nature genetics

Perspective

https://doi.org/10.1038/s41588-025-02121-5

The Farm Animal Genotype–Tissue Expression (FarmGTEx) Project

Received: 21 May 2024

Accepted: 6 February 2025

Published online: 17 March 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

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. For each species, we aim to collect multiomics data, particularly genomics and transcriptomics, from 50 tissues of 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 investigating context-specific regulatory effects, which will deepen our understanding of molecular mechanisms underlying complex phenotypes. The knowledge and insights provided by FarmGTEx will contribute to improving sustainable agriculture-based food systems, comparative biology and eventual human biomedicine.

The genomes of modern farmed animals reflect over 10,000 years of coevolution 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 (https://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 greater similarities to humans in anatomical size and structure, development, physiology and immunology than the widely adopted rodent model[2–8]. 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 disorders[7], cattle for viral infections (for example, 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 (that is, both single point mutations and structural variations) and their roles in shaping complex phenotypes[9]. The Animal Quantitative Trait Loci (QTL) database (https://www.animalgenome.org/cgi-bin/QTLdb/index, release 54) has cataloged 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 noncoding regions of the genome. Therefore, the underlying biological mechanisms

e-mail: lingzhao.fang@qgg.au.dk; lijy03@g.ucla.edu; aap@ucsd.edu; laurent.frantz@lmu.de; hzhou@ucdavis.edu; zhezhang@scau.edu.cn; george.liu@usda.gov

https://doi.org/10.1038/s41588-025-02121-5

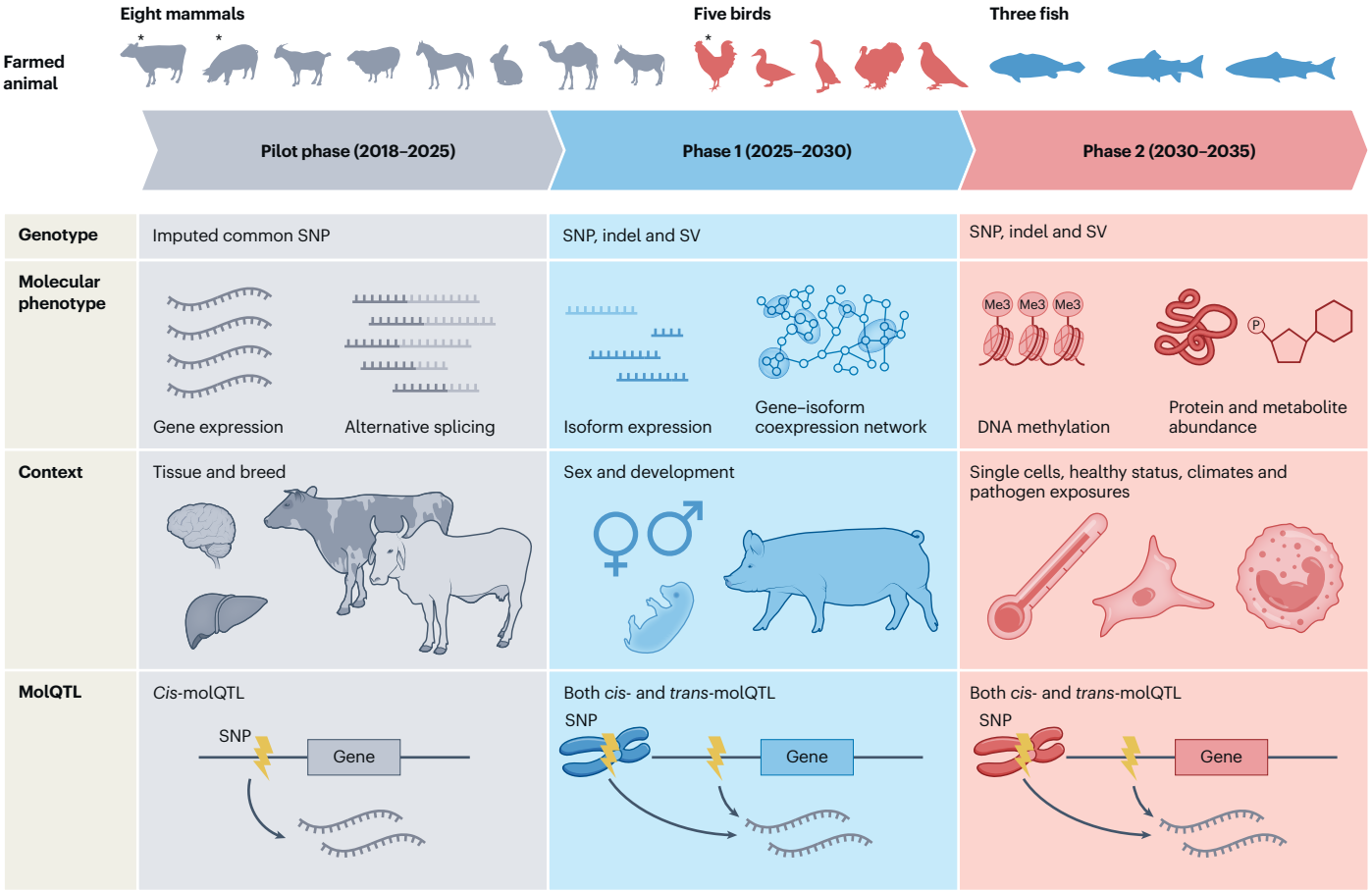

Fig. 1 | Overview of the FarmGTEx Project. To date, we are considering 16 farmed animal species, including eight mammals (that is, cattle, pigs, goats, sheep, horses, rabbits, camels and donkeys), five birds (that is, chickens, ducks, geese, turkeys and pigeons) and three fish (that is, large yellow 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*-molQTL and *trans*-molQTL mapping. Different goals and milestones for the three phases of FarmGTEx are outlined. Asterisks indicate that the pilot phases for these species have been completed. Indel, short insertion and deletion (length less than 50 nucleotides); SNP, single nucleotide polymorphism; SV, structural variants.

(for example, causal variants, genes, pathways, cell types and tissues) by which they affect complex phenotypes remain largely unknown.

Population-based molecular quantitative trait locus (molQTL) mapping, statistically associating genomic variants with molecular phenotypes (for example, 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 nondiseased 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 nonhuman primates (NHPs)[14], rats[15] and fruit flies[16]. Given that many trait-associated noncoding 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 (for example, 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 (for example, embryonic developmental stages in chickens), (4) exposure to various environmental conditions (for example, diet, climate and pathogen exposure) at the population scale, making them ideal for studying gene–environment interactions, (5) the existence of wild progenitors and congeners for understanding artificial selection and domestication (for example, 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 in vivo, and (7) several farmed species possess physiological characteristics that are similar to those of humans (for example, the cardiovascular and respiratory systems of pigs and in utero development of the immune system of cattle)[2,4].

Here, we present the FarmGTEx Project, a global initiative (including 90 universities and research institutes worldwide to date) that aims to generate a comprehensive catalog 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 QTL (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 loci colocalized with at least one of five types of molQTL in pigs[23]. Transcriptome-wide association studies (TWAS) revealed that expression of *ABCD4* in the brain and alternative splicing of *MYO7B* in the small intestine had been associated with back fat thickness in pigs[23], and 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, the 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 back fat thickness versus human body mass index[23,24].

Capitalizing on the substantial 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 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 women[27]. Pregnant sheep have been used to study the potential impacts of analgesics on maternal and fetal well being[28]. Pigs are key models for cardiovascular diseases[2], while chickens play a critical role in studying influenza A virus transmission and its pandemic risks[29].

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 sequencing (RNA-seq) and whole-genome-sequencing (WGS) data, we have been able to publish multi-tissue and multi-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 in an additional six mammals (that is, goats, sheep, horses, rabbits, camels and donkeys), four birds (that is, ducks, geese, turkeys and pigeons) and three fish (that is, 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, in which we aim to progressively incorporate broader biological contexts and diverse environmental conditions by collecting 20–50 tissues from 200 additional animals in each context or condition (Fig. 1).

Phase 1 (2025–2030) is dedicated to unraveling 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; 100 male and 100 female). 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; 150 male, 150 female). To investigate postnatal development-specific regulatory effects and to 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 (that is, 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 substantial 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 and/or single-nucleus RNA-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) project[32]. 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 omic sequencing technologies. For instance, future efforts to generate population-level single-cell and spatial omics data, alongside the use of in vitro models (for example, organoids: self-organized three-dimensional 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[33]. This involves thorough documentation of the attributes of animals used in a project (for example, species, breed, sex, age and health status), specimens (for example, tissue, cell, time of collection, storage and histological images) and experimental assays (for example, 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[33], 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[33].

https://doi.org/10.1038/s41588-025-02121-5

Table 1 | Summary of sample analysis in the pilot phase of FarmGTEx

Organ system	Tissue	CattleGTEx			PigGTEx			ChickenGTEx		
		n	eGene	sGene	*n*	eGene	sGene	*n*	eGene	sGene
Cardiovascular	Artery	–	–	–	59	673	357	–	–	–
	Heart	–	–	–	164	688	398	258	768	735
Digestive	Liver	576	7,292	4,909	501	6,154	2,318	741	5,233	4,985
	Rumen	202	3,488	2,361	–	–	–	–	–	–
	Jejunum	105	1,348	1,558	75	1,027	633	65	168	165
	Ileum	43	173	188	128	1,729	665	62	201	196
	Salivary gland	40	564	4	–	–	–	–	–	–
	Colon	–	–	–	67	1,007	555	–	–	–
	Duodenum	–	–	–	49	371	261	50	53	52
	Large intestine	–	–	–	68	1,008	582	–	–	–
	Small intestine	–	–	–	270	3,518	1,836	169	835	797
	Cecum	–	–	–	-	-	-	56	92	90
Embryonic	Blastocyst	–	–	–	56	3,651	199	–	–	–
	Blastomere	–	–	–	76	1	17	–	–	–
	Embryo	281	3,428	166	536	1,643	411	470	629	618
	Morula	–	–	–	150	46	21	–	–	–
Endocrine	Adipose	151	2,073	2,973	285	4,078	2,080	115	479	448
Female reproductive	Uterus	359	5,192	4,132	213	2,127	1,103	–	–	–
	Mammary	175	4,825	3,554	–	–	–	–	–	–
	Ovary	139	2,039	3,263	204	1,437	852	96	657	635
	Oviduct	85	2,478	4,928	–	–	–	68	81	79
	Oocyte	–	–	–	98	147	57	–	–	–
	Placenta	–	–	–	61	111	87	–	–	–
Immune	Blood	698	10,157	6,693	386	6,076	2,217	224	2,549	2,476
	Bursa	–	–	–	–	–	–	115	474	453
	Milk cell	173	893	3,327	63	1,129	119	–	–	–
	Lymph node	87	3,633	2,495	50	48	77	–	–	–
	Leukocyte	63	554	237	–	–	–	122	1,197	1,162
	Spleen	–	–	–	91	1,030	808	383	4,833	4,604
	Thymus	–	–	–	–	–	–	68	123	110
	Macrophage	295	8,793	7,913	84	1,298	331	60	922	869
	Monocytes	113	3,806	1,985	–	–	–	–	–	–
	Fetal thymus	–	–	–	48	1,629	440	–	–	–
Integumentary	Skin	41	423	33	-	–	–	163	392	373
Male reproductive	Testis	60	809	1,573	184	6,175	3,482	44	120	119
Muscular	Muscle	699	7,164	4,849	1,321	9,724	3,833	517	2,170	2,088
Nervous	Pituitary	134	1,794	1,751	53	356	68	135	216	199
	Hypothalamus	112	1,403	1,481	73	2,099	695	107	551	528
	Brain	–	–	–	419	5,815	2,605	479	4,953	4,783
	Frontal cortex	–	–	–	75	928	128	–	–	–
	Cerebellum	–	–	–	–	–	–	52	87	82
	Retina	–	–	–	–	–	–	119	253	235
Respiratory	Lung	87	4,779	3,778	149	1,820	1,455	89	302	288
	Trachea	–	–	–	–	–	–	92	427	403
Skeletal	Cartilage	–	–	–	65	232	54	–	–	–
	Synovial membrane	–	–	–	88	1,051	168	–	–	–
Urinary	Kidney	–	–	–	44	139	164	78	211	202

n, sample size; –, not applicable; eGene, a gene with at least one significant eQTL; sGene, a gene with at least one significant sQTL; brain, all brain regions; CattleGTEx[22]; PigGTEx[23]; ChickenGTEx[24].

https://doi.org/10.1038/s41588-025-02121-5

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). As 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[34], 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 > 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 genome assemblies to detect rare and somatic mutations (minor allele frequencies < 0.05)[35] as well as long-read sequencing technologies coupled with graph-based pangenome approaches to genotype short insertions and deletions, short tandem repeats and structural variants[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 (for example, 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 coexpression 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 (for example, DNA methylation and chromatin accessibility), RNA modifications (for example, N^6-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, by integrating long-read WGS data, we will incorporate short tandem repeats and structural variants in the genotype reference panels to impute them in the target populations[39]. We will also develop and benchmark methods to infer missing or mislabeled molecular phenotypes and metadata (for example, sex and developmental stage), particularly for those in the public domain, such as hypergraph factorization for imputing gene expression from missing tissues[40]. To account for batch effects caused by local and global population stratification, we employ genotype principal components and genetic relationship matrices[23]. Additionally, we use phenotype principal components to account for nongenetic unmeasured confounders (for example, cell type or state composition)[41–43].

MolQTL mapping

In the pilot phase of FarmGTEx, owing 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). 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 use 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 to simple linear models for mapping molQTL, particularly *trans*-molQTL, in farmed animal populations with complex familial relatedness[46]. We are developing the Omics Genetic Analysis (OmiGA) toolkit to efficiently and effectively implement linear mixed models in molQTL mapping (https://omiga.farmgtex.org)[47]. Ultimately, we will integrate molQTL mapping with other computational approaches, such as allelic imbalance analyses and large language DNA models (for example, pre-trained bidirectional encoder representations from transformers model for DNA language in genome (DNABERT) and DNA generative pre-trained transformer (DNAGPT))[48] to fully understand the regulatory landscape of genomic variants.

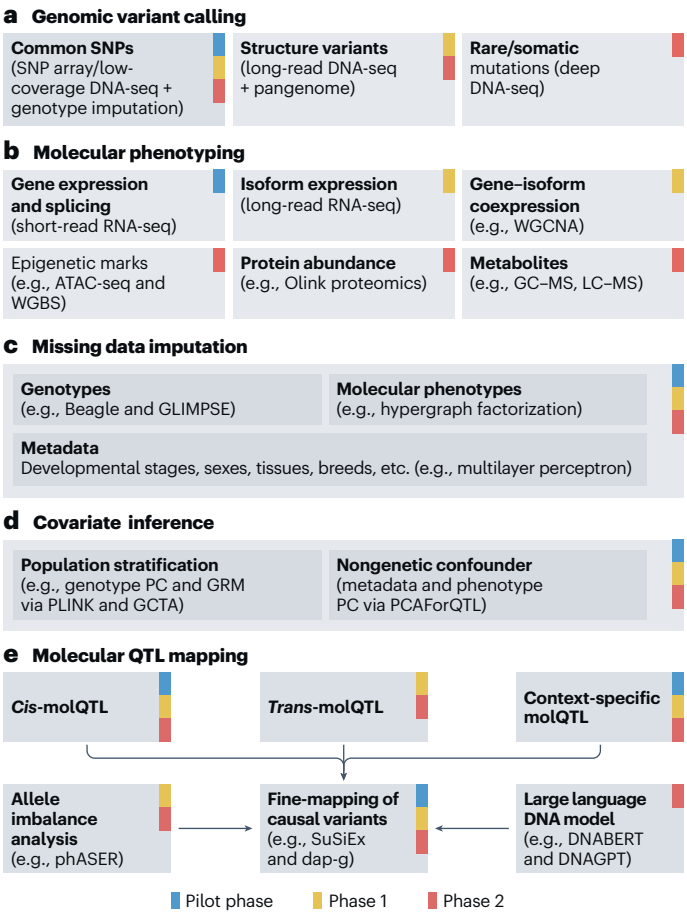

Fig. 2 | The overall data analysis pipeline of FarmGTEx. a–e, The flowchart highlights the key stages and methodologies of data generation and analyses across the three phases of FarmGTEx, including genomic variant calling (**a**), definition and quantification of molecular phenotypes (**b**), missing data imputation (**c**), batch effect inference and removal (**d**), and molQTL mapping via, for example, OmiGA (https://omiga.farmgtex.org) and ClipperQTL (**e**)[47,68]. Examples of approaches and methods used for each stage are shown in parentheses. ATAC-seq, assay for transposase-accessible chromatin with sequencing; dap-g, deterministic approximation of posteriors: genetics[69]; DNABERT[70]; DNAGPT[71]; DNA-seq, DNA sequencing; GC–MS, gas chromatography–mass spectrometry; GCTA, a tool for genome-wide complex trait analysis[72]; GRM, genetic relationship matrix; GLIMPSE, genotype likelihoods imputation and phasing method[73]; LC–MS, liquid chromatography–mass spectrometry; PC, principal component; PLINK, a tool for whole-genome association and linkage analyses[74]; phASER, phasing and allele-specific expression from RNA-seq[75]; SuSiEx, an enhanced tool for cross-ancestry fine-mapping of causal variants based on 'sum of single effects'[76]; WGBS, whole-genome bisulfite sequencing; WGCNA, weighted gene coexpression network analysis.

https://doi.org/10.1038/s41588-025-02121-5

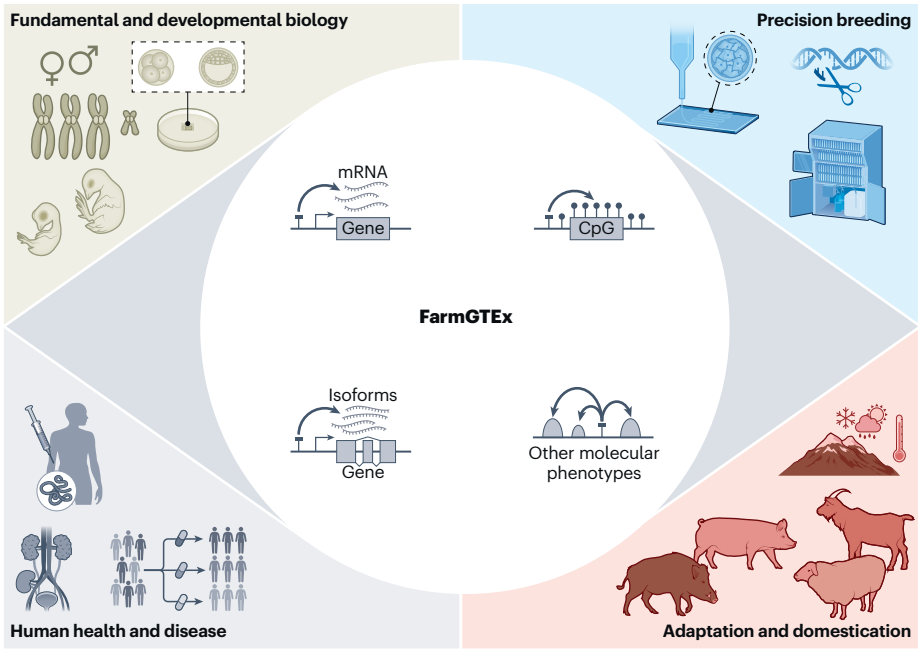

Fig. 3 | Potential applications of FarmGTEx resources. Insights into the dynamic landscape of the regulatory effects of genomic variants across diverse biological contexts (for example, sex and developmental stage) can advance our understanding of how genetics, development and sex interact with complex phenotypes. Identified causal variants and genes for complex traits and diseases can be used in precision animal breeding via gene editing and synthetic biology. Determining 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 that underlie 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.

Applications of FarmGTEx

FarmGTEx resources have potential scientific and social impacts on fundamental biology (for example, 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 genome-wide association study results via various integrative statistical approaches, including TWAS, colocalization, Mendelian randomization and artificial intelligence (AI)-based approaches (for example, 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 by creating 'ultimate' genotypes efficiently and effectively by integrating AI, gene editing and synthetic biology[51].

By incorporating large-scale genotypes and ancient DNA that extend back in time[52], 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 genomic studies of farmed and wild animals can also help to advance our understanding of recent human adaptation and evolution[53,54].

Aside from their impacts on agriculture, the rich and unique resources provided by FarmGTEx will also hold promise for comparative genomics by 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 linkage disequilibrium 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 back fat thickness and human body weight revealed

novel genes associated with human body weight[23]. 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 genetic research in farmed animals. It will thus adhere to data-sharing principles of making data findable, accessible, interoperable and reusable[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 (https://www.farmgtex.org) summarizes all resources from the project, including publications, computer code and public servers with specific functions. For instance, the PigGTEx (https://piggtex.farmgtex.org) and ChickenGTEx (https://chicken.farmgtex.org) servers allow any researcher to explore and download all processed data and results. The FarmGTEx TWAS-server (https://twas.farmgtex.org) supports online interactive TWAS analyses, while PigBiobank (https://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 (such as 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 Snakemake[59] and Nextflow[60], and code-sharing platforms, such as GitHub and Bioconductor (https://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 the 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, genomic 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 (https://www.farmgtex.org/) and the FAANG–FarmGTEx Task Force (https://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[36], Bovine Long Read[39] 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 (for example, the US Council on Dairy Cattle Breeding, Hendrix Genetics and Guangxi Yangxiang) and policymakers (for example, the US National Pork Board and the 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[44], 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, analyzing 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 (for example, between breeds), and long-term evolutionary processes (for example, within species and subspecies), (2) a comprehensive catalog of context-specific regulatory effects for detecting causal variants, genes and mechanisms involved in complex phenotypes, local environmental adaption and domestication, (3) uniformly processed multiomics datasets for developing advanced integrative genomic methods (for example, 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 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.

References

1. Frantz, L. A. F., Bradley, D. G., Larson, G. & Orlando, L. Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* **21**, 449–460 (2020).
2. Lunney, J. K. et al. Importance of the pig as a human biomedical model. *Sci. Transl. Med.* **13**, eabd5758 (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**, 46999–47023 (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* **11**, 1251 (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. *Nat. Rev. Methods Primers* **3**, 4 (2023).
13. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **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. *Nat. Genet.* **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. Preprint at *bioRxiv* https://doi.org/10.1101/2023.05.20.541576 (2023).
17. Umans, B. D., Battle, A. & Gilad, Y. Where are the disease-associated eQTLs? *Trends Genet.* **37**, 109–124 (2021).
18. Soskic, B. et al. Immune disease risk variants regulate gene expression dynamics during $CD4^+$ T cell activation. *Nat. Genet.* **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 (2019).
20. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2016).
21. Lin, W. et al. Genetic regulatory effects in response to a high-cholesterol, high-fat diet in baboons. *Cell Genom.* **4**, 100509 (2024).
22. Liu, S. et al. A multi-tissue atlas of regulatory variants in cattle. *Nat. Genet.* **54**, 1438–1447 (2022).
23. Teng, J. et al. A compendium of genetic regulatory effects across pig tissues. *Nat. Genet.* **56**, 112–123 (2024).
24. Guan, D. et al. The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.27.546670 (2023).
25. Xiang, R. et al. Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle. *Cell Genom.* **3**, 100385 (2023).
26. Zhao, R. et al. The conservation of human functional variants and their effects across livestock species. *Commun. Biol.* **5**, 1003 (2022).
27. Varcoe, T. J. et al. Considerations in selecting postoperative analgesia for pregnant sheep following fetal instrumentation surgery. *Anim. Front.* **9**, 60–67 (2019).
28. Abedal-Majed, M. A. & Cupp, A. S. Livestock animals to study infertility in women. *Anim. Front.* **9**, 28–33 (2019).
29. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* **17**, 67–81 (2019).
30. Chen, L. et al. Construction of a multi-tissue cell atlas reveals cell-type-specific regulation of molecular and complex phenotypes in pigs. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.12.544530 (2023).
31. Fang, L. et al. Cattle Cell Atlas: a multi-tissue single cell expression repository for advanced bovine genomics and comparative biology. Preprint at *Research Square* https://doi.org/10.21203/RS.3.RS-4631710/V1 (2024).
32. Giuffra, E. & Tuggle, C. K. Functional Annotation of Animal Genomes (FAANG): current achievements and roadmap. *Annu. Rev. Anim. Biosci.* **7**, 65–88 (2019).
33. 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).
34. Langer, B. E. et al. Empowering bioinformatics communities with Nextflow and nf-core. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.10.592912 (2024).
35. Cameron, D. L., Di Stefano, L. & Papenfuss, A. T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat. Commun.* **10**, 3240 (2019).
36. 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**, 139 (2023).
37. Wang, T. et al. The Human Pangenome Project: a global resource to map genomic diversity. *Nature* **604**, 437–446 (2022).
38. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190097 (2019).
39. 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. *Genet. Sel. Evol.* **55**, 9 (2023).
40. Viñas, R. et al. Hypergraph factorization for multi-tissue gene expression imputation. *Nat. Mach. Intell.* **5**, 739–753 (2023).
41. 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**, 210 (2022).
42. 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. *Nat. Protoc.* **7**, 500–507 (2012).
43. Mostafavi, S. et al. Normalizing RNA-sequencing data by modeling hidden covariates with prior knowledge. *PLoS ONE* **8**, e68141 (2013).
44. Engreitz, J. M. et al. Deciphering the impact of genomic variation on function. *Nature* **633**, 47–57 (2024).
45. van der Wijst, M. G. P. et al. The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).
46. Sul, J. H., Martin, L. S. & Eskin, E. Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* **14**, e1007309 (2018).
47. Teng, J. et al. OmiGA: a toolkit for ultra-efficient molecular trait analysis in complex populations. Preprint at *bioRxiv* https://doi.org/10.1101/2024.12.19.629424 (2024).
48. Liu, J. et al. Large language models in bioinformatics: applications and perspectives. Preprint at arxiv.org/abs/2401.04155 (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. USA* **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**, 2310–2317 (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 Biol.* **17**, 98 (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**, 60–71 (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* **587**, 240–245 (2020).
56. Lewin, H. A. et al. The Earth BioGenome Project 2020: starting the clock. *Proc. Natl Acad. Sci. USA* **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 Genom.* **3**, 100390 (2023).
58. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (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. *Nat. Biotechnol.* **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 Genom.* **2**, 100085 (2022).
62. Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
63. Denoyelle, L. et al. VarGoats project: a dataset of 1159 whole-genome sequences to dissect *Capra hircus* global diversity. *Genet. Sel. Evol.* **53**, 86 (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**, 3 (2022).
66. Clark, E. L. et al. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol.* **21**, 285 (2020).
67. Kalbeisch, T. et al. The Ruminant Telomere-to-Telomere (RT2T) Consortium. *Nat. Genet.* **56**, 1566–1573 (2024).
68. Zhou, H. J., Ge, X. & Li, J. J. ClipperQTL: ultrafast and powerful eGene identification method. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.28.555191 (2023).
69. 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).
70. 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).
71. Zhang, D. et al. DNAGPT: a generalized pre-trained tool for multiple DNA sequence analysis tasks. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.11.548628 (2024).
72. 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**, 76–82 (2011).
73. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
74. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
75. 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).
76. Yuan, K. et al. Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nat. Genet.* **56**, 1841–1850 (2024).

Acknowledgements

L. Fang is supported by Agriculture and Food Research Initiative Competitive grant 2022-67015-36215 from the USDA National Institute of Food and Agriculture, and seed funding from CellFood Hub (Aarhus University Foundation). Y. Luo is supported by the Innovative Health Initiative Joint Undertaking under grant agreement 101165643. Z. Zhang is supported by funding from the China Agriculture Research System (CARS-35). 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 I. Kaplow at the Department of Biological Sciences and the Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University for valuable comments on the manuscript.

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. made 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. Gong, Z.P., Y.Y., E.L.C., J. Smith., K.R., R.X., A.J.C., M.E.G., M. Littlejohn, G.L., D.E.M., J.F.O'G., P.S., G.S., M.S.L., Z.J., X.P., W.G., Haihan Zhang, X. He, 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., C.j.L., C.P.V.T., B.D.R., N.B., J. Lunney., W. Liu., L.G., X. Zhao., E.M.I.-A., Y. Luo, L. Lin, O.C.-X., M.F.L.D., R.P.M.A.C., M. Gòdia, 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., Jing Li, C.F., M.F., Qishan Wang, Z.H., Qin Wang, F.Z., L. Jiang, G.Z., Z. Zhou, R.Z., H.L., J.D., L. Jin, M. Li, D.M., Xiaohong Liu, Y.C., X. Yuan, Jiaqi Li, S.Z., Yi Zhang, X.D., D.S., H.-Z.S., C.L., Y. Wang, Y.J., D.W., W. Wang, X.F., Q.Z., K.L., Hao Zhang, N.Y., X. Hu, W.H., J. Song., Y. Wu, J.Y., W. Wu, C.K., Xinfeng Liu, X. Yu., L.C., X. Zhou, S.K., W. Li, H.K.I., E.S.B., B.R. and M.C.S. have provided 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.

Additional information

Correspondence should be addressed to Lingzhao Fang, Jingyi Jessica Li, Abraham A. Palmer, Laurent Frantz, Huaijun Zhou, Zhe Zhang or George E. Liu.

Peer review information *Nature Genetics* thanks Ben Hayes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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[1,7], Zhangyuan Pan[7], Ying Yu[8], Emily L. Clark[9], Jacqueline Smith[9], Konrad Rawlik[10], Ruidong Xiang[11,12,13], Amanda J. Chamberlain[11,14], Michael E. Goddard[11,13], 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],

https://doi.org/10.1038/s41588-025-02121-5

Siqian Chen[8], Robert S. Young[29,30], Xia Shen[29,31,32,33], Charley Xia[34,35], Hao Cheng[5], Li Ma[36], John B. Cole[37,38,39], Ransom L. Baldwin VI[40], Cong-jun Li[40], Curtis P. Van Tassell[40], Benjamin D. Rosen[40], Nayan Bhowmik[40], Joan Lunney[41], Wansheng Liu[42], Leluo Guan[43,44], Xin Zhao[45], Eveline M. Ibeagha-Awemu[46], Yonglun Luo[47,48], Lin Lin[47,48], Oriol Canela-Xandri[49], Martijn F. L. Derks[50], Richard P. M. A. Crooijmans[50], Marta Gòdia[50], Ole Madsen[50], Martien A. M. Groenen[50], James E. Koltes[51], Christopher K. Tuggle[51], Fiona M. McCarthy[52], Dominique Rocha[53], Elisabetta Giuffra[53], Marcel Amills[54,55], Alex Clop[54,56], Maria Ballester[57], Gwenola Tosser-Klopp[58], Jing Li[1,59], Chao Fang[60], Ming Fang[61], Qishan Wang[62], Zhuocheng Hou[8], Qin Wang[8], Fuping Zhao[7], Lin Jiang[7], Guiping Zhao[7], Zhengkui Zhou[7], Rong Zhou[7], Hehe Liu[63], Juan Deng[63], Long Jin[63], Mingzhou Li[63], Delin Mo[64], Xiaohong Liu[64], Yaosheng Chen[64], Xiaolong Yuan[2], Jiaqi Li[2], Shuhong Zhao[25,26], Yi Zhang[8], Xiangdong Ding[8], Dongxiao Sun[8], Hui-Zeng Sun[65], Cong Li[66], Yu Wang[66], Yu Jiang[66], Dongdong Wu[67], Wenwen Wang[68], Xinzhong Fan[68], Qin Zhang[68], Kui Li[23], Hao Zhang[8], Ning Yang[8], Xiaoxiang Hu[69], Wen Huang[70], Jiuzhou Song[36], Yang Wu[71], Jian Yang[3,4], Weiwei Wu[72], Claudia Kasper[73], Xinfeng Liu[1,74], Xiaofei Yu[75], Leilei Cui[76,77], Xiang Zhou[78], Seyoung Kim[79], Wei Li[80], Hae Kyung Im[81], Edward S. Buckler[82,83,84], Bing Ren[85], Michael C. Schatz[86], Jingyi Jessica Li[87] ✉, Abraham A. Palmer[88,89] ✉, Laurent Frantz[90,91] ✉, Huaijun Zhou[5] ✉, Zhe Zhang[2] ✉, George E. Liu[40] ✉ & The FarmGTEx Consortium*

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, 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, China. [3]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. [4]School of Life Sciences, Westlake University, Hangzhou, China. [5]Department of Animal Science, University of California, Davis, Davis, CA, USA. [6]Department of Animal and Veterinary Sciences, Scotland's Rural College, Midlothian, UK. [7]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China. [8]National Engineering Laboratory for Animal Breeding, State Key Laboratory of Animal Biotech Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction of the Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, China Agricultural University, Beijing, China. [9]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, UK. [10]Baillie Gifford Pandemic Science Hub, Centre for Inflammation Research, Institute for Regeneration and Repair, the University of Edinburgh, Edinburgh, UK. [11]Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, 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, Victoria, Australia. [14]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia. [15]Research and Development, Livestock Improvement Corporation, Hamilton, New Zealand. [16]AL Rae Centre for Genetics and Breeding, Massey University, Palmerston North, New Zealand. [17]The Palaeogenomics and 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, Ireland. [19]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin, Ireland. [20]UCD One Health Centre, University College Dublin, Belfield, Dublin, Ireland. [21]Department of Animal Sciences and Center for Reproductive Biology, Washington State University, Pullman, WA, USA. [22]College of Animal Science and Technology, Hunan Agricultural University, Changsha, China. [23]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [24]Hainan Institute, Zhejiang University, Yongyou Industry Park, Yazhou Bay Sci–Tech City, Sanya, China. [25]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of the Ministry of Education, Huazhong Agricultural University, Wuhan, China. [26]Yazhouwan National Laboratory, Sanya, China. [27]Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan, China. [28]Poultry Institute, Shandong Academy of Agricultural Sciences, Jinan, China. [29]Usher Institute, University of Edinburgh, Edinburgh, UK. [30]Zhejiang University–University of Edinburgh Institute, Zhejiang University, Haining, P. R. China. [31]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, 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]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [34]Lothian Birth Cohort studies, University of Edinburgh, Edinburgh, UK. [35]Department of Psychology, University of Edinburgh, Edinburgh, UK. [36]Department of Animal and Avian Sciences, University of Maryland, College Park, MD, USA. [37]Council on Dairy Cattle Breeding, Bowie, MD, USA. [38]Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program and the Genetics Institute, University of Florida, Gainesville, FL, USA. [39]Department of Animal Science, North Carolina State University, Raleigh, NC, USA. [40]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, MD, USA. [41]Animal Parasitic Diseases Laboratory, BARC, NEA, ARS, USDA, Beltsville, MD, USA. [42]Department of Animal Science, Center for Reproductive Biology and Health, College of Agricultural Sciences, the Pennsylvania State University, University Park, PA, USA. [43]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada. [44]Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, Canada. [45]Department of Animal Science, McGill University, Sainte-Anne-de-Bellevue, Quebec, Canada. [46]Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada, Sherbrooke, Quebec, Canada. [47]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [48]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark. [49]MRC Human Genetics Unit at the Institute of Genetics and Cancer, the University of Edinburgh, Edinburgh, UK. [50]Animal Breeding and Genomics, Wageningen University & Research, Wageningen, the Netherlands. [51]Department of Animal Science, Iowa State University, Ames, IA, USA. [52]University of Arizona, Tucson, AZ, USA. [53]GABI, AgroParisTech, INRAE, Paris-Saclay University, Jouy-en-Josas, France. [54]Department of Animal Genetics, Centre for Research in Agricultural Genomics, CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [55]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain. [56]Consejo Superior de Investigaciones Científicas, Barcelona, Spain. [57]Animal Breeding and Genetics Programme, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Torre Marimon, Caldes de Montbui, Spain. [58]GenPhySE, INRAE, ENVT, Université de Toulouse, Castanet-Tolosan, France. [59]School of Agriculture and Life Sciences, Kunming University, Kunming, China. [60]LC-Bio Technologies, Co., Ltd, Hangzhou, China. [61]Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture and Rural Affairs, Jimei University, Xiamen, China. [62]College of Animal Sciences, Zhejiang University, Hangzhou, China. [63]College of Animal Science and

Technology, Sichuan Agricultural University, Chengdu, China. [64]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China. [65]Key Laboratory of Dairy Cow Genetic Improvement and Milk Quality Research of Zhejiang Province, College of Animal Sciences, Zhejiang University, Hangzhou, China. [66]College of Animal Science and Technology, Northwest A&F University, Yangling, China. [67]Key Laboratory of Genetic Evolution and Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [68]Shandong Provincial Key Laboratory for Livestock Germplasm Innovation and Utilization, College of Animal Science, Shandong Agricultural University, Tai'an, China. [69]State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, China. [70]Department of Animal Science, Michigan State University, East Lansing, MI, USA. [71]Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China. [72]Institute of Animal Science, Xinjiang Academy of Animal Science, Ürümqi City, China. [73]Animal GenoPhenomics, Animal Production Systems and Animal Health, Agroscope Posieux, Fribourg, Switzerland. [74]State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystem, College of Ecology, Lanzhou University, Lanzhou, China. [75]College of Marine Life Sciences, Ocean University of China, Qingdao, China. [76]School of Life Sciences, Nanchang University, Nanchang, China. [77]Jiangxi Province Key Laboratory of Aging and Disease, Human Aging Research Institute and School of Life Science, Nanchang University, Jiangxi, China. [78]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [79]Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [80]Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, CA, USA. [81]Department of Medicine and Human Genetics, the University of Chicago, Chicago, IL, USA. [82]Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA. [83]Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA. [84]Agricultural Research Service, United States Department of Agriculture, Ithaca, NY, USA. [85]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, USA. [86]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [87]Department of Statistics and Data Science, University of California, Los Angeles, Los Angeles, CA, USA. [88]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [89]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [90]Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany. [91]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: lingzhao.fang@qgg.au.dk; lijy03@g.ucla.edu; aap@ucsd.edu; laurent.frantz@lmu.de; hzhou@ucdavis.edu; zhezhang@scau.edu.cn; george.liu@usda.gov

The FarmGTEx Consortium

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[1,7], Zhangyuan Pan[7], Ying Yu[8], Emily L. Clark[9], Jacqueline Smith[9], Konrad Rawlik[10], Ruidong Xiang[11,12,13], Amanda J. Chamberlain[11,14], Michael E. Goddard[11,13], 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[29,31,32,33], Charley Xia[34,35], Hao Cheng[5], Li Ma[36], John B. Cole[37,38,39], Ransom L. Baldwin VI[40], Cong-jun Li[40], Curtis P. Van Tassell[40], Benjamin D. Rosen[40], Nayan Bhowmik[40], Joan Lunney[41], Wansheng Liu[42], Leluo Guan[43,44], Xin Zhao[45], Eveline M. Ibeagha-Awemu[46], Yonglun Luo[47,48], Lin Lin[47,48], Oriol Canela-Xandri[49], Martijn F. L. Derks[50], Richard P. M. A. Crooijmans[50], Marta Gòdia[50], Ole Madsen[50], Martien A. M. Groenen[50], James E. Koltes[51], Christopher K. Tuggle[51], Fiona M. McCarthy[52], Dominique Rocha[53], Elisabetta Giuffra[53], Marcel Amills[54,55], Alex Clop[54,56], Maria Ballester[57], Gwenola Tosser-Klopp[58], Jing Li[1,59], Chao Fang[60], Ming Fang[61], Qishan Wang[62], Zhuocheng Hou[8], Qin Wang[8], Fuping Zhao[7], Lin Jiang[7], Guiping Zhao[7], Zhengkui Zhou[7], Rong Zhou[7], Hehe Liu[63], Juan Deng[63], Long Jin[63], Mingzhou Li[63], Delin Mo[64], Xiaohong Liu[64], Yaosheng Chen[64], Xiaolong Yuan[2], Jiaqi Li[2], Shuhong Zhao[25,26], Yi Zhang[8], Xiangdong Ding[8], Dongxiao Sun[8], Hui-Zeng Sun[65], Cong Li[66], Yu Wang[66], Yu Jiang[66], Dongdong Wu[67], Wenwen Wang[68], Xinzhong Fan[68], Qin Zhang[68], Kui Li[23], Hao Zhang[8], Ning Yang[8], Xiaoxiang Hu[69], Wen Huang[70], Jiuzhou Song[36], Yang Wu[71], Jian Yang[3,4], Weiwei Wu[72], Claudia Kasper[73], Xinfeng Liu[1,74], Xiaofei Yu[75], Leilei Cui[76,77], Xiang Zhou[78], Seyoung Kim[79], Wei Li[80], Hae Kyung Im[81], Edward S. Buckler[82,83,84], Bing Ren[85], Michael C. Schatz[86], Jingyi Jessica Li[87], Abraham A. Palmer[88,89], Laurent Frantz[90,91], Huaijun Zhou[5], Zhe Zhang[2] & George E. Liu[40]