
This is the **accepted version** of the journal article:

Fang, Lingzhao [et al.]. «The Farm Animal Genotype-Tissue Expression (FarmGTEx) Project». *Nature genetics*, Vol. 57 (April 2025), p. 786-796 DOI 10.1038/s41588-025-02121-5

This version is available at <https://ddd.uab.cat/record/329398>

under the terms of the  ^{IN}
COPYRIGHT license.

The Farm Animal Genotype-Tissue Expression (FarmGTEx) Project

Lingzhao Fang^{1*}, Jinyan Teng², Qing Lin^{1,2}, Zhonghao Bai¹, Shuli Liu^{3,4}, Dailu Guan⁵, Bingjie Li⁶, Yahui Gao², Yali Hou⁷, Mian Gong⁷, Zhangyuan Pan⁷, Ying Yu⁸, Emily Clark⁹, Jacqueline Smith⁹, Konrad Rawlik¹⁰, Ruidong Xiang^{11,12,13}, Amanda J. Chamberlain^{11,14}, Michael E. Goddard¹¹, Mathew Littlejohn^{15,16}, Greger Larson¹⁷, David E. MacHugh^{18,19,20}, John F. O'Grady¹⁸, Peter Sørensen¹, Goutam Sahana¹, Mogens Sandø Lund¹, Zhihua Jiang²¹, Xiangchun Pan², Wentao Gong², Haihan Zhang²², Xi He²², Yuebo Zhang²², Ning Gao²², Jun He²², Guoqiang Yi²³, Yuwen Liu²³, Zhonglin Tang²³, Pengju Zhao²⁴, Yang Zhou^{25,26}, Liangliang Fu²⁵, Xiao Wang²⁷, Dan Hao²⁸, Lei Liu²⁶, Siqian Chen⁸, Robert S. Young^{29,30}, Xia Shen^{29,31,32}, Charley Xia^{33,34}, Hao Cheng⁵, Li Ma³⁵, John B. Cole^{36,37,38}, Ransom L. Baldwin IV³⁹, Cong-jun Li³⁹, Curtis P. Van Tassell³⁹, Benjamin D. Rosen³⁹, Nayan Bhowmik³⁹, Joan Lunney⁴⁰, Wansheng Liu⁴¹, Leluo Guan^{42,43}, Xin Zhao⁴⁴, Eveline M. Ibeagha-Awemu⁴⁵, Yonglun Luo^{46,47}, Lin Lin^{46,47}, Oriol Canela-Xandri⁴⁸, Martijn Derks⁴⁹, Richard P.M.A. Crooijmans⁴⁹, Marta Gòdia⁴⁹, Ole Madsen⁴⁹, Martien A.M. Groenen⁴⁹, James E. Koltes⁵⁰, Christopher K. Tuggle⁵⁰, Fiona M. McCarthy⁵¹, Dominique Rocha⁵², Elisabetta Giuffra⁵², Marcel Amills^{53,54}, Alex Clop^{53,55}, Maria Ballester⁵⁶, Gwenola Tosser-Klopp⁵⁷, Jing Li^{1,58}, Chao Fang⁵⁹, Ming Fang⁶⁰, Qishan Wang⁶¹, Zhuocheng Hou⁸, Qin Wang⁸, Fuping Zhao⁷, Lin Jiang⁷, Guiping Zhao⁷, Zhengkui Zhou⁷, Rong Zhou⁷, Hehe Liu⁶², Juan Deng⁶², Long Jin⁶², Mingzhou Li⁶², Delin Mo⁶³, Xiaohong Liu⁶³, Yaosheng Chen⁶³, Xiaolong Yuan², Jiaqi Li², Shuhong Zhao^{25,26}, Yi Zhang⁸, Xiangdong Ding⁸, Dongxiao Sun⁸, Huizeng Sun⁶¹, Cong Li⁶⁴, Yu Wang⁶⁴, Yu Jiang⁶⁴, Dongdong Wu⁶⁵, Wenwen Wang⁶⁶, Xinzhong Fan⁶⁶, Qin Zhang⁶⁶, Kui Li²³, Hao Zhang⁸, Ning Yang⁸, Xiaoxiang Hu⁶⁷, Wen Huang⁶⁸, Jiuzhou Song³⁵, Yang Wu⁶⁹, Jian Yang^{3,4}, Weiwei Wu⁷⁰, Claudia Kasper⁷¹, Xinfeng Liu^{1,72}, Xiaofei Yu⁷³, Leilei Cui^{74,75}, Xiang Zhou⁷⁶, Seyoung Kim⁷⁷, Wei Li⁷⁸, Hae Kyung Im⁷⁹, Edward S. Buckler^{80,81,82}, Bing Ren⁸³, Michael C. Schatz⁸⁴, 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

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

²State 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

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

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

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

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

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

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

⁹The Roslin Institute and Royal (Dick) School of Veterinary Studies R(D)SVS, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK

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

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

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

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

49 ¹⁴School of Applied Systems Biology, La Trobe University, Bundoora, VIC, 3083, Australia

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

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

52 ¹⁷The Palaeogenomics & Bio-Archaeology Research Network, School of Archaeology, University of Oxford,
53 Oxford, UK

54 ¹⁸UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin, D04 V1W8, Ire-
55 land

56 ¹⁹UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dub-
57 lin, D04 V1W8, Ireland

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

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

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

61 ²³Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124,
62 China

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

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

66 ²⁶Yazhouwan National Laboratory, Sanya City, Hainan Province, 572024, China

67 ²⁷Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan,
68 250100, China

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

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

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

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

75 ³²Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fu-
76 dan University, Guangzhou, China

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

78 ³⁴Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, UK

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

80 ³⁶Council on Dairy Cattle Breeding, Bowie, Maryland 20716, USA

81 ³⁷Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Pro-
82 gram, and the Genetics Institute, University of Florida, Gainesville, Florida 32611-0910, USA

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

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

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

- 87 ⁴¹Department of Animal Science, Center for Reproductive Biology and Health (CRBH), College of Agricultural
88 Sciences, The Pennsylvania State University, 311 AVBS Building, University Park, PA, 16802, USA
- 89 ⁴²Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G
90 2P5, Canada
- 91 ⁴³Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, V6T 1Z4,
92 Canada
- 93 ⁴⁴Department of Animal Science, McGill University, Sainte-Anne-de-Bellevue, Quebec, H9X 3V9, Canada
- 94 ⁴⁵Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada, Sherbrooke, Quebec, J1M
95 0C8, Canada
- 96 ⁴⁶Department of Biomedicine, Aarhus University, Aarhus, 8000, Denmark
- 97 ⁴⁷Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark
- 98 ⁴⁸MRC Human Genetics Unit at the Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh
99 EH4 2XU, UK
- 100 ⁴⁹Animal Breeding and Genomics, Wageningen University and Research, Wageningen, The Netherlands
- 101 ⁵⁰Department of Animal Science, Iowa State University, Ames, IA, 50011, USA
- 102 ⁵¹University of Arizona, Tucson, AZ, USA
- 103 ⁵²Paris-Saclay University, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France
- 104 ⁵³Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-
105 UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra 08193, Spain
- 106 ⁵⁴Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain
- 107 ⁵⁵Consejo Superior de Investigaciones Científicas, Barcelona, Catalonia, Spain
- 108 ⁵⁶Animal Breeding and Genetics Programme, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre
109 Marimon, Caldes de Montbui, Spain
- 110 ⁵⁷GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France
- 111 ⁵⁸School of Agriculture and Life Sciences, Kunming University, Kunming, 650214, China
- 112 ⁵⁹LC-Bio Technologies, Co., Ltd., Hangzhou, China
- 113 ⁶⁰Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture and Rural Affairs,
114 Jimei University, Xiamen, China
- 115 ⁶¹Department of Animal Science, College of Animal Sciences, Zhejiang University, Hangzhou 310058, China
- 116 ⁶²College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
- 117 ⁶³State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China
- 118 ⁶⁴College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China
- 119 ⁶⁵Key Laboratory of Genetic Evolution and Animal Models, Kunming Institute of Zoology, Chinese Academy
120 of Sciences, Kunming, 650223, China
- 121 ⁶⁶College of Animal Science and Technology, Shandong Agricultural University, Taian 271018, China
- 122 ⁶⁷State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural
123 University, Beijing, 100193, China
- 124 ⁶⁸Department of Animal Science, Michigan State University, East Lansing, MI, USA
- 125 ⁶⁹Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China

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

128 ⁷¹Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope Posieux, 1725,
129 Switzerland

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

132 ⁷³College of Marine Life Sciences, Ocean university of China

133 ⁷⁴School of Life Sciences, Nanchang University, Nanchang, China

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

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

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

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

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

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

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

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

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

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

148 ⁸⁵Department of Statistics and Data Science, University of California, Los Angeles, Los Angeles, CA, 90095,
149 USA

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

151 ⁸⁷Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich 80539,
152 Germany

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

155 ***Corresponding Authors:**

156 **Lingzhao Fang:** Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark
157 E-mail: lingzhao.fang@qgg.au.dk;

158 **Jingyi Jessica Li:** Department of Statistics and Data Science, University of California, Los Angeles, Los
159 Angeles, CA, 90095, USA
160 E-mail: lijy03@g.ucla.edu;

161 **Abraham Palmer:** Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
162 E-mail: aap@ucsd.edu;

163 **Laurent Frantz:** Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History
164 of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany
165 E-mail: laurent.frantz@lmu.de;

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

167 E-mail: hzhou@ucdavis.edu;
168 **Zhe Zhang**: College of Animal Science, South China Agricultural University, Guangzhou, China
169 E-mail: zhezhang@scau.edu.cn
170 **George E. Liu**: Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD 20705, USA.
171 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 humans¹. 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 transportation¹. 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 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 model²⁻⁸. For examples, 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 disorders⁷, cattle for viral infections (e.g., respiratory syncytial virus and papillomavirus^{4,5}), and chickens for spontaneous ovarian cancer and embryonic development^{3,8}. These examples highlight the pivotal role of farm animal research in advancing our understanding of human health.

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 phenotypes⁹. 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 trout¹⁰. Similar to observations made in humans¹¹, 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 diseases¹³. Similar but relatively small-scale efforts have been undertaken in non-human primates (NHPs)¹⁴, rats¹⁵, and fruit flies¹⁶. Given that many trait-associated non-coding variants likely exert regulatory effects only within specific biological contexts or environmental conditions¹⁷, 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 models⁵, 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 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 cattle²⁵. Additionally, around 80% of genome-wide association study (GWAS) loci colocalized with at least one of five types of molQTL in pigs²³. 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 pigs²³, and the expression of *KPNA3* in the retina had been linked to body weight in chickens²⁴. 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 polycystic ovarian syndrome in women²⁷. Pregnant sheep have been used to study the potential impacts of analgesics on maternal and foetal well-being²⁸. Pigs are key models for cardiovascular diseases², while chickens play a critical role in studying influenza A virus transmission and its pandemic risks²⁹.

Overview of the FarmGTEx project

In the Pilot Phase of the FarmGTEx project (2018-2025), the aim is 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 cattle²², pigs²³ and chickens²⁴ (**Table 1**). Related efforts in other farmed animal species are underway, including in 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 information in 1,000 healthy adult animals per species. Here, we outline the future developments of the project, where we aim 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^{30,31}. By leveraging these cell atlases, we will deconvolute the cell-type or cell-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) project³². 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 project³³. 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 ontology³³, 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 metadata³³.

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 community³⁴, 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 and Telomere-to-Telomere (T2T) genome assemblies to detect rare and somatic mutations ($MAF < 0.05$)³⁵, 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)^{36,37}.

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)³⁸. 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 populations³⁹. 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 tissues⁴⁰. To account for batch effects caused by local and global population stratification, we employ genotype principal components (PCs) and genetic relationship matrices²³. Additionally, we use phenotype PCs to account for non-genetic unmeasured confounders (e.g., cell type/state composition)^{41–43}.

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^{44,45}. 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 relatedness⁴⁶. We are developing the Omics Genetic Analysis (OmiGA) toolkit to efficiently and effectively implement LMMs in molQTL mapping (omiga.farmgtex.org)⁴⁷. 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)⁴⁸, to fully understand the regulatory landscape of genomic variants.

Applications of FarmGTEx

FarmGTEx resources have potential scientific and social impacts on 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 context-specific molQTL obtained from FarmGTEx with GWAS results via various integrative statistical approaches, including TWAS, colocalization, Mendelian randomization, and artificial intelligence (AI) based approaches (e.g., large language models). This will allow us to systematically link molecular phenotypes to complex phenotypes 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 determine 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 in farmed animals, particularly for those of economic importance. 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 integrating AI, gene editing, and synthetic biology⁵¹.

Through incorporating large-scale genotypes and ancient DNA that extend back in time⁵², the per-gene predictive models of molecular phenotypes obtained through 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 that are adaptable

to climate change and disease resistant. Furthermore, the comparative information delivered by these population-level functional genomics studies of farmed and wild animals can also help to 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 Zoonomia⁵⁵ and Earth BioGenome Projects⁵⁶. 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 DeepGCF⁵⁷, 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 among 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 weight²³. 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, such as Zenodo and Figshare, for wider accessibility and application. **3)** We will post manuscripts on preprint servers (e.g., bioRxiv) 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 Snakemake⁵⁹, Nextflow⁶⁰, 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 AnVIL⁶¹, 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 Genomes⁶², VarGoats⁶³, Bovine Pangenome³⁶, Bovine Long Read³⁹, and Chicken Genomic Diversity⁶⁴ Consortia, focusing on specific species, and with AG2PI⁶⁵, FAANG⁶⁶, and RT2T⁶⁷ 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 GTEx¹³ and IGVF⁴⁴, 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 (e.g., AI-based approaches) 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 analyses 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^{47,68}. 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 mEthod⁶⁹; PLINK: a tool for whole-genome association and linkage analyses⁷⁰; GCTA: A Tool for Genome-wide Complex Trait Analysis⁷¹; phASER: phasing and allele specific expression from RNA-seq⁷². 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 – Genetics⁷⁴; DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome⁷⁵; DNAGPT: DNA Generative Pre-trained Transformer⁷⁶.

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 sample analysis in the Pilot Phase of FarmGTEx. *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, see ²²; PigGTEx, see ²³; ChickenGTEx, see 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 $\gamma\delta$ 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).

- 648 18. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during CD4+
649 T cell activation. *Nature Genetics* 2022 54:6 **54**, 817–826 (2022).
- 650 19. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain In-
651 forms Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
- 652 20. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci
653 in whole blood. *Nature Genetics* 2016 49:1 **49**, 139–145 (2016).
- 654 21. Lin, W. *et al.* Genetic regulatory effects in response to a high-cholesterol, high-fat diet in ba-
655 boons. *Cell Genomics* **4**, 100509 (2024).
- 656 22. Liu, S. *et al.* A multi-tissue atlas of regulatory variants in cattle. *Nature Genetics* 2022 54:9 **54**,
657 1438–1447 (2022).
- 658 23. Teng, J. *et al.* A compendium of genetic regulatory effects across pig tissues. *Nature Genetics*
659 2024 56:1 **56**, 112–123 (2024).
- 660 24. Guan, D. *et al.* The ChickenGTEx pilot analysis: a reference of regulatory variants across 28
661 chicken tissues. *bioRxiv* 2023.06.27.546670 (2023) doi:10.1101/2023.06.27.546670.
- 662 25. Xiang, R. *et al.* Gene expression and RNA splicing explain large proportions of the heritability
663 for complex traits in cattle. *Cell Genomics* **3**, 100385 (2023).
- 664 26. Zhao, R. *et al.* The conservation of human functional variants and their effects across live-
665 stock species. *Communications Biology* 2022 5:1 **5**, 1–13 (2022).
- 666 27. Varcoe, T. J. *et al.* Considerations in selecting postoperative analgesia for pregnant sheep fol-
667 lowing fetal instrumentation surgery. *Animal Frontiers* **9**, 60–67 (2019).
- 668 28. Abedal-Majed, M. A. & Cupp, A. S. Livestock animals to study infertility in women. *Animal*
669 *Frontiers* **9**, 28–33 (2019).
- 670 29. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza
671 A virus species specificity. *Nat Rev Microbiol* **17**, 67–81 (2019).
- 672 30. Chen, L. *et al.* Construction of a multi-tissue cell atlas reveals cell-type-specific regulation of
673 molecular and complex phenotypes in pigs. *bioRxiv* 2023.06.12.544530 (2023)
674 doi:10.1101/2023.06.12.544530.
- 675 31. Fang, L. *et al.* Cattle Cell Atlas: a multi-tissue single cell expression repository for advanced
676 bovine genomics and comparative biology. (2024) doi:10.21203/RS.3.RS-4631710/V1.
- 677 32. Giuffra, E. & Tuggle, C. K. Functional Annotation of Animal Genomes (FAANG): Current
678 Achievements and Roadmap. *Annu Rev Anim Biosci* **7**, 65–88 (2019).
- 679 33. Harrison, P. W. *et al.* FAANG, establishing metadata standards, validation and best practices
680 for the farmed and companion animal community. *Anim Genet* **49**, 520–526 (2018).
- 681 34. Langer, B. E. *et al.* Empowering bioinformatics communities with Nextflow and nf-core. *bio-*
682 *Rxiv* **18**, 2024.05.10.592912 (2024).
- 683 35. Cameron, D. L., Di Stefano, L. & Papenfuss, A. T. Comprehensive evaluation and characterisa-
684 tion of short read general-purpose structural variant calling software. *Nature Communica-*
685 *tions* 2019 10:1 **10**, 1–11 (2019).

- 686 36. Smith, T. P. L. *et al.* The Bovine Pangenome Consortium: democratizing production and acces-
687 sibility of genome assemblies for global cattle breeds and other bovine species. *Genome Biol*
688 **24**, 1–13 (2023).
- 689 37. Wang, T. *et al.* The Human Pangenome Project: a global resource to map genomic diversity.
690 *Nature* 2022 604:7906 **604**, 437–446 (2022).
- 691 38. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcrip-
692 tome sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**,
693 (2019).
- 694 39. Nguyen, T. V. *et al.* In it for the long run: perspectives on exploiting long-read sequencing in
695 livestock for population scale studies of structural variants. *Genetics Selection Evolution* 2023
696 55:1 **55**, 1–15 (2023).
- 697 40. Viñas, R. *et al.* Hypergraph factorization for multi-tissue gene expression imputation. *Nature*
698 *Machine Intelligence* 2023 5:7 **5**, 739–753 (2023).
- 699 41. Zhou, H. J., Li, L., Li, Y., Li, W. & Li, J. J. PCA outperforms popular hidden variable inference
700 methods for molecular QTL mapping. *Genome Biol* **23**, 1–17 (2022).
- 701 42. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expres-
702 sion residuals (PEER) to obtain increased power and interpretability of gene expression anal-
703 yses. *Nature Protocols* 2012 7:3 **7**, 500–507 (2012).
- 704 43. Mostafavi, S. *et al.* Normalizing RNA-Sequencing Data by Modeling Hidden Covariates with
705 Prior Knowledge. *PLoS One* **8**, e68141 (2013).
- 706 44. Engreitz, J. M. *et al.* Deciphering the impact of genomic variation on function. *Nature* 2024
707 633:8028 **633**, 47–57 (2024).
- 708 45. van der Wijst, M. G. P. *et al.* The single-cell eQTLGen consortium. *Elife* **9**, (2020).
- 709 46. Sul, J. H., Martin, L. S. & Eskin, E. Population structure in genetic studies: Confounding factors
710 and mixed models. *PLoS Genet* **14**, e1007309 (2018).
- 711 47. Teng, J. *et al.* OmiGA: A Toolkit for Ultra-efficient Molecular Trait Analysis in Complex Popula-
712 tions. *bioRxiv* 2024.12.19.629424 (2024) doi:10.1101/2024.12.19.629424.
- 713 48. Liu, J. *et al.* Large language models in bioinformatics: applications and perspectives. *ArXiv*
714 arXiv:2401.04155v1 (2024).
- 715 49. Xiang, R. *et al.* Quantifying the contribution of sequence variants with regulatory and evolu-
716 tionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A* **116**, 19398–19408
717 (2019).
- 718 50. Fang, L. *et al.* Use of biological priors enhances understanding of genetic architecture and ge-
719 nomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics*
720 **18**, 604 (2017).
- 721 51. Hayes, B. J. *et al.* Potential approaches to create ultimate genotypes in crops and livestock.
722 *Nat Genet* **56**, (2024).
- 723 52. McHugo, G. P., Dover, M. J. & MacHugh, D. E. Unlocking the origins and biology of domestic
724 animals using ancient DNA and paleogenomics. *BMC Biology* 2019 17:1 **17**, 1–20 (2019).

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

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

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

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

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

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

737 59. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinform-*
738 *atics* **28**, 2520–2522 (2012).

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

741 61. Schatz, M. C. *et al.* Inverting the model of genomics data sharing with the NHGRI Genomic
742 Data Science Analysis, Visualization, and Informatics Lab-space. *Cell Genomics* **2**, 100085
743 (2022).

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

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

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

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

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

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

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

758 69. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation
759 of low-coverage sequencing data using large reference panels. *Nature Genetics* 2021 53:1 **53**,
760 120–126 (2021).

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

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

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

767 73. Yuan, K. *et al.* Fine-mapping across diverse ancestries drives the discovery of putative causal
768 variants underlying human complex traits and diseases. *Nature Genetics* **56**, 1841–
769 1850 (2024).

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

772 75. Ji, Y., Zhou, Z., Liu, H. & Davuluri, R. V. DNABERT: pre-trained Bidirectional Encoder Represen-
773 tations from Transformers model for DNA-language in genome. *Bioinformatics* **37**, 2112–2120
774 (2021).

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

777