The Farm Animal Genotype-Tissue Expression (FarmGTEx) Project

Lingzhao Fang[1]*, Jinyan Teng[2], Qing Lin[1,2], Zhonghao Bai[1], Shuli Liu[3,4], Dailu Guan[5], Bingjie Li[6], Yahui Gao[2], Yali Hou[7], Mian Gong[7], Zhangyuan Pan[7], Ying Yu[8], Emily Clark[9], Jacqueline Smith[9], Konrad Rawlik[10], Ruidong Xiang[11,12,13], Amanda J. Chamberlain[11,14], Michael E. Goddard[11], Mathew Littlejohn[15,16], Greger Larson[17], David E. MacHugh[18,19,20], John F. O'Grady[18], Peter Sørensen[1], Goutam Sahana[1], Mogens Sandø Lund[1], Zhihua Jiang[21], Xiangchun Pan[2], Wentao Gong[2], Haihan Zhang[22], Xi He[22], Yuebo Zhang[22], Ning Gao[22], Jun He[22], Guoqiang Yi[23], Yuwen Liu[23], Zhonglin Tang[23], Pengju Zhao[24], Yang Zhou[25,26], Liangliang Fu[25], Xiao Wang[27], Dan Hao[28], Lei Liu[26], Siqian Chen[8], Robert S. Young[29,30], Xia Shen[29,31,32], Charley Xia[33,34], Hao Cheng[5], Li Ma[35], John B. Cole[36,37,38], Ransom L. Baldwin IV[39], Cong-jun Li[39], Curtis P. Van Tassell[39], Benjamin D. Rosen[39], Nayan Bhowmik[39], Joan Lunney[40], Wansheng Liu[41], Leluo Guan[42,43], Xin Zhao[44], Eveline M. Ibeagha-Awemu[45], Yonglun Luo[46,47], Lin Lin[46,47], Oriol Canela-Xandri[48], Martijn Derks[49], Richard P.M.A. Crooijmans [49], Marta Gòdia[49], Ole Madsen[49], Martien A.M. Groenen[49], James E. Koltes[50], Christopher K. Tuggle[50], Fiona M. McCarthy[51], Dominique Rocha[52], Elisabetta Giuffra[52], Marcel Amills[53,54], Alex Clop[53,55], Maria Ballester[56], Gwenola Tosser-Klopp[57], Jing Li[1,58], Chao Fang[59], Ming Fang[60], Qishan Wang[61], Zhuocheng Hou[8], Qin Wang[8], Fuping Zhao[7], Lin Jiang[7], Guiping Zhao[7], Zhengkui Zhou[7], Rong Zhou[7], Hehe Liu[62], Juan Deng[62], Long Jin[62], Mingzhou Li[62], Delin Mo[63], Xiaohong Liu[63], Yaosheng Chen[63], Xiaolong Yuan[2], Jiaqi Li[2], Shuhong Zhao[25,26], Yi Zhang[8], Xiangdong Ding[8], Dongxiao Sun[8], Huizeng Sun[61], Cong Li[64], Yu Wang[64], Yu Jiang[64], Dongdong Wu[65], Wenwen Wang[66], Xinzhong Fan[66], Qin Zhang[66], Kui Li[23], Hao Zhang[8], Ning Yang[8], Xiaoxiang Hu[67], Wen Huang[68], Jiuzhou Song[35], Yang Wu[69], Jian Yang[3,4], Weiwei Wu[70], Claudia Kasper[71], Xinfeng Liu[1,72], Xiaofei Yu[73], Leilei Cui[74,75], Xiang Zhou[76], Seyoung Kim[77], Wei Li[78], Hae Kyung Im[79], Edward S. Buckler[80,81,82], Bing Ren[83], Michael C. Schatz[84], Jingyi Jessica Li[85]*, Abraham Palmer[86]*, Laurent Frantz[87,88]*, Huaijun Zhou[5]*, Zhe Zhang[2]*, George E. Liu[39]* and the FarmGTEx Consortium

Affiliations

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, 8000, Denmark

[2]State Key Laboratory of Swine and Poultry Breeding Industry, National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

[3]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China

[4]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[5]Department of Animal Science, University of California, Davis, CA, 95616, USA

[6]Department of Animal and Veterinary Sciences, Scotland's Rural College (SRUC), Roslin Institute Building, Midlothian EH25 9RG, UK

[7]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China

[8]National Engineering Laboratory for Animal Breeding, State Key Laboratory of Animal Biotech Breeding, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

[9]The Roslin Institute, University of Edinburgh, Edinburgh, Midlothian, EH25 9RG, UK

[10]Baillie Gifford Pandemic Science Hub, Centre for Inflammation Research, Institute for Regeneration and Repair, The University of Edinburgh, EH16 4UU Edinburgh, UK

[11]Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, 5 Ring Rd, Bundoora, VIC, 3083, Australia

[12]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

[13]School of Agriculture, Food and Ecosystem Sciences, The University of Melbourne, Parkville, VIC, 3052, Australia

[14]School of Applied Systems Biology, La Trobe Uiversity, Bundoora, VIC, 3083, Australia

[15]Research & Development, Livestock Improvement Corporation, Ruakura Road, Hamilton, New Zealand

[16]AL Rae Centre for Genetics and Breeding, Massey University, Palmerston North, New Zealand

[17]The Palaeogenomics & Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford, UK

[18]UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland

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

[20] UCD One Health Centre, University College Dublin, Belfield, Dublin, D04 V1W8, Ireland

[21]Department of Animal Sciences, Washington State University, Pullman, WA 99164-7620, USA

[22]College of Animal Science and Technology, Hunan Agricultural University, Changsha 410128, China

[23]Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

[24]Hainan Institute, Zhejiang University, Yongyou Industry Park, Yazhou Bay Sci-Tech City, Sanya 572000, China

[25]College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China

[26]Yazhouwan National Laboratory, Sanya City, Hainan Province, 572024, China

[27]Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan, 250100, China

[28]Poultry Institute, Shandong Academy of Agricultural Sciences, Jinan, 250100, China

[29]Usher Institute, University of Edinburgh, 5-7 Little France Road, Edinburgh BioQuarter - Gate 3, Edinburgh, EH16 4UX

[30]Zhejiang University - University of Edinburgh Institute, Zhejiang University, 718 East Haizhou Road, 314400 Haining, PR China

[31]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

[32]Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China

[33]Lothian Birth Cohort studies, University of Edinburgh, 7 George Square, Edinburgh, UK

[34]Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, UK

[35]Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA

[36]Council on Dairy Cattle Breeding, Bowie, Maryland 20716, USA

[37]Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, Florida 32611-0910, USA

[38]Department of Animal Science, North Carolina State University, Raleigh, North Carolina 27695-7621, USA

[39]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, Maryland 20705, USA

[40]Animal Parasitic Diseases Laboratory, BARC, NEA, ARS, USDA, Beltsville, MD 20705, USA

[41]Department of Animal Science, Center for Reproductive Biology and Health (CRBH), College of Agricultural Sciences, The Pennsylvania State University, 311 AVBS Building, University Park, PA, 16802, USA

[42]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2P5, Canada

[43]Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

[44]Department of Animal Science, McGill University, Sainte-Anne-de-Bellevue, Quebec, H9X 3V9, Canada

[45]Agriculture and Agri-Food Canada, Sherbrooke Research and Development Centre, Sherbrooke, QC, Canada

[46]Department of Biomedicine, Aarhus University, Aarhus, 8000, Denmark

[47]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark

[48]MRC Human Genetics Unit at the Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh EH4 2XU, UK

[49]Animal Breeding and Genomics, Wageningen University and Research, Wageningen, The Netherlands

[50]Department of Animal Science, Iowa State University, Ames, IA, 50011, USA

[51]University of Arizona, Tucson, AZ, USA

[52]Paris-Saclay University, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France

[53]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

[54]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

[55]Consejo Superior de Investigaciones Científicas, Barcelona, Catalonia, Spain

[56]Animal Breeding and Genetics Programme, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre Marimon, Caldes de Montbui, Spain

[57]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France

[58]School of Agriculture and Life Sciences, Kunming University, Kunming, 650214, China

[59]LC-Bio Technologies, Co., Ltd., Hangzhou, China

[60]Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture and Rural Affairs, Jimei University, Xiamen, China

[61]Department of Animal Science, College of Animal Sciences, Zhejiang University, Hangzhou 310058, China

[62]College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

[63]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

[64]College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China

[65]Key Laboratory of Genetic Evolution and Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China

[66]College of Animal Science and Technology, Shandong Agricultural University, Taian 271018, China

[67]State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, 100193, China

[68]Department of Animal Science, Michigan State University, East Lansing, MI, USA

[69]Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China

[70]Institute of Animal Science, Xinjiang Academy of Animal Science. No. 468, Alishan Street, Toutunhe District, Urumqi City, Xinjiang Uygur Autonomous Region, China.

[71]Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope Posieux, 1725, Switzerland

[72]State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystem, College of Ecology, Lanzhou University, Lanzhou, 730000, China

[73]College of Marine Life Sciences, Ocean university of China

[74]School of Life Sciences, Nanchang University, Nanchang, China

[75]Human Aging Research Institute and School of Life Science, Nanchang University, and Jiangxi Province Key Laboratory of Aging and Disease, Jiangxi, China

[76]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

[77]Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA 15260, USA

[78]Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, CA 92697, USA

[79]Department of Medicine and Human Genetics, The University of Chicago, Chicago, IL 60637, USA

[80]Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA 14853

[81]Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA 14853

[82]Agricultural Research Service, United States Department of Agriculture, Ithaca, NY, USA 14853

[83]Department of Cellular and Molecular Medicine, Center for Epigenomics, Moores Cancer Center and Institute of Genomic Medicine, University of California, San Diego School of Medicine, La Jolla, CA, 92093-0653, USA

[84]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

[85]Department of Statistics and Data Science, University of California, Los Angeles, Los Angeles, CA, 90095, USA

[86]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA, USA

[87]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich 80539, Germany

[88]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4DQ, United Kingdom

***Corresponding Authors:**

Lingzhao Fang: Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

E-mail: lingzhao.fang@qgg.au.dk;

Jingyi Jessica Li: Department of Statistics and Data Science, University of California, Los Angeles, Los Angeles, CA, 90095, USA

E-mail: lijy03@g.ucla.edu;

Abraham Palmer: Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

E-mail: aap@.ucsd.edu;

Laurent Frantz: Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany

E-mail: laurent.frantz@lmu.de;

Huaijun Zhou: Department of Animal Science, University of California, Davis, CA, 95616, USA

E-mail: hzhou@ucdavis.edu;

Zhe Zhang: College of Animal Science, South China Agricultural University, Guangzhou, China

E-mail: zhezhang@scau.edu.cn

George E. Liu: Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD 20705, USA.

E-mail: George.Liu@usda.gov

Abstract

Genetic mutation and drift, coupled with natural and human-mediated selection and migration, have produced a wide variety of genotypes and phenotypes in farmed animals. We here introduce the Farm animal Genotype-Tissue Expression (FarmGTEx) Project, which aims to elucidate the genetic determinants of gene expression across 16 terrestrial and aquatic domestic species under diverse biological and environmental contexts. In each species, we aim to collect 50 tissues from 1,000 healthy adults and 200 additional animals representing a specific context. This Perspective provides an overview of the priorities of FarmGTEx and advocates for coordinated strategies of data analysis and resource-sharing initiatives. FarmGTEx aims to serve as a platform for fundamental discoveries of context-specific regulatory effects, which will deepen our understanding of molecular mechanisms underlying complex phenotypic variation within and between species. The knowledge and insights provided by FarmGTEx will contribute to improving sustainable agriculture-based food system, comparative biology, and eventual human biomedicine.

Keywords

FarmGTEx; Animal genetics; molQTL; Context-specific gene regulation; Complex trait

Main

The genomes of modern farmed animals reflect over 10,000 years of co-evolution with humans[1]. Over this time frame of livestock domestication, these species have adapted to a myriad of both natural and human-modified environments across the globe and have been selectively bred by humans for specific needs, such as food, clothing, and transportation[1]. This rich history of both natural and human-mediated selection has led to the establishment of thousands of genetically distinct populations, genetic lines, and breeds. According to the FAO Domestic Animal Diversity Information System (www.fao.org/dad-is), there are currently over 8,800 recognized breeds, representing 38 farmed animal species. These diverse genetic resources thus provide unparalleled opportunities to address fundamental gaps in our biological knowledge, such as the intricate pathways linking genome to phenome, environmental adaptation and domestication within and across species. In addition, several farmed animals have substantial potential as biomedical models for *in vivo* elucidation of human biology and diseases, due to their higher similarities to humans in anatomical size and structure, development, physiology, and immunology, compared to the widely adopted rodent model[2–8]. For instance, pigs have been well-recognized as human biomedical models for the identification of effective therapeutics for a range of diseases and as xenotransplant organ donors[2,7]. Sheep have been proposed as a model for cardiovascular and neurodegenerative disorders[7], cattle for viral infections (e.g., respiratory syncytial virus and papillomavirus[4,5]), and chickens for spontaneous ovarian cancer and embryonic development[3,8].

Recent developments in quantitative genetics and molecular genomics, coupled with advances in cost-effective sequencing technologies, are transforming our understanding of individual differences in DNA sequences (i.e., both single point mutations and structural variations) and their roles in shaping complex phenotypes[9]. The Animal Quantitative Trait Loci database (QTLdb) (www.animalgenome.org/cgi-bin/QTLdb/index, release 54) has catalogued 192,247 trait-associated loci in cattle, 55,688 in pigs, 18,602 in chickens, 4,743 in sheep, 2,216 in horses, and 2,201 in rainbow trout[10]. Similar to observations made in humans[11], the majority of these trait-associated variants have small to medium effects and lie in non-coding regions of the genome. Therefore, the underlying biological mechanisms (e.g., causal variants, genes, pathways, cell types, and tissues) through which they affect complex phenotypes remain largely unknown.

Population-based molecular QTL (molQTL) mapping, statistically associating genomic variants with molecular phenotypes (e.g., gene expression and epigenetic modifications), is a key approach to better understanding the biological mechanisms underlying complex phenotypes *in vivo*[11,12]. For example, the human Genotype-Tissue Expression (GTEx) project has explored the genetic control of gene expression across 54 non-diseased postmortem tissues from nearly 1,000 adults, highlighting the intricate interplay among variants, genes, tissues, and diseases[13]. Similar but relatively small-scale efforts have been undertaken in non-human primates (NHPs)[14], rats[15], and fruit flies[16]. Given that many trait-associated non-coding variants likely exert regulatory effects only within specific biological contexts or environmental conditions[17], several molQTL mapping projects have been initiated to study the dynamics of regulatory effects across distinct biological contexts, including ancestry, diet, developmental stage, and pathogen exposure[18–21].

Compared to laboratory animals (e.g., mice and fruit flies) and NHPs that are often used as surrogate and organism-level models[5], farmed animals offer multiple advantages for exploring context-specific regulatory effects at the natural population level. These advantages include: 1) large and outbred populations with ample genetic and phenotypic variations; 2) species, such as Holstein cattle and Duroc pigs, have well-defined pedigrees and selection histories; 3) the possibility of collecting numerous pan-tissue samples across diverse biological contexts (e.g., embryonic developmental stages in chickens); 4) exposure to various environmental conditions (e.g., diet, climate, and pathogen exposure) at the population scale, making them ideal for studying gene-environment interactions; 5) existence of wild progenitors and congeners for understanding artificial selection and domestication (e.g., wild boar and jungle fowl); 6) species, such as pigs, sheep and chickens, offer the potential for functional validation using gene editing or transgenesis *in vitro* and *vivo*; and 7) several farmed species possess physiological characteristics that are similar to humans (e.g., the cardiovascular and respiratory system of pigs, and in utero development of the immune system of cattle)[2,4].

Here, we present the Farm animal GTEx (FarmGTEx) project, a global initiative (including 88 universities and research institutes worldwide to date) that aims to generate a comprehensive catalogue of context-specific regulatory effects in farmed animals, including both terrestrial and aquatic species (**Fig. 1**). Since its inception in 2018, the FarmGTEx Consortium has completed several milestones in cattle, pigs, and chickens using public datasets[22–24], which have already offered valuable insights into the genetic and molecular

basis of complex phenotypes. For instance, expression (eQTL) and splicing QTL (sQTL) in 16 tissues together explained approximately 70% of the heritability across 37 complex traits in cattle[25]. Additionally, around 80% of genome-wide association study (GWAS) loci colocalized with at least one of five types of molQTL in pigs[23]. Transcriptome-wide association studies (TWAS) revealed that the expression of *ABCD4* in the brain and the alternative splicing of *MYO7B* in the small intestine had been associated with backfat thickness in pigs[23], and the expression of *KPNA3* in the retina had been linked to body weight in chickens[24]. These resources also accelerated research in comparative genomics and biology. For example, regulatory effects of one-to-one orthologous genes were found to be generally conserved across species[23,24,26]. Furthermore, cross-species meta-TWAS based on orthologous genes has revealed that conserved gene regulation underlies complex phenotypes of physiological similarity between species, such as pig backfat thickness versus human body mass index[23,24].

Capitalizing on the significant reduction in the costs of high-throughput sequencing, the FarmGTEx Consortium has begun to expand this pioneering work into the next decade (**Fig. 1**). It will generate paired genomic and functional genomic data at the population scale, with a focus on the transcriptomes, across various biological contexts and environmental conditions in a range of both farmed species and their wild progenitors and congeners. To this end, we aim to systematically annotate context-specific effects of genomic variants on various molecular phenotypes through molQTL mapping. Moreover, the comparative analyses facilitated by the FarmGTEx resource are anticipated to deepen our understanding of the evolutionary processes shaping molecular and phenotypic diversity across species. Such insights into animal biology and diseases, including reproductive and developmental biology, metabolic syndromes, and cardiovascular diseases, will also aid in addressing human diseases through the development of better animal models for understanding their pathogenesis and testing drug safety. For example, beef cows with excess intrafollicular androstenedione have been proposed as an animal model for elucidating the molecular mechanisms underpinning polycystic ovarian syndrome in women[27]. Pregnant sheep have been used to study the potential impacts of analgesics on maternal and foetal well-being[28].

Overview of the FarmGTEx project

The Pilot Phase of the FarmGTEx project (2018-2025) aims to explore tissue- and breed-specific regulatory effects on gene average expression and alternative splicing. By

leveraging public RNA-seq and whole-genome sequencing (WGS) data, we have been able to publish multi-tissue/breed atlases of molQTL (mainly for eQTL and sQTL) for cattle[22], pigs[23] and chickens[24] (**Table 1**). Related efforts in other farmed animal species are underway, including an additional six mammals (i.e., goats, sheep, horses, rabbits, camels, and donkeys), four birds (i.e., ducks, geese, turkeys and pigeons), and three fishes (i.e., Atlantic salmon, rainbow trout and large yellow croaker). The addition of other farmed animal species to the project will be considered. To comprehend tissue-specific effects, we aim to generate RNA-seq data from 50 tissues with paired WGS in 1,000 healthy adult animals per species. Here, we outline the future development of the project, which aims to progressively incorporate broader biological contexts and diverse environmental conditions through collecting 20-50 tissues from 200 additional animals in each context or condition (**Fig. 1**).

Phase 1 (2025-2030) is dedicated to unravelling sex- and development-specific regulatory effects. Such knowledge and insights would allow us to better understand how sex, development, and genetics interact to affect complex phenotypes. A substantial body of multi-tissue RNA-seq data has been or is being collected with paired WGS data and rich metadata, spanning both sexes and multiple developmental stages. For instance, to explore sex-specific regulatory effects, we have generated 8,000 RNA-seq samples from 40 tissues in 200 healthy rabbits (a commercial breed at the age of 75 days), including 100 male and 100 female animals. Similarly, we collected 9,000 RNA-seq samples from 30 tissues in 300 healthy chickens (a Chinese village breed at the age of 90 days), including 150 male and 150 female animals. To investigate postnatal development-specific regulatory effects and complement human and NHP developmental GTEx resources, we generated 30,000 RNA-seq samples from 50 tissues across four developmental stages (150 animals per stage) in a commercial pig population. Similarly, we produced 18,000 RNA-seq samples from 20 tissues across three developmental stages (300 animals per stage) in a commercial chicken breed. To study embryonic development-specific regulatory effects, we plan to collect RNA-seq samples from 20-50 tissues at three embryonic stages (i.e., early, middle, and late stages) in 200 animals per stage per species. Furthermore, we are expanding similar study designs to include other farmed species, such as sheep, goats, and ducks.

To elucidate the effects of environment and disease on gene expression, Phase 2 (2030-2035) will undertake a significant expansion to encompass multiple environmental conditions (50 tissues in 200 animals per condition), including alternative diets, climates, healthy status and pathogen exposures, at single-cell resolution and in a spatially resolved

manner. We have started building the farmed animal cell atlas (systemic annotation of cell types and states across the whole body) by collecting single-cell/nuclei RNA-seq (sc/snRNA-seq) samples from 50 tissues in several species, including cattle, sheep, goats, pigs, rabbits, and chickens[29,30]. By leveraging these cell atlases, we will deconvolute the cell type/state composition of bulk RNA-seq samples to study the regulatory effects at single-cell resolution and further resolve regulatory element annotation initiated in the Functional Annotation of Animal Genomes (FAANG) project[31]. Beyond the transcriptome, we will consider additional molecular phenotypes, starting with DNA methylation, chromatin accessibility, protein abundance, and metabolite profile. Moreover, we aim to maintain flexibility within the FarmGTEx framework to adapt to the rapidly evolving landscape of omics sequencing technologies. For instance, future efforts to generate population-level single-cell and spatial omics data, alongside the use of *in vitro* models (e.g., organoids — self-organized 3D tissue cultures derived from stem cells), will provide valuable opportunities to probe the spatiotemporal specificity of regulatory effects directly. This future single-cell refinement of FarmGTEx will align with our core mission of elucidating fundamental biological mechanisms, while also developing practical applications in farmed animals and human biomedical research.

Metadata collection

To enhance the quality and reusability of the data resources generated by FarmGTEx, we are implementing rigorous standards for samples and experimental metadata, as outlined by the FAANG project[32]. This involves the thorough documentation of the attributes of animals used in a project (e.g., species, breed, sex, age, and health status), specimens (e.g., tissue, cell, time of collection, storage, and histological images), and experimental assays (e.g., sample storage, assay type, extraction protocol, RNA integrity, library preparation, and sequencing strategy). These metadata descriptions will use standardized ontology terms, such as the experimental factor ontology and the uber-anatomy ontology[32], to ensure consistency and interoperability across datasets and projects. In addition, detailed experimental protocols will be stored separately but referenced in the metadata submissions to maintain reproducibility. Beyond generating data directly from groups participating in FarmGTEx, we will also integrate any existing datasets that meet our metadata standards into a centralized data resource portal to benefit the broader research community. To promote 'best practices' in data archiving, we plan to continuously advance metadata standards, support data

submissions, and develop user-friendly tools to facilitate the deposition and validation of metadata[32].

Analytical approaches used in FarmGTEx

In FarmGTEx, the primary data analysis tasks include genomic variant calling, molecular phenotyping, missing data imputation, batch effect inference, and molQTL mapping (**Fig. 2**). Since sequencing technologies and analytical methodologies are constantly evolving and improving, we will maintain and update FarmGTEx computational and statistical approaches with systematic version control, as outlined by the nf-core community[33], to promote 'best practices' in data analysis, integration and interpretation in this field.

Genomic variant calling: Within FarmGTEx, different technologies are used for genotyping animals. The Pilot Phase focuses on common SNPs (with minor allele frequencies, MAF > 0.05) that are derived from RNA-seq, SNP arrays or low-coverage sequencing technology, followed by genotype imputation[22–24]. In FarmGTEx Phase 1 and 2, we plan to use high-coverage WGS to detect rare and somatic mutations (MAF < 0.05)[34], as well as long-read sequencing technologies coupled with graph-based pangenome approaches to genotype short insertion and deletion (InDel), short tandem repeats (STRs), and structural variants (SVs)[35,36].

Molecular phenotyping: Gene expression is a complex biological process that encompasses transcription, post-transcriptional modifications, and translation. In the FarmGTEx Pilot Phase, we focus on average gene expression and alternative splicing owing to the limitations of short-read RNA-seq (e.g., being insufficient to quantify individual isoforms)[37]. In Phase 1, we will apply long-read RNA-seq to quantify the abundance of individual isoforms, haplotype-specific transcription, and gene-isoform co-expression networks. Finally, in Phase 2, we will consider additional layers of gene expression information at both bulk tissue and single-cell levels, such as epigenetic modifications (e.g., DNA methylation and chromatin accessibility), RNA modifications (e.g., N6-methyladenosine and 7-methylguanosine), protein abundance, and metabolite levels.

Imputation of missing data and removal of batch effects: To increase statistical power and reduce false-positive rates in molQTL mapping, it is critical to impute missing data, including genotypes, molecular phenotypes, and metadata information. In the subsequent FarmGTEx phases, through integrating long-read WGS data, we will incorporate STRs and SVs in the genotype reference panels to impute them in the target populations[38]. We will also develop

and benchmark methods to infer missing or mislabelled molecular phenotypes and metadata (e.g., sex and developmental stage), particularly for those in the public domain, such as hypergraph factorization for imputing gene expression from missing tissues[39]. To account for batch effects caused by local and global population stratification, we employ genotype principal components (PCs) and genetic relationship matrices[23]. Additionally, we use phenotype PCs to account for non-genetic unmeasured confounders (e.g., cell type/state composition)[40–42].

***MolQTL mapping*:** In the Pilot Phase of FarmGTEx, due to the limited sample size (n < 1,000), we mainly considered *cis*-effects of common SNPs (SNPs within 1 Mb of a transcript start site, TSS). In Phases 1 and 2, once the sample size exceeds 1,000, we will map *trans*-molQTL and explore how they interact with *cis*-molQTL to affect molecular and complex phenotypes. In addition, we will systematically explore context-specific effects of both rare and common variants, and utilize available cell-type specific expression and epigenetic data to resolve the cell type in which these variants function and potentially interact[43,44]. Although it is computationally intensive, we prefer linear mixed models (LMMs) to simple linear models for mapping molQTL, particularly *trans*-molQTL, in farmed animal populations with complex familial relatedness[45]. We are developing the Omics Genetic Analysis (OmiGA) toolkit to efficiently and effectively implement LMMs in molQTL mapping (omiga.farmgtex.org)[46]. Ultimately, we will integrate molQTL mapping with other computational approaches, such as allelic imbalance analyses and large language DNA models (e.g., DNABERT and DNAGPT)[47,48], to comprehend the regulatory landscape of genomic variants.

Applications of FarmGTEx

Potential scientific and social impacts of FarmGTEx resources include fundamental biology (e.g., sex- and development-specific gene regulation), animal precision breeding, local environmental adaptation, domestication, and human health (**Fig. 3**). To improve our understanding of the genetic and molecular architecture underpinning complex phenotypes in farmed animals, we will integrate FarmGTEx context-specific molQTL with GWAS results via various integrative statistical approaches, including TWAS, colocalization, and Mendelian randomization. This will allow us to systematically link molecular phenotypes to complex traits in appropriate biological contexts and environmental conditions, enabling the detection of genes, tissues, and mechanisms involved in phenotypic determination. Additionally, the rich selection history and population structures of farmed animals will

enable us to comprehend the genetic architectures underlying both molecular and complex phenotypes. For example, systematic breeding of farmed animals has resulted in directional selection for specific phenotypes, such as milk production in dairy cattle and egg production in laying hens. FarmGTEx will develop new methodologies tailored specifically to reconstruct the regulatory mechanisms of complex traits, particularly for those of economic importance, in farmed animals. We propose that knowledge and insights into these critical genes and mechanisms will have a positive impact on breeding strategies. For instance, advanced statistical models can be developed to use the functional annotation generated by FarmGTEx for genetic evaluation, particularly in predicting performance for animals from genetically distant populations or those several generations away from the training set. Previous studies showed that incorporating regulatory variants into genomic prediction models has made the genomic selection of cattle more accurate and efficient[49,50]. Furthermore, the identified causal variants and genes can be exploited to improve productivity, reproductivity and disease resistance through creating 'ultimate' genotypes efficiently and effectively via gene editing or synthetic biology[51].

Through incorporating large-scale genotypes and ancient DNA that extend back in time[52], the per-gene predictive models of molecular phenotypes provided by FarmGTEx will allow us to predict molecular phenotypes in global farmed animal breeds and their wild progenitors and congeners. This will then enable us to explore the polygenic adaptation of molecular phenotypes in distinct environmental conditions and the common patterns of functional genomic alterations caused by domestication across breeds and species. The knowledge and insights obtained in this way will allow us to breed animals to be more climate smart and disease resistant. Furthermore, the comparative information delivered by these population-level functional genomics studies of farmed and wild animals will advance our understanding of recent human adaptation and evolution[53,54].

Besides their impacts on agriculture, the rich and unique resources provided by FarmGTEx will also hold promise for comparative genomics through integrating resources from biodiversity genome sequencing projects, such as the Zoonomia[55] and Earth BioGenome Projects[56]. Furthermore, farmed animals offer a unique opportunity to study complex traits and mechanisms at a population scale in a context where ethical and/or regulatory concerns may be restrictive in humans and NHPs, such as early development and gene editing. Large language and deep-learning models, such as DeepGCF[57], can be used to explore the conservation and evolution of regulatory variants and elements, which will

provide extensive insights into the evolutionary mechanisms underlying shared and specialized traits across species. Given that synteny and LD patterns of genes are not conserved among species, cross-species meta-TWAS analysis can help discriminate which of the multiple linked genes is the most promising candidate for physiologically similar complex traits between species. For instance, we have demonstrated that meta-TWAS analysis between pig backfat thickness and human body weight revealed novel genes associated with human body weight[23]. The further insights into the molecular mechanisms of animal complex phenotypes provided by FarmGTEx thus can help to advance our understanding of context-specific regulation and origin of diseases relevant to humans, potentially leading to diagnostic and therapeutic advances.

Resource sharing and outreach

The FarmGTEx Project is dedicated to promoting open science and international collaboration to advance both fundamental and applied genetics research in farmed animals. It will thus adhere to data sharing principles of making data Findable, Accessible, Interoperable, and Reusable (FAIR)[58]. To that end, four key strategies have been implemented to share all the datasets, resources, and results generated by the project: **1)** FarmGTEx will use public repositories, such as NCBI-SRA, CNCB-NGDC-GSA, and EMBL-ENA, to disseminate raw sequencing and metadata, allowing direct access for further investigations. **2)** The central FarmGTEx Web Portal (www.farmgtex.org) summarizes all resources from the project, including publications, computer code, and public servers with specific functions. For instance, the PigGTEx (piggtex.farmgtex.org) and ChickenGTEx (chicken.farmgtex.org) servers allow any researcher to explore and download all processed data and results. The FarmGTEx-TWAS server (twas.farmgtex.org) supports online interactive TWAS analyses, while PigBiobank (pigbiobank.farmgtex.org) facilitates online queries for results from integrative dissection of various complex phenotypes in pigs. Additionally, processed datasets will be shared on public repositories like Zenodo and Figshare for wider accessibility and application. **3)** We will post manuscripts on preprint servers (e.g., bioRχiv) before submission, and release all computer codes under an open-source license. These workflows will be made available using containerized shareable pipeline framework tools, such as Snakemake[59], Nextflow[60], and code-sharing platforms, such as GitHub and Bioconductor (github.com/FarmGTEx). **4)** We plan to support interactive real-time analyses to conduct self-defined analyses using cloud-based platforms, such as AnVIL[61], and fine-tune ChatGPT application programming interface to assist users.

Understanding genomic variation and downstream functional impacts is a complex challenge that demands global and interdisciplinary collaboration. Therefore, FarmGTEx fosters partnerships with research consortia, genomics companies and service providers, breeding industry leaders, and the broader scientific community to advance our understanding. Researchers interested in FarmGTEx can seek interactions and collaborations (50 researchers have joined already) *via* the FarmGTEx web portal (www.farmgtex.org/) and the FAANG-FarmGTEx Task Force (www.faang.org/tf?name=FarmGTEx) to actively participate in working groups and other initiatives within the consortium. FarmGTEx has established close collaborations with several existing farm animal research consortia, such as the 1000 Bull Genomes[62], VarGoats[63], Bovine Pangenome[35], Bovine Long Read[38], and Chicken Genomic Diversity[64] Consortia, focusing on specific species, and with AG2PI[65], FAANG[66], and RT2T[67] aiming to address multiple farmed species. We have also established collaborations and communications with breeding industry partners (e.g., the US Council on Dairy Cattle Breeding, Hendrix Genetics, and Guangxi Yangxiang Co., Ltd) and policymakers (e.g., the US National Pork Board and Danish Dairy Board) toward the shared goal of integrating FarmGTEx resources with additional phenotypic and genotypic information in large breeding populations to improve the sustainability of farmed animal production. Finally, we will closely collaborate and coordinate with similar efforts in humans, such as GTEx[13] and IGVF[43], to enhance the application of FarmGTEx resources in human genetic and biomedical research.

Concluding remarks

Farmed animals are integral to global food production and security. Identifying the causal genetic and molecular factors governing economically important performance, sustainability, and welfare traits is key to optimizing precision breeding strategies. To accelerate this, the FarmGTEx consortium aims to pursue a coordinated strategy for resource generation and data analyses across multiple research groups that would not be possible through individual efforts alone. These collective efforts by FarmGTEx members from various disciplines are crucial in developing best practice recommendations for identifying, analysing and interpreting meaningful genomic effects.

The key outcomes from FarmGTEx are anticipated to include: 1) novel insights into the genetic mechanisms that control gene expression across a range of biological contexts, environmental conditions, and evolutionary timescales, including recent artificial selection (e.g., between breeds), and long-term evolutionary processes (e.g., within species and

subspecies); 2) a comprehensive catalogue of context-specific regulatory effects for detecting causal variants, genes and mechanisms involved in complex phenotypes, local environmental adaption and domestication; 3) uniformly processed multi-omics datasets for developing advanced integrative genomics methods to improve gene prioritization and genomic prediction of complex phenotypes; and 4) powerful web platforms for querying, visualizing, and downloading results, as well as accelerating the *in vitro* and/or *in vivo* functional follow-ups, such as massively parallel reporter assays and CRISPR-based editing. Looking ahead, the FarmGTEx consortium will be instrumental in developing precision breeding strategies, leading to the selection of farmed animals that are efficient, healthy, and environmentally friendly. This multidisciplinary team effort will also pave the way for innovative applications of farm animal models in human biomedical fields through in-depth comparative genomics and biology analyses.

Acknowledgements

We thank all the researchers who have contributed to the generation and analyses of both publicly available and newly generated data used in the pilot phase of FarmGTEx. We thank Dr. Irene Kaplow at the Department of Biological Sciences and Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University for valuable comments on the manuscript. L.F. was supported by Agriculture and Food Research Initiative Competitive grants nos. 2022-67015-36215 from the USDA National Institute of Food and Agriculture, and seed-funding from CellFood Hub (Aarhus University Foundation, AUFF).

Author contributions

L.Fang. took the lead in writing the manuscript with input from all authors. J.J.L., A.A.P., L.Frantz, H.Zhou, Z.Zhang, and G.E.L. performed critical revisions to the manuscript and made major contributions to the pilot phase of FarmGTEx. L.Fang., J.T., Q.L., Z.B., Y.H., X.P., and W.G., contributed to figure generation. S.L., D.G., B.L., Y.G., Y.H., M.G., Z.P., Y.Y., E.L.C., J.S., K.R., R.X., A.J.C., M.E.G., M.L., G.L., D.E.M., J.F.O., P.S., G.S., M.S.L., Z.J., X.P., W.G., Haihai Zhang., X.H., Yuebo Zhang., N.G., J.H., G.Y., Y.Liu, Z.T., P.Z., Y.Zhou., L.Fu., X.W., D.H., L.Liu, S.C., R.S.Y., X.S., C.X., H.C., L.M., J.B.C., R.L.B., Congjun L., C.P.V.T., B.D.R., N.B., J.L., W.L., L.G., X.Z., E.I.A., Y.Luo, L.Lin, O.C.X., M.D., R.P.M.A.C., M.G., O.M., M.A.M.G., J.E.K., C.K.T., F.M.M., D.R., E.G., M.A., A.C., M.B., G.T.K., J.L., C.F., M.F., Qishan W., Z.Hou., Qin W., F.Z., L.J., G.Z., Z.Zhou, R.Z., H.L., J.D., L.J., M.L., D.M., X.L., Y.C., X.Y., J.L., S.Z., Yi Zhang, X.D., D.S., H.S., Cong

L., Y.W., Y.J., D.W., W.Wang., X.F., Q.Z., K.L., Hao Zhang, N.Y., X.H., W.H., J.S., Y.W., J.Y., W.Wu., C.K., X.L., X.Z., S.K., W.L., H.K.I., E.S.B., B.R., and M.C.S. have given valuable comments and edits to the manuscript. All authors provided critical feedback and helped shape the manuscript.

Competing interests

The authors declare no competing interests.

Figure legends

Fig. 1. Overview of the FarmGTEx Project. To date, we are considering 16 farmed animal species, including eight mammals (i.e., cattle, pigs, goats, sheep, horses, rabbits, camels, and donkeys), five birds (i.e., chickens, ducks, geese, turkeys, and pigeons), and three fishes (i.e., large yellow croaker, Atlantic salmon, and rainbow trout). Additional farmed species will be considered as the project progresses. The regulatory effects of genomic variants on different molecular phenotypes across biological contexts and environmental conditions will be systematically exploited using *cis*- and *trans*-molQTL mapping. Different goals and milestones for the three phases of FarmGTEx are outlined. * indicates that the Pilot Phases for these species have been completed. SNP: single nucleotide polymorphism; InDel: short insertion and deletion (length less than 50 nucleotides); SV: structural variants.

Fig. 2 The overall data analysis pipeline of FarmGTEx. The flowchart highlights the key stages and methodologies of data generation and analyses across the three phases of FarmGTEx, including: **1**) Genomic variant calling; **2**) Definition and quantification of molecular phenotypes; **3**) Missing data imputation; **4)** Batch effect inference and removal; and **5)** Molecular quantitative trait loci (molQTL) mapping via e.g., OmiGA (omiga.farmgtex.org) and ClipperQTL[46,68]. Examples of approaches and methods used for each stage are shown in brackets. WGCNA: weighted gene co-expression network analysis; ATAC-seq: assay for transposase-accessible chromatin with sequencing; WGBS: whole-genome bisulfite sequencing; GC-MS: gas chromatography-mass spectrometry; LC-MS: liquid chromatography-mass spectrometry; MPRA: massively parallel reporter assay; CRISPR: clustered regularly interspaced short palindromic repeats; PC: principal component. GRM: genetic relationship matrix; GLIMPSE: Genotype Likelihoods IMputation and PhaSing mEthod[69]; PLINK: a tool for whole-genome association and linkage analyses[70]; GCTA: A Tool for Genome-wide Complex Trait Analysis[71]; phASER: phasing and allele specific expression from RNA-seq[72]. SuSiEx; an enhanced tool for cross-ancestry fine-mapping of causal variants based on "sum of single effects"[73]; dap-g: Deterministic Approximation of Posteriors – Genetics[74]; DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome[75]; DNAGPT: DNA Generative Pre-trained Transformer[76].

Fig. 3. Potential applications of FarmGTEx resources. Insights into the dynamic landscape of the regulatory effects of genomic variants across diverse biological contexts (e.g., sex and

developmental stage) can advance our understanding of how genetics, development and sex interact with complex phenotypes. Identified causal variants and genes for complex traits and diseases can be used in precision animal breeding via gene editing and synthetic biology. Elucidating the genetic control of molecular phenotypes in global farmed animal breeds and their wild progenitors and congeners can enable us to understand the molecular mechanisms underlying domestication and local environmental adaptation. Comparative analyses of farmed animals and humans across multiple biological layers can help to develop suitable animal models for studying human biology, disease, and xenotransplantation.

Table1. Summary of tissue sample size, eGene and sGene in the Pilot Phase of CattleGTEx, PigGTEx and ChickenGTEx. *N,* sample size; - not applicable; eGene, a gene with at least one significant expression quantitative trait loci (eQTL); sGene, a gene with at least one significant splicing QTL (sQTL); Brain, all the brain regions; CattleGTEx ref.[22]; PigGTEx ref.[23]; ChickenGTEx ref.[24].

References:

1. Frantz, L. A. F., Bradley, D. G., Larson, G. & Orlando, L. Animal domestication in the era of ancient genomics. *Nature Reviews Genetics 2020 21:8* **21**, 449–460 (2020).

2. Lunney, J. K. *et al.* Importance of the pig as a human biomedical model. *Sci Transl Med* **13**, (2021).

3. Davey, M. G. & Tickle, C. The chicken as a model for embryonic development. *Cytogenet Genome Res* **117**, 231–239 (2007).

4. Baldwin, C. L. & Telfer, J. C. The bovine model for elucidating the role of γδ T cells in controlling infectious diseases of importance to cattle and humans. *Mol Immunol* **66**, 35–47 (2015).

5. Saied, A. R. A., Metwally, A. A., Mohamed, H. M. A. & Haridy, M. A. M. The contribution of bovines to human health against viral infections. *Environ Sci Pollut Res Int* **28**, 46999 (2021).

6. Spencer, T. E. *et al.* Future of biomedical, agricultural, and biological systems research using domesticated animals. *Biol Reprod* **106**, 629–638 (2022).

7. Banstola, A. & Reynolds, J. N. J. The Sheep as a Large Animal Model for the Investigation and Treatment of Human Disorders. *Biology (Basel)* **11**, (2022).

8. Hawkridge, A. M. The chicken model of spontaneous ovarian cancer. *Proteomics Clin Appl* **8**, 689–699 (2014).

9. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

10. Hu, Z. L., Park, C. A. & Reecy, J. M. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res* **50**, D956–D961 (2022).

11. Lappalainen, T., Li, Y. I., Ramachandran, S. & Gusev, A. Genetic and molecular architecture of complex traits. *Cell* **187**, 1059–1075 (2024).

12. Aguet, F. *et al.* Molecular quantitative trait loci. *Nature Reviews Methods Primers 2023 3:1* **3**, 1–22 (2023).

13. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979)* **369**, 1318–1330 (2020).

14. Jasinska, A. J. *et al.* Genetic variation and gene expression across multiple tissues and developmental stages in a nonhuman primate. *Nature Genetics 2017 49:12* **49**, 1714–1721 (2017).

15. Munro, D. *et al.* The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. *Nucleic Acids Res* **50**, 10882–10895 (2022).

16. Pallares, L. F. *et al.* Saturating the eQTL map in Drosophila melanogaster: genome-wide patterns of cis and trans regulation of transcriptional variation in outbred populations. *bioRxiv* 2023.05.20.541576 (2023) doi:10.1101/2023.05.20.541576.

17. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends in Genetics* **37**, 109–124 (2021).

18. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. *Nature Genetics 2022 54:6* **54**, 817–826 (2022).

19. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).

20. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nature Genetics 2016 49:1* **49**, 139–145 (2016).

21. Lin, W. *et al.* Genetic regulatory effects in response to a high-cholesterol, high-fat diet in baboons. *Cell Genomics* **4**, 100509 (2024).

22. Liu, S. *et al.* A multi-tissue atlas of regulatory variants in cattle. *Nature Genetics 2022 54:9* **54**, 1438–1447 (2022).

23. Teng, J. *et al.* A compendium of genetic regulatory effects across pig tissues. *Nature Genetics 2024 56:1* **56**, 112–123 (2024).

24. Guan, D. *et al.* The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues. *bioRxiv* 2023.06.27.546670 (2023) doi:10.1101/2023.06.27.546670.

25. Xiang, R. *et al.* Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle. *Cell Genomics* **3**, 100385 (2023).

26. Zhao, R. *et al.* The conservation of human functional variants and their effects across livestock species. *Communications Biology 2022 5:1* **5**, 1–13 (2022).

27. Varcoe, T. J. *et al.* Considerations in selecting postoperative analgesia for pregnant sheep following fetal instrumentation surgery. *Animal Frontiers* **9**, 60–67 (2019).

28. Abedal-Majed, M. A. & Cupp, A. S. Livestock animals to study infertility in women. *Animal Frontiers* **9**, 28–33 (2019).

29. Chen, L. *et al.* Construction of a multi-tissue cell atlas reveals cell-type-specific regulation of molecular and complex phenotypes in pigs. *bioRxiv* 2023.06.12.544530 (2023) doi:10.1101/2023.06.12.544530.

30. Fang, L. *et al.* Cattle Cell Atlas: a multi-tissue single cell expression repository for advanced bovine genomics and comparative biology. (2024) doi:10.21203/RS.3.RS-4631710/V1.

31. Giuffra, E. & Tuggle, C. K. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu Rev Anim Biosci* **7**, 65–88 (2019).

32. Harrison, P. W. *et al.* FAANG, establishing metadata standards, validation and best practices for the farmed and companion animal community. *Anim Genet* **49**, 520–526 (2018).

33. Langer, B. E. *et al.* Empowering bioinformatics communities with Nextflow and nf-core. *bioRxiv* **18**, 2024.05.10.592912 (2024).

34. Cameron, D. L., Di Stefano, L. & Papenfuss, A. T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nature Communications 2019 10:1* **10**, 1–11 (2019).

35. Smith, T. P. L. *et al.* The Bovine Pangenome Consortium: democratizing production and accessibility of genome assemblies for global cattle breeds and other bovine species. *Genome Biol* **24**, 1–13 (2023).

36. Wang, T. *et al.* The Human Pangenome Project: a global resource to map genomic diversity. *Nature 2022 604:7906* **604**, 437–446 (2022).

37. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).

38. Nguyen, T. V. *et al.* In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. *Genetics Selection Evolution 2023 55:1* **55**, 1–15 (2023).

39. Viñas, R. *et al.* Hypergraph factorization for multi-tissue gene expression imputation. *Nature Machine Intelligence 2023 5:7* **5**, 739–753 (2023).

40. Zhou, H. J., Li, L., Li, Y., Li, W. & Li, J. J. PCA outperforms popular hidden variable inference methods for molecular QTL mapping. *Genome Biol* **23**, 1–17 (2022).

41. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols 2012 7:3* **7**, 500–507 (2012).

42. Mostafavi, S. *et al.* Normalizing RNA-Sequencing Data by Modeling Hidden Covariates with Prior Knowledge. *PLoS One* **8**, e68141 (2013).

43. Engreitz, J. M. *et al.* Deciphering the impact of genomic variation on function. *Nature 2024 633:8028* **633**, 47–57 (2024).

44. van der Wijst, M. G. P. *et al.* The single-cell eQTLGen consortium. *Elife* **9**, (2020).

45. Sul, J. H., Martin, L. S. & Eskin, E. Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genet* **14**, e1007309 (2018).

46. Teng, J. *et al.* OmiGA: A Toolkit for Ultra-efficient Molecular Trait Analysis in Complex Populations. *bioRxiv* 2024.12.19.629424 (2024) doi:10.1101/2024.12.19.629424.

47. Chen, K. M., Wong, A. K., Troyanskaya, O. G. & Zhou, J. A sequence-based global map of regulatory activity for deciphering human genetics. *Nature Genetics 2022 54:7* **54**, 940–949 (2022).

48. Liu, J. *et al.* Large language models in bioinformatics: applications and perspectives. *ArXiv* arXiv:2401.04155v1 (2024).

49. Xiang, R. *et al.* Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A* **116**, 19398–19408 (2019).

50. Fang, L. *et al.* Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics* **18**, 604 (2017).

51. Hayes, B. J. *et al.* Potential approaches to create ultimate genotypes in crops and livestock. *Nat Genet* **56**, (2024).

52. McHugo, G. P., Dover, M. J. & MacHugh, D. E. Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology 2019 17:1* **17**, 1–20 (2019).

53. Chen, M. *et al.* Evidence of Polygenic Adaptation in Sardinia at Height-Associated Loci Ascertained from the Biobank Japan. *Am J Hum Genet* **107**, 60 (2020).

54. Bouwman, A. C. *et al.* Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet* **50**, 362–367 (2018).

55. Genereux, D. P. *et al.* A comparative genomics multitool for scientific discovery and conservation. *Nature 2020 587:7833* **587**, 240–245 (2020).

56. Lewin, H. A. *et al.* The Earth BioGenome Project 2020: Starting the clock. *Proc Natl Acad Sci U S A* **119**, e2115635118 (2022).

57. Li, J. *et al.* Learning functional conservation between human and pig to decipher evolutionary mechanisms underlying gene expression and complex traits. *Cell Genomics* **3**, 100390 (2023).

58. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1* **3**, 1–9 (2016).

59. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

60. DI Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature Biotechnology 2017 35:4* **35**, 316–319 (2017).

61. Schatz, M. C. *et al.* Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space. *Cell Genomics* **2**, 100085 (2022).

62. Daetwyler, H. D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics 2014 46:8* **46**, 858–865 (2014).

63. Denoyelle, L. *et al.* VarGoats project: a dataset of 1159 whole-genome sequences to dissect Capra hircus global diversity. *Genetics Selection Evolution* **53**, 1–14 (2021).

64. Smith, J. *et al.* Fourth Report on Chicken Genes and Chromosomes 2022. *Cytogenet Genome Res* **162**, 405–528 (2022).

65. Tuggle, C. K. *et al.* The Agricultural Genome to Phenome Initiative (AG2PI): creating a shared vision across crop and livestock research communities. *Genome Biol* **23**, 1–11 (2022).

66. Clark, E. L. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol* **21**, 1–9 (2020).

67. Kalbeisch, T. *et al.* RT2T: A Global Collaborative Project to Study Chromosomal Evolution in the Suborder Ruminantia.

68. Zhou, H. J., Ge, X. & Li, J. J. ClipperQTL: ultrafast and powerful eGene identification method. *bioRxiv* 2023.08.28.555191 (2023) doi:10.1101/2023.08.28.555191.

69. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics 2021 53:1* **53**, 120–126 (2021).

70. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559 (2007).

71. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet* **88**, 76 (2011).

72. Castel, S. E., Mohammadi, P., Chung, W. K., Shen, Y. & Lappalainen, T. Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat Commun* **7**, 12817 (2016).

73. Yuan, K. *et al.* Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nature Genetics 2024 56:9* **56**, 1841–1850 (2024).

74. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. *Am J Hum Genet* **98**, 1114–1129 (2016).

75. Ji, Y., Zhou, Z., Liu, H. & Davuluri, R. V. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics* **37**, 2112–2120 (2021).

76. Zhang, D. *et al.* DNAGPT: A Generalized Pre-trained Tool for Multiple DNA Sequence Analysis Tasks. *bioRxiv* 2023.07.11.548628 (2024) doi:10.1101/2023.07.11.548628.