ANNUAL REVIEWS

Annual Review of Animal Biosciences

Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap

Elisabetta Giuffra,[1] Christopher K. Tuggle,[2] and the FAANG Consortium*

[1]Génétique Animale et Biologie Intégrative (GABI), Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris Saclay, 78350 Jouy-en-Josas, France; email: elisabetta.giuffra@inra.fr

[2]Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA; email: cktuggle@iastate.edu

ANNUAL REVIEWS **CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

domesticated animals, epigenetics, functional genomics, genotype-to-phenotype, genomic selection, genome editing

Abstract

Functional annotation of genomes is a prerequisite for contemporary basic and applied genomic research, yet farmed animal genomics is deficient in such annotation. To address this, the FAANG (Functional Annotation of Animal Genomes) Consortium is producing genome-wide data sets on RNA expression, DNA methylation, and chromatin modification, as well as chromatin accessibility and interactions. In addition to informing our understanding of genome function, including comparative approaches to elucidate constrained sequence or epigenetic elements, these annotation maps will improve the precision and sensitivity of genomic selection for animal improvement. A scientific community–driven effort has already created a coordinated data collection and analysis enterprise crucial for the success of this global effort. Although it is early in this continuing process, functional data have already been produced and application to genetic improvement reported. The functional annotation delivered by the FAANG initiative will add value and utility to the greatly improved genome sequences being established for domesticated animal species.

WHAT IS FAANG? TOOLS FOR PREDICTING BIOLOGICAL OUTCOMES

A major goal of modern biology is to move from a descriptive science toward predicting biological outcomes at the pathway, cell, tissue, or even organismal level. One paradigm for such prediction is the systems biology approach, which attempts to measure multiple levels of biological components at several states to develop a quantitative model that can predict the effect of perturbation on the system (1–3). This approach can greatly increase our biological understanding, yet it requires substantial investment in collection of genome component information and function for modeling. Systems biology also links data-driven or inductive and hypothesis-led approaches in a virtuous cycle to advance knowledge (4) and places biology on a par with physics as big data science. Another paradigm is predicting organism-level phenotype through linking to variation present in large numbers of different genomes. Genomic prediction is a technology that uses genome-wide sets of DNA variants to estimate the genetic merit of farmed animals and crop plants for target traits (5). With training through associating whole-genome genotypes with phenotypes for thousands of individuals, genomic prediction has sufficient accuracy to be effective in genomic selection to improve quantitative traits in domesticated animals (6–8). However, the reliance on imputation of associated genomic regions decreases prediction accuracy when applied to distant populations with different linkage disequilibrium structures (9). The use of functional information to reduce the search space for causative variation will improve cross-population accuracy (10). In both paradigms, understanding the functional components of the genome under study is required to accurately predict future outcomes.

Functional components of genomes are identified and annotated by molecular assays targeting RNA expression and structure, as well as chromatin accessibility, modifications, and interactions (**Figure 1**). Put simply, such annotation involves finding out which regions of the genome are expressed; when and to what level they are expressed; whether they are alternately spliced and whether they code for proteins or are noncoding; where the transcription start sites (TSSs) are and their usage; which regions of the genome have regulatory functions; and what the 3D structure of the genome is and how this influences expression and regulation. Enormous steps forward have been made in the past 10 years by ENCODE (Encyclopedia of DNA Elements) (**https://www.encodeproject.org/**) and several subsequent projects (i.e., **http://ihec-epigenomes.org/**) in human and model organisms, largely driven by the evidence that most trait-associated loci, including ones that contribute to diseases and susceptibility, lie outside protein-coding regions (reviewed in 11). The insights these studies produce have pointed to the value of effective models to integrate diverse 'omics information to predict phenotypic traits and outcomes, elucidating biomarkers and generating insights into the genetic architecture of complex traits (12, 13), and have facilitated personalized diagnostics, disease management, and biomarker discovery (14).

However, the availability of such open-source data sets has limited utility for animals other than humans and mice, because understanding the phenotypes of interest in the authentic biological context requires organism-specific information. Although sequence similarity across species is a strong indication of genomic functionality (15), several comparative genomics studies have shown that conservation is low at the regulatory sequence level. For example, analysis of two liver-specific transcription factor (TF) sites found that interspecies differences in DNA binding by TFs could be explained by sequence changes to the bound motif (16). This group also showed that TF binding highly conserved across species could not predict regulation of the gene near the conserved site (16). A genome-wide study from the mouse ENCODE consortium has found a large degree of divergence of sequences involved in transcriptional regulation, chromatin state, and higher-order
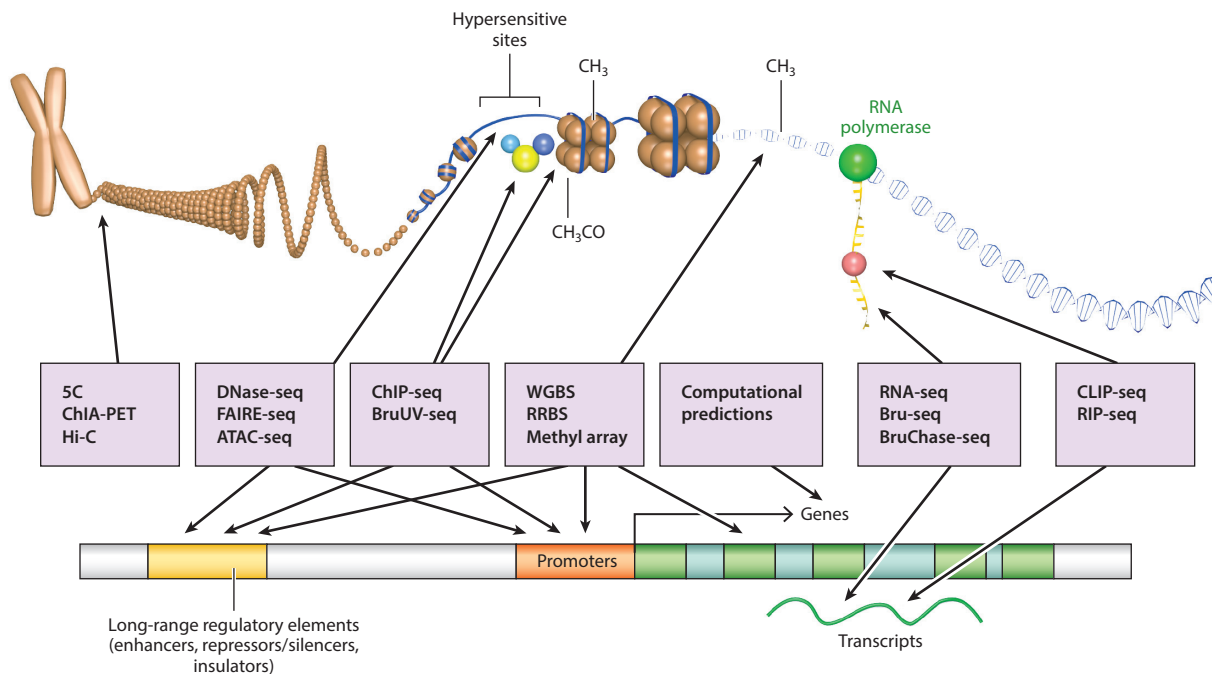
Figure 1

Functional annotation of genomes. The analysis of data sets obtained by different assays-by-sequence allows construction of comprehensive genome annotation maps, including elements that act at the protein and RNA levels and regulatory elements that control cells and circumstances in which a gene is active. The assays-by-sequence being used for the annotation of farmed animal genomes are described in text. Adapted from original image by Darryl Leja (NHGRI), Ian Dunham (EBI), and Michael Pazin (NHGRI) and kindly provided by Forrest Tanaka, Data Coordinating Center, ENCODE project (**https://www.encodeproject.org/**).

chromatin organization (17). Rapid evolution of enhancers is a universal feature of mammalian genomes, and recently evolved enhancers can be associated with genes under positive selection (18). Moreover, important genes involved in speciation and the differences between species, including genes involved in immune and reproduction functions, are not conserved (19, 20). Thus, not only do we require functional genomic information relevant to the biology/species context, but we are interested in how species differ and what makes a carp, chicken, or cow function.

The overarching goal of the FAANG Consortium is to support and coordinate the international community in creating reference functional maps of domesticated animal genomes, with an initial focus on farmed and companion species. As with human and model organisms, such maps can unravel which bits of each genome are functional, at what time, in what context, and in which cells and tissues and thus contribute to predictive biology.

An early aspiration of the FAANG Consortium was to create a framework for organizing data standardization, collection, and sharing from many groups. A major common thread was strong organization and communication among groups to maximize overall results. Substantial achievements have been reached; for example, a major FAANG accomplishment in this arena is the Data Coordination Center (DCC), described below.

The need for a project equivalent to ENCODE for farmed and companion animals was first mooted at the 33rd Conference of the International Society for Animal Genetics in 2012 (21).
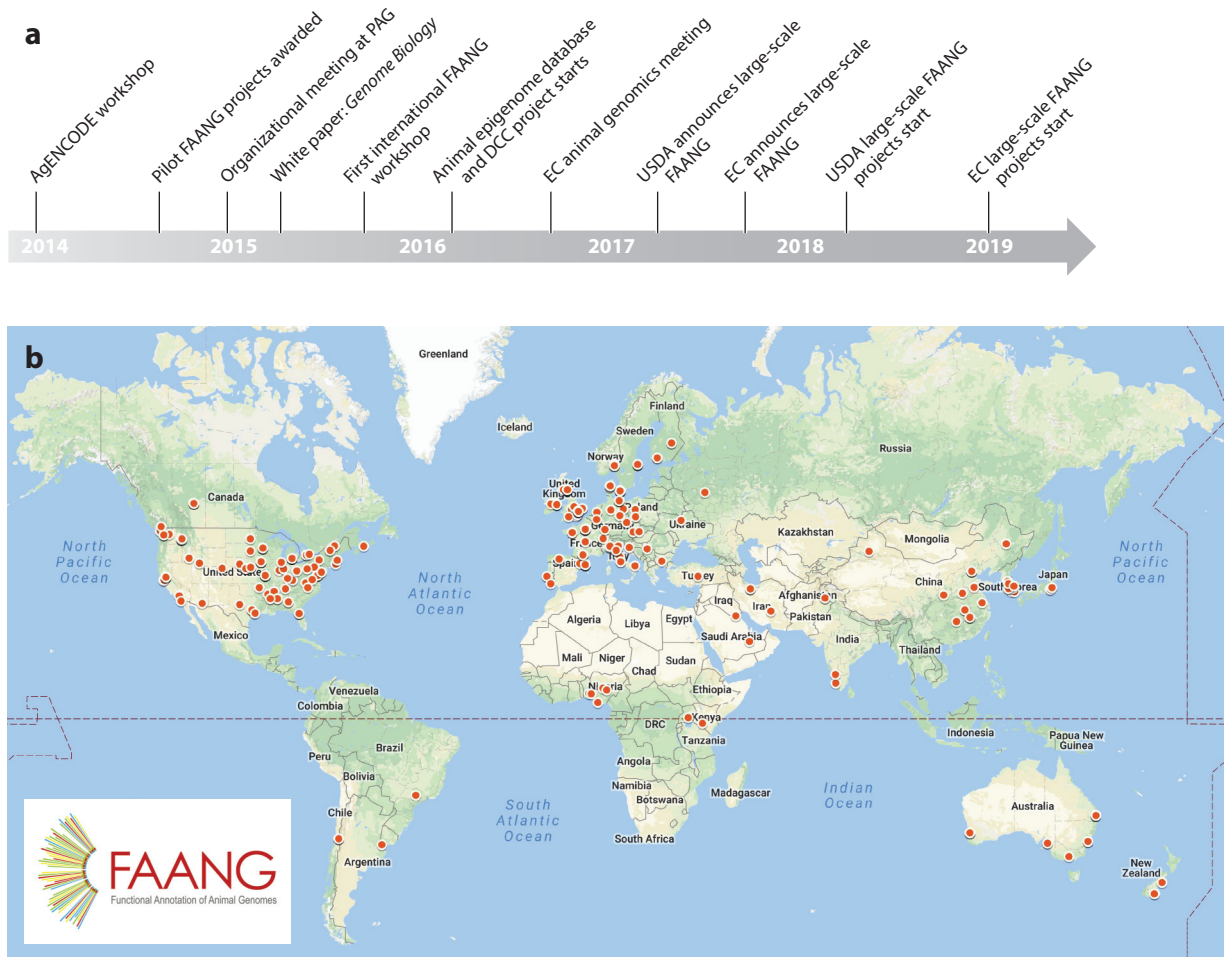
Figure 2

(*a*) Timeline of selected Functional Annotation of Animal Genomes (FAANG) events. (*b*) Worldwide map of FAANG participation (*red dots* indicate FAANG contributors as of April 2018). Figure adapted from **http://data.faang.org/**. Abbreviations: DCC, Data Coordination Center; EC, European Commission; PAG, Plant and Animal Genome Conference; USDA, US Department of Agriculture.

Figure 2*a* shows a timeline of the major events occurring during the initiation of FAANG. Development of the consortium has been rapid since the AgENCODE Workshop and Livestock Genomics meetings in 2014. That year, the US Department of Agriculture–National Institute of Food and Agriculture (USDA-NIFA) and the French National Institute for Agricultural Research (INRA) awarded pilot projects to develop FAANG infrastructure and initiate epigenetic data creation for adult tissues, and a project to create the FAANG DCC was awarded in 2016. A FAANG white paper describing the purpose and goals of the consortium was published in 2015 (22) and stressed the importance of moving "from sequence to consequence" in fundamental and applied research in domesticated animals. This landmark paper was followed by the first International FAANG Workshop, whose purpose was to bring together researchers and funding agencies to further nurture the FAANG enterprise (23). This workshop, as well as a 2016 Animal

Genomics workshop, developed agency support for FAANG, culminating in requests from the United States and European Union for proposals for expanding FAANG data resources. By 2019, we predict that the equivalent of more than $40 million worldwide will be supporting the creation and analysis of functional genomic data that will be integrated within the FAANG data storage and analysis structure for public use. **Figure 2*b*** shows the location of current FAANG contributors (totaling 387 as of October 2018), demonstrating the global reach of and interest in FAANG.

To date, most FAANG efforts have been devoted to farmed species, in part prompted by direct application to genomic selection (see below). Therefore, this review has a specific focus on farmed species.

REFERENCE GENOME SEQUENCES: THE FRAMEWORK FOR FUNCTIONAL ANALYSES

High-quality reference genome sequences are a key framework for functional analyses, including the discovery and exploitation of sequence variation. Draft reference genome sequences have been established for a wide range of domesticated and farmed animal species since the publication of the chicken genome in 2004 (24). The methodical hierarchical shotgun sequence strategy of sequencing bacterial artificial chromosome (BAC) clones selected from clone-based physical maps used by the public Human Genome Project, with its price tag of $3 billion, could clearly not be applied to a wide range of organisms. Thus, the draft reference genome sequences established for other animals have employed a range of strategies, including whole-genome shotgun sequencing (WGS) with Sanger sequencing technology, as used by the Venter human genome sequencing group (e.g., dog, horse) (25, 26); hybrid assemblies built on WGS and BAC clones sequenced with Sanger technology (e.g., cattle, pig) (19, 20); and WGS using next-generation sequencing technology (e.g., sheep, goat, rainbow trout) (27–29). Although these draft reference genome sequences have proven to be valuable resources for research and applications, they are incomplete and are an inadequate framework on which to build comprehensive genome annotation (see, e.g., Reference 30).

More recently, substantially improved reference genome sequences have been established for several animal species using long-read sequencing technologies, such as those by Pacific Biosciences (**https://www.pacb.com/**). Although such long-read sequencing technology suffers from high raw error rates, the errors can be corrected through a mix of high sequence depth and the use of more accurate short-sequence data. A key strength of long-read technology for characterizing complex animal genomes is its capability to traverse the highly repetitive sequences found in such genomes. The first of these improved animal reference sequences is the goat genome (31). The new goat genome is approximately 400-fold more contiguous than the earlier assembly built upon WGS short-read data. The improved assembly delivers better coverage of the genome, improved resolution of repeat structures, and more accurate gene models. Although other improved reference genomes based on long-read technology have been released into the public sequence databases in the spirit of prepublication data sharing (32), as espoused by the FAANG Consortium, as yet these genomes are unpublished.

ASSAYS-BY-SEQUENCE IN THE FAANG CONSORTIUM

Following the example of ENCODE and other consortia (**Figure 1**), FAANG has proposed a set of core FAANG assays (22), along with recommendations to prioritize species with genome assemblies of suitable quality. As high-quality assemblies have been achieved for several domesticated animal genomes, FAANG is expanding to additional species and fostering new

consortia [e.g., the Functional Annotation of All Salmonid Genomes (FAASG) Consortium (**https://www.faasg.org/**)] (33). The assay-by-sequencing technologies for identifying functional sequences continue to evolve. Thus, FAANG can progressively adopt technological improvements and benefit from interactions with ENCODE and other communities but also apply new techniques instead of, rather than in addition to, those used by ENCODE. In this context, the Animals, Samples, and Assays (ASA) Committee coordinates the optimization and standardization of sample collection and tissue/species assay protocols to obtain suitable data sets for the work of the FAANG Metadata and Data Sharing and Bioinformatics and Data Analysis (BDA) committees (23).

Transcribed Loci

The annotation of the transcriptome helps define biological pathways and transcriptional regulatory relationships and is thus an essential resource to help understand the impact of genetic variants associated with diseases and agronomic traits. A complete structural gene annotation catalogs the genomic loci with evidence of being transcribed in an organism's cells (tissues, developmental stages). A single gene can have different TSSs, transcription termination sites, and splicing to create alternative transcripts with unique functional roles (**Figure 3**). Representations of these transcripts as projected onto the genome are called transcript models. The identification of accurate transcript models is the primary objective of transcriptome annotation, and accurate estimates of transcript abundance improve estimates of gene expression and differential analysis (34).

RNA sequencing (RNA-seq) based on short reads has been a mainstay for many transcriptome annotation projects (35, 36), although in repeat regions and large gene families it can be difficult to define the exact transcript from which a short read originates (37).

Recently, advances in sequencing technologies have opened up exciting opportunities to build transcriptome annotations for almost any species on par with the best-studied model organisms. Each technology comes with its own strengths and weaknesses (see sidebar titled New Sequencing Technologies Allow Accurate Transcript Models). In sheep, a comprehensive catalog of gene
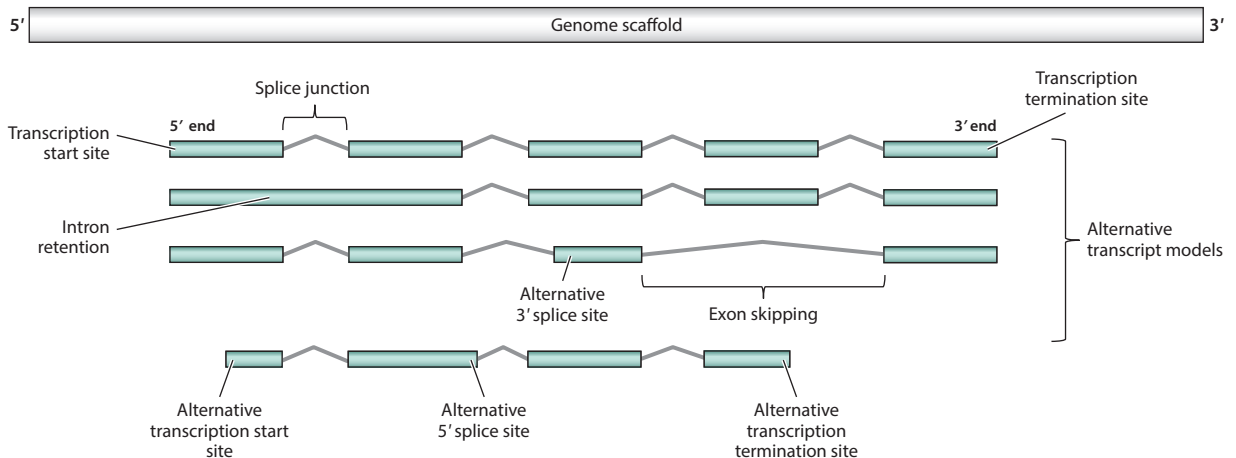

Figure 3

Depiction of alternative transcription events. Transcript models are shown compared with a scaffold in a genome assembly. This depiction is similar to how transcript models are displayed in genome browsers such as the Ensembl genome browser (**https://www.ensembl.org**). Exon to intron sizes are not to scale of typical eukaryotic transcripts.

NEW SEQUENCING TECHNOLOGIES ALLOW ACCURATE TRANSCRIPT MODELS

Long-read RNA sequencing methods (135), sometimes combined with a capture step (136, 137), have the ability to acquire the full-length structure of transcripts and to potentially capture all RNA species present in a cell (long and short, coding and regulatory). There are currently two long-read sequencing technologies for transcriptome sequencing: Pacific Biosciences Iso-Seq and Oxford Nanopore. The Nanopore technology (138) is still nascent and requires further development before widespread adoption.

Other technologies have been developed to resolve different aspects of transcriptome annotation. CAGEseq (139) and RAMPAGE (140) sequencing serve similar functions of identifying the transcription start sites (TSSs) and thus providing information on promoter regions. RAMPAGE provides a link between the TSS and an internal region of the transcript that can be used to look at the relationship between TSS and alternative splicing. TAIL-seq (141) is a sequencing method designed to capture the transcription termination sites (TTSs). Unlike some other methods, TAIL-seq does not use oligo-dT primers and is thus able to provide information on non-polyadenylated transcripts.

expression in more than 50 tissues from 6 animals has been generated via short-read RNA-seq. Using coexpression patterns across tissues and cells, Clark et al. (38) analyzed these data to functionally annotate genes with no assigned gene name. In chicken, the long-read sequencing technique Iso-Seq (Pacific Biosciences) was used to improve transcriptome annotation, revealing that transcriptional complexity can prevent accurate transcript annotation from short-read data (39).

These results indicate that diverse sampling and specialized sequencing methods have the potential to vastly improve transcriptome annotations. Full-length transcript sequencing using third-generation long-read sequencing technology is an example in which FAANG can substitute new technology to achieve equivalent or improved annotation relative to the earlier work on humans and mice. Annotation of the protein-coding content of the human and mouse genomes has relied heavily on full-length cloned complementary DNA sequence data (40).

Understanding of transcript expression is also incomplete, as can be demonstrated by the complexity of allele-specific expression (ASE). In cattle, most genes display tissue-specific ASE, with varying degrees of allelic imbalance, which is even reversed in some cases (41). These patterns of ASE vary greatly among individuals (41), indicating a high dependency on individual genomes and the need to profile ASE in additional animals and species to improve our understanding of transcriptional regulation.

Because noncoding transcripts have important regulatory roles (42) and are much less evolutionarily conserved than protein-coding transcripts, in terms of primary sequence, structure, and expression (43), it is essential to generate transcriptomic data on both short (RNAs with size <200 nucleotides) and long (size >200 nucleotides) noncoding RNAs across farmed species. Regarding small noncoding transcripts, the community has focused mainly on the identification of micro-RNAs (miRNAs) using miRNA-seq; however, more general protocols are needed to get a broader picture of the small RNA landscape (44). Such protocols are starting to be used in farmed animals; for example, Anthon et al. (45) identified a conserved set of 3,556 structured RNA loci in pigs that revealed mostly novel noncoding transcripts, including small nuclear RNA (snRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA), miRNA, ribosomal RNA (rRNA), and ribozymes.

Long noncoding RNAs (lncRNAs) have been a particular focus of the FAANG initiative, with lncRNA transcript sets reported for cattle, sheep, horse, chicken, pig, and goat (**Supplemental Table 1**). Cataloging of lncRNAs across multiple species is critical to improve understanding of

this class of transcripts and pinpoint intergenic regions where causative mutations might occur. Across these species, lncRNAs demonstrated similar characteristics to those observed in human and mouse. Weak sequence conservation, but strong positional conservation, was commonly observed (39, 46–50). Indeed, this conservation of synteny was harnessed to identify consensus lncRNAs in ruminants, demonstrating that combining RNA-seq from related species can improve lncRNA detection by reducing the effect of stochastic sampling (46). The positional conservation of lncRNAs seems likely to reflect a functional role, as data suggest that proximal lncRNA–mRNA pairs with divergent orientation may share promoters (39, 49). This conclusion is further supported by the observation that lncRNAs are enriched close to coexpressed genes, and that many intergenic lncRNAs are on the same strand or divergent from the closest protein-coding gene (48).

The computational methods for the identification of noncoding, especially lncRNA, genes are continuing to develop rapidly, and as yet there is no consensus around the optimum method. The research groups that have established initial lncRNA catalogs for farmed animal species have each used their own custom lncRNA identification pipeline (49, 51, 52). Thus, these initial noncoding transcriptomes may not be directly comparable. Additionally, RNA-seq methods varied across studies, with some using poly-A selection prior to RNA-seq (51, 52). Efforts to identify lncRNAs in the horse revealed that using poly-A selection may impair lncRNA detection (52), implying that certain RNA-seq methods may be more suited for the identification of lncRNAs than others. These discrepancies highlight the need to standardize the process of identifying lncRNAs within FAANG in future studies.

Iso-Seq has been used within the context of the FAANG Consortium in a study on chickens (39), where the authors were able to identify four times the number of transcripts as compared with the Ensembl chicken annotation, including ∼20,000 lncRNA transcripts. It should be noted that long-read technologies may not always represent full-length transcripts owing to RNA degradation. Thus, 5′ end capture strategies are beneficial in identification of full-length transcript models using long-read sequencing technologies (39).

Modified Histones

Histone modifications epigenetically regulate gene expression by altering the chromatin structure of DNA, affecting the ability of molecular machinery such as TFs and polymerases to bind. By profiling multiple histone modifications at a genome-wide level using chromatin immunoprecipitation sequencing (ChIP-seq), regulatory elements such as promoters, enhancers, and insulators and their state (e.g., active, inactive, poised) can be annotated. The ENCODE projects in human and mouse have studied a wide range of histone modifications with various functions, many of which often co-occur with other modifications (**Figure 1**). The FAANG Consortium initially selected four modified histones that were shown by ENCODE to be the most informative in identifying three key regulatory elements: promoters, enhancers, and insulators. The H3K4me3 modification is associated with accessible chromatin at TSSs, making it an important mark to identify active promoters. H3K27me3 is associated with the polycomb-group proteins that modify chromatin to silence genes and is therefore used to identify inactive promoters. H3K27ac is associated with active regulatory elements, including promoters and enhancers. Finally, H3K4me1 is associated with a dynamic chromatin structure that can identify both active and inactive enhancer locations as well as regions flanking promoters (22).

Histone modifications have been previously studied in farmed animals (53–56). Examples are the H3K27me3 modification in cattle lymphocytes (57) and several modifications in pig pluripotent stem cells (58), chicken polychromatic erythrocytes (59), and immune-related tissues using a

chicken Marek's disease model (60). Efforts are ongoing in the FAANG community to obtain functional maps of histone modifications from different tissues and species (see example data set in **Supplemental Figure 1**). However, a comprehensive genome-wide annotation of histone modifications across multiple tissues to produce a chromatin state map has not yet been produced for any farmed species.

DNA Methylation

DNA methylation is one of the best-studied epigenetic mechanisms. In vertebrates, DNA methylation is generally restricted to CpG sites established during early embryological development by methylation reprogramming involving demethylation and de novo remethylation (61). The majority of nonmethylated CpG sites are found in CpG islands near promoter regions of genes, and distinct distributions of methylation at different genome features (e.g., gene bodies, TSSs) can be observed (62, 63). Methylation levels at CpG islands generally determine gene expression, as high methylation levels prevent binding of TFs. Methylation in gene bodies is high overall and often correlates with transcriptional activity. Studies of numerous species have demonstrated that the silencing of transposable elements and repetitive sequences is a common function of DNA methylation (64, 65).

In farmed species, DNA methylation analyses have been used to generate methylation atlases (66) and by comparing different breeds or different developmental and physiological conditions (67), with most studies focused on single or limited panels of tissues. Methylation status appears to be sensitive and responsive to internal or environmental cues (e.g., nutrition, pathogens, and toxins) (68). For example, Lan et al. (69) investigated the impact of supplementing different energy sources during pregnancy on the expression and methylation of imprinted genes and DNA methyltransferases in sheep fetuses, finding that the methylation levels of the IGF2R and H19 imprinted genes were higher in fetuses whose mothers were fed supplements rich in methyl donors compared with those fed low-methyl supplements. In a subsequent study (70), whole-genome DNA methylation analysis revealed 60 differentially methylated regions between the two maternal diets. Integrative methylome and transcriptome analysis revealed an association between gene expression and inter-/intragenic methylated regions (70).

Recent studies have established that DNA methylation changes in sperm are strongly associated with infertility in humans (71) and with low reproductive performance in farm species. In cattle, 76 regions were found to be differentially methylated between sires of differing fertility status (72). Similarly, methylation analysis by microarray of high-fertility and subfertile spermatozoa in buffalos revealed that 73 genes in high-fertility and 78 genes in subfertile spermatozoa were hypermethylated (73).

To date, only a few genome-wide multi-tissue methylation studies have been performed in farmed species, namely in the horse (74) and pig (75–77). Thus, there is still a great need for further multi-tissue studies to obtain full methylation atlases in farmed animals.

Chromatin Accessibility

Profiling chromatin accessibility is essential for the identification and characterization of regulatory elements, because open chromatin facilitates DNA–protein interactions. Although DNase I hypersensitive sites sequencing (DNase-seq) (78) has been used extensively to detect regions of open chromatin, the Assay for Transposase Accessible Chromatin (ATAC-seq) is simpler and requires less sample input (79, 80). ATAC-seq is another example of FAANG adopting new alternative technologies rather than following ENCODE methods. ATAC-seq was used to profile

the accessible chromatin of both tissue-dissociated (liver) cells and primary cells (sorted CD4+ and CD8+ cells from blood or spleen) in cattle, goats, chickens, and pigs. By combining these data with RNA-seq data, Foissac et al. (48) provided evidence that ATAC-seq peaks can be used to support TSS prediction.

It should be noted that although ChIP-seq protocols for snap-frozen tissues exist (**https://www.diagenode.com/en/protocols**), the original ATAC-seq protocol relies on fresh cells. This represents a significant limitation to the potential of this technique for the screening of several tissue collections already existing or being realized for farm animals (e.g., **https://www.crb-anim.fr/crb-anim_eng/The-CRB-Anim**). However, a recent report has proposed a modified ATAC-seq protocol (Omni-ATAC) suited to profile open chromatin from snap-frozen tissues (81). Efforts are ongoing at FAANG laboratories to optimize the collection of cryopreserved nuclei of quality suited for ATAC-seq, with community exchanges promoted by the ASA Committee on key protocol steps and adequate validation of chromatin profiles of fresh versus frozen cell preparations. Of note, sequencing only the subnucleosomal fraction of ATAC-seq libraries significantly improved signal, leading to the detection of most of the same sites observed by DNase-seq analyses in chicken lung. This technique was subsequently applied to multiple tissues in pig and cattle, yielding similarly high-quality chromatin accessibility data (M. Halstead, unpublished observations).

Spatial Conformation of Chromatin

Nuclear genome organization is not random (82), with chromatin architecture playing a crucial role in regulating gene expression and cell phenotypes. Interactions/proximities between different genomic regions of a chromosome or between regions located on different chromosomes have been visualized by local analyses using confocal microscopy and Chromosome Conformation Capture (3C) technologies (83). These *cis-* or *trans-*interactions enable distal genes and regulatory elements to engage in chromosomal contacts, which are strongly correlated with their transcriptional activity (84, 85). High-throughput 3C technologies (86) allow the deciphering of these interactions either for a specific locus or genomic region (87, 88) or for the whole genome (Hi-C) (89). At megabase resolution, Hi-C is an accurate tool to identify active and inactive genomic compartments (86) and topologically associating domains (TADs) (90, 91), whereas high-resolution Hi-C enables precise mapping of active chromatin subdomains, loops between promoter and enhancer, and loops through transcriptional repressor CTCF response elements for *cis*-interaction (89).

To date, few Hi-C data sets are available for farmed species, although this technique was used to verify the long-read assembly of the goat reference genome (31). The FAANG pilot project FR-AgENCODE produced Hi-C maps for liver tissues from four livestock species (chicken, pigs, cows, and goats) (48). Using four replicates per species (two adult males and two adult females), the authors used genomic interaction matrices to identify active (A) and inactive (B) genomic compartments and TADs. Results confirmed the global conservation of genome organization between animals of the same species and verified that A compartments are transcriptionally more active than B compartments. In addition, TAD boundaries of all species were enriched for CTCF binding sites (48).

An important limitation of the Hi-C method is the sequence depth required to produce high-resolution interaction maps (more than 2,000 M of read pairs for a mammalian genome). To overcome this limitation, improvements to the Hi-C technology have been developed to capture genome-wide interactions from specific genomic regions of interest. This allows sequencing at a lower depth and improves resolution of the interaction maps (see sidebar titled Capturing Genome-Wide Interactions from Specific Genomic Regions of Interest). For example, Hi-C

CAPTURING GENOME-WIDE INTERACTIONS FROM SPECIFIC GENOMIC REGIONS OF INTEREST

Capture Hi-C, or CHI-C (142), combines Hi-C with capture of a targeted genomic region of interest, but these regions must be defined beforehand. Additional approaches aim to capture regulatory elements without any a priori assumption. The first one uses DNase-I instead of restriction endonucleases to cleave DNA during the Hi-C process. DNase-I preferentially cleaves DNA at the level of open, highly accessible chromatin and allows enrichment of Hi-C libraries for active regulatory elements (143). The second one combines Hi-C with chromatin immunoprecipitation to enrich for genomic regions with regulatory activities. Two recent studies demonstrated the efficiency of such an approach by using antibodies that recognize proteins known to be enriched in promoters, amplifiers, or insulators [HiChIP (144), PLAC-seq (145)]. This made it possible to produce Hi-C interaction maps at high resolutions for many tissues and experimental conditions and allowed further studies to evaluate how noncoding genetic variants located in enhancers are associated with specific genome architecture (92, 146).

Capture performed on 31,322 human promoters in 17 human primary hematopoietic cell types showed that interacting regions are enriched in genetic variants linked with altered expression of genes they contact. These data were then used to connect noncoding disease variants to putative target promoters among thousands of disease-candidate genes and pathways (92).

Overall, studying spatial genome organization can be a powerful tool to better understand the link between genotype and phenotype. Importantly, the functional annotations provided by FAANG will support the design of ad hoc capture strategies for livestock species at affordable costs.

DATA ANALYSIS AND DATA INTEGRATION

The role of the BDA group within FAANG is to define standard workflows. Once applied to appropriate data sets, such workflows ensure that the results generated are interoperable and comparable and that work from different groups around the world can be integrated to form a cohesive and coherent whole. To achieve this, the group must agree upon and define standards and workflows early in the project, and these workflows must be adhered to by all participants. The BDA group takes as input the assay-by-sequence data, runs primary analyses on each data type, runs integrative analyses on all data types, and produces a functional and regulatory annotation of our target species in unprecedented detail.

The BDA group is split into six subgroups on RNA, chromatin and regulatory genomics, epigenetics/methylation, and data integration. The RNA and lncRNA groups focus on both long- and short-read RNA-seq, using the reference genome as a guide to build the reference transcriptome using long-read sequencing (as in 39) and then the short-read data to quantify the transcripts (reviewed in 93), taking into account multi-mapped reads (37, 94).

The ChIP and Structural groups focus on data from ChIP-seq (95) and ATAC-seq (79) experiments, which, although different from a laboratory perspective, owing to their sequence-based nature, are very similar in terms of quality control, mapping, and (differential) peak calling analyses. The Methylation group focuses on Meth-seq data from whole-genome and reduced-representation bisulfite data (96), calling methylated cytosines and CpG islands from the resulting data. Finally, the Network group takes in all of these data inputs and builds an annotation of the expressed and regulatory regions within each tissue of each genome (see, e.g., 97) for integrated/network approaches.

THE FAANG DATA COORDINATION CENTER

The FAANG DCC implements one of the consortium's key aims: to ensure high-quality and rich supporting metadata to describe its farmed and companion animals, samples, and data sets (98). The DCC applies the standards set by the FAANG Metadata and Data Sharing Committee (**https://github.com/FAANG/faang-metadata**; **http://www.ebi.ac.uk/vg/faang/rule_sets/**), through provision of validation tools, support for submission to public data archives (**https://www.ebi.ac.uk/vg/faang**), and guidance and help desk support. The FAANG standards are version controlled and incorporate ontologies to accurately record diverse characteristics from around the world and integrate them so that cross–data set analyses can be performed. FAANG members regularly contribute improvements to the ontology databases, enhancing the quality of farmed and companion animals' specific terminology for the benefit of the entire community, as existing records can often be human or mouse focused. The DCC supports FAANG's aims through promoting best practices in data descriptions, deposition, openness, and reusability of its data sets and conforms to the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles (99) (**https://fairsharing.org/bsg-s000672**; **https://fairsharing.org/bsg-s000673**). All submissions must include detailed sampling, experimental, and bioinformatic protocols, and members have thus far provided 37 sample protocols, 16 experiment protocols, and 2 analysis protocols (**ftp://ftp.faang.ebi.ac.uk/ftp/protocols/**). **Figure 4** illustrates the number of registered
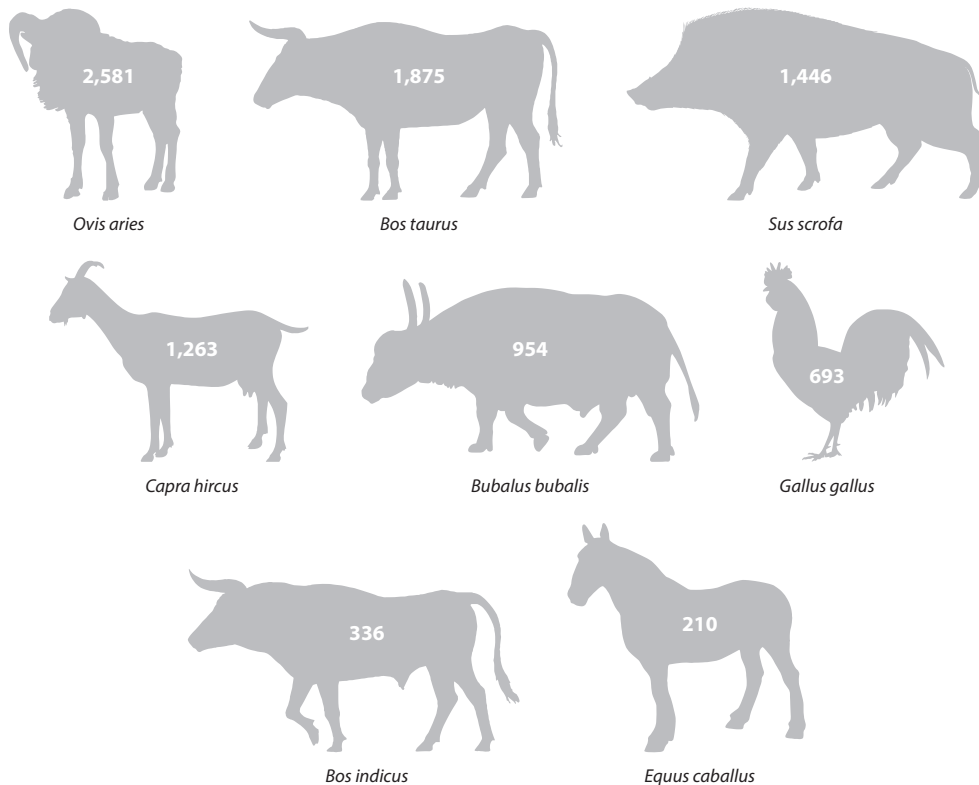

Figure 4

Registered specimens in the Functional Annotation of Animal Genomes (FAANG) data portal by species (as of October 2018). Current data are available from **http://data.faang.org/home**.

specimens in the FAANG DCC. Established FAANG committees and working groups develop recommended protocols for different types of assay preparation and data analysis to ensure that the resulting data are as intercomparable and high quality as possible. These standards and analysis pipelines will be made available through the FAANG GitHub collaborative workspace repository (**https://github.com/FAANG**) and FAANG data portal (**http://data.faang.org/**).

Another aim of the project is to collate data produced outside of the FAANG community. The FAANG data portal collates all appropriate publicly available farm and companion animal data sets into its single portal interface to allow their use in comparative analyses. To ensure that these data sets are useful for downstream analyses, they are all validated against a more permissive FAANG legacy standard, which although less strict than the full FAANG standard still ensures that the data meet minimal requirements set by the community. All data are clearly marked on the portal as to what standard they meet, with easy filtering options to allow inclusion or exclusion of the legacy data from analyses.

The FAANG data portal, which provides a single collated interface to identify validated high-quality farm and companion animal samples and data sets, complete with rich supporting metadata, enables researchers to identify, filter, and download complete comparable data sets for downstream analyses.

The FAANG DCC is continuing to seek usability and visualization improvements to assist the community in their genome-to-phenome research efforts. For example, track hubs (100) will enable researchers to visualize FAANG data in the context of public genomic annotations and in relation to their own data sets that they load into the University of California, Santa Cruz (101), or Ensembl (102) genome browsers (e.g., **Supplemental Figure 1**). Track hubs are powerful and highly customizable tools for the exploration of complex data across the genome, and they have proved highly effective in facilitating the community to access and process large consortium data sets, for example, the ENCODE (103) and Blueprint (104) projects.

GENOMIC SELECTION USING FUNCTIONAL INFORMATION

As introduced previously, genomic selection is a two-step process: (*a*) The effects of the genome-wide DNA markers on the target trait or traits are estimated in a large reference population, with the assumption that there are enough markers that there will be linkage disequilibrium between the markers and the causative mutations affecting the trait, and (*b*) the resulting equation is used to predict genomic estimated breeding values (GEBV) for selection candidates in an independent population, and the animals with the best GEBV are selected for breeding. An alternative, but equivalent, implementation of genomic selection uses relationships among the animals derived from the DNA markers in a single-step approach (105). Genomic selection has been widely adopted by the dairy cattle industry, with over 4 million animals genotyped with SNP arrays for this purpose. Although genomic selection is based on the use of evenly spaced genome-wide SNP markers and works well within narrow populations, such as Holstein-Friesian dairy cattle, the technology has been less successful in more diverse, multi-breed populations. The success of genomic selection is based on the extent of linkage disequilibrium between causative mutations and the SNP markers (typically ∼50,000 SNPs) used in the genome predictions (5). However, as previously mentioned, in the case of multi-breed predictions this linkage disequilibrium may not persist across breeds.

One way to overcome the problems associated with imperfect linkage disequilibrium between SNP and causal mutations would be to base the genomic predictions on whole-genome sequence data rather than DNA markers. If whole-genome sequence data are used, the causal mutations will be among the set of SNPs used for predictions, so there is no reliance on linkage disequilibrium.

Improvements in GEBV accuracy as a result of using sequence data to date have been mixed, with between a 0% (106) and a 5% improvement in the accuracy of genomic predictions (107, 108). The challenge with using sequence data is that millions of parameters (e.g., millions of SNP effects) must now be estimated from reference populations with 10,000 or, in the best case, 100,000 records. Therefore, information to prioritize SNPs for inclusion in the model, or identify classes of SNPs that are more likely to include causal mutations affecting the traits, would be extremely valuable. Simulations indeed indicated a dramatic increase in accuracy if causal variants were included in the model (109) but also showed that this is highly sensitive to misspecification of the causal variants.

The use of FAANG information in genomic selection could take several different forms. As there is increasing evidence that many of the mutations affecting complex traits are in regulatory regions (e.g., 110), the simplest way would be to use in genome prediction only SNPs and other genetic variants mapping to enhancer, promoter, and other regulatory elements, as well as variants in exons. However, this assumes that the annotation is perfect, with no regulatory elements unidentified. Alternatively, FAANG data could be combined with sequence conservation data (both between and within species conservation) to calculate functionality scores for every nucleotide position in the genome. In human, a variety of methods have been proposed to integrate these different annotations into a single measure of functional importance using both supervised (111, 112) and unsupervised (112) approaches. The potential power of conservation scores was illustrated in bovine, in which regulatory regions that could be predicted based on information from human and mouse were found enriched for variants associated with multiple production traits in beef and dairy cattle (113).

Another possibility is to use a Bayesian model that uses all SNPs but groups them into different classes based on their genome annotation and includes different proportions of SNPs in each class in the final prediction model. A model implementing an initial approach to this method, termed BayesRC, has been used for genomic prediction of traits such as milk production in dairy cattle, using classes based on whether SNPs were in, or close to, genes identified as important for lactation in gene expression experiments (108).

Thus, the next step toward genomic prediction from causal mutations will be to include FAANG information in models such as BayesRC. There is already some evidence that this type of information may help identify causal mutations affecting complex traits. In an experiment with 16,581 dairy cattle, SNPs in candidate bovine enhancer regions marked by bovine-liver H3K4me3 and H3K27ac histone modifications were found enriched for associations with milk production traits (114). A similar finding (enrichment of associations in putative enhancer regions) was reported for sequence variants associated with stature based on a meta-analysis of 58,265 cattle (115). Finally, it was found that sequence variants in target sites for DNA methylation (i.e., genomic regions that are found to be highly methylated in bovine placentae) captured a significant proportion of the variance in milk production traits in dairy cattle. Per sequence variant, splice site variants explained the highest proportion of variance in the investigated traits (116).

These results suggest that including FAANG information in genomic prediction of complex traits is a promising strategy for increasing the accuracy of these predictions, particularly when the genomic predictions are across breeds or in multi-breed populations. As described below, the accuracy of genome function predictions based on FAANG data will be crucial for implementing this integrated strategy.

VALIDATION OF PREDICTED FUNCTIONAL SEQUENCES

The molecular classes of candidate genomic elements being generated by FAANG provide a starting point for testing how these signatures relate to molecular and cellular function, as well

as organismal phenotype. As for ENCODE, assay-by-sequence data identify very large numbers of sequence elements of differing sizes and signal strengths, thus potentially generating hundreds of thousands of hypotheses (11). The requirements to test such hypotheses are (*a*) technology to modify the sequence(s) of interest and (*b*) a biological system in which the effect(s) of the modification can be measured.

Genome Editing of Putative Regulatory Elements

Gene-editing technologies, such as CRISPR/Cas9, enable selective modification of the putative functional sequence and are widely used for gene knockouts and edits of coding sequence. Many groups have used the CRISPR/Cas9 system for DNA sequence–specific mutagenesis to demonstrate the function of regulatory elements (enhancers, promoters) predicted by chromatin- and sequence-based assays (117, 118). For example, genome editing of putative distant regulatory sites validated such regulator-target relationships at the human *FTO* gene (119). Chromatin interaction prediction data or CTCF-binding data can also be verified by mutagenesis of these sites (118, 120). Such verification strategies have direct relevance for the proposed use of FAANG results as described in the previous section, i.e., for filtering genetic variation associated with traits in genetic improvement. Further, an integrated analysis has shown that pleiotropic association of distant SNPs on transcriptome and DNA methylation differences on target genes can be linked to control of complex traits (121).

Editing of the epigenome via addition or removal of epigenetic marks, such as DNA methylation, histone modifications, and noncoding RNAs, is also now feasible (122). Fusion proteins composed of an effector domain that adds or removes epigenetic marks and a deactivated Cas9 (dCas9), along with guide RNA, can activate or silence specific target genes. Thus, in contrast to genome editing, epigenome editing changes the chromatin state or interactions without changing the DNA sequence (118). A proof of principle of epigenome editing has recently been demonstrated in human cell lines, in which targeted acetylation of both promoters and enhancers resulted in a successful transcriptional activation of several genes (123). Sophisticated tools are now available, including inducible/reversible tools to reorganize chromatin architecture and bring specific chromatin regions together to test predicted regulatory relationships (124), as well as to create novel genomic regulation at target genes.

Thus, both genome and epigenome editing have the potential to manipulate gene expression in the organs of living animals, as well as to test predicted genome function. Ultimately, we envision that such editing tools will provide further characterization and validation of putative regulatory elements in livestock genomes that will contribute to the improvement of animal health and production traits.

New Tools to Model Cell and Tissue Function In Vitro

Validating the predictions, including assessing the impact of sequence variation within the predicted functional sequences, also requires experimental work in an appropriate biological system. The ideal biological system within which to validate a functional sequence would be the whole animal, but this is not feasible given the large numbers of putative functional sequences and the related economic and ethical concerns. Although cultured transformed cell lines, or less commonly cultured primary cells, can be used to test the effect of modifying a putative function sequence, such cells often provide a poor test for the effects in a whole animal.

Over the past decade, major advances have been made in the exploitation of stem cell technology for generating self-renewing and -organizing models of different tissues in a 3D matrix, otherwise

known as organoids. Organoid culture relies on the provision of tissue-specific factors that sustain long-term growth of the stem cells and differentiation into the main cell lineages found in the tissue. To date, the vast majority of organoid technology has been established using mouse or human stem cells, although recent work on large animals has shown potential to adapt the methods for generating intestinal organoids in farmed and companion animals (125–127). Organoids provide near-physiological models to dissect complex traits into molecular phenotypes using FAANG and other assays in highly reproducible conditions. One of the most attractive aspects of organoid technology is the possibility to derive multiple tissue models from individual humans or animals from relatively small amounts of tissue material or from iPSCs (induced pluripotent stem cells).

Organoid models are amenable to high-throughput, cost-effective, large-scale phenotyping and thus could be used in domesticated species to bring a better understanding of genotype–phenotype interactions in relation to important traits such as resilience, feed efficiency, and susceptibility/resistance to disease (e.g., 128). Organoids are amenable to gene editing by CRISPR-Cas9 technologies, thus making powerful genotype-to-phenotype systems for testing candidate causal mutations (129). Importantly, given the ease of biobanking, the organoid technology has a strong ethical benefit in reducing the number of animals used in experimentation, with screening tests in vitro providing ground data for less- and/or better-defined tests and validations of hypotheses in vivo.

CONCLUSIONS AND FUTURE DIRECTIONS

The current achievements of FAANG have been the creation of a coordinated network and of the related support infrastructure (DCC) to host and harmonize results from pilot and/or biologically oriented projects worldwide. To date, funding has been secured from USDA-NIFA to establish biological reference data sets for the main tissues and developmental phases for four farmed animal species (chickens, pigs, cattle, and sheep). Applications for EU funding for FAANG-centric projects are currently under review. Although there is already evidence of the value of such data for genotype-to-phenotype research and applications in genomic selection, due to limited levels of funding information generated in the foreseeable future for farmed as well as companion animals, FAANG is likely to remain less comprehensive compared with the human and mouse (113). However, we anticipate that the annotation of the most relevant tissues and developmental stages in domesticated species will deliver sufficient reference information required by emerging approaches and technologies in genotype-to-phenotype research. Furthermore, comparative functional genomics insights will be informed by FAANG data across several domesticated species. As outlined previously (22), hypotheses of function can be tested through evaluating microevolution of regulatory sequence elements across space and time. Population genetics of domestic animals is rapidly transitioning to high-resolution population genomics using whole-genome data sets from both extant and extinct domestic animal populations and their wild ancestors (130). Analyses of spatiotemporal genomic variation in domestic animals are most advanced in the horse and related equids (131–134); however, during the coming years, large surveys will be completed of whole-genome variation and paleogenomes from many domestic animal populations. Analysis of these data will reveal patterns of regulatory sequence microevolution owing to domestication, migration, and adaptive introgression from wild populations, as well as systematic human-mediated genetic improvement for production, health, and behavioral phenotypes.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

FAANG CO-AUTHORS AND THEIR CONTRIBUTIONS

Co-authors (manuscript section): E.G., C.K.T., and A.L.A. (What is FAANG? Tools for Predicting Biological Outcomes); A.L.A. (Reference Genome Sequences: The Framework for Functional Analyses); S.D., R.I.K., P.J.R., C.K., M.H., H.Z., H.K., O.M., H.A., S.F., E.G., and C.K.T. (Assays-by-Sequence in the FAANG Consortium); M.W., L.E., and J.M.R. (Data Analysis and Data Integration); P.W.H. and G.C. (The FAANG Data Coordination Center); B.J.H., A.J.C., I.M.M., H.D.D., M.E.G., and M.A.M.G. (Genomic Selection Using Functional Information); H.K., J.M.W., A.L.A., E.G., and C.K.T. (Validation of Predicted Functional Sequences); E.G., C.K.T., D.E.M., and A.L.A. (Conclusions and Future Directions). E.G. and C.K.T. (cocorresponding authors) and A.L.A. structured and finalized the manuscript.

Alan L. Archibald,[1] Herve Acloque,[2] Amanda J. Chamberlain,[3] Guy Cochrane,[4] Hans D. Daetwyler,[3,5] Sarah Djebali,[6] Lel Eory,[1] Sylvain Foissac,[6] Michael E. Goddard,[3,7] Martien A.M. Groenen,[8] Michelle Halstead,[9] Peter W. Harrison,[4] Benjamin J. Hayes,[3,10] Colin Kern,[9] Hasan Khatib,[11] Richard I. Kuo,[1] David E. MacHugh,[12,13] Iona M. Macleod,[3] Ole Madsen,[8] James M. Reecy,[14] Pablo J. Ross,[9] Mick Watson,[1] Jerry M. Wells,[15] Huaijun Zhou[9]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh EH29 9RG, UK; email: alan.archibald@roslin.ed.ac.uk; Lel.Eory@roslin.ed.ac.uk; Richard.kuo@roslin.ed.ac.uk; mick.watson@roslin.ed.ac.uk

[2]GABI, AgroParisTech, INRA, Université Paris Saclay, 78350 Jouy-en-Josas, France; email: herve.acloque@inra.fr

[3]AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria 3083, Australia; email: amanda.chamberlain@ecodev.vic.gov.au, hans.daetwyler@ecodev.vic.gov.au, meg@unimelb.edu.au, ben.hayes@ecodev.vic.gov.au, iona.macleod@ecodev.vic.gov.au

[4]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK; email: cochrane@ebi.ac.uk, peter@ebi.ac.uk

[5]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

[6]GenPhySE, Université de Toulouse, INRA, INPT, ENVT, F-31326 Castanet-Tolosan Cedex, France; email: sarah.djebali-quelen@inra.fr, sylvain.foissac@inra.fr

[7]Faculty of Land and Food Resources, University of Melbourne, Parkville, Victoria 3010, Australia

[8]Animal Breeding and Genomics Centre, Wageningen University, 6700 AH Wageningen, The Netherlands; email: martien.groenen@wur.nl, ole.madsen@wur.nl

[9]Department of Animal Science, University of California, Davis, California 95616, USA; email: mhalstead226@gmail.com, colin.kern@gmail.com, pross@ucdavis.edu, hzhou@ucdavis.edu

[10]Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, St. Lucia 4072, Australia

[11]Department of Animal Sciences, University of Wisconsin, Madison, WI 53706, USA; email: hkhatib@wisc.edu

[12]Animal Genomics Laboratory, School of Agriculture and Food Science, University College Dublin, Dublin D04 V1W8, Ireland; email: david.machugh@ucd.ie; **https://orcid.org/ 0000-0002-8112-4704**

[13]Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin D04 V1W8, Ireland

[14]Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA; email: jreecy@ iastate.edu

[15]Host-Microbe Interactomics, Animal Sciences Group, Wageningen University, 6700 Wageningen, The Netherlands; email: jerry.wells@wur.nl

LITERATURE CITED

1. Tuggle CK, Towfic F, Honavar V. 2011. Introduction to systems biology for animal scientists. In *Systems Biology and Livestock Science*, ed. MFW te Pas, H Woelders, A Bannick, pp. 1–30. Malden, MA: John Wiley & Sons

2. Suravajhala P, Kogelman LJ, Kadarmideen HN. 2016. Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. *Genet. Sel. Evol.* 48:38

3. Loor JJ, Vailati-Riboni M, McCann JC, Zhou Z, Bionaz M. 2015. Triennial Lactation Symposium: nutrigenomics in livestock: systems biology meets nutrition. *J. Anim. Sci.* 93:5554–74

4. Kell DB, Oliver SG. 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26:99–105

5. Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–29

6. Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. R. Soc. B Biol. Sci.* 283:20160569

7. Hayes BJ, Lewin HA, Goddard ME. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29:206–14

8. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92:433–43

9. Lund MS, Su G, Janss L, Guldbrandtsen B, Brøndum RF. 2014. Genomic evaluation of cattle in a multi-breed context. *Livest. Sci.* 166:101–10

10. Wang M, Hancock TP, Chamberlain JA, Vander Jagt CJ, Pryce JE, et al. 2018. Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. *BMC Genom.* 19:395

11. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, et al. 2014. Defining functional DNA elements in the human genome. *PNAS* 111:6131–38

12. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* 16:85–97

13. Brookes AJ, Robinson PN. 2015. Human genotype-phenotype databases: aims, challenges and opportunities. *Nat. Rev. Genet.* 16:702–15

14. Chakravorty S, Hegde M. 2017. Gene and variant annotation for Mendelian disorders in the era of advanced sequencing technologies. *Annu. Rev. Genom. Hum. Genet.* 18:229–56

15. Meadows JRS, Lindblad-Toh K. 2017. Dissecting evolution and disease using comparative vertebrate genomics. *Nat. Rev. Genet.* 18:624–36

16. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328:1036–40

17. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515:355–64

18. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* 160:554–66

19. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, et al. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324:522–28

20. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491:393–98

21. Archibald AL, Flicek P, Birney E. 2012. *Enabling the reading of genome sequences for farmed and companion animals—a proposal for ENCODE consortia*. Presented at the 33rd Conference of the International Society for Animal Genetics, Cairns, Aust., July 15–20. **https://www.isag.us/2012/docs/ISAG_2012_Abstracts.pdf**

22. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, et al. 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16:57

23. Tuggle CK, Giuffra E, White SN, Clarke L, Zhou H, et al. 2016. GO-FAANG meeting: a Gathering On Functional Annotation of Animal Genomes. *Anim. Genet.* 47:528–33

24. Int. Chick. Genome Seq. Consort. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716

25. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–19

26. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865–67

27. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, et al. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344:1168–73

28. Dong Y, Xie M, Jiang Y, Xiao N, Du X, et al. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31:135–41

29. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5:3657

30. Warr A, Robert C, Hume D, Archibald AL, Deeb N, Watson M. 2015. Identification of low-confidence regions in the pig reference genome (Sscrofa10.2). *Front. Genet.* 6:338

31. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49:643–50

32. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, et al. 2009. Prepublication data sharing. *Nature* 461:168–70

33. Macqueen DJ, Primmer CR, Houston RD, Nowak BF, Bernatchez L, et al. 2017. Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genom.* 18:484

34. Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* 4:1521

35. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621–28

36. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10:57–63

37. Robert C, Watson M. 2015. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol.* 16:177

38. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, et al. 2017. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genet.* 13:e1006997

39. Kuo RI, Tseng E, Eory L, Paton IR, Archibald AL, Burt DW. 2017. Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genom*. 18:323

40. Imanishi T, Itoh T, Suzuki Y, O'Donovan C, Fukuchi S, et al. 2004. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLOS Biol*. 2:e162

41. Chamberlain AJ, Vander Jagt CJ, Hayes BJ, Khansefid M, Marett LC, et al. 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genom*. 16:993

42. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, et al. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539:452–55

43. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*. 11:1110–22

44. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, et al. 2012. Landscape of transcription in human cells. *Nature* 489:101–8

45. Anthon C, Tafer H, Havgaard JH, Thomsen B, Hedegaard J, et al. 2014. Structured RNAs and synteny regions in the pig genome. *BMC Genom*. 15:459

46. Bush SJ, Muriuki C, McCulloch MEB, Farquhar IL, Clark EL, Hume DA. 2018. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet. Sel. Evol.* 50:20

47. Bush SJ, Muriuki C, McCulloch MEB, Farquhar IL, Clark EL, Hume DA. 2018. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet. Sel. Evol.* 50:20

48. Foissac S, Djebali S, Munyard K, Villa-Vialaneix N, Rau A, et al. 2018. Livestock genome annotation: transcriptome and chromatin structure profiling in cattle, goat, chicken and pig. *bioRxiv*. **https://doi.org/10.1101/316091**

49. Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, et al. 2017. Long noncoding RNA repertoire in chicken liver and adipose tissue. *Genet. Sel. Evol.* 49:6

50. Weikard R, Hadlich F, Hammon HM, Frieten D, Gerbert C, et al. 2018. Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. *Oncotarget* 9:21052–69

51. Koufariotis LT, Chen YP, Chamberlain A, Vander Jagt C, Hayes BJ. 2015. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLOS ONE* 10:e0141225

52. Scott EY, Mansour T, Bellone RR, Brown CT, Mienaltowski MJ, et al. 2017. Identification of long non-coding RNA in the horse transcriptome. *BMC Genom*. 18:511

53. Wang X, Zhang FX, Wang ZM, Wang Q, Wang HF, et al. 2016. Histone H3K9 acetylation influences growth characteristics of goat adipose-derived stem cells *in vitro*. *Genet. Mol. Res*. 15:gmr15048954

54. Kociucka B, Stachecka J, Szydlowski M, Szczerbal I. 2017. Rapid communication: the correlation between histone modifications and expression of key genes involved in accumulation of adipose tissue in the pig. *J. Anim. Sci*. 95:4514–19

55. Byrne K, McWilliam S, Vuocolo T, Gondro C, Cockett NE, Tellam RL. 2014. Genomic architecture of histone 3 lysine 27 trimethylation during late ovine skeletal muscle development. *Anim. Genet*. 45:427–38

56. Li C, Guo S, Zhang M, Gao J, Guo Y. 2015. DNA methylation and histone modification patterns during the late embryonic and early postnatal development of chickens. *Poult. Sci*. 94:706–21

57. He Y, Yu Y, Zhang Y, Song J, Mitra A, et al. 2012. Genome-wide bovine H3K27me3 modifications and the regulatory effects on genes expressions in peripheral blood lymphocytes. *PLOS ONE* 7:e39094

58. Xiao S, Xie D, Cao X, Yu P, Xing X, et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* 149:1381–92

59. Jahan S, Xu W, He S, Gonzalez C, Delcuve GP, Davie JR. 2016. The chicken erythrocyte epigenome. *Epigenet. Chromatin* 9:19

60. Mitra A, Luo J, He Y, Gu Y, Zhang H, et al. 2015. Histone modifications induced by MDV infection at early cytolytic and latency phases. *BMC Genom*. 16:311

61. Messerschmidt DM, Knowles BB, Solter D. 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev*. 28:812–28

62. Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet*. 13:484–92

63. Schübeler D. 2015. Function and information content of DNA methylation. *Nature* 517:321–26
64. Coleman-Derr D, Zilberman D. 2012. DNA methylation, H2A.Z, and the regulation of constitutive expression. *Cold Spring Harb. Symp. Quant. Biol.* 77:147–54
65. Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9:465–76
66. Li M, Wu H, Luo Z, Xia Y, Guan J, et al. 2012. An atlas of DNA methylomes in porcine adipose and muscle tissues. *Nat. Commun.* 3:850
67. Bang WY, Kim SW, Kwon SG, Hwang JH, Kim TW, et al. 2013. Swine liver methylomes of Berkshire, Duroc and Landrace breeds by MeDIPS. *Anim. Genet.* 44:463–66
68. Ibeagha-Awemu EM, Zhao X. 2015. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Front. Genet.* 6:302
69. Lan X, Cretney EC, Kropp J, Khateeb K, Berg MA, et al. 2013. Maternal diet during pregnancy induces gene expression and DNA methylation changes in fetal tissues in sheep. *Front. Genet.* 4:49
70. Namous H, Peñagaricano F, Del Corvo M, Capra E, Thomas DL, et al. 2018. Integrative analysis of methylomic and transcriptomic data in fetal sheep muscle tissues in response to maternal diet during pregnancy. *BMC Genom.* 19:123
71. Jenkins TG, Carrell DT. 2012. The sperm epigenome and potential implications for the developing embryo. *Reproduction* 143:727–34
72. Kropp J, Carrillo JA, Namous H, Daniels A, Salih SM, et al. 2017. Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. *BMC Genom.* 18:280
73. Verma A, Rajput S, De S, Kumar R, Chakravarty AK, Datta TK. 2014. Genome-wide profiling of sperm DNA methylation in relation to buffalo (*Bubalus bubalis*) bull fertility. *Theriogenology* 82:750–59.e1
74. Lee JR, Hong CP, Moon JW, Jung YD, Kim DS, et al. 2014. Genome-wide analysis of DNA methylation patterns in horse. *BMC Genom.* 15:598
75. Schachtschneider KM, Madsen O, Park C, Rund LA, Groenen MA, Schook LB. 2015. Adult porcine genome-wide DNA methylation patterns support pigs as a biomedical model. *BMC Genom.* 16:743
76. Choi M, Lee J, Le MT, Nguyen DT, Park S, et al. 2015. Genome-wide analysis of DNA methylation in pigs using reduced representation bisulfite sequencing. *DNA Res.* 22:343–55
77. Schachtschneider KM, Liu Y, Rund LA, Madsen O, Johnson RW, et al. 2016. Impact of neonatal iron deficiency on hippocampal DNA methylation and gene transcription in a porcine biomedical model of cognitive development. *BMC Genom.* 17:856
78. Song L, Crawford GE. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* 2010:pdb.prot5384
79. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10:1213–18
80. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109:21.9.1–9
81. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14:959–62
82. Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T. 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8:104–15
83. Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295:1306–11
84. Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. 2013. Chromosomal contact permits transcription between coregulated genes. *Cell* 155:606–20
85. Pombo A, Dillon N. 2015. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* 16:245–57
86. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–93

87. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485:381–85

88. Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D. 2014. Temporal dynamics and developmental memory of 3D chromatin architecture at *Hox* gene loci. *eLife* 3:e02557

89. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–80

90. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–80

91. Pope BD, Ryba T, Dileep V, Yue F, Wu W, et al. 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature* 515:402–5

92. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, et al. 2016. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 167:1369–84.e19

93. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13

94. Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, et al. 2012. Fast computation and applications of genome mappability. *PLOS ONE* 7:e30377

95. Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316:1497–502

96. Krueger F, Kreck B, Franke A, Andrews SR. 2012. DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods* 9:145–51

97. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. 2015. The Ensembl regulatory build. *Genome Biol.* 16:56

98. Harrison PW, Fan J, Richardson D, Clarke L, Zerbino D, et al. 2018. FAANG, establishing metadata standards, validation and best practice for the farmed and companion animal community. *Anim. Genet.* 49:520–26

99. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018

100. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, et al. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30:1003–5

101. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, et al. 2018. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* 46:D762–D69

102. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, et al. 2018. Ensembl 2018. *Nucleic Acids Res.* 46:D754–D61

103. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, et al. 2016. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* 44:D726–32

104. Bujold D, Morais DAL, Gauthier C, Côté C, Caron M, et al. 2016. The International Human Epigenome Consortium Data Portal. *Cell Syst.* 3:496–9.e2

105. Misztal I, Legarra A. 2017. Invited review: efficient computation strategies in genomic selection. *Animal* 11:731–36

106. Veerkamp RF, Bouwman AC, Schrooten C, Calus MP. 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. *Genet. Sel. Evol.* 48:95

107. Brøndum RF, Su G, Janss L, Sahana G, Guldbrandtsen B, et al. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98:4107–16

108. MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, et al. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genom.* 17:144

109. Pérez-Enciso M, Rincón JC, Legarra A. 2015. Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised. *Genet. Sel. Evol.* 47:43

110. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. 2012. Linking disease associations with regulatory information in the human genome. *Genome Res.* 22:1748–59

111. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46:310–15

112. Gulko B, Hubisz MJ, Gronau I, Siepel A. 2015. A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.* 47:276–83

113. Nguyen QH, Tellam RL, Naval-Sanchez M, Porto-Neto LR, Barendse W, et al. 2018. Mammalian genomic regulatory regions predicted by utilizing human genomics, transcriptomics, and epigenetics data. *Gigascience* 7:1–17

114. Wang M, Hancock TP, MacLeod IM, Pryce JE, Cocks BG, Hayes BJ. 2017. Putative enhancer sites in the bovine genome are enriched with variants affecting complex traits. *Genet. Sel. Evol.* 49:56

115. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* 50:362–67

116. Koufariotis LT, Chen YP, Stothard P, Hayes BJ. 2018. Variance explained by whole genome sequence variants in coding and regulatory genome annotations for six dairy traits. *BMC Genom.* 19:237

117. Klann TS, Black JB, Gersbach CA. 2018. CRISPR-based methods for high-throughput annotation of regulatory DNA. *Curr. Opin. Biotechnol.* 52:32–41

118. Lau CH, Suh Y. 2018. CRISPR-based strategies for studying regulatory elements and chromatin structure in mammalian gene control. *Mamm. Genome* 29:205–28

119. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, et al. 2015. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373:895–907

120. Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, et al. 2017. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function *in vivo*. *Nat. Cell Biol.* 19:952–61

121. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, et al. 2018. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* 9:918

122. Kungulovski G, Jeltsch A. 2016. Epigenome editing: state of the art, concepts, and perspectives. *Trends Genet.* 32:101–13

123. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, et al. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33:510–17

124. Morgan SL, Mariano NC, Bermudez A, Arruda NL, Wu F, et al. 2017. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* 8:15993

125. Powell RH, Behnke MS. 2017. WRN conditioned media is sufficient for *in vitro* propagation of intestinal organoids from large farm and small companion animals. *Biol. Open* 6:698–705

126. Khalil HA, Lei NY, Brinkley G, Scott A, Wang J, et al. 2016. A novel culture system for adult porcine intestinal crypts. *Cell Tissue Res.* 365:123–34

127. van der Hee B, Loonen LMP, Taverne N, Taverne-Thiele JJ, Smidt H, Wells JM. 2018. Optimized procedures for generating an enhanced, near physiological 2D culture system from porcine intestinal organoids. *Stem Cell Res.* 28:165–71

128. Meijerink E, Neuenschwander S, Fries R, Dinter A, Bertschinger HU, et al. 2000. A DNA polymorphism influencing $\alpha(1,2)$fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to *Escherichia coli* F18 adhesion. *Immunogenetics* 52:129–36

129. Driehuis E, Clevers H. 2017. CRISPR/Cas 9 genome editing and its applications in organoids. *Am. J. Physiol. Gastrointest. Liver Physiol.* 312:G257–G65

130. MacHugh DE, Larson G, Orlando L. 2017. Taming the past: ancient DNA and the study of animal domestication. *Annu. Rev. Anim. Biosci.* 5:329–51

131. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *PNAS* 111:E5661–69

132. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, et al. 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 360:111–14

133. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, et al. 2017. Ancient genomic changes associated with domestication of the horse. *Science* 356:442–45

134. Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, et al. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *PNAS* 112:E6889–97

135. Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, et al. 2015. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* 33:736–42

136. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, et al. 2011. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30:99–104

137. Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, et al. 2017. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat. Genet.* 49:1731–40

138. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. 2015. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* 12:351–56

139. Takahashi H, Lassmann T, Murata M, Carninci P. 2012. 5′ End-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* 7:542–61

140. Batut P, Gingeras TR. 2013. RAMPAGE: promoter activity profiling by paired-end sequencing of 5′-complete cDNAs. *Curr. Protoc. Mol. Biol.* 104:25B.11.1–16

141. Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. *Mol. Cell* 53:1044–52

142. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, et al. 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47:598–606

143. Ramani V, Cusanovich DA, Hause RJ, Ma W, Qiu R, et al. 2016. Mapping 3D genome architecture through *in situ* DNase Hi-C. *Nat. Protoc.* 11:2104–21

144. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, et al. 2016. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13:919–22

145. Fang R, Yu M, Li G, Chee S, Liu T, et al. 2016. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* 26:1345–48

146. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, et al. 2017. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49:1602–12