more suitable datasets will become available in the next five years, improving and adapting these methods to enhance genomic prediction accuracy, whilst conserving genetic diversity, across farmed animal species will be a priority for FAANG.

FAANG$_{\mathrm{GTEx}}$ - Linking genetic variation to genome function

The first phase of FAANG is using a specific set of transcriptomic and epigenomic assays to define functional regions of the genome in tissues [2,3]. Due to the significant investment per sample, this phase was limited to only a few individuals and ascribed function was averaged across these replicates [2]. Progress has been made in defining functional regions, and this should be built upon to ascertain the effect of genetic variation on genome function [4]. Collecting functional genomic data across many genetically diverse animals lends itself to the application of statistical genomics to detect quantitative trait loci (QTL) controlling molecular phenotypes. This is particularly powerful when done at sequence level resolution to directly relate molecular phenotypes (e.g. gene expression or methylation information) to variants associated with complex traits. The GTEx consortium has achieved this very effectively across human tissues, enabling expression QTL (eQTL) studies linking gene expression to genetic variation [8], and providing a framework for FAANG to develop a similar project for farmed animals (FAANG$_{\mathrm{GTEx}}$). Large farmed animal cohorts in controlled and well characterised environments with extensive pedigree information and molecular phenotypes would allow researchers, in partnership with industry, to: 1) build better predictive models of genotype-to-phenotype, 2) better understand genotype-by-environment interactions, and 3) prioritise functional variants for inclusion in breeding programmes [5]. Hundreds of thousands of farmed animals currently have imputed genotypes and extended pedigrees with deep phenotypic records [7]. A project analysing the relationship between GWAS SNPs and gene expression for cattle, mining publicly available sequence data, was published earlier this year, demonstrating the feasibility, timeliness and potential of a GTEx approach for farmed animals [9].

Beyond genomic selection: towards genome-enabled management

Beyond its use in genomic prediction [5], the functional data produced by FAANG will provide new perspectives for informed management decisions. Epigenetic and expression information for individual animals could be combined with microbiome data and high-throughput phenotypes from new management technologies (e.g. wearables, GPS, *in-vivo* imaging systems) [10]. These datasets from large cohorts of animals would enhance prediction of adaptive capacity at the individual level through integration of prior environmental data with individual genome information. Thus, providing new opportunities for informed animal management decisions during an animal's lifetime (e.g. to optimize diets or for steering animals into the most appropriate production systems). A genome enabled management approach (providing animals, within a production system, with their specific needs during their lifetime) will be beneficial to improving animal health and welfare, facilitate adaptation to changing environments and contribute to addressing public concerns related to animal production. Achieving this within the next ten years may be possible, but the challenge will be to ensure it is practical and affordable for animal breeders and producers.

Understanding and conserving genomic diversity – the power of pangenomes

Through large-scale sequencing efforts by the farmed animal genomics community data are now accumulating that characterise the sequence diversity of farmed animals including locally adapted breeds/populations. As a consequence, future genetic management is likely to include the use of pangenomes, that will capture all available population-level genomic information for a given farmed animal species. Using graph-based frameworks, we can more accurately genotype and annotate the genomic diversity present in any given individual [11]. In this way, pangenomes can reveal population- or breed- specific adaptations that could be used to tailor the genotypes chosen in future farming systems in order to conserve biodiversity whilst improving production efficiency and animal health [1]. Furthermore, the highly annotated genomes produced by FAANG allows evolutionary conservation across species to be defined for all genomic features [3,12]. Ongoing FAANG projects involve comparative analysis which will reveal the functional basis of phenotypes present in one species that are desirable in others. Such projects contribute to addressing the major opportunity that exists to enhance the sustainable production of a wider diversity of animal species, including numerous and diverse aquaculture species that are poised to exploit functional genomics to expedite genetic improvement, where tailored and cost-efficient approaches will be required [13]. Current FAANG-related projects already extend to several major farmed finfish species in Europe and North America. We envisage an increased representation of aquatic species, including shellfish, and further expansion to include invertebrates, within FAANG projects during the next five to ten years.

FAANG$_{\text{SingleCell}}$ – deconvoluting transcriptional and regulatory complexity

The use of bulk tissue samples in the FAANG studies performed to date captures regulatory elements and expression signals averaged across all represented cell types but fails to reveal the cell-specific basis of the molecular phenotypes of interest. In order to more accurately link genotype to phenotype, data at the level of individual cell types are required. Single-cell sequencing technologies enable the deconvolution of the transcriptional and regulatory complexity in tissues made up of multiple cell types. New technologies to detect gene expression as well as chromatin accessibility, structure, and interactions within single cells provide more comprehensive data to predict function and interaction partners for regulatory elements. As a consequence, one of the main priorities for FAANG within the next five to ten years is to create single-cell atlases for the key tissues of farmed animal species

($\text{FAANG}_{\text{SingleCell}}$). The organisational processes, standardisation and data sharing infrastructure established by the community for the first stages of FAANG [4] will provide a strong foundation for $\text{FAANG}_{\text{SingleCell}}$ to progress quickly and efficiently. The $\text{FAANG}_{\text{SingleCell}}$ project should build on existing functional tissue maps for other species e.g. [14] and will enable the identification of genomic variants underpinning trait-linked cell types/factors and causal variants [13]. In the $\text{FAANG}_{\text{GTEx}}$ project described above, single-cell atlases will provide a powerful layer of resolution including cell-specific molecular phenotypes, enabling the fine-scale dissection of complex traits of interest.

In vitro systems - bridging the gaps between cell, tissue and whole animal scale knowledge

Single-cell sequencing technologies can also be used to deeply characterise cell and tissue complexity of *in vitro* systems such as organoids. Over the last five years organoids for many different organ systems and for multiple farmed animal species have been developed [4]. Organoids provide *ex vivo/in vitro* systems for testing candidate causal variants by genome editing technologies, and potentially a system for high-throughput, cost-effective, large-scale *in vitro* phenotyping. Importantly, given the ease of biobanking, organoids have a strong ethical benefit in reducing the number of animals used in experimentation [4]. Multiple organoid models can be derived from very small quantities of tissue or from induced pluripotent stem cells (iPSCs). They provide the potential to generate and test multiple phenotypes to unravel when, and under what conditions, a putative causal variant has an effect. Therefore, farm animal organoids will be valuable over the coming decade, providing information about fundamental biology to model the effects of changing environmental conditions and supporting immunology, vaccinology, physiology, nutritional, and biodiversity conservation studies. The ability to decompose complex phenotypes into key processes will provide a means to robustly relate the deep phenotypes measured in these systems with the traits used for selection, opening to the possibility of using organoids for breeding purposes.

Genome editing - a route to application for FAANG data

The application of genome editing to farmed animals is advancing rapidly, mainly due to development of CRISPR/Cas technologies [13, 15]. The CRISPR toolbox has expanded to improve precision, allow modulation of gene expression and epigenetic modifications, and now forms an integral part of the future FAANG roadmap [4]. CRISPR-mediated modification of putative genomic elements can confirm their functionality and reveal their roles in cellular (and organoid) function. Genome-wide multiplexed CRISPR approaches now enable the simultaneous interrogation of thousands of genomic features in cell lines, increasing the feasibility of this approach for genome-scale annotation [16]. These high-throughput approaches can also be used in combination with single-cell sequencing technologies to obtain high-resolution molecular phenotypes. In addition, genome editing represents a potential major route for the application of FAANG research in farmed animal breeding programmes via: 1) detection and utilisation of causative variants affecting important traits, 2) targeted introgression, or ‘introgression-by-editing’, of favourable alleles from other strains or species into a closed breeding population, or 3) creation of *de novo* alleles with favourable effects, either predicted from unbiased genome-wide screens or from *a priori* knowledge of the biology of the trait in question. Public perception and regulatory hurdles remain and ongoing discussion through stakeholder engagement must continue and evolve to keep pace with technological advances. While the use of genome editing for the improvement of farmed animals may currently only be possible in some countries, its use in *in vitro* models, such as organoids, is not subject to the same legislation and ethical considerations as the use of whole animals and thus represents a new frontier for FAANG research.

Data recording, computation and integration to support the emerging objectives of FAANG

As a scientific community FAANG continues to develop a coordinated analysis and data collection infrastructure crucial for its success [4]. The FAANG bioinformatics community, including the centralised Data Coordination Centre (DCC), is focused on open reproducible science, the FAANG data portal [17] as the focal point for this activity. Technological development, coordination and standardisation by the DCC will continue to be crucial for the shift towards population scale studies, single cell datasets, cell atlases and pangenomes, across a growing number of species. This will require new reproducible analysis pipelines and infrastructure, metadata validation services, data portal features such as a centralised atlas browser and online training resources. Single cell atlases and *in vitro* systems for farmed animal species will be accompanied by high quality metadata, archiving and visualisations across species, organ systems, tissues and cell types. As FAANG datasets continue to increase in complexity, there is a growing need for new methods of data visualisation and integration to be made available. These future developments, and the distributed data and analysis infrastructure, will be crucial for the successful application of functional data to farmed animal breeding programmes.

Priorities for the future of FAANG

The research priorities we have outlined for FAANG for the coming decade are depicted in Figure 1. The uptake by the farmed animal production industry and the expected outcomes of each prioritised action are summarised in Figure 2. FAANG will improve our ability to more accurately use genotype to predict phenotype. This will directly contribute to addressing the challenges faced for sustainable and responsible global food production in the next decade (Box 1). However, whilst the molecular assays used to enable functional annotation can now be delivered at much lower cost, the costs for the research priorities outlined above remain substantial, especially considering the rapid increase in number and diversity of target species in the aquaculture sector. As such a strong commitment to invest in research is needed. Persuading the European Commission to include FAANG projects in Horizon 2020 was a major success for the first stage of FAANG and its leadership. Current funding for FAANG supports the research community to improve the functional annotation of key farmed animal species and to facilitate more refined genomics-enabled animal breeding / genetic improvement. The research priorities outlined here are already strategically aligned to the objectives of the European Green Deal and current USDA National Institute for Food and Agriculture programmes, and international cooperation will be essential to secure funding for their achievement. Given the scale and cost of the research involved it will likely be necessary to initially prioritise the development of *in vitro* systems and the enhancement of data infrastructure to provide a solid foundation for $\text{FAANG}_{\text{SingleCell}}$ and $\text{FAANG}_{\text{GTEx}}$.

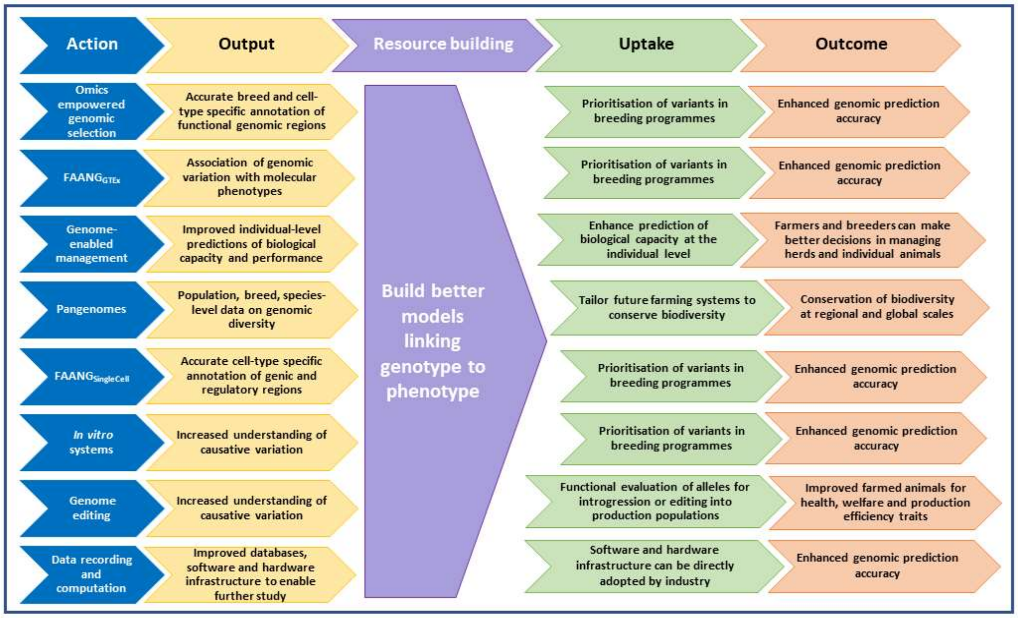

Figure 2: How implementation of FAANG research priorities over the next decade will benefit farmed animal production.

The timely achievement of all of the research priorities we outline here for the next stages of FAANG will together increase the capacity of the farmed animal production industry to face the challenges of the future, empowering genomic selection, enhancing adaptation to changing environments, conserving biodiversity, and bridging the knowledge gaps between cellular and whole animal scale knowledge.

Authors' Contributions

All co-authors contributed to the writing of the manuscript. ELC assembled and curated the manuscript in collaboration with the co-authors, with joint editing by MW and FAANG co-coordinators (CKT and EG).

Acknowledgements

The work presented in this manuscript was initiated at discussions held at three scientific meetings held in early 2020. The authors would like to thank the FAANG Scientific Advisory Board for useful advice and the following people for providing comments on earlier drafts of this manuscript: Amanda J. Chamberlain (Agriculture Victoria; La Trobe University), Appolinaire Djikeng (Centre for Tropical Livestock Genetics and Health, University of Edinburgh), Denis J. Headon (The Roslin Institute, University of Edinburgh), Ian Johnston (Xelect Ltd), Andreas Kranis (Aviagen; The Roslin Institute, University of Edinburgh), Michèle Tixier-Boichard (Université Paris Saclay, INRAE), Stephen N. White (USDA-ARS; Washington State University), Ruidong Xiang (University of Melbourne), Daniel Zerbino (EMBL-EBI).

Funding

In the EU, FAANG has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817923 (AQUA-FAANG), 817998 (GENE-SWitCH), and 815668 (BovReg) and from European COST Action CA15112: Functional Annotation of Animal Genomes - European network, FAANG- Europe. At EMBL-EBI the work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/N019563/1 and BB/N019202/1), Wellcome (WT108749/Z/15/Z) and the European Molecular Biology Laboratory. At the Roslin Institute the work was supported through BBSRC institute strategic programme grants (BB/P013732/1 and BB/P013759/1). In the US, FAANG has received funding from the US National Science Foundation (IOS-1548275) and the US Department of Agriculture (2015-68004-24104, 2018-67015-27501). In Australia, FAANG has received funding from the DairyBio project (a joint venture between Agriculture Victoria, Dairy Australia, and the Gardiner Foundation) and through the University of Queensland.

Competing Interests

The authors declare that they have no competing interests.

References

1. Rexroad C, Vallet J, Matukumalli LK, Reecy J, Bickhart D, Blackburn H, et al. Genome to Phenome: Improving Animal Health, Production, and Well-Being – A New USDA Blueprint for Animal Genome Research 2018–2027. Front. Genet. 2019;10:327. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.00327

2. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated

international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015;16:57. Available from: https://doi.org/10.1186/s13059-015-0622-4

3. Macqueen DJ, Primmer CR, Houston RD, Nowak BF, Bernatchez L, Bergseth S, et al. Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. BMC Genomics. 2017;18:484. Available from: https://doi.org/10.1186/s12864-017-3862-8

4. Giuffra E, Tuggle CK. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Annu Rev Anim Biosci. 2019;7:65. Available from: https://doi.org/10.1146/annurev-animal-020518-114913

5. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. Nat Rev Genet. 2019;20:135. Available from: https://doi.org/10.1038/s41576-018-0082-2

6. Xiang R, Berg I van den, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. Proc Natl Acad Sci. 2019;116:19398. Available from: http://www.pnas.org/content/116/39/19398

7. Hayes BJ, Daetwyler HD. 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. Annu Rev Anim Biosci . Annual Reviews; 2019;7:89–102. Available from: https://doi.org/10.1146/annurev-animal-020518-115024

8. Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. Science. 2020;369:eaaz5900. Available from: https://science.sciencemag.org/content/369/6509/eaaz5900

9. Fang L, Cai W, Liu S, Canela-Xandri O, Gao Y, Jiang J, et al. Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle.

Genome Res. 2020;5:790. Available from: https://genome.cshlp.org/content/30/5/790.abstract

10. Koltes JE, Cole JB, Clemmens R, Dilger RN, Kramer LM, Lunney JK, et al. A Vision for Development and Utilization of High-Throughput Phenotyping and Big Data Analytics in Livestock. Front. Genet. 2019;10:1197. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.01197

11. Crysnanto D, Pausch H. Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. Genome Biol. 2020;21:184. Available from: https://doi.org/10.1186/s13059-020-02105-0

12. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol. 2019;17:108. Available from: https://doi.org/10.1186/s12915-019-0726-5

13. Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020;21:389. Available from: https://doi.org/10.1038/s41576-020-0227-y

14. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature. 2018;562:367. Available from: https://doi.org/10.1038/s41586-018-0590-4

15. Tait-Burkard C, Doeschl-Wilson A, McGrew MJ, Archibald AL, Sang HM, Houston RD, et al. Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. Genome Biol. 2018;19:204. Available from: https://doi.org/10.1186/s13059-018-1583-1

16. Borys SM, Younger ST. Identification of functional regulatory elements in the human genome using pooled CRISPR screens. BMC Genomics. 2020;21:107. Available from: https://doi.org/10.1186/s12864-020-6497-0

17. The FAANG Data Portal. https://data.faang.org/home. Accessed 18 September 2020.