Title Page:

The enhanced multi-tissue atlas of regulatory effects in cattle

Houcheng Li[1†], Huicong Zhang[1†], Di Zhu[1†], Pengju Zhao[2†], Zhenyu Wei[3†], Jingsheng Lu[4,5,6†], Mian Gong[1,7], Qi Zhang[8], Weijie Zheng[8], Xinfeng Liu[1,9], Dailu Guan[10,11], Jinyan Teng[11], Qin Lin[11], Yongjie Tang[8], Yahui Gao[11], Shanjiang Zhao[7], Zhe Zhang[11], Junting Du[12], Chao Fang[12], Bingxing An[1,7], Bingjin Lin[1,13], Haihan Zhang[13], Min Tian[1,3], Jingjing Tian[1,7], Siqian Chen[8], Wansheng Liu[14], Yanan Wang[4,5,6], Mingshan Wang[4], Eveline M. Ibeagha-Awemu[15], Richard P. M. A. Crooijmans[16], Martijn F. L. Derks[16], Marta Gòdia[16], Ole Madsen[16], Hubert Pausch[17], Alexander S. Leonard[17], Laurent Frantz[18,19], David E. MacHugh[20,21,22], John F. O'Grady[17], Iuliana Ionita-Laza[23], Xin Zhao[24], Leluo Guan[25], Huaijun Zhou[10], Emilio Mármol-Sánchez[26], Monique G. P. van der Wijst[27,28], Xubin Lu[29], Hui Jiang[29], Zhangping Yang[29], Qien Yang[30], Qinyou Liu[31], Chuang Xu[32], Moli Li[32], Yali Hou[7], Zhangyaun Pan[7], Yan Chen[7], Ruidong Xiang[33,34,35,36], Mathew Littlejohn[37,38], Emily L. Clark[39], Wenfa Lyu[40,41], Yuwen Liu[42,43], Lin Jiang[44], Peng Su[44], Xuexue Liu[45], Senlin Zhu[46], Minghui Jia[46], Huizeng Sun[46], Bo Han[8], Yi Zhang[8], Ying Yu[8], Dongxiao Sun[8], Yaokun Li[31], Dewu Liu[31], Goutam Sahana[1], Zexi Cai[1], Mogens Sandø Lund[1], John B. Cole[47,48,49,50], Li Ma[51], Jicai Jiang[52], Wenjian Li[53], Yang Wu[53], Jianbin Li[54], Jun Teng[55], Qin Zhang[55], Xiao Wang[54]*, Xuemei Lu[4,5,6]*, Yu Jiang[3]*, Yang Zhou[39,56]*, Yu Wang[3]*, Bingjie Li[1,57]*, Peter Sørensen[1]*, Lingzhao Fang[1]*

Affiliations

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

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

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

[4]Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[5]Biodiversity Data Center of Kunming Institute of Zoology, Chinese Academy of sciences, Kunming, Yunnan 650201, China

[6]Yunnan Key Laboratory of Biodiversity Information, Kunming Institute of Zoology, The Chinese Academy of Sciences, Kunming 650201, China

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

[8]State Key Laboratory of Animal Biotech Breeding, National Engineering Laboratory for Animal 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.

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

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

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

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

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

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

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

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

[17]Animal Genomics, ETH Zurich, Universitatstrasse 2, 8092, Zurich, Switzerland

[18]Chair of Animal Systems Genomics, Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany

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

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

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

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

[23]Department of Biostatistics, Columbia University, New York, NY, USA.

[24]Department of Animal Science, McGill University, Quebec, H9X3V9, Canada.

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

[26]Center for Evolutionary Hologenomics, The GlobeInstitute, University of Copenhagen, Copenhagen,1353, Denmark

[27]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[28]Oncode Institute, Utrecht, The Netherlands

[29]Animal Science and Technology, Yangzhou University, Yangzhou 225009, China

[30]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China

[31]College of Animal Science, South China Agricultural University (SCAU), Guangzhou, China

[32]Department of Clinical Veterinary Medicine, College of Veterinary Medicine, China Agricultural University, Beijing 100193, China

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

[34]School od Applied Systems Biology, La Trobe University, Bundoora, VIC, 3083, Australia

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

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

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

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

[39]European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, Cambridge, UK

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

[41]Key Laboratory of Animal Production, Product Quality, and Security, Ministry of Education, College of Animal Science and Technology, Jilin Agricultural University, Changchun, 130118, China

[42]State Key Laboratory of Genome and Multi-omics Technologies, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

[43]Innovation Group of Pig Genome Design and Breeding, Research Centre for Animal Genome, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

[44]National Germplasm Center of Domestic Animal Resources, Ministry of Science and Technology of the People's Republic of China, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China

[45]Centre for Anthropobiology and Genomics of Toulouse (CNRS/Université de Toulouse), Faculté de Santé, Bâtiment A, 37 Allées Jules Guesde, 31000 Toulouse, France

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

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

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

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

[50]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario N1G 2W1, Canada

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

[52]Department of Animal Science, North Carolina State University, Raleigh, NC, USA

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

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

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

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

[57]Department of Animal and Veterinary Sciences, Scotland's Rural College (SRUC), King's Buildings, Edinburgh, EH9 3JG, UK

[†]Co-first authors; *Co-corresponding authors

Emails for all authors:
Houcheng Li: Houcheng.li@qgg.au.dk; Huicong Zhang: huicongz@qgg.au.dk; Di Zhu: zhudi@qgg.au.dk; Pengju Zhao: zhaopengju2014@gmail.com; Zhenyu Wei: weizhenyu@nwafu.edu.cn; Jingsheng Lu: lujingsheng@mail.kiz.ac.cn; Mian Gong: gongmian2767@qgg.au.dk; Qi Zhang: yiguizhang_666@163.com; Weijie Zheng: weijiezheng426@163.com; Xinfeng Liu: liuxinfeng@qgg.au.dk; Dailu Guan: dguan@ucdavis.edu; Jinyan Teng: jinyan.teng@scau.edu.cn; Qin Lin: qing_lin1996@126.com; Yongjie Tang: cautyj@163.com; Yahui Gao: yahui.gao@scau.edu.cn; Shanjiang Zhao: zhaoshanjiang@caas.cn; Zhe Zhang: zhezhang@scau.edu.cn; Junting Du: jtdu@lc-bio.com; Chao Fang: cfang@lc-bio.com; Bingxing An: abx@qgg.au.dk; Bingjin Lin: bingjinlin1127@163.com; Haihan Zhang: hhzhang@hunau.edu.cn; Min Tian: mintian@qgg.au.dk; Jingjing Tian: jingjingtian@qgg.au.dk; Siqian Chen: sqchen1028@163.com; Wansheng Liu: wul12@psu.edu; Yanan Wang: wangyanan@mail.kiz.ac.cn; Mingshan Wang: wangmingshan@mail.kiz.ac.cn; Eveline M. Ibeagha-Awemu: eveline.ibeagha-awemu@agr.gc.ca; Richard P. M. A. Crooijmans: richard.crooijmans@wur.nl; Martijn F. L. Derks: martijn.derks@wur.nl; Marta Gòdia: marta.godia@wur.nl; Ole Madsen: ole.madsen@wur.nl; Hubert Pausch: hubert.pausch@usys.ethz.ch; Alexander S. Leonard: alexander.leonard@usys.ethz.ch; Laurent Frantz: laurent.frantz@gmail.com; David E. MacHugh: david.machugh@ucd.ie; John F. O'Grady: johnfrancis.ogrady@usys.ethz.ch; Iuliana Ionita-Laza: iuliana.laza@gmail.com; Xin Zhao: xin.zhao@mcgill.ca; Leluo

Guan: leluo.guan@ubc.ca; Huaijun Zhou: hzhou@ucdavis.edu; Emilio Mármol Sánchez: emilio.marmol.sanchez@gmail.com; Monique G. P. van der Wijst: m.g.p.van.der.wijst@umcg.nl; Xubin Lu: lxb@yzu.edu.cn; Hui Jiang: dx120210160@stu.yzu.edu.cn; Zhangping Yang: yzp@yzu.edu.cn; Qien Yang: yangqien@nwipb.cas.cn, Qinyou Liu: qingyouliu@scau.edu.cn; Chuang Xu: xuchuang@cau.edu.cn; Moli Li: lml94520@cau.edu.cn; Yali Hou: houyali1210@gmail.com; Zhangyuan Pan: zhypan01@163.com; Yan Chen: chenyan@caas.cn; Ruidong Xiang: ruidong.xiang@agriculture.vic.gov.au; Mathew Littlejohn: mathew.littlejohn@lic.co.nz; Emily L. Clark: eclark@ebi.ac.uk; Wenfa Lyu: wenfa2004@163.com; Yuwen Liu: liuyuwen@caas.cn; Lin Jiang: jianglin@caas.cn; Peng Su: super96110@gmail.com; Xuexue Liu: Xuexue.liu@utoulouse.fr; Senlin Zhu: senlin_zhu@zju.edu.cn; Minghui Jia: jmhui@zju.edu.cn; Huizeng Sun: huizeng@zju.edu.cn; Bo Han: bohan@cau.edu.cn; Yi Zhang: yizhang@cau.edu.cn; Ying Yu: yuying@cau.edu.cn; Dongxiao Sun: sundx@cau.edu.cn; Yaokun Li: ykli@scau.edu.cn; Dewu Liu: dwliu@scau.edu.cn; Goutam Sahana: goutam.sahana@qgg.au.dk; Zexi Cai: zexi.cai@qgg.au.dk; Mogens Sandø Lund: mogens.lund@qgg.au.dk; John B. Cole: john.cole@uscdcb.com; Li Ma: lima@umd.edu; Jicai Jiang: jicai_jiang@ncsu.edu; Wenjian Li: liwenjian@wchscu.edu.cn; Yang Wu: yang.wu@wchscu.cn; Jianbin Li: msdljb@163.com; Jun Teng: tengjun0520@163.com; Qin Zhang: qzhang@sdau.edu.cn; Xiao Wang: xiaowangzntc@163.com; Xuemei Lu: xuemeilu@mail.kiz.ac.cn; Yu Jiang: yu.jiang@nwafu.edu.cn; Yang Zhou: yangzhou@mail.hzau.edu.cn; Yu Wang: wang_yu@nwafu.edu.cn; Bingjie Li: bingjie.li@sruc.ac.uk; Peter Sorenson: pso@qgg.au.dk; Lingzhao Fang: lingzhao.fang@qgg.au.dk

Abstract

Cattle are integral to global food security, yet the molecular architecture of their complex traits remains poorly understood. Here, we present the **Cattle Genotype–Tissue Expression (CattleGTEx) Phase 1** resource (https://cattlegtex.farmgtex.org/), a substantial expansion of the pilot study. By leveraging 12,422 RNA-seq profiles across 43 tissues and 82 breeds, we characterized 433,972 primary and 161,428 non-primary regulatory effects spanning seven molecular phenotypes. This high-resolution atlas resolves 75% of GWAS signals for 44 complex traits, significantly addressing the "missing regulation" in livestock. We propose a genetic regulatory model demonstrating how variants across multiple biological layers interact with specific biological contexts to shape phenotypic variation. Furthermore, CattleGTEx elucidates mechanisms underlying adaptive evolution between *Bos taurus* and *Bos indicus*, as well as artificial selection in dairy and beef breeds. Finally, by mapping evolutionary constraints on these regulatory effects, we demonstrate the translational value of this resource for prioritizing causal variants in human complex diseases. Together, Phase 1 of CattleGTEx provides a transformative framework for functional genomics, precision breeding, and comparative genetics.

Introduction

Since their domestication approximately 10,000-8,000 years ago, cattle have played a crucial role in human societal development and global food security, providing high-quality animal protein, labour, economic value, and cultural significance[1]. As they spread across the globe alongside human migration, cattle adapted genetically to diverse local environments ranging from cold, high-altitude mountains (e.g., Tibetan cattle) to tropical climates (e.g., Zebu)[2]. Modern breeding further intensified selection for production-related traits, resulting in specialized breeds with distinct physiological characteristics, such as the Holstein for dairy and Angus for beef[3]. Today, rising demand for sustainable food systems places new environmental and welfare pressures on cattle industry, necessitating improved production efficiency via precision breeding[4-6]. Beyond agriculture, the substantial genetic similarities between cattle and humans, particularly in immune, metabolic, and reproductive pathways, position cattle as valuable human biomedical models. These include conditions such as polycystic ovarian syndrome (PCOS), tuberculosis[7,8,9,10], and other conditions such as Charcot Marie Tooth disease where the pathology of the disease is difficult to model in small animals[11]. In addition, bovine embryonic genome activation (EGA) has been proposed as a useful model for human EGA, supported by comparative studies[12]. To address these agricultural challenges and leverage cattle as biomedical models, it is essential to decipher the genetic and molecular architecture underlying complex phenotypes of economic, ecological, and biomedical importance in cattle.

Deciphering this architecture requires a comprehensive characterization of the regulatory genome, as it drives phenotypic variation among individuals, populations, and species. The importance of regulatory elements has been well documented in humans and model organisms[13], where large-scale initiatives like the Encyclopedia of DNA Elements (ENCODE) and Genotype-Tissue Expression (GTEx) projects have provided fundamental insights into the molecular mechanisms of complex traits and diseases[14,15]. Paralleling these efforts, the ongoing FAANG and FarmGTEx initiatives aim to systematically annotate the regulatory landscapes across farmed animals[16,17], such as cattle, pigs, sheep, and chickens[18-26]. Although the pilot CattleGTEx study demonstrated the utility of expression quantitative trait loci (eQTL) for dissecting complex traits in cattle[18], its scope was constrained by limited sample size, reduced tissue coverage, and incomplete molecular phenotypes. Consequently, the majority (60-95%) of genome-wide association study (GWAS) loci could not be colocalized with detected eQTL across complex traits, giving rise to the so-called 'missing regulation' phenomenon similar to what has been previously observed in humans[27,28]. This limited overlap likely stems from trait-associated variants exerting regulatory effects on specific molecular phenotypes (e.g., alternative

splicing and RNA stability) in specific tissues[28,29], or possessing small effect sizes that previous datasets lack the statistical power to detect[30,31,32,33].

To address these limitations, we report here the Phase 1 of CattleGTEx, where we have substantially expanded sample size (from 4,889 to 12,422), tissue coverage (from 23 to 43), molecular phenotype (from two to seven), and breeds (from 46 to 82) relative to the pilot study (**Fig.** 1). Our dataset achieves an average of 489 individuals per tissue (range: 52 in adrenal gland to 2,693 in immune tissues) (**Fig.** 1a-c). To enable high-resolution genetic analyses, we constructed a multi-breed genotype imputation reference panel of 3,530 individuals worldwide[34], whose population structure closely matches that of the RNA-seq cohort (**Fig.** 1d). Using this resource, we systematically associated 4,317,531 genomic variants with seven classes of molecular phenotypes derived from RNA-seq data, followed by conditional and fine-mapping analyses to disentangle primary and non-primary regulatory effects (**Fig.** 1e). We further characterized sharing patterns of these regulatory effects across molecular phenotypes, tissues, and breeds (**Fig.** 1f). By integrating GWAS signals of 44 bovine complex traits, and signatures of selection between *Bos indicus* and *Bos taurus* as well as between dairy and beef breeds, we show the power of CattleGTEx Phase 1 for elucidating complex trait architecture, environmental adaptation, and selective breeding (**Fig.** 1h-i). Ultimately, by mapping evolutionary constraints of orthologous genes and variants between cattle and humans, we highlight the translational potential of CattleGTEx for dissecting complex traits in humans (**Fig.** 1j). All uniformly processed data, results, and computational codes are publicly available through the CattleGTEx web portal (https://cattlegtex.farmgtex.org/) (**Fig.** 1g).

Results

Data summary and molecular phenotyping

In total, we collected 27,681 RNA-seq samples spanning 103 bovine tissues (**Supplementary Table** 1-2). After quality control (**Fig.** 2a**; Supplementary Fig.** 1**; Supplementary Note**), 19,789 high-quality RNA-seq samples of 88 tissues were retained for further analysis, and their genotypes were imputed using the population structure matched reference panel described above[34]. Following extensive quality control and independent validation of genotype imputation accuracy (**Supplementary Notes**), we obtained 4,317,531 common variants (minor allele frequency, MAF > 0.05), with a median concordance rate of 0.97 and genotype Pearson's correlation-r^2 of 0.91 **(Fig.** 2b**)**. On average, 4,576 variants were available within the *cis*-window (±1 Mb from the transcription start site, TSS) of expressed genes (Transcript per million, TPM > 0.1 and read count > 6) **(Supplementary Fig.**

2a-d**)**. Compared with all 28,958,211 variants in the full reference panel, RNA-seq imputed variants were more enriched in gene bodies and *cis*-regulatory elements (e.g., promoter and enhancer) but depleted in intergenic regions **(Fig.** 2c**; Supplementary Fig.** 2e**)**. After removing duplicated samples within each tissue based on identity-by-state (IBS ≥ 0.9), the final dataset comprised 12,422 RNA-seq samples from 8,956 individuals across 43 tissues, each represented by more than 40 samples **(Supplementary Fig.** 2f-g**)**. For 5,170 samples lacking breed information, their breed identity was inferred using imputed genotypes by using a linear support vector machine (Linear SVM)[35], achieving an overall prediction accuracy of 97.7% with 10-fold stratified cross validation **(Supplementary Fig.** 3; **Supplementary Table** 5-6**)**.

Using these RNA-seq datasets, we defined and quantified seven classes of molecular phenotypes, including four abundance-based phenotypes (expression levels for 20,875 protein-coding and lncRNA genes, 102,033 exons, 20,690 isoforms, and 27,653 enhancers), and three structure-based phenotypes (68,258 alternative splicing, 17,995 RNA stability, and 18,387 3'UTR alternative poly-adenylation-3'UTR APA) **(Fig.** 2a**)**. Across all molecular phenotypes, samples clustered primarily by tissue type, and then reflected developmental origin, with samples derived from the same embryonic germ layer tending to cluster together **(Fig.** 2d, **Supplementary Fig.** 4**)**. Genes exhibiting high inter-tissue variability in expression were significantly enriched in tissue-specific biological processes, such as nervous system processes (FDR = 8.81×10^{-30}) and visual perception (FDR = 6.82×10^{-6}), whereas genes with low inter-tissue variability were enriched in fundamental cellular functions such as protein localization (FDR = 3.15×10^{-32}) and translation (FDR = 2.09×10^{-27}) **(Supplementary Fig.** 5a, c; **Supplementary Table** 7**)**. As expected, lncRNA genes exhibited higher tissue-specificity compared to protein-coding genes[22] **(Supplementary Fig.** 5b**)**. Functional enrichment analysis of tissue-specific genes (defined by FDR < 0.05 and $|\log_2$ fold change$|$ > 2 using the Wilcoxon rank-sum test[36]) recapitulated well-established tissue biology **(Fig.** 2d**)**. For example, genes upregulated in immune tissues were enriched in regulation of immune system processes (FDR = 1.77×10^{-70}), while genes upregulated in the digestive system were enriched in monocarboxylic acid transport (FDR = 4.71×10^{-8}) **(Fig.** 2d; **Supplementary Table** 8**)**. Altogether, these results demonstrated that the derived molecular phenotypes from publicly available RNA-seq data can robustly capture tissue-specific biological programs, providing a foundation for dissecting the regulatory architecture underlying tissue-specific gene regulation.

Fine-mapping and validation of regulatory variants

Following rigorous quality control and the mitigation of confounding factors via a linear mixed model implemented in OmiGA[37] (**Methods**; **Supplementary Fig.** 6), we estimated the *cis*-heritability (*cis*-h^2) of seven molecular phenotypes (**Supplementary Fig.** 7), and subsequently mapped molecular quantitative trait loci (molQTL) in the *cis*-windows across 43 tissues. These molQTL included four types of abundance-based signals: eQTL for gene expression, eeQTL for exon expression, enQTL for enhancer expression, isoQTL for isoform expression; and three types of structure-based signals: sQTL for alternative splicing, 3'aQTL for 3'UTR APA, and stQTL for RNA stability (**Fig.** 3a). We detected 19,180 (91.9% of tested genes) eGenes (with at least one significant variant) for gene expression, 16,358 (78.4%) eeGenes for exon expression, 7,529 (36.1%) isoGenes for isoform expression, 5,903 (28.3%) enGenes for enhancer expression, 8,454 (40.5%) sGenes for alternative splicing, 15,205 (72.9%) stGenes for RNA stability, and 7,851 (37.6%) 3'aGenes for 3'UTR APA in 35 tissues, collectively termed molGenes (**Fig.** 3a; **Supplementary Table** 9). Of them, there were 18,192 (82.9% of tested genes) genes associated with at least two types of molQTL, and 747 (3.4%) genes that were simultaneously associated with all seven molQTL. In addition, 2,203 (10%) genes were associated with a single molQTL only, including 865 eGenes, 412 eeGenes, 298 sGenes, 226 isoGenes, 183 enGenes, 147 stGenes, and 72 3'aGenes. Further statistical fine-mapping analyses for molGenes identified 771,007 credible sets (variants with cumulative PIP ≥ 95%), among which a substantial fraction contained only a single variant, such as 37.7% (85,604) of eQTL, 41.8% (75,769) of eeQTL, 29.2% (14,547) of enQTL, 35.2% (26,774) of isoQTL, 44.0% (35,710) of sQTL, 40.8% (46,077) of stQTL, and 52.3% (21,679) of 3'aQTL (**Fig**. 3b**; Extended Data Fig.** 1a-b).

To assess the reliability of molQTL identified above, we conducted internal and external validations, allele-specific expression (ASE) analysis at individual level, and functional enrichment (**Extended Data Fig.** 2). Internal validation across 28 tissues with sample size over 100 yielded a high replication rate (π_1, mean = 0.93) and strong effect-size consistency (mean Pearson's r = 0.89) (**Extended Data Fig**. 2a). External validation using blood eQTL in an independent population of 99 Holstein cattle achieved a decreased but remain robust replication rate of 0.76 and effect size concordance of r = 0.6 (**Extended Data Fig.** 2c). At the individual level, lead eQTL effect sizes were significantly correlated with ASE in whole blood (r = 0.66) and muscle (r = 0.57) (**Extended Data Fig.** 2b). Beyond statistical validation, the identified molQTL recapitulated known regulatory logic for corresponding molecular phenotypes: abundance-based signals (e.g., eQTL and isoQTL) clustered near TSS, whereas structure-based signals (e.g., sQTL and 3'aQTL) were polarized toward the

transcription end site (TES) (**Extended Data Fig.** 1c-d). Notably, isoQTL exhibited the highest enrichment in start-gain and 5'UTR, while 3'aQTL and enQTL specifically targeted 3'UTR/stop-codons and enhancer-like chromatin states, respectively (**Extended Data Fig.** 1c). Finally, we verified the robustness of our approach by comparing molQTL mapping using RNA-seq imputed versus whole-genome sequencing (WGS)-called genotypes in an independent population with paired RNAseq and 30x WGS data (n = 99) (**Extended Data Fig.** 2d-m). Both approaches obtained comparable statistical power for eGene discovery and identified the same causal signals, as evidenced by lead variants being in high LD with consistent effect sizes and genomic distributions (**Extended Data Fig.** 2d-m).

Functional and evolutionary signatures of primary and non-primary effects

To further explore the complexity of the bovine regulatory landscape beyond the lead association signals, we performed conditional analysis to identify independent variants. Across all seven molecular phenotypes, an average of 21.0–24.3% of genes harboured at least two independent molQTL (**Fig**. 3c; **Extended Data Fig.** 3). Within each molGene, we defined the variant with the strongest association as the primary effect and subsequent independent variants as non-primary effects. Notably, compared to genes with single signals, genes with multiple independent signals exhibited significantly higher *cis-h²* and lower evolutionary constraint at the sequence level (**Fig**. 3d; **Supplementary Fig.** 9), a pattern consistent with a regime of relaxed purifying selection[32]. This observation suggests that genes with more "robust" or redundant regulatory architectures might be better positioned to accumulate and maintain diverse genetic regulatory variation. Compared to primary molQTL, variants with non-primary effects were located at greater distances to the TSS and displayed lower MAF (**Fig**. 3e; **Supplementary Fig.** 9). While primary variants typically dominated core regulatory elements (e.g., 5′UTR and active promoter), non-primary molQTL showed specialized enrichments. For instance, compared to their primary effects, non-primary enQTL and 3′aQTL showed significantly higher enrichments in bivalent enhancers, while non-primary isoQTL, sQTL, and stQTL were enriched in non-coding exon variants, poised enhancer, and start-gained codon, respectively (**Extended Data Fig.** 4). These findings suggested that while primary variants govern the baseline molecular output, non-primary variants might provide a secondary layer of "fine-tuning" through distal or context-specific regulatory elements.

Finally, to evaluate the power to detect independent regulatory signals, we performed a down-sampling analyses using the whole blood samples (n = 2,195) to model the relationship between discovery power of molQTL and sample size. We observed that the detection of both primary and non-primary signals follows a non-saturated trajectory at the current sample size, with the discovery of non-primary effects showing a particularly steep dependence on sample size (**Fig**. 3f). These results highlight that a substantial portion of the bovine "fine-tuning" regulatory architecture remains hidden, and further expansion of the CattleGTEx resource will be essential to achieving a saturating map of gene expression regulation.

Pleiotropic effects of genomic variants on molecular phenotypes

To further examine whether individual genomic variants act as multi-layer regulators or site-specific modulators (vertical pleiotropic effect), we performed colocalization analyses between eQTL and the other six molQTL. Our results revealed a fundamental decoupling between abundance-based and structure-based regulation (**Extended Data Fig.** 5a). More specifically, 39.5% of sQTL, 35.1% of 3aQTL, and 39.2% stQTL showed distinct signals (PP.H3 > 0.8) from their corresponding eQTL (**Extended Data Fig.** 5a), a finding corroborated by the presence of low LD between lead variants (LD range: 0.06–0.22) (**Extended Data Fig.** 5b). These results suggest that different layers of the transcriptomic lifecycle, such as total abundance versus isoform structure, are frequently governed by distinct genetic determinants. Examples of this decoupling include genes such as *VWA3B* in the brain and *MRPS25* in adipose tissue, where eQTL and independent structural molQTL coexist within the same tissue but operate through independent genomic variants (**Extended Data Fig.** 5d).

Despite the prevalence of layer-specific regulation, 40.0%, 20.7%, 6.2%, 5.1%, 16.6% and 10.7 % of eGenes were colocalized (PP.H4 > 0.8) with eeGenes, isoGenes, sGenes, 3aGenes, stGenes and enGenes, respectively (**Extended Data Fig.** 5c). We identified 41,425 pleiotropic eQTL (including 31,044 primary and 10,381 non-primary eQTL) that modulated at least one additional molecular phenotype, with some variants (2.4%) acting as master regulators for up to six molecular phenotypes (**Fig**. 3g). Crucially, non-primary eQTL contributed a disproportionately higher fraction of pleiotropic effects when coordinating gene expression with enhancer activity (enQTL) or structural phenotypes, whereas primary eQTL dominated the regulation of abundance-based phenotypes such as exon and isoform expression (**Fig**. 3g). Furthermore, the contribution of non-primary eQTL increased progressively with the degree of pleiotropy (**Fig**. 3h). This suggest that while primary eQTL drive the ma-

jority of total expression variance, non-primary eQTL likely play a key role in coordinating the interplay between expression and post-transcriptional modulation (such as splicing and RNA stability). Genes regulated by genomic variants with higher pleiotropic effects were characterized by elevated expression levels, higher *cis-h²*, and reduced evolutionary constraint, mirroring the "relaxed selection" profile observed for complex regulatory architectures (**Fig**. 3i). The respective pleiotropic variants displayed higher MAF, larger effect sizes on gene expression, and stronger enrichment in regulatory regions (e.g., 5′UTRs, 3′UTRs, promoters, and enhancers) (**Fig**. 3j; **Extended Data Fig.** 5e).

Tissue-sharing patterns of regulatory effects

In general, tissues with similar biological functions (e.g., within the nervous or digestive systems) tended to cluster together based on their molQTL effect profiles (**Fig.** 4a; **Supplementary Fig**. 10). However, rumen, a ruminant-specific digestive organ, clustered separately from other intestinal regions across all molQTL types, underscoring their specialized cellular composition, developmental origin and biological roles in plant-based digestion[38]. These regulatory effects tended to be either tissue-specific or ubiquitous across tissues, consistent with previous findings in other species[18-20,22] (**Fig**. 4b). The eQTL that were active in more tissues (stronger horizontal pleiotropic effects) exhibited smaller effect sizes, higher MAF, shorter distances to TSS, and more enrichment in regulatory elements (e.g., 3'UTR and 5'UTR) (**Supplementary Fig**. 11). Notably, primary regulatory effects showed a slightly greater tissue specificity than non-primary ones (**Fig**. 4c; **Supplementary Fig**. 12a). Furthermore, eQTL with stronger vertical pleiotropic effects tended to be more tissue-specific (weaker horizontal pleiotropic effects) (**Fig**. 4d; **Supplementary Fig**. 12b), indicating that complex regulatory mechanisms are often confined to specific tissue contexts.

To systematically characterize tissue specificity of eQTL between any two tissues, we first identified three classes of tissue-specific eGenes (**Fig**. 4e) : **1) Specific-expression**, defined as genes expressed in only one of each of the tissue pairs (e.g., *INMT* between liver and brain, **Fig**. 4f); **2) Specific-eGenes**, defined as detected eGenes in only one tissue despite being expressed in both tissues (e.g., *C21of140* between testis and muscle, **Fig**. 4g); **3) Specific-eQTL**, defined as lead variants of shared eGenes were weakly linked (LD < 0.2), such as *OTOR* between muscle and testis (LD = 0.02; **Fig**. 4h). We defined shared-eQTL when lead variants of same eGenes were highly correlated (LD ≥ 0.8) (**Fig**. 4e). These shared-eQTL were further divided into three classes according to their direction and magnitude of effect sizes between tissues: **1) Opposite**, eQTL exhibiting opposite effect directions between tissues (e.g., *GPR39* between mammary gland and adipose, **Fig**. 4i); **2) Heterogeneous**,

eQTL with concordant effect directions but distinct effect size (non-overlapping 95% confidence intervals for effect size estimates), such as *TMEM106C* between lung and other tissues (**Fig**. 4j); **3) Homogeneous**, eQTL with overlapping confidence intervals (e.g., *RCAN1* across immune system tissue and lung, leukocyte, and testis (**Fig**. 4k).

Among all eGenes, **Specific-eGenes** constituted the largest proportion (64%) across tissues, followed by **Specific-expression** (7%)**,** and **Specific-eQTL** (2%, **Fig**. 4l). Among shared-eQTL, **Homogeneous** constituted the largest proportion across tissues (63%), followed by **Heterogeneous** (33%) and **Opposite** (4%)**.** These shared-eQTL showed shorter distances to TSS, higher MAF, larger effect sizes, and stronger enrichment in active regulatory elements compared with other eQTL (**Fig**. 4m; **Supplementary Fig**. 13). Notably, among shared-eQTL, **opposite** showed a strongest enrichment in enhancers and flanking TTS regions (**Fig**. 4m), indicating that these variants may participate in context-dependent or antagonistic regulatory mechanisms that fine-tune gene expression in a tissue-specific manner.

Given the widespread tissue-sharing of eQTL, particularly those with small effect size, we conducted a joint *cis*-eQTL mapping analysis by pooling 8,743 non-redundant samples across all tissues (see **Supplementary Notes**). This integrative approach identified 30,880 additional fine-mapped eQTL (variants with cumulative PIP ≥ 95%) that were previously undetected in single-tissue analyses. Compared to eQTL detected via the single-tissue framework, these newly discovered variants were characterized by smaller effect sizes, lower MAF, greater distances to TSS, and higher enrichment in enhancers (**Supplementary Fig**. 14).

Deciphering the regulatory architecture underlying population selection signatures

To explore the regulatory conservation between *B. taurus* and *B. indicus*, we conducted population-stratified eQTL mapping across eight tissues, each with at least 40 individuals for both two subspecies (**Methods**; **Supplementary Fig**. 15a-b). In general, effect sizes of population-stratified eQTL clustered primarily by tissue types and then by subspecies (**Fig**. 5a). The eQTL showed higher replication rates (mean π_1 = 0.73) and effect-size concordance (mean Spearman's ρ = 0.76) between subspecies than between tissues (**Fig**. 5a; **Extended Data Fig.** 6a). Even though they diverged at least 300,000 years ago, these results highlighted the conservation of regulatory effects between the two subspecies[39-41].

To further explore the regulatory adaptation of *B. indicus* to heat, humidity and parasites challenges in tropical environments, we scanned the genome-wide selection sweeps between 200 *B. taurus* and 142 *B. indicus* with WGS data (> 10×) via an F_{ST}-based approach[34] (**Supplementary Table** 11; **Methods**). The 248,679 genomic regions tested were evenly divided into 10 deciles (F_{ST}1-10) based on their F_{ST} values from the smallest (no selection) to the largest (strongest selection). For example, the F_{ST}10 genomic region contained *DNAJC18* (F_{ST} = 0.86), *HOXC4* (F_{ST} = 0.77), and *FAAP20* (F_{ST} = 0.77), which were previously reported to be implicated in environmental adaptation between cattle populations[42-44] (**Supplementary Fig**. 16a). The enrichment of fine-mapped molQTL in each of 10 deciles was then tested. Notably, tissue-specific molQTL showed a significantly higher enrichment in regions with larger F_{ST} values (**Fig**. 5b; **Supplementary Fig**. 16b). As expected, tissues with higher enrichment of eQTL in strongly selected regions showed lower correlation of effect-size between *B. taurus* and *B. indicus* in the population-stratified analysis (**Extended Data Fig.** 6b). Among all tissues, molQTL specific in skin, adipose, immune and digestive tissues displayed the highest enrichment in the F_{ST}10 group (**Fig**. 5c; **Supplementary Fig**. 16c). The eGenes of these tissues located in strongly selected regions were significantly involved in immune-related functions such as choline metabolism (FDR = 0.02) and bacterial invasion (FDR = 0.01; **Fig**. 5c). For example, the lead eQTL (rs135226326) of *FUT2*, a skin-specific eGene, resided in a strong selection region (F_{ST} = 0.54; **Fig.** 5d). The G allele of rs135226326 was associated with increased *FUT2* expression and occurred at a higher frequency in *B. indicus*, potentially enhancing immune responsiveness in extreme environments and thereby contributing to adaptive differentiation between *B. taurus* and *B. indicus* due to the role of *FUT2* in mediating antiviral immune defence[45]. These results suggest that regulatory variants in immune and metabolic tissues have played important roles in the divergence of *B. taurus* and *B. indicus*, likely reflecting selection pressures related to environmental, metabolism, and immune defence adaption[2].

To elucidate the regulatory mechanisms underlying intensive artificial selection between dairy and beef breeds in more recent timescales, we conducted population-stratified eQTL in 11 tissues **(**sample size > 40 in each of breed**; Supplementary Fig.** 15c-d**)**. Similar to findings between *B. taurus* and *B. indicus*, effect sizes of eQTL were generally conserved between dairy and beef (**Fig**. 5f; **Extended Data Fig.** 7). Compared to tissue-shared eQTL, tissue-specific ones also showed significantly higher enrichment in strongly selected regions between 651 dairy and 552 beef cattle detected using WGS data (**Supplementary Fig**. 17a; **Supplementary Table** 12). In contrast to the divergence between *B. taurus* and *B. indicus*, tissues with highest enrichment of eQTL in strongly selected regions

between dairy and beef were brain-related tissues, particularly hypothalamus, rather than skin and immune tissues (**Fig**. 5g; **Supplementary Fig**. 17b). Brain-specific eGenes located in the strongly selected regions (F_{ST}10) were significantly enriched in pathways related to growth and developmental regulation, including glutathione metabolism (FDR = 0.03) and the GnRH signalling pathway (FDR = 0.03; **Fig**. 5g). This indicated that artificial selection between dairy and beef cattle has preferentially targeted neuroendocrine and growth-regulatory mechanisms. For example, the lead eQTL rs383563305 for *HTR2B*, a brain-specific eGene, resided in a region under strong selection (F_{ST} = 0.58; **Fig**. 5h). The allele G in rs383563305 was associated with upregulated expression of *HTR2B* and showed a higher frequency in beef compared to dairy cattle (**Fig**. 5i). As *HTR2B* mediates serotonergic regulation of hypothalamic appetite control, selection on its eQTL likely reflects divergent breeding pressures on feed intake, energy allocation and growth-related traits between beef and dairy cattle[46]. Altogether, these results suggest that both long-term adaptive evolution and intensive artificial breeding tended to target on variants with tissue-specific effects rather than those with ubiquitous effects.

The genetic and molecular architecture underlying bovine complex traits

To investigate the regulatory mechanisms underlying complex traits in cattle, we examined 2,070 GWAS loci of 44 complex traits of economic importance, encompassing 22 body conformation, 6 milk production, 8 reproduction, and 8 health traits (**Supplementary** Table 13). We systematically explored enrichment patterns of these GWAS loci against all seven types of molQTL and different categories of regulatory variants based on effect size, tissue-sharing pattern, and pleiotropy. Overall, eQTL showed the strongest enrichment with GWAS loci, followed by eeQTL, 3'aQTL, sQTL, isoQTL, and stQTL (**Fig**. 6a). Notably, compared to primary effects, non-primary effects exhibited higher enrichment in GWAS loci, consistent across all types of molQTL (**Fig**. 6a). Compared to tissue-shared effects, tissue-specific effects were more enriched in GWAS loci (**Fig**. 6a; **Extended Data Fig.** 8a). The eQTL with stronger pleiotropy showed higher enrichment in GWAS loci (**Fig**. 6a). These observations were further supported by their similarity to GWAS loci in terms of distance to TSS, population selection pressure (measured by MAF-effect size relationship), and enrichment in regulatory elements (**Fig**. 6b-d; **Extended Data Fig.** 8b-c). We thereby propose a genetic regulatory model in which genetic effects spanning multiple layers of the central dogma (vertical pleiotropic effects) are more likely to propagate to complex traits, whereas variants acting across multiple biological contexts of a single biological layer (horizontal pleiotropic effects) may incur greater fitness costs and are therefore less likely to translate their effects to higher-order phenotypes (**Fig**. 6e).

To pinpoint putative causal variants and genes governing complex traits, we performed colocalization analyses of molQTL and GWAS loci. Overall, 75% (1,545) of GWAS loci, ranging from 33% (5 loci) for calving-to-conception interval to 96% (96 loci) for milk protein percentage, were colocalized (PP4 > 0.8) with at least one molQTL (**Supplementary Fig.** 18a). Each molQTL type also made unique contributions to GWAS colocalization. Among them, eeQTL accounted for the largest number of unique colocalizations (n = 64), followed by eQTL (29), stQTL (24), sQTL (21), isoQTL (17), enQTL (15), and 3′aQTL (12; **Supplementary Fig.** 18b). For instance, an isoQTL of *PHACTR2* in whole blood was exclusively associated with the udder cleft trait, while an sQTL of *NUDT6* in the corpus luteum was exclusively associated with rump width trait (**Extended Data Fig.** 9). Together, these findings underscored the importance of integrating multiple molecular phenotypes to elucidating the molecular basis of complex traits. Consistent with enrichment analyses, non-primary, tissue-specific, and pleiotropic eQTL showed higher colocalization proportions (PP4 > 0.8) with GWAS loci compared to their counterparts (**Fig**. 7a). Each category also made distinct contributions to explaining GWAS signals: non-primary eQTL contributed 57 additional colocalizations beyond primary eQTL, tissue-specific eQTL contributed 79 unique colocalizations, and pleiotropic eQTL accounted for 34 unique colocalizations (**Fig**. 7b). For example, a colocalization (PP4 = 0.93) between rump width and *EPN3* in the liver was detected for a non-primary eQTL (rs378415691) after conditioning on the primary eQTL (rs379113614). As an endocytic adaptor protein, *EPN3* regulates lipid transport and signal transduction, potentially influencing growth and development in cattle[47]. For pleiotropic eQTL, *GIGYF1*, which is related to adverse metabolic health related to higher glycemic levels and fat mass[48], showed high colocalization with the net merit trait through both eQTL (PP4 = 0.96) and stQTL (PP4 = 0.93) effects driven by the same lead variant (rs109407080). In addition, the lead eQTL rs135173214 of *ENSBTAG00000050359* exhibited strong colocalization with somatic cell score (SCS) specifically in the mammary gland (PP4 = 0.87) but not in other tissues (e.g., adipose PP4 = 2.4×10^{-2}), highlighting tissue-specific regulatory control of this gene in mastitis-related traits (**Fig**. 7c).

To further prioritize tissues relevance to complex traits, we adjusted the number of colocalized genes by the total number of molGenes detected in each tissue. Granulosa cells, which play a central role in follicular growth and hormonal signalling[49], showed high relevance with multiple complex traits, including body conformation traits (e.g., stature and rear udder attachment) and reproductive traits (e.g., sire calving ease and gestation length). Mammary gland and skin exhibited high relevance with milk production and health traits, such as livability and SCS. Colon and rumen showed high

relevance with milk production traits, including milk yield and fat percentage (**Fig**. 7d; **Supplementary Fig.** 19). These findings were generally consistent across all types of molQTL (**Supplementary Fig.** 18c).

To explore the molecular architecture underlying pleiotropic effects of GWAS loci, we detected 151 loci associated with at least two complex traits (**Supplementary Table**. 14). Overall, eQTL with stronger pleiotropic effects and broader tissue-sharing patterns showed higher colocalization rate with GWAS loci associated with more complex traits (**Fig**. 7e). For example, rs109421300 was associated with the gene and exon expression of *DGAT1* in whole blood and liver, a gene with major effect on milk yield and composition in cattle[50], and showed strong colocalization with multiple complex traits, including protein percentage, net merit, and SCS (**Fig**. 7f). In addition, the lead SNP (rs110878911) that was associated with both gene and isoform expression of *LSP1* in leukocytes and liver, showed strong colocalization with multiple complex traits, including calving ease and body strength (**Supplementary Fig.** 20). *LSP1* encodes a protein that regulates immune cell activity and is involved in lymphocyte migration, activation, and intercellular interactions[51]. This gene has also been linked to the development of multiple diseases in humans, including rheumatoid arthritis, Crohn's disease, and various cancers[52].

Evolutionary constraint of regulatory effects between cattle and humans

To explore the evolutionary constraint of regulatory effects, we aligned 198,897 fine-mapped cattle eQTL to the human genome *via* LiftOver (referred to as orthologous variants) and then systematically evaluated the impact of these orthologous variants on genome function and complex traits in humans **(Fig.** 8a; See **Methods)**. Compared to MAF-matched random variants (51.27%) in cattle, fine-mapped eQTL exhibited a higher mapping rate to the human genome, with 53.25~74.45% across different groups of eQTL, indicating cross-species conservation of regulatory variants at the DNA sequence level. Primary eQTL (57.53%) were more evolutionarily constrained than non-primary ones (54.34%). Notably, variants with higher pleiotropic effect were more evolutionarily constrained (**Fig.** 8b). We then used *AlphaGenome* to explore the functional impact of these orthologous variants in the human genome[53]. Compared with MAF-matched random variants, orthologous variants of cattle eQTL exhibited significantly higher predicted quantile scores, indicative of the functional conservation of regulatory variants between cattle and humans. Orthologous variants of tissue-shared eQTL showed higher *AlphaGenome*-predicted effects than those of tissue-specific variants, and variants with higher pleiotropic effect on molecular phenotypes also displayed higher *AlphaGenome*-predicted

effects (**Fig.** 8c). To test whether orthologous variants of cattle eQTL tended to overlap the same human regulatory elements as eQTL detected in human GTEx[54], we assessed their enrichment across 15 human chromatin states predicted from ENCODE data[55]. Notably, human-orthologous variants of cattle eQTL exhibited an enrichment profile across these chromatin states that closely resembled that of human GTEx eQTL (Pearson's r = 0.809, P-value = 2.62e-4) (**Fig. 8d; Supplementary Fig.** 22a-d). Notably, variants with higher pleiotropic effect exhibited a higher enrichment in CTCF and enhancers (**Supplementary Fig.** 21d).

To further investigate whether cattle eQTL can provide evolutionary evidence for dissecting complex traits in humans, we employed S-LDSC to partition heritability of 133 human complex traits using orthologous variants of cattle eQTL[56] (**Supplementary Table** 15). These orthologous variants were significantly enriched for heritability across all complex traits (**Supplementary Table** 16-17). Orthologous variants of tissue-specific cattle eQTL showed significantly higher enrichment than those of the remaining cattle eQTL, consistent with reports from human studies[57] (**Fig.** 8e). For instance, compared to immune-specific eQTL detected in human GTEx, orthologous variants of cattle immune-specific eQTL displayed significantly higher enrichment in human immune-related traits (**Fig 8e; Extended Data Fig.** 10), supporting the idea that immune-related regulatory programs showed substantial conservation between cattle and humans[5,6]. Furthermore, we performed colocalization between human GTEx eQTL and 133 human GWAS traits, yielding 508,191 SNP–gene–tissue trios (PP.H4 > 0.8). Among them, 344 SNP–gene–tissue trios were tested in the CattleGTEx (**Extended Data Fig.** 10c**; Supplementary Table** 18), and 158 of them were also eQTL in cattle (FDR < 0.05), which were significantly more than that of MAF-matched random variants (P-value = 0.0021) (**Extended Data Fig.** 10d). Moreover, trios that were significant in the CattleGTEx exhibited stronger colocalization (higher PP.H4 values) with human complex traits than non-significant trios (P-value = 4.36×10^{-13}; **Fig. 8f**) and were associated with more complex traits (**Fig. 8g**, P-value = 2.22×10^{-2}). For instance, a variant (rs11559147 in humans and rs210096698 in cattle) showed conserved regulatory effects on *ZDHHC4* expression in whole blood in both humans and cattle, and it was significantly colocalized with a GWAS locus of neutrophil count in humans (PP.H4 = 0.99; **Fig. 8h**). The *ZDHHC* family encodes palmitoyltransferases that catalyze reversible protein S-palmitoylation[58], thereby modulating membrane localization and signalling complex assembly of immune regulators and fine-tuning key immune signalling cascades[59,60]. Altogether, these results suggest the presence of evolutionary constraint regulatory variants between humans and cattle, which can facilitate the prioritization of causal variants and genes associated to complex traits and diseases in humans.

Discussions

The CattleGTEx Phase1, built on 12,422 RNA-seq samples of 43 tissues from 8,956 animals worldwide, provided a comprehensive multi-tissue catalogue of regulatory effects of 4,317,531 genomic variants in cattle. Compared with the pilot phase, we extended the number of targeted tissues from 23 to 43, revealing more tissue-specific regulatory effects. In addition, reaching an average increase of 378 samples per tissue allow us to investigate non-primary variants. We also systematically explored pleiotropic regulatory effects of genomic variants on seven molecular phenotypes, revealing the distinct genetic architecture underlying molecular phenotypes, particularly between abundance and structure-based phenotypes. By examining selection signatures of *B. taurus* vs. *B. indicus* and dairy vs. beef, we demonstrated that both long-term adaptive evolution and short-term intensive artificial breeding were more likely to act on genomic variants with tissue-specific regulatory effects, which is also agreed with findings in sheep[15]. During divergence between *B. taurus* and *B. indicus*, selection signals were most pronounced in immune tissues, adipose, and skin, with enriched pathways related to immune responses such as bacterial invasion, consistent with adaptation of *B. indicus* to heat, humidity, and parasite exposure in tropical environments[2]. In contrast, dairy-beef divergence is not driven by environmental pressures but by strong artificial selection for production traits, including milk and meat yield[3]. Accordingly, selection signals in this comparison were concentrated in nervous and digestive tissues, with enriched pathways related to growth, metabolism, and production. This finding agreed with previous studies, highlighting the importance of the nervous system in responses to artificial selection for dairy cattle and associated milk production traits[18,61].

The CattleGTEx Phase1 provided novel insights into the regulatory mechanisms underlying 44 complex traits in cattle. This enabled us to link 1,545 (75%) of GWAS loci to molecular phenotypes, genomic variants, target genes, and relevant tissues, representing a substantial improvement over the 260 (13%) colocalized GWAS loci reported in the pilot CattleGTEx study for the same complex traits. The different categories of molQTL exhibited distinct contributions to complex traits, with non-primary effects, pleiotropic eQTL, and tissue-specific molQTL showing higher enrichment for GWAS loci compared to the rest of regulatory variants. For the strong enrichment of non-primary and tissue-specific molQTL with GWAS loci, one possible explanation is that variants with large regulatory effects or effects shared across many tissues are more likely to perturb multiple biological processes and therefore incur stronger fitness costs. In contrast, regulatory variants that act in more restricted contexts, such as secondary regulatory layers or specific tissues, may avoid these strong selection pressures, allowing them to persist in populations and contribute to phenotypic variation

detected by GWAS[62]. In addition, the enrichment between vertical pleiotropic eQTL and GWAS loci may reflect buffering mechanisms that operate across diverse biological processes or layers along the central dogma to main the stability of the whole system. Many genetic perturbations at a single molecular layer might be compensated through regulatory feedback, network redundancy, or post-transcriptional control. However, genetic variants that simultaneously affect multiple layers of molecular regulation may be more likely to bypass such buffering mechanisms. Their effects can therefore propagate more effectively through molecular networks and ultimately manifest at the organismal level, increasing the probability that they contribute to complex trait variation detected by GWAS. These findings also underscore the importance of using larger sample sizes to detect more variants with small, additive effects, profiling more molecular phenotypes, and considering additional tissue/cell types to better understand the genetic and molecular architecture underlying complex traits[30-33]. By aligning fine-mapped cattle eQTL to the human genome, we revealed evolutionary constraints on regulatory effects, facilitating the prioritization of causal variants and genes associated with complex traits and diseases in humans. For instance, a variant exhibited conserved regulatory effects on ZDHHC4 expression in blood in both humans and cattle and was significantly colocalized with neutrophil count in humans. Previous studies have reported that ZDHHC4 enhances the efficacy of immunotherapy, a role that has been experimentally validated[63].

Despite these advances, we noted several limitations that will be addressed through future development of the CattleGTEx Project: **1**) only common SNPs were considered in this phase. More RNA-seq samples with paired WGS, improved reference genomes (i.e., telomere-to-telomere genomes[64] and pangenomes[65]), and enhanced functional annotations (e.g., by integrating CAGE data to refine transcription start site resolution and mitigate tissue-specific annotation artifacts[66]) will enable us to explore regulatory effects of rare variants, transposable elements, short tandem repeats and complex structural variants. **2**) The current dataset remains biased toward commercial breeds, limiting the representation of global genetic diversity in cattle. Future efforts should prioritize the inclusion of local and indigenous breeds worldwide, which are shaped by distinct ecological and environmental conditions, to better capture regulatory mechanisms underlying local environmental adaptation and conservation-relevant traits[67]. **3**) This study focused on molecular phenotypes that were derived from short-read RNA-seq. Long-read RNA-seq data will be essential for more accurate quantification of molecular phenotypes, particularly isoform expression and alternative splicing[68]. Moreover, integrating additional molecular layers, such as profiles from epigenomics, proteomics, metabolomics, and

gene regulatory networks[69-72], will be essential to capture downstream and complex regulatory mechanisms that cannot be resolved at the transcriptomic level alone. **4**) We only conducted molQTL mapping at bulk-tissue resolution and therefore did not fully capture regulatory heterogeneity across cell types and states[73-75]. Although we explored cell-type-dependent regulation using computational deconvolution approaches in a companion study[21], population-scale single-cell molQTL mapping[73,76], together with spatially resolved transcriptomic data[77,78], will be required to dissect regulatory mechanisms operating at finer cellular and spatiotemporal scales. **5**) This study considered a limited set of biological contexts, primarily focusing on different tissue types and breeds. Future development will incorporate additional contexts, including sex, developmental stages, and diverse environmental conditions[79-81], which will require the establishment of a FarmGTEx tissue biobank with standardized sampling protocols, controlled environmental conditions, and comprehensive metadata collection, to characterize context-dependent regulatory mechanisms in cattle and further advance our understanding of the genetic architecture of complex traits. **6**) The regulatory variants prioritized through computational analyses in this study remain largely inferential. Experimental validation, including *in vitro* assays and *in vivo* functional studies[82,83], will be crucial to confirm causal regulatory effects and to translate these findings into mechanistic understanding of complex traits.

Author contribution statement

L.Fang conceived and designed the project. H.Li., H.Zhang and Z.Wei performed bioinformatic analyses of RNA-Seq data analysis. P.Zhao conducted whole-genome sequence data analysis. H.Zhang performed multi-omics analysis. H.Li and H.Zhang conducted molQTL mapping. H.Li performed GWAS integrative analysis. H.Li and P.Zhao performed selection region analysis. D.Zhu performed comparative analysis between human and cattle. L.Fang., H.Li., S.Peter., Y.Zhou., X.Wang., X.Lu., Q.Zhang., J.Li., Y.Wu., W.Li., J.Cole., Z.Cai., G.Sahana., D.Liu., Y.Li., D.Sun., Y.Yu., Y.Zhang., B.Han., H.Sun., M.Jia., S.Zhu., X.Liu., P.Su., L.Jiang., Y.Liu., W.Lyu., E.Clark., M. Littlejohn., R.Xiang., Y.Chen., Z.Pan., Y.Hou., M.Li., C.Xu., Q.Liu., Q.Yang., Z.Yang., H.Jiang., X.Lu., M. Wijst., E. Mármol-Sánchez., H.Zhou., L.Guan., X.Zhao., I. Ionita-Laza., J.Grady., D.MacHugn., L.Frantz., A.Leonard., H.Pausch., O.Madsen., M.Gòdia., M.Derks., R.Crooijmans., E.Ibeagha-Awemu., M.Wang., Y.Wang., W.Liu., S.Chen., J.Tian., M.Tian., H.Zhang., B.Lin., B.An., C.Fang., J.Du., Z.Zhang., S.Zhao., Y.Gao., Y.Tang., Q.Lin., J.Teng., D.Guan., X.Liu., W.Zheng., Q.Zhang and M.Gong contributed to the critical interpretation of analytical results before and during manuscript preparation. Y.Wang., J.Sheng and M.Wang built the Cattle-GTEx Phase1 web portal. L.Fang., M.Lund., B.Li., Y.Wang and J.Jiang. contributed to the data and computational resources. H.Li, H.Zhang, D.Zhu., and L.Fang drafted the manuscript. All authors read, edited, and approved the final manuscript.

Online Methods

Ethics and sample collection

The vast majority of RNA-seq samples and all WGS are publicly available. No Ethics of animal sampling are not applicable. For the newly generated data, all animal procedures and experimental protocols were approved by the Institutional Animal Care and Use Committee of Northwest A&F University following the recommendations of the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

Bulk RNA-seq data processing

We downloaded 24,975 RNA-seq datasets from the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) and 1,385 datasets from the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa) using the Aspera command-line tool (**Supplementary Table** 1-2). In addition, we newly generated 1,268 RNA-seq datasets from 3 tissues of 533 Holstein cattle, including 514 samples in lung, 459 sample in rumen and 295 samples in lymph node (See **Supplementary Notes**). After raw RNA sequencing data were obtained, the adapters were removed along with low quality reads using Fastp (v0.23.4) with parameters: --trim_front 3, --trim_tail 3, --length_required 36, --cut_right, --cut_window_size 4 and --cut_mean_quality 15[84]. Clean reads were then aligned to the B. taurus reference genome (ARS-UCD1.2, Ensembl v108[85]) using STAR (v 2.7.10b) with parameters: --chimSegmentMin 10, --twopassMode Basic, --outFilterMismatchNmax 10, --chimOutJunctionFormat 1, --outFilterType BySJout, --alignSJoverhangMin 8 and -- alignSJDBoverhangMin 1[86]. In total, 19,790 samples with >10M mapped reads, >40% uniquely mapped reads, and >40% expressed genes were retained for downstream analyses. The detailed mapping quality of all the RNA-seq samples is summarized in **Supplementary Table** 4.

Genotype calling and imputation of RNA-seq samples

PCR duplicates were removed in bam files produced by STAR aligner previously using the MarkDuplicates module of Genome Analysis Toolkit (GATK, v4.3.0.0)[87], and then SplitNCigarReads module was engaged to split reads that contained Ns in their cigar string. Next, GATK BaseRecalibrator module and ApplyBQSR module were applied to recalibrate base quality score based on the Ensembl dbSNP database (v108). The recalibrated BAM files were used for direct genotype imputation with GLIMPSE2 (v2.0.0)[88], based on a multi-breed genotype reference panel[31]. Variants with minor allele frequency (MAF) < 0.05 or imputation quality score (INFO) < 0.75 were subsequently filtered out

using BCFtools (v1.17)[89], resulting in a total of 4,317,531 SNPs. After obtaining the genotype files, WASP (v0.3.4) was used to remove mapping bias of RNA-seq data[90]. SNPs from each chromosome were extracted separately using BCFtools (v1.17)[89], and the snp2h5 script provided in the WASP package was used to convert the resulting VCF files into HDF5 format[90]. These HDF5 files were subsequently used to correct mapping bias for each sample. The find_intersecting_snps.py script was applied to extract reads overlapping with SNPs, and STAR (v2.7.10b) aligner was then employed to remap these reads using the same parameters as in the initial mapping[86]. Reads for which one or more allelic versions failed to remap to the same genomic location were filtered out using filter_remapped_reads.py. Finally, duplicate reads were removed using rmdup.py or rmdup_pe.py for single-end and paired-end RNA-seq data, respectively.

Quantification of molecular phenotypes

After mapping bias removal, the corrected bam file was used for molecular phenotype definition and quantification. Normalized expression (TPM) of 23,103 protein-coding and lncRNA genes (Bos_taurus. ARS-UCD1.2, Ensembl v108) were generated using Stringtie (v.2.2.1)[91]. Raw read counts for all 20,875 genes and 102,033 exons were obtained with featureCounts (v2.0.3)[92]. Transcribed enhancer regions were annotated, and a custom SAF file was prepared after excluding those overlapping with gene bodies[93], then the raw counts were also quantified using featureCounts (v.2.0.3)[92]. For transcript-level expression, Salmon (v1.10.1)[94] was used in mapping-based mode with clean reads as input to quantify transcript expression, generating both TPM and raw count outputs. The index for Salmon was built with the decoy sequences generated using SalmonTools[94]. For alternative splicing, Leafcutter (v.0.2.9)[95] was used to quantify the splicing events as following: 1) mapping bias–corrected BAM files from the previous step were converted into junction files using the bam2junc.sh script provided by Leafcutter; (2) intron clusters were identified with leafcutter_cluster.py and annotated to genes using map_clusters_to_genes.R; (3) introns with zero read counts in more than 50% of samples or with fewer than max(10 or 0.1n) unique values in each tissue (where n is the sample size) were removed; (4) low-complexity introns were filtered out based on the z score of cluster read fractions across individuals; (5) intron excision ratios were calculated, and the filtered counts were converted to BED format using prepare_phenotype_table.py. Finally, excision levels of introns were normalized as percent spliced-in (PSI) values. For the quantification of 3'UTR alternative polyadenylation (3'UTR APA), Dapars (v2.1) was used with the following steps[96]: 1) distal polyadenylation sites from the Ensembl annotation (v108) were extracted using the script DaPars_Extract_Anno.py provided by Dapars (v2.1) package, 2) genomecov function in BEDTools (v2.31.0) was utilized to obtain wiggle

alignment files[97], 3) all the wiggle files were utilized to quantify APA usage within each tissue[98,99]. The output represents the percentage of distal polyA site usage index (PDUI) for each gene in each sample. Loci with more than 50% missing values were excluded, and the remaining entries were imputed using the K nearest neighbor (KNN) method[22,100]. For RNA stability estimation, genomic coordinates of exon and intron were extracted, and HTSeq was used to quantify raw exon and intron read counts[101,102]. The constitutive exon-to-intron ratio was then calculated for each gene. Genes with more than 50% missing values were excluded, and the remaining missing entries were imputed using the KNN approach[22]. Allele specific expression analysis was conducted using phASER (v1.1.1)[103]. Imputed genotype VCF file together with bam file from WASP software were input to phASER tool for phasing variants with options of "--mapq 2 --baseq 10"[103]. Then, gene-level haplotype counts were obtained using the script phaser_gene_ae.py. Then all the individual haplotypic counts file were aggregated by tissue using the script phaser_expr_matrix.py.

molQTL mapping

Within each tissue, duplicated samples (IBS > 0.9) were removed. The molQTL mapping was performed using a linear mixed model implemented in OmiGA (v1.0.3)[37]. SNPs with MAF < 0.05 or minor allele count (MAC) < 6 were excluded with "--mac-threshold 6 --maf-threshold 0.05" in each of tissue. We defined the *cis*-window of genes as ±1 Mb of TSS and obtained the nominal *P* values of *cis*-molQTL with the option of "--mode cis", while selecting the suitable phenotypic PCs as covariables using the optional of "--dprop-pc-covar 0.001"[14,37]. Covariables with high co-linearity were removed with the option of "--rm-collinear-covar 0.95". We then employed two layers of multiple testing corrections using OmiGA[37]. In the first layer, we applied an adaptive permutation approach to calculate the empirical *P* values of variants within each gene and obtained the permutation *P* value of the lead variant for each gene. In the second layer, we conducted the clipper correction for the nominal *P* values and computed the FDR of lead variants across all tested genes using the option of "--multiple-testing clipper"[104]. A gene was considered as an molGene if at least one SNP had a corrected $P < 0.05$ and a *P*-value below the gene-level permutation-derived threshold calculated by OmiGA (v1.0.3)[37]; and molQTL were defined as SNPs with nominal P values below the gene-level significance threshold.

Conditional and fine-mapping analysis of molQTL

To detect multiple independent *cis*-molQTL signals of each molGene, we employed two complementary approaches. First, conditionally independent molQTL were identified using a stepwise regression

procedure implemented in OmiGA (v1.0.1.3) with "--mode cis_independent", which iteratively identifies the most significant SNPs while conditioning on previously selected variants[37]. We further applied "Sum of Single Effects" (SuSiE) model (v.1.0)[105] to fine-map causal QTL and define credible sets of variants, which uses a Bayesian multiple regression framework to estimate the posterior inclusion probability (PIP) for each variant. Variants with PIP ≥ 0.8 were considered putative causal variants, and 95% credible sets were constructed to capture the most likely causal variants within each signal.

Colocalization analysis between molQTL

To investigate whether molecular phenotypes share genetic regulatory mechanisms with eQTL, we conducted colocalization analysis between each molecular phenotype with eQTL. Colocalization was performed using the *coloc.abf* function in the *coloc* (v5.2.3) R package[106], which implements an approximate Bayes factor framework to identify shared causal variants. For each pair of molecular phenotypes, the method computes posterior probabilities for five hypotheses: (H0) no association with either trait; (H1) association only with the molecular phenotype; (H2) association only with the QTL; (H3) association with both traits but with independent causal variants; and (H4) association with both traits and a shared causal variant. As the results, we defined the colocalized signals between two molQTL by PP.H4 > 0.8, whereas the distinct signals by PP.H3 < 0.25[21]. In addition, LD between the lead SNPs of colocalized signals was calculated using PLINK (v1.9) to further characterize shared regulatory loci[107].

Tissue-sharing/specificity analysis of molQTL

We used multivariate adaptive shrinkage (MashR, v0.2.57)[108] to characterize cross-tissue sharing patterns of molQTL. For the overall estimation, we extracted the z scores (slope / slope_se) of the top molQTL for each molGene across all tissues in which it was detected and used these top-effect molQTL–gene pairs as input to MashR. In addition, one million molQTL-gene pairs were randomly sampled from all nominal associations tested by OmiGA across tissues[37], and their corresponding z-scores were obtained. Pairs with missing values in more than five tissues were removed to avoid inflating tissue-specific signals driven by sparse observations rather than true biological effects; for the remaining pairs, missing z scores were assigned a standard error of 1×10^{6}. MashR was then applied to estimate the local false sign rate (LFSR) for each tissue, and molQTL with LFSR < 0.05 were considered active in that tissue. Tissue similarity was assessed by computing pairwise Spearman correlations of effect sizes among active molQTL.

Functional enrichment of molQTL

To assess whether molQTL were enriched in functional regions, we obtained functional annotations for all tested SNPs using SnpEff (v5.2)[109] and, separately, chromatin-state annotations from the Cattle FAANG project[110]. Enrichment analyses were performed independently for each annotation type. For each tissue and each annotation type, a 2×2 contingency table was constructed comparing (i) SNPs identified as molQTL and (ii) SNPs carrying the given annotation. Odds ratios (ORs), 95% confidence intervals, and *P*-values were computed for each QTL annotation pair. ORs and confidence intervals were derived from the 2×2 tables, with continuity adjustments applied when necessary. *P*-values were obtained using a chi square test of independence. In each tissue, all SNPs tested by OmiGA were used as the background for enrichment calculations.

GWAS summary statistics of complex traits

To investigate the regulatory mechanisms underlying complex traits in cattle, we systematically integrated identified eQTL with GWAS summary statistics for 44 complex traits in Holstein cattle[111], including 22 body conformation, 6 milk production, 8 reproduction, and 8 health traits. Independent SNPs were identified for each trait using the *GCTA-cojo* function in GCTA (v1.94.1)[112], based on the Holstein population in the reference genotype imputation panel[31]. For colocalization analysis, independent SNPs with *p*-values less than 1×10^{-5} were then selected as suggestive significant, and corresponding QTL regions were defined by extracting all SNPs located within ±1 Mb of each lead SNP[21].

Enrichment analysis between *cis*-molQTL and GWAS loci

QTLenrich (v2.0), the software assesses enrichment of common variant associations for a given complex trait among molQTL, was used for enrichment analysis between our seven molQTL and GWAS loci from 44 complex traits[54]. Three input files were prepared, 1). Top pairs were generated with lead SNPs of molGenes in each tissue; 2). Significant pairs were generated with SNPs in which the nominal *P*-value larger than the permutation *P*-value in each gene; 3). All pairs were generated with all SNPs used in molQTL mapping. All parameters were set as default. We next performed multiple testing by using p.adjust function in R (v4.0.2) with the parameter *Method = "BH"* across P-values from all 1,540 tissue-trait pairs and calculated FDR values[113]. Significant enrichment results (FDR < 0.05) were used for statistical plot.

***cis*-molQTL-GWAS colocalization**

To identify shared genetic variants between GWAS loci and molQTL, we conducted a colocalization analysis between each molGene and overlapped GWAS loci using the *coloc.abf* function in the *coloc* package (v5.2.3)[106], which is an approximate Bayes factor colocalization analysis. The package computed posterior probabilities for (1) no association with either GWAS loci and molQTL (*H*0); (2) association only with GWAS loci (*H*1); (3) association only with molQTL (*H*2); (4) association with both GWAS loci and molQTL but two independent signals (*H*3); and (5) association with both molecular phenotype and shared signals (*H*4). The colocalized signals between GWAS loci and molQTL were defined by PP.H4 > 0.8[21,22]. Moreover, we calculated the LD of two lead SNPs for a pair of GWAS variants and molQTL using PLINK (v.1.9)[107].

Function enrichment analysis of gene lists

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using the clusterProfiler (v.4.0)[114]. The GO terms and KEGG pathways of selected genes were enriched in the org.Bt.eg.db and KEGG-bta databases, using the erichGO and enrichKEGG functions, respectively, with a threshold parameter of "pvalueCutoff = 0.05". We considered a GO term or a KEGG pathway significant if it had FDR < 0.05.

Population-stratified molQTL analysis

To investigate the difference in *cis*-eQTL between breeds, global ancestry of all samples within each tissue was estimated by ADMIXTURE (v1.3.0) with K=2[115]. To ensure clear genetic assignments of samples to four breed groups (i.e., *B. taurus, B. indicus,* dairy, and beef cattle), we matched the population labels with ancestral lineage composition, and considered the label correct if the lineage purity was greater than 80%. Subsequently, other samples with ancestral purity greater than 80% (even if the original label was not) were defined as belonging to this breed. We then conducted *cis*-eQTL mapping separately in *B. taurus*, *B. indicus*, dairy and beef cattle populations using the same procedures as described in **molQTL mapping** above**.** Replications of *cis*-eQTL between populations within a tissue and tissues within a population were quantified using Storey's π_1 statistic and Pearson's correlation of normalized eQTL effect sizes and compared to replication between tissues within a population and populations within a tissue[22]. To further characterize sharing patterns across populations, we conducted a meta-analysis of *cis*-eQTL across tissues and population groups using the MashR package (v0.2.57)[108], following the same pipeline used for **Tissue-sharing/specificity analysis of molQTL**.

Selective sweep and molQTL enrichment analysis

To detect selection signatures between *B. taurus* and *B. indicus,* we selected 200 *B. taurus* (50 Holstein, 50 Jersey, 50 Simental and 50 Angus) and 142 *B. indicus* (20 Leiqiong, 20 Butana, 11 Horro, 11 Boran, 10 Arsi, 10 Goffa, 10 Mursi, 9 Barka, 9 Fogera, 9 Afar, 9 Sheko, 7 Baoule and 7 Nelore) with WGS data (>10x) from the genotype imputation panel[31] (**Supplementary Table 3**) with 28,958,211 SNPs. To detect selection signatures between beef and dairy cattle, we selected 651 dairy (Holstein) and 552 beef (326 Simmental, 64 Angus, 60 Charolais, 58 Limousin and 44 Hereford) cattle from the genotype imputation panel[31] (**Supplementary Table 4**) with 28,958,211 SNPs. We analyzed genetic differentiation (F_{ST}) using the vcftools (v0.1.16) tool[116], employing a sliding window of 30 Kb with a step size of 10 Kb. The windows were divided into deciles based on their F_{ST} values, forming ten groups ($F_{ST}1$–$F_{ST}10$) with increasing selection intensity. Enrichment analysis was then performed between each group and the molQTL from credible sets derived through fine mapping. Odds ratio (OR) was calculated to evaluate enrichment levels between selected regions and molQTL with the following formula[21]:

$$OR = \frac{a/b}{c/d}$$

In which *a* represents the numbers of molQTL located in genomic regions of each F_{ST} groups, *b* represents the number of molQTL, *c* represents the number of SNPs located in genomic regions of each F_{ST} group, *d* represents the number of all SNPs. The frequency of alleles in each SNP was calculated using the option of "--freq" in Plink (v1.9)[107].

Cross-species conservation analysis between cattle and humans

To evaluate cross-species conservation between cattle and humans, we mapped fine-mapped cattle eQTL to the human reference genome (hg38) using the UCSC liftOver tool[117]. SNPs that were successfully converted were considered as orthologous (sequence-level conserved) variants between cattle and humans. To assess conservation at the functional level, we tested whether enrichment of eQTL within regulatory elements was preserved between species. Specifically, we performed: (i) enrichment analysis of cattle eQTL in cattle regulatory elements[110]; (ii) enrichment analysis of liftOvered cattle eQTL in human regulatory elements[55]; and (iii) enrichment analysis of human eQTL from GTEx v8 in human regulatory elements. We then quantified concordance in enrichment patterns across the three analyses using Pearson's correlation to assess cross-species functional conservation.

In addition, we used the *AlphaGenome* model with *in silico* mutagenesis to predict the regulatory effects of orthologous variants of cattle eQTL in the human genome[53]. Only the tissue-level gene expression track in the *AlphaGenome* was considered in this study. Predicted effects were then compared across different eQTL groups and MAF-matched random SNPs.

We next collected GWAS summary statistics for 133 human complex traits (**Supplementary Table 7**) from the Pan-UK Biobank[118] and applied LD score regression (*LDSC* v1.0.1)[56] to estimate the proportion of SNP-heritability attributable to each predefined eQTL group. We conducted colocalization analysis between human GTEx v8 eQTL and GWAS of human complex traits using coloc v5.2.3[91], and considered a posterior probability threshold of PP.H4 > 0.8 as high-confidence shared signals. To ensure biologically interpretable cross-species comparisons, we retained only tissue-gene-SNP trios that satisfied all of the following criteria: (i) genes have identifiable human–cattle orthologs; (ii) the tissue is matched between humans and cattle; and (iii) the orthologous locus is segregating in both the human GTEx and cattle GTEx datasets. As a control, we randomly selected an equal number of trios. We then tested whether SNPs from the retained trios were also associated with gene expression in CattleGTEx using a single-SNP linear mixed model, in which the genomic relationship matrix (GRM) was included as a random effect to account for population structure and relatedness; this analysis was performed using GCTA (v1.93.2)[112]. Fixed-effect covariates were kept identical to those used in the cattle eQTL mapping. SNPs with FDR < 0.05 were considered significant.

Statistics and reproducibility

No statistical method was used to predetermine the sample size. The details of data exclusions for each specific analysis are available in the Methods section. For all the boxplots, the horizontal lines inside the boxes show the medians. Box bounds show the lower quartile (Q1, the 25th percentile) and the upper quartile (Q3, the 75th percentile). Whiskers are minima (Q1 − 1.5× IQR) and maxima (Q3 + 1.5× IQR), where IQR is the interquartile range (Q3–Q1). Outliers are shown in the boxplots unless otherwise stated. The experiments were not randomized, as all the datasets are publicly available from observational studies. The investigators were not blinded to allocation during experiments and outcome assessment, as the data were not from controlled randomized studies.

Data availability

All raw data analyzed in this study are publicly available for download without restrictions from SRA (https://www.ncbi.nlm.nih.gov/sra/) and GSA (https://ngdc.cncb.ac.cn/gsa). The ARS-UCD1.2

(v.110) cattle reference genome is available at Ensembl (https://www.ensembl.org). Details of RNA-Seq samples, WGS samples, molQTL summary results, selection analysis results, GWAS information of human diseases, colocalization, TWAS and SMR results can be found in Supplementary Tables, respectively. All processed data and the full summary statistics of molQTL mapping and genotype imputation reference panel are available at https://cattlegtex.kiz.ac.cn/.

Code availability

All the computational scripts and codes for RNA-Seq and WGS analyses, as well as the ene_expression_analyses, molQTL mapping, fine mapping, conditional analysis, tissue sharing analysis, functional enrichment, colocalization, TWAS, SMR, selection signal analysis and LDSC are available at the FarmGTEx GitHub website (https://github.com/FarmGTEx/CattleGTEx_Phase1_Pipeline_v0).

Acknowledgments

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, seed-funding from CellFood Hub (AUFF), Danmarks Frie Forskningsfond nos. 5254-00069B. B. L. acknowledges funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) with grant BB/X009505/1.

1. Frantz, L.A.F., Bradley, D.G., Larson, G. & Orlando, L. Animal domestication in the era of ancient genomics. *Nature Reviews Genetics* **21**, 449-460 (2020).
2. Utsunomiya, Y.T. *et al.* Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics* **50**, 557-568 (2019).
3. Xia, X. *et al.* Global dispersal and adaptive evolution of domestic cattle: a genomic perspective. *Stress Biology* **3**, 8 (2023).
4. Van Eenennaam, A.L. Current and future uses of genetic improvement technologies in livestock breeding programs. *Animal Frontiers* **15**, 80-90 (2025).
5. Yu, X. *et al.* Role of multi-omics in aquaculture genetics and breeding: current status and future perspective. *Science China Life Sciences* **68**, 2591-2604 (2025).
6. Georges, M., Charlier, C. & Hayes, B. Harnessing genomic information for livestock improvement. *Nature Reviews Genetics* **20**, 135-156 (2019).
7. Elsik Cg Fau - Tellam, R.L. *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution.
8. Van de Walle, G.R. & Harman, R.M. Contributions of large and agricultural animal models to immunology. *The Journal of Immunology* **214**, 2494-2503 (2025).
9. Hamernik, D.L. Farm animals are important biomedical models.
10. Abedal-Majed, M.A. & Cupp, A.S. Livestock animals to study infertility in women. *Animal Frontiers* **9**, 28-33 (2019).
11. Reynolds, E.G.M. *et al.* Non-additive association analysis using proxy phenotypes identifies novel cattle syndromes. *Nature Genetics* **53**, 949-954 (2021).
12. Halstead, M.M., Ma, X., Zhou, C., Schultz, R.M. & Ross, P.J. Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *Nature Communications* **11**, 4654 (2020).
13. Making sense of the regulatory genome. *Nature Reviews Genetics* **26**, 651-652 (2025).
14. The GTEx Consortium atlas of genetic regulatory effects across human tissues.
15. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
16. Fang, L. *et al.* The Farm Animal Genotype–Tissue Expression (FarmGTEx) Project. *Nature Genetics* **57**, 786-796 (2025).
17. Clark, E.A.-O. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production.
18. Liu, S. *et al.* A multi-tissue atlas of regulatory variants in cattle. *Nature Genetics* **54**, 1438-1447 (2022).
19. Teng, J. *et al.* A compendium of genetic regulatory effects across pig tissues. *Nature Genetics* **56**, 112-123 (2024).
20. Guan, D. *et al.* Genetic regulation of gene expression across multiple tissues in chickens. *Nature Genetics* **57**, 1298-1308 (2025).
21. Li, H. *et al.* An atlas of cell type specific regulatory effects in cattle. *bioRxiv*, 2025.06.23.661035 (2025).
22. Gong, M. *et al.* A multi-tissue atlas of genetic regulatory effects in sheep. *bioRxiv*, 2025.11.03.686346 (2025).
23. Pan, Z. *et al.* An atlas of regulatory elements in chicken: A resource for chicken genetics and genomics. *Science Advances* **9**, eade1204.
24. Pan, Z. *et al.* Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nature Communications* **12**, 5848 (2021).
25. Fang, L. *et al.* Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. *BMC Biology* **17**, 68 (2019).
26. Zhang, D. *et al.* Comprehensive multi-tissue epigenome atlas in sheep: A resource for complex traits, domestication, and breeding. *iMeta* **3**, e254 (2024).
27. Connally, N.A.-O. *et al.* The missing link between genetic association and regulatory function. LID - 10.7554/eLife.74970 [doi] LID - e74970.
28. Mostafavi, H., Spence, J.P., Naqvi, S. & Pritchard, J.K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics* **55**, 1866-1875 (2023).
29. Huang, E. *et al.* Genetic variants affecting RNA stability influence complex traits and disease risk. *Nature Geneti* ***Regulatory genomic circuitry of human disease loci by integrative epigenomics***

cs **57**, 2578-2588 (2025).
30. Lee, S. & Mostafavi, H. Missing regulatory effects on complex traits: Contribution of distal variants. *Cell Genomics* **5**, 100809 (2025).
31. Yao, D.W., O'Connor, L.J., Price, A.L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics* **52**, 626-633 (2020).
32. Brotman, S.M. *et al.* Adipose tissue eQTL meta-analysis highlights the contribution of allelic heterogeneity to gene expression regulation and cardiometabolic traits. *Nature Genetics* **57**, 180-192 (2025).
33. Orchard, P.A.-O. *et al.* Cross-cohort analysis of expression and splicing quantitative trait loci in TOPMed. LID - 2025.02.19.25322561 [pii] LID - 10.1101/2025.02.19.25322561 [doi].
34. Zhao, P. The pangenome graphs and SNP-SV joint reference panel in cattle. (2024).
35. Murty, M.N. & Raghava, R. Linear Support Vector Machines. in *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks* (eds. Murty, M.N. & Raghava, R.) 41-56 (Springer International Publishing, Cham, 2016).
36. Li, Y., Ge, X., Peng, F., Li, W. & Li, J.J. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology* **23**, 79 (2022).
37. Teng, J. *et al.* OmiGA for ultra-efficient molecular quantitative trait loci mapping. *Nature Communications* (2026).
38. Chen, L. *et al.* Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* **364**, eaav6202 (2019).
39. Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M. & Cunningham, P. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences* **91**, 2757-2761 (1994).
40. Ward, J.A. & MacHugh, D.E. Cattle Genomics: Aurochs Admixture in East Asia. *Animal Research and One Health* **3**, 341-343 (2025).
41. Chen, N. *et al.* Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications* **9**, 2337 (2018).
42. Huang, N. *et al.* Signatures of selection in indigenous Chinese cattle genomes reveal adaptive genes and genetic variations to cold climate. LID - 10.1093/jas/skad006 [doi] LID - skad006.
43. Ayalew, W. *et al.* Signatures of positive selection for local adaptation of African native cattle populations: A review. *Journal of Integrative Agriculture* **22**, 1967-1984 (2023).
44. Xu, L. *et al.* Genome-wide scanning reveals genetic diversity and signatures of selection in Chinese indigenous cattle breeds. *Livestock Science* **216**, 100-108 (2018).
45. Iqbal, M.W. *et al.* Exploring deleterious non-synonymous SNPs in FUT2 gene, and implications for norovirus susceptibility and gut microbiota composition. *Scientific Reports* **15**, 10395 (2025).
46. Yadav, V.K. *et al.* A Serotonin-Dependent Mechanism Explains the Leptin Regulation of Bone Mass, Appetite, and Energy Expenditure. *Cell* **138**, 976-989 (2009).
47. Li, Y. *et al.* Lowering expression of Epsin-3 inhibits migration and invasion of lung adenocarcinoma cells by inhibiting the epithelial–mesenchymal transition. *Scientific Reports* **14**, 17069 (2024).
48. Zhao, Y. *et al.* GIGYF1 loss of function is associated with clonal mosaicism and adverse metabolic health. *Nature Communications* **12**, 4178 (2021).
49. Cavalcanti, G.A.-O. *et al.* Granulosa cells and follicular development: a brief review.
50. Grisart, B. *et al.* Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition.
51. Schäringer, K. *et al.* LSP1-myosin1e bimolecular complex regulates focal adhesion dynamics and cell migration.
52. Xu, P. *et al.* Upregulation of LSP1 in AML cells enhances AML resistance to sorafenib and promotes apoptosis in CD8+ T cells. *International Immunopharmacology* **150**, 114255 (2025).
53. Avsec, Ž. *et al.* Advancing regulatory variant effect prediction with AlphaGenome. *Nature* **649**, 1206-1218 (2026).
54. The, G.C. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
55. Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307 (2021).
56. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
57. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics* **50**, 956-967 (2018).

58. Zaręba-Kozioł, M., Figiel, I., Bartkowiak-Kaczmarek, A. & Włodarczyk, J. Insights Into Protein S-Palmitoylation in Synaptic Plasticity and Neurological Disorders: Potential and Limitations of Methods for Detection and Analysis.
59. Mukai, K. *et al.* Activation of STING requires palmitoylation at the Golgi. *Nature Communications* **7**, 11932 (2016).
60. Lu, Y.A.-O. *et al.* Palmitoylation of NOD1 and NOD2 is required for bacterial sensing.
61. Han, B. *et al.* A multi-tissue single-cell expression atlas in cattle. *Nature Genetics* (2025).
62. Housman, G. *et al.* Addressing missing context in regulatory variation across primate evolution. *Current Opinion in Genetics & Development* **93**, 102374 (2025).
63. Zhou, L. *et al.* Palmitoylation of GPX4 via the targetable ZDHHC8 determines ferroptosis sensitivity and antitumor immunity. *Nature Cancer* **6**, 768-785 (2025).
64. Luo, L.-Y. *et al.* Telomere-to-telomere sheep genome assembly identifies variants associated with wool fineness. *Nature Genetics* **57**, 218-230 (2025).
65. 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 Biology* **24**, 139 (2023).
66. Salavati, M.A.-O. *et al.* Improving the annotation of the cattle genome by annotating transcription start sites in a diverse set of tissues and populations using Cap Analysis Gene Expression sequencing. LID - 10.1093/g3journal/jkad108 [doi] LID - jkad108.
67. Liu, S. *et al.* Genome assembly and structural variations of Guyuan cattle. *Scientific Data* **12**, 1863 (2025).
68. Pardo-Palacios, F.J. *et al.* Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nature Methods* **21**, 1349-1363 (2024).
69. Yang, C. *et al.* European and African ancestry-specific plasma protein-QTL and metabolite-QTL analyses identify ancestry-specific T2D effector proteins and metabolites. *Nature Communications* **16**, 7412 (2025).
70. Yin, B. *et al.* Dynamic metabolite QTL analyses provide novel biochemical insights into kernel development and nutritional quality improvement in common wheat. *Plant Communications* **5**, 100792 (2024).
71. Kumasaka, N., Knights, A.J. & Gaffney, D.J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature Genetics* **48**, 206-213 (2016).
72. Cote, A.C., Young, H.E. & Huckins, L.M. Critical reasoning on the co-expression module QTL in the dorsolateral prefrontal cortex. *Human Genetics and Genomics Advances* **5**, 100311 (2024).
73. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120-128 (2022).
74. Sheng, X.A.-O. *et al.* Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments.
75. Kim-Hellmuth, S. *et al.* Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
76. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science* **376**, eabf3041.
77. Song, L., Chen, W., Hou, J., Guo, M. & Yang, J. Spatially resolved mapping of cells associated with human complex traits. *Nature* **641**, 932-941 (2025).
78. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
79. Coorens, T.H.H. *et al.* The human and non-human primate developmental GTEx projects. *Nature* **637**, 557-564 (2025).
80. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
81. Valencia, C. *et al.* Modeling heterogeneity in single-cell perturbation states enhances detection of response eQTLs. *Nature Genetics* **57**, 2882-2890 (2025).
82. Freimer, J.W. *et al.* Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nature Genetics* **54**, 1133-1144 (2022).
83. Weber, J., Braun, C.J., Saur, D. & Rad, R. In vivo functional screening for systems-level integrative cancer genomics. *Nature Reviews Cancer* **20**, 573-593 (2020).
84. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
85. Rosen, B.D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* **9**, giaa021 (2020).
86. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

87. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
88. 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).
89. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2021).
90. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J.K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods* **12**, 1061-1063 (2015).
91. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290-295 (2015).
92. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
93. Chen, H. *et al.* A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* **173**, 386-399.e12 (2018).
94. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417-419 (2017).
95. Li, Y.I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nature Genetics* **50**, 151-158 (2018).
96. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nature Communications* **5**, 5274 (2014).
97. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11.12.1-34 (2014).
98. Li, L. *et al.* An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nature Genetics* **53**, 994-1005 (2021).
99. Feng, X., Li, L., Wagner, E.J. & Li, W. TC3A: The Cancer 3' UTR Atlas. *Nucleic Acids Res* **46**, D1027-d1030 (2018).
100. Ji, G. *et al.* stAPAminer: Mining Spatial Patterns of Alternative Polyadenylation for Spatially Resolved Transcriptomic Studies. *Genomics, Proteomics & Bioinformatics* **21**, 601-618 (2023).
101. Alkallas, R., Fish, L., Goodarzi, H. & Najafabadi, H.S. Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. *Nature Communications* **8**, 909 (2017).
102. Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943-2945 (2022).
103. Castel, S.E., Mohammadi, P., Chung, W.K., Shen, Y. & Lappalainen, T. Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nature Communications* **7**, 12817 (2016).
104. Zhou, H.J., Ge, X. & Li, J.J. ClipperQTL: ultrafast and powerful eGene identification method. *Genome Biology* **26**, 207 (2025).
105. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**, 1273-1300 (2020).
106. Wallace, C. *et al.* Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum Mol Genet* **21**, 2815-24 (2012).
107. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses.
108. Urbut, S.M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187-195 (2019).
109. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
110. Guan, D. *et al.* An integrated multi-tissue atlas of epigenomic landscapes and regulatory elements in the bovine genome. *bioRxiv*, 2025.08.21.671512 (2025).
111. Jiang, J. *et al.* Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Communications Biology* **2**, 212 (2019).
112. 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).
113. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300 (1995).
114. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

115. Lawson, D.J., van Dorp, L. & Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* **9**, 3258 (2018).
116. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
117. Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53**, D1243-d1249 (2025).
118. Karczewski, K.J. *et al.* Pan-UK Biobank genome-wide association analyses enhance discovery and resolution of ancestry-enriched effects. *Nature Genetics* **57**, 2408-2417 (2025).
119. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies.

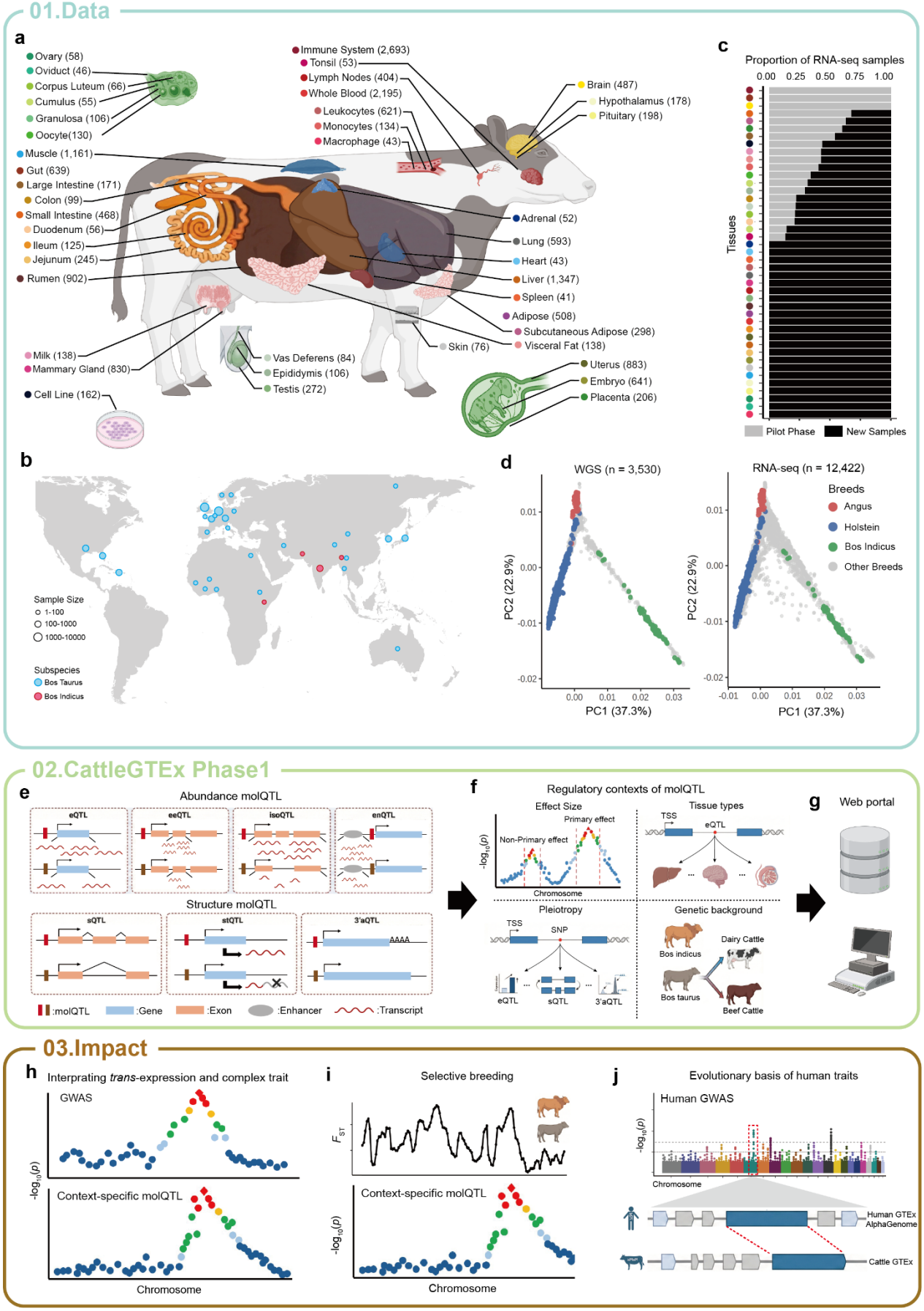

01.Data
a
Ovary (58)
Oviduct (46)
Corpus Luteum (66)
Cumulus (55)
Granulosa (106)
Oocyte(130)
Muscle (1,161)
Gut (639)
Large Intestine (171)
Colon (99)
Small Intestine (468)
Duodenum (56)
Ileum (125)
Jejunum (245)
Rumen (902)
Milk (138)
Mammary Gland (830)
Cell Line (162)
Immune System (2,693)
Tonsil (53)
Lymph Nodes (404)
Whole Blood (2,195)
Leukocytes (621)
Monocytes (134)
Macrophage (43)
Brain (487)
Hypothalamus (178)
Pituitary (198)
Adrenal (52)
Lung (593)
Heart (43)
Liver (1,347)
Spleen (41)
Adipose (508)
Subcutaneous Adipose (298)
Visceral Fat (138)
Skin (76)
Vas Deferens (84)
Epididymis (106)
Testis (272)
Uterus (883)
Embryo (641)
Placenta (206)
c
Proportion of RNA-seq samples
0.00 0.25 0.50 0.75 1.00
Tissues
Pilot Phase
New Samples
b
Sample Size
1-100
100-1000
1000-10000
Subspecies
Bos Taurus
Bos Indicus
d
WGS (n = 3,530)
RNA-seq (n = 12,422)
Breeds
Angus
Holstein
Bos Indicus
Other Breeds
PC2 (22.9%)
PC1 (37.3%)
02.CattleGTEx Phase1
e
Abundance molQTL
eQTL
eeQTL
isoQTL
enQTL
Structure molQTL
sQTL
stQTL
3'aQTL
AAAA
:molQTL
:Gene
:Exon
:Enhancer
:Transcript
f
Regulatory contexts of molQTL
Effect Size
Primary effect
Non-Primary effect
Chromosome
Tissue types
TSS
eQTL
Pleiotropy
SNP
sQTL
3'aQTL
Genetic background
Bos indicus
Dairy Cattle
Bos taurus
Beef Cattle
g
Web portal
03.Impact
h
Interpreting trans-expression and complex trait
GWAS
Context-specific molQTL
Chromosome
i
Selective breeding
Context-specific molQTL
Chromosome
j
Evolutionary basis of human traits
Human GWAS
Chromosome
Human GTEx
AlphaGenome
Cattle GTEx

Fig. 1 | Overview of the CattleGTEx Phase1 project. a, Summary of 43 tissues profiled for molecular quantitative trait loci (molQTL) mapping. Sample sizes (n ≥ 40) are indicated in the respective parentheses, representing a total of 8,956 animals. The tissue colour code is used consistently throughout the manuscript. **b,** Geographic distribution of the 12,422 cattle RNA-seq samples analysed in this study. **c**, Comparison of sample sizes between the pilot study and Phase 1. Grey bars represent sample sizes in the pilot CattleGTEx project, while black bars represent newly collected samples in Phase 1. **d**, Principal component analysis (PCA) of 3,530 whole-genome sequencing (WGS) data and 12,422 RNA-seq samples. **e**, Overview of molQTL mapping and resource generation. We mapped seven types of molQTL and performed four types of regulatory context analyses. eQTL, expression QTL; eeQTL, exon expression QTL; isoQTL, isoform expression QTL; enQTL, enhancer expression QTL; sQTL, splicing QTL; stQTL, QTL for RNA stability of genes; 3'aQTL, QTL for 3' UTR alternative polyadenylation (APA). All data are accessible at https://cattlegtex.farmgtex.org/. **f**, Applications of the CattleGTEx Phase1 resource. Analyses include deciphering *trans*- gene expression, interpreting genome-wide association studies (GWAS) results of 44 bovine complex traits, exploring population divergence (between *B. taurus* and *B. indicus*, as well as between dairy and beef cattle), and investigating evolutionary constraints on orthologous genes and regulatory variants between cattle and humans.

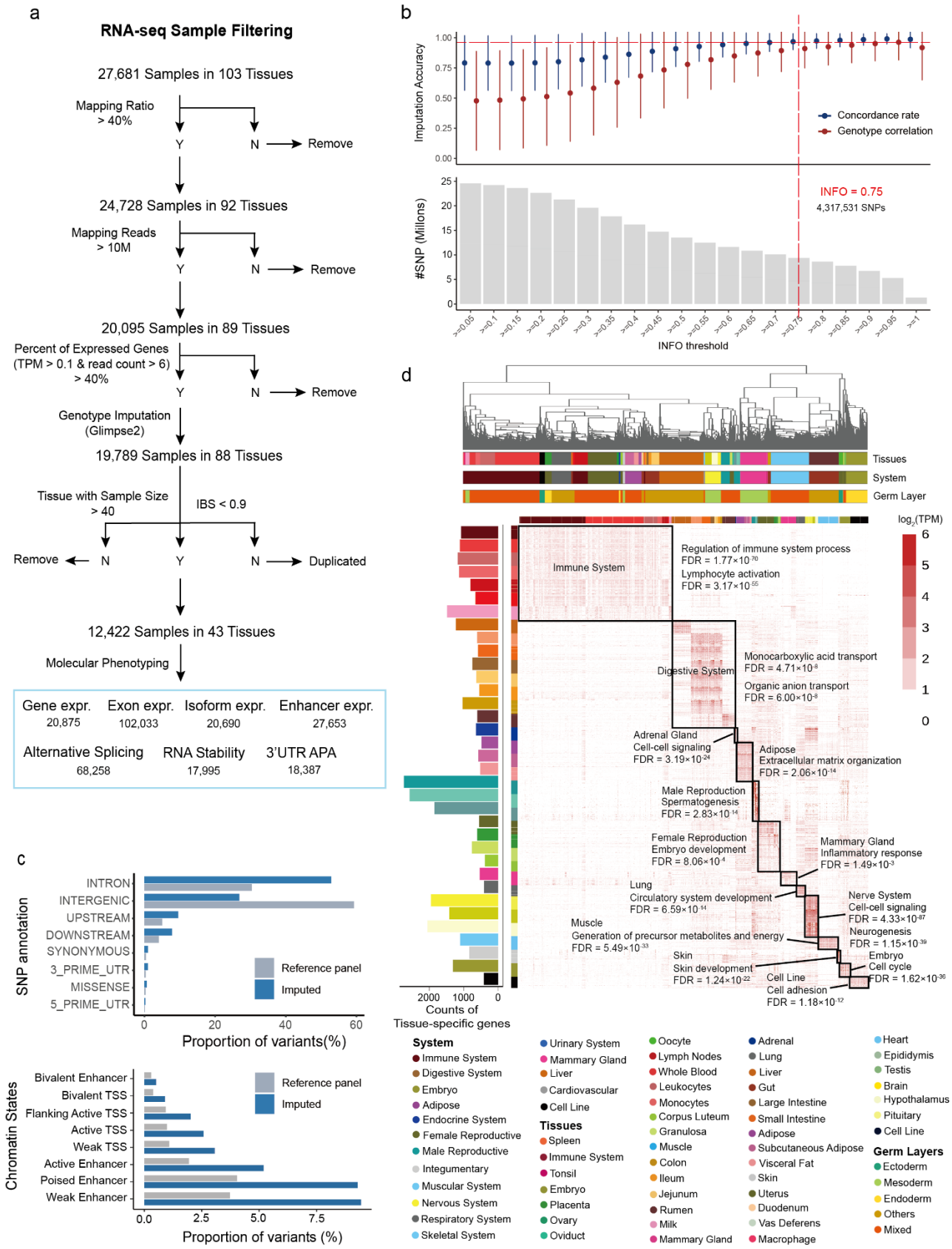

a
RNA-seq Sample Filtering
27,681 Samples in 103 Tissues
Mapping Ratio > 40%
Y
N
Remove
24,728 Samples in 92 Tissues
Mapping Reads > 10M
Y
N
Remove
20,095 Samples in 89 Tissues
Percent of Expressed Genes (TPM > 0.1 & read count > 6) > 40%
Y
N
Remove
Genotype Imputation (Glimpse2)
19,789 Samples in 88 Tissues
Tissue with Sample Size > 40
IBS < 0.9
Remove
N
Y
N
Duplicated
12,422 Samples in 43 Tissues
Molecular Phenotyping
Gene expr. 20,875
Exon expr. 102,033
Isoform expr. 20,690
Enhancer expr. 27,653
Alternative Splicing 68,258
RNA Stability 17,995
3'UTR APA 18,387
b
Imputation Accuracy
Concordance rate
Genotype correlation
#SNP (Millons)
INFO = 0.75
4,317,531 SNPs
INFO threshold
c
SNP annotation
INTRON
INTERGENIC
UPSTREAM
DOWNSTREAM
SYNONYMOUS
3_PRIME_UTR
MISSENSE
5_PRIME_UTR
Reference panel
Imputed
Proportion of variants(%)
Chromatin States
Bivalent Enhancer
Bivalent TSS
Flanking Active TSS
Active TSS
Weak TSS
Active Enhancer
Poised Enhancer
Weak Enhancer
Reference panel
Imputed
Proportion of variants (%)
d
Tissues
System
Germ Layer
log2(TPM)
Immune System
Regulation of immune system process
FDR = 1.77×10-70
Lymphocyte activation
FDR = 3.17×10-55
Digestive System
Monocarboxylic acid transport
FDR = 4.71×10-8
Organic anion transport
FDR = 6.00×10-8
Adrenal Gland
Cell-cell signaling
FDR = 3.19×10-24
Adipose
Extracellular matrix organization
FDR = 2.06×10-14
Male Reproduction
Spermatogenesis
FDR = 2.83×10-14
Female Reproduction
Embryo development
FDR = 8.06×10-4
Mammary Gland
Inflammatory response
FDR = 1.49×10-3
Lung
Circulatory system development
FDR = 6.59×10-14
Nerve System
Cell-cell signaling
FDR = 4.33×10-87
Neurogenesis
FDR = 1.15×10-39
Muscle
Generation of precursor metabolites and energy
FDR = 5.49×10-33
Skin
Skin development
FDR = 1.24×10-22
Embryo
Cell cycle
FDR = 1.62×10-36
Cell Line
Cell adhesion
FDR = 1.18×10-12
Counts of Tissue-specific genes
System
Immune System
Digestive System
Embryo
Adipose
Endocrine System
Female Reproductive
Male Reproductive
Integumentary
Muscular System
Nervous System
Respiratory System
Skeletal System
Urinary System
Mammary Gland
Liver
Cardiovascular
Cell Line
Tissues
Spleen
Immune System
Tonsil
Embryo
Placenta
Ovary
Oviduct
Oocyte
Lymph Nodes
Whole Blood
Leukocytes
Monocytes
Corpus Luteum
Granulosa
Muscle
Colon
Ileum
Jejunum
Rumen
Milk
Mammary Gland
Adrenal
Lung
Liver
Gut
Large Intestine
Small Intestine
Adipose
Subcutaneous Adipose
Visceral Fat
Skin
Uterus
Duodenum
Vas Deferens
Macrophage
Heart
Epididymis
Testis
Brain
Hypothalamus
Pituitary
Cell Line
Germ Layers
Ectoderm
Mesoderm
Endoderm
Others
Mixed

Fig. 2 | Genotype imputation and molecular phenotyping. a, Workflow for sample quality control (QC), duplicate removal, and molecular phenotyping. Y: Sample passed the filtering, N: Sample didn't pass the filtering, TPM: Transcript per million, IBS: Identity-by-state, expr.: expression, APA: Alternative polyadenylation. **b**, Imputation accuracy assessed by concordance rate and genotype correlation using an external dataset, which include 15,952 samples with paired RNA-seq and whole-genome sequencing data (WGS, 30×). The top panel shows the distribution of concordance rates and genotype correlations across various INFO score cutoffs (reported by Glimpse2); the bottom panel displays the number of SNPs retained at each cutoff. **c**, Distribution of 4,317,531 imputed SNPs (from RNA-seq) and 28,958,211 reference panel SNPs across eight genomic features and eight chromatin states. TSS: Transcription start site. **d**, **Top:** Hierarchical clustering of 43 tissues based on tissue-specific gene expression, annotated by tissue type, organ system, and germ layer origin. **Bottom:** Number of tissue-specific genes (left), corresponding expression heatmap across the 43 tissues (centre), and enriched Gene Ontology (GO) terms (right).

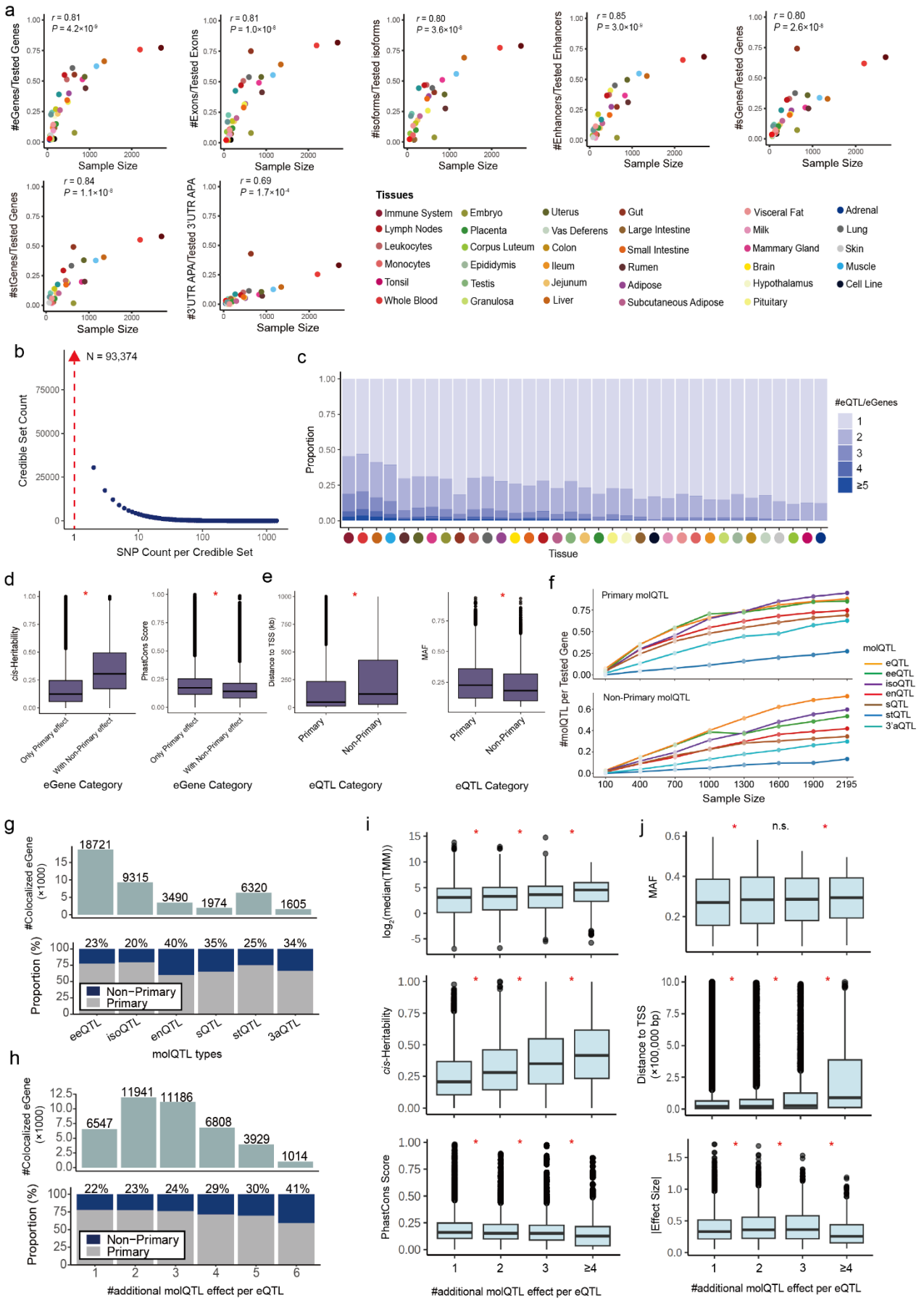

a
#eGenes/Tested Genes
r = 0.81
P = 4.2×10⁻⁹
#Exons/Tested Exons
r = 0.81
P = 1.0×10⁻⁸
#isoforms/Tested isoforms
r = 0.80
P = 3.6×10⁻⁸
#Enhancers/Tested Enhancers
r = 0.85
P = 3.0×10⁻⁹
#sGenes/Tested Genes
r = 0.80
P = 2.6×10⁻⁸
#stGenes/Tested Genes
r = 0.84
P = 1.1×10⁻⁸
#3'UTR APA/Tested 3'UTR APA
r = 0.69
P = 1.7×10⁻⁴
Sample Size
Tissues
Immune System
Lymph Nodes
Leukocytes
Monocytes
Tonsil
Whole Blood
Embryo
Placenta
Corpus Luteum
Epididymis
Testis
Granulosa
Uterus
Vas Deferens
Colon
Ileum
Jejunum
Liver
Gut
Large Intestine
Small Intestine
Rumen
Adipose
Subcutaneous Adipose
Visceral Fat
Milk
Mammary Gland
Brain
Hypothalamus
Pituitary
Adrenal
Lung
Skin
Muscle
Cell Line
b
N = 93,374
Credible Set Count
SNP Count per Credible Set
c
Proportion
Tissue
#eQTL/eGenes
1
2
3
4
≥5
d
cis-Heritability
PhastCons Score
Only Primary effect
With Non-Primary effect
eGene Category
e
Distance to TSS (kb)
MAF
Primary
Non-Primary
eQTL Category
f
Primary molQTL
Non-Primary molQTL
#molQTL per Tested Gene
Sample Size
molQTL
eQTL
eeQTL
isoQTL
enQTL
sQTL
stQTL
3'aQTL
g
#Colocalized eGene (×1000)
18721
9315
3490
1974
6320
1605
Proportion (%)
23%
20%
40%
35%
25%
34%
Non-Primary
Primary
eeQTL
isoQTL
enQTL
sQTL
stQTL
3aQTL
molQTL types
h
#Colocalized eGene (×1000)
6547
11941
11186
6808
3929
1014
Proportion (%)
22%
23%
24%
29%
30%
41%
Non-Primary
Primary
#additional molQTL effect per eQTL
i
log2(median(TMM))
cis-Heritability
PhastCons Score
#additional molQTL effect per eQTL
j
n.s.
MAF
Distance to TSS (×100,000 bp)
|Effect Size|
#additional molQTL effect per eQTL

Fig. 3 | Regulatory architecture of seven molecular phenotypes across 35 tissues.

a, Pearson's correlation (*r*) between the proportion of molGenes (i.e., genes with significant molecular quantitative trait loci (molQTL) / all tested genes) and sample size across 35 tissues. *P* values were calculated using a two-sided Student's *t*-test. eQTL, expression QTL; eeQTL, exon expression QTL; isoQTL, isoform expression QTL; enQTL, enhancer expression QTL; sQTL, splicing QTL; stQTL, QTL for RNA stability of genes; 3'aQTL, QTL for 3' UTR alternative polyadenylation (APA). **b,** The size distribution of fine-mapped credible sets for all identified *cis*-eQTL across 35 tissues. The red arrow indicates 93,374 credible sets containing only one single candidate variant. **c,** The distribution of the number of conditionally independent *cis*-eQTL per gene is shown as blue stacked bars across 28 tissues. Tissues are ordered by decreasing sample size. **d,** Characteristics of genes with and without non-primary *cis*-eQTL. (Left) cis-heritability of mRNA and lncRNA expression; and (Right) sequence constraint (PhastCons score). Statistical significance was determined using a two-sided Student's t-test. *: $P < 0.05$. **e,** Characteristics of primary and non-primary *cis*-eQTL. (Left) distance to the target gene; and (Right) Minor Allele Frequency (MAF). Statistical significance was determined using a two-sided Student's t-test. *: $P < 0.05$. **f,** Numbers of detected primary eQTL and non-primary eQTL per tested gene across different sample size in whole blood. **g,** eQTL colocalized with other molQTL (top) and the proportion of primary versus non-primary eQTL among them (bottom). Exact numbers and proportions are shown above each bar. **h,** Pleiotropic eQTL regulating varying numbers of molecular phenotypes (top) and the proportion of primary versus non-primary eQTL among them (bottom). Exact numbers and proportions are shown above each bar. **i,** Characterization of genes harboring pleiotropic cis-eQTL. The box plots show the relationship between the number of additional molQTL regulated per lead cis-eQTL (X-axis, excluding the eQTL itself) and three properties: (Top) mRNA and lncRNA expression level, measured as $\log_2$(median TMM) (bp); (Middle) cis-heritability of mRNA and lncRNA expression; and (Down) sequence constraint, quantified by PhastCons scores. In each box plot, the median is shown as the central line, the 25th – 75th percentiles define the box limits, and whiskers extend 1.5×the interquartile range. Statistical significance was determined using a two-sided Student's t-test. *: $P < 0.05$; n.s.: not significant. **j,** Characterization of pleiotropic cis-eQTL. The box plots show the relationship between the number of additional molQTL regulated per lead cis-eQTL (X-axis, excluding the eQTL itself) and three properties: (Top) Minor Allele Frequency (MAF); (Middle) distance to the target gene (bp); and (Down) absolute effect size on mRNA and lncRNA expression. In each box plot, the median is shown as the central line, the

25th – 75th percentiles define the box limits, and whiskers extend 1.5×the interquartile range. Statistical significance was determined using a two-sided Student's t-test. *: $P < 0.05$; n.s.: not significant.

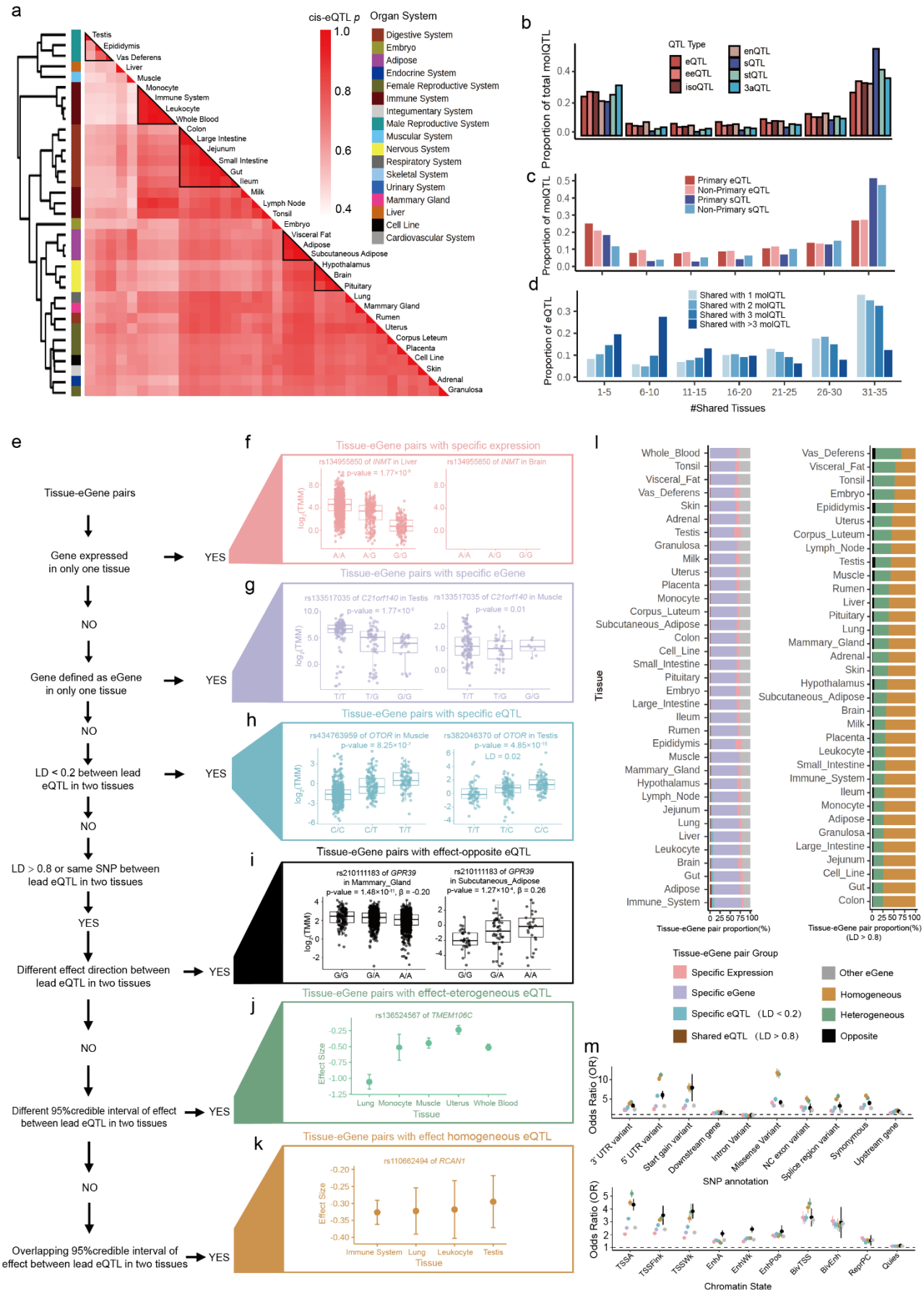

a
cis-eQTL p
1.0
0.8
0.6
0.4
Organ System
Digestive System
Embryo
Adipose
Endocrine System
Female Reproductive System
Immune System
Integumentary System
Male Reproductive System
Muscular System
Nervous System
Respiratory System
Skeletal System
Urinary System
Mammary Gland
Liver
Cell Line
Cardiovascular System
Testis
Epididymis
Vas Deferens
Liver
Muscle
Monocyte
Immune System
Leukocyte
Whole Blood
Colon
Large Intestine
Jejunum
Small Intestine
Gut
Ileum
Milk
Lymph Node
Tonsil
Embryo
Visceral Fat
Adipose
Subcutaneous Adipose
Hypothalamus
Brain
Pituitary
Lung
Mammary Gland
Rumen
Uterus
Corpus Leteum
Placenta
Cell Line
Skin
Adrenal
Granulosa
b
QTL Type
eQTL
eeQTL
isoQTL
enQTL
sQTL
stQTL
3aQTL
Proportion of total molQTL
c
Primary eQTL
Non-Primary eQTL
Primary sQTL
Non-Primary sQTL
Proportion of molQTL
d
Shared with 1 molQTL
Shared with 2 molQTL
Shared with 3 molQTL
Shared with >3 molQTL
Proportion of eQTL
1-5
6-10
11-15
16-20
21-25
26-30
31-35
#Shared Tissues
e
Tissue-eGene pairs
Gene expressed in only one tissue
YES
NO
Gene defined as eGene in only one tissue
YES
NO
LD < 0.2 between lead eQTL in two tissues
YES
NO
LD > 0.8 or same SNP between lead eQTL in two tissues
YES
Different effect direction between lead eQTL in two tissues
YES
NO
Different 95%credible interval of effect between lead eQTL in two tissues
YES
NO
Overlapping 95%credible interval of effect between lead eQTL in two tissues
YES
f
Tissue-eGene pairs with specific expression
rs134955850 of INMT in Liver
p-value = 1.77×10⁻⁸
rs134955850 of INMT in Brain
log2(TMM)
A/A
A/G
G/G
g
Tissue-eGene pairs with specific eGene
rs133517035 of C21orf140 in Testis
p-value = 1.77×10⁻⁶
rs133517035 of C21orf140 in Muscle
p-value = 0.01
T/T
T/G
G/G
h
Tissue-eGene pairs with specific eQTL
rs434763959 of OTOR in Muscle
p-value = 8.25×10⁻⁷
rs382046370 of OTOR in Testis
p-value = 4.85×10⁻¹⁵
LD = 0.02
C/C
C/T
T/T
T/C
i
Tissue-eGene pairs with effect-opposite eQTL
rs210111183 of GPR39 in Mammary_Gland
p-value = 1.48×10⁻¹¹, β = -0.20
rs210111183 of GPR39 in Subcutaneous_Adipose
p-value = 1.27×10⁻⁴, β = 0.26
log2(TMM)
G/G
G/A
A/A
j
Tissue-eGene pairs with effect-eterogeneous eQTL
rs136524567 of TMEM106C
Effect Size
Lung
Monocyte
Muscle
Uterus
Whole Blood
Tissue
k
Tissue-eGene pairs with effect homogeneous eQTL
rs110662494 of RCAN1
Effect Size
Immune System
Lung
Leukocyte
Testis
Tissue
l
Tissue
Tissue-eGene pair proportion(%)
Tissue-eGene pair proportion(%) (LD > 0.8)
Tissue-eGene pair Group
Specific Expression
Specific eGene
Specific eQTL (LD < 0.2)
Shared eQTL (LD > 0.8)
Other eGene
Homogeneous
Heterogeneous
Opposite
m
Odds Ratio (OR)
3' UTR variant
5' UTR variant
Start gain variant
Downstream gene
Intron Variant
Missense Variant
NC exon variant
Splice region variant
Synonymous
Upstream gene
SNP annotation
TSSA
TSSFlnk
TSSWk
EnhA
EnhWk
EnhPos
BivTSS
BivEnh
ReprPC
Quies
Chromatin State

Fig. 4 | Tissue-specific regulatory effects.

a, Heatmap showing pairwise Spearman's correlation (ρ) of *cis*-eQTL effect sizes across 51 tissues. Tissues were hierarchically clustered using the complete linkage method based on the maximum distance of ρ. **b,** Proportion of molQTL active across tissues, measured by the number of tissues with a *mashr* local false sign rate (LFSR, equivalent to FDR) < 0.05. **c,** Proportion of primary and non-primary *cis*-eQTL and *cis*-sQTL active across different numbers of tissues. **d,** Proportion of *cis*-eQTL shared with different numbers of other molQTL active across different numbers of tissues. **e,** Pipeline for classifying different tissue-eGene pairs in each tissue. LD means linkage disequilibrium. **f,** Example of expression-specific tissue-eGene pairs, including gene name, lead eQTL, p-value and three genotypes in gene-expressed tissue (left) and non-expressed tissue (right). The two-sided P value of SNPs was calculated by the linear mixed regression model. **g,** Example of eGene-specific tissue-eGene pairs, including gene name, lead eQTL, p-value and three genotypes in eGene-defined tissue (left) and non-eGene tissue (right). The two-sided P value of SNPs was calculated by the linear mixed regression model. **h,** Example of eQTL-specific tissue-eGene pairs, including gene name, lead eQTL, p-value, three genotypes and LD between lead eQTL from two tissues. The two-sided P value of SNPs was calculated by the linear mixed regression model. **i,** Example of effect-opposite tissue-eGene pairs, including gene name, lead eQTL, p-value, three genotypes and effect sizes from lead eQTL in two tissues. The two-sided P value of SNPs was calculated by the linear mixed regression model. **j,** Example of effect-heterogeneous tissue-eGene pairs. The title represents gene name and lead eQTL. Y axis represents the 95% credible interval of effect size in lead eQTL. The first tissue in X axis represents the tested tissue and others represent tissues with non-overlapped interval with the tested tissue. **k,** Example of effect-homogeneous tissue-eGene pairs, The title represents gene name and lead eQTL. Y axis represents the 95% credible interval of effect size in lead eQTL. The first tissue in X axis represents the tested tissue and others represent tissues with overlapped interval with the tested tissue. **l,** Proportion of each tissue-eGene classification. Bar plot in the left represents proportion of expression-specific, eGene-specific, eQTL-specific (LD < 0.2), eQTL-shared (LD > 0.8) and other tissue-eGene pairs, whereas the right one represents proportion of eQTL-shared (LD > 0.8) tissue-eGene pairs, including effect-opposite, effect-heterogeneous and effect-homogeneous. **m,** Functional enrichment between each group of tissue-eGene and regulatory elements (top: SNP classification, bottom: chromatin states).

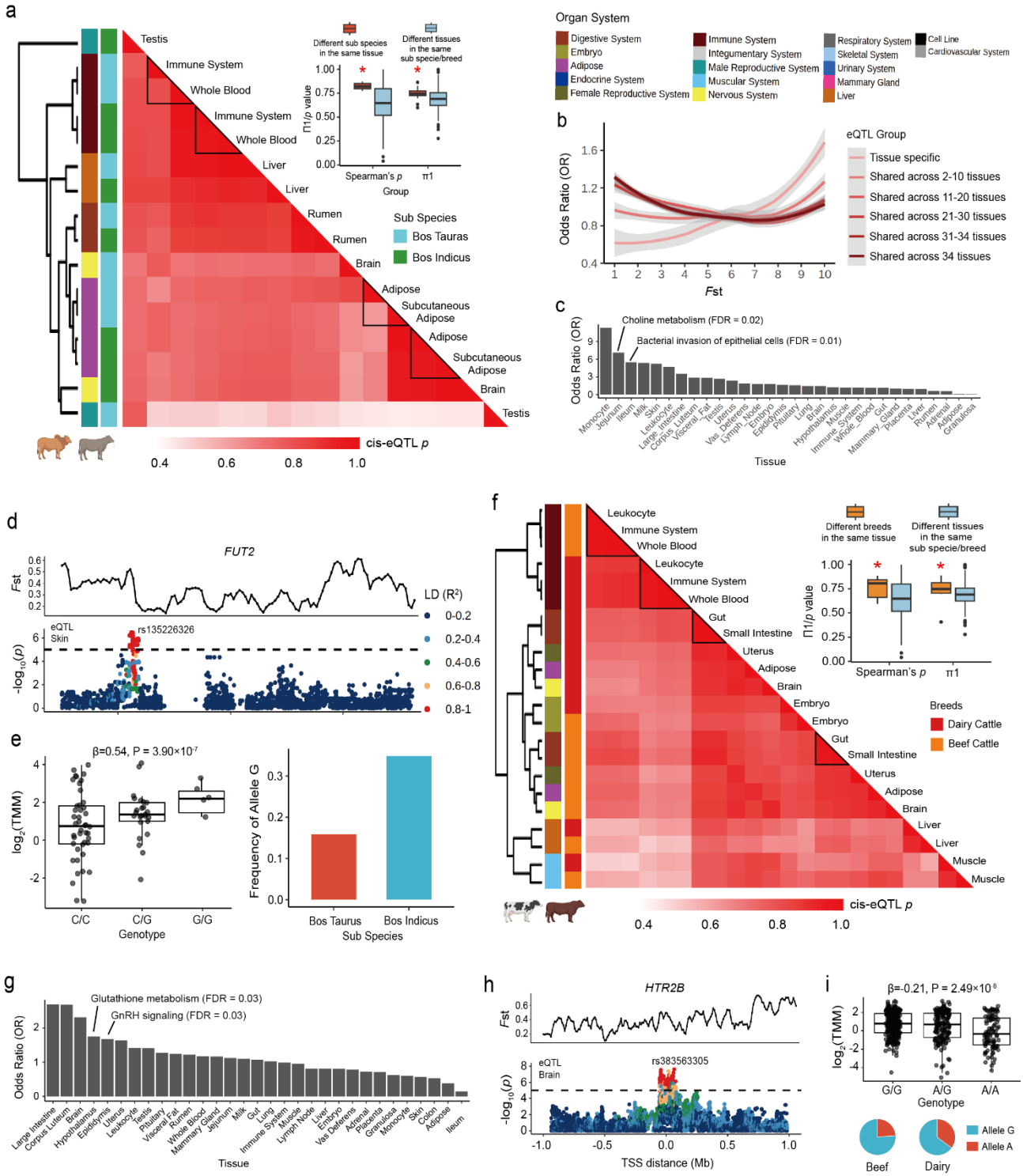


Fig. 5 | Liking molQTL to selection signatures between subspecies and breeds.

a, Heatmap showing pairwise Spearman's correlation (ρ) of population-stratified *cis*-eQTL effect sizes across 8 tissues between *B. taurus* and *B. indicus*. Tissues were hierarchically clustered using the complete linkage method based on the maximum distance of ρ. Statistical significance of $\Pi 1$ and

p-value between two groups were determined using a two-sided Student's t-test. *: P < 0.05. **b,** Enrichment (odds ratio) of shared and tissue-specific molQTL across 10 F_{ST} deciles. The lines were fitted by using the geom_smooth function from ggplot2 (v3.3.2) in R (v4.0.2). **c,** Enrichment (odds ratio) of tissue-specific eQTL and strongly selected regions (top 5% F_{ST}) between *B. taurus* and *B. indicus* across 35 tissues. Associated KEGG pathways related to immune response are indicated. **d,** Manhattan plots illustrating an example of a selected eGene (*FUT2*) in skin. The eQTL rs135226326 lies in a high-F_{ST} region (F_{ST} = 0.54). The two-sided P value of lead SNP was calculated by the linear mixed regression model. **e,** Functional interpretation of rs135226326: boxplot (left) shows expression of *FUT2* across three genotypes (C/C, C/G and G/G), while the barplot (right) shows the frequency distribution of the allele G between *B. taurus* and *B. indicus*. The two-sided P value of lead SNP was calculated by the linear mixed regression model. **f,** Heatmap showing pairwise Spearman's correlation (ρ) of population-stratified *cis*-eQTL effect sizes across 11 tissues between dairy and beef cattle. Tissues were hierarchically clustered using the complete linkage method based on the maximum distance of ρ. Statistical significances of $\Pi 1$ and p-value between two groups were determined using a two-sided Student's t-test. *: P < 0.05. **g,** Enrichment (odds ratio) of tissue-specific eQTL and strongly selected regions (top 5% F_{ST}) between dairy and beef cattle across 35 tissues. Associated KEGG pathways related to growth and development are indicated. **h,** Manhattan plots illustrating an example of a selected eGene (*HTR2B*) in brain. The eQTL rs383563305 lies in a high-F_{ST} region (F_{ST} = 0.58). The two-sided P value of lead SNP was calculated by the linear mixed regression model. **i,** Functional interpretation of rs383563305: boxplot (top) shows expression of *HTR2B* across three genotypes (G/G, G/A and A/A), while the pie chart (bottom) shows the frequency distribution of the allele G and allele A between dairy and beef cattle. The two-sided P value of lead SNP was calculated by the linear mixed regression model.

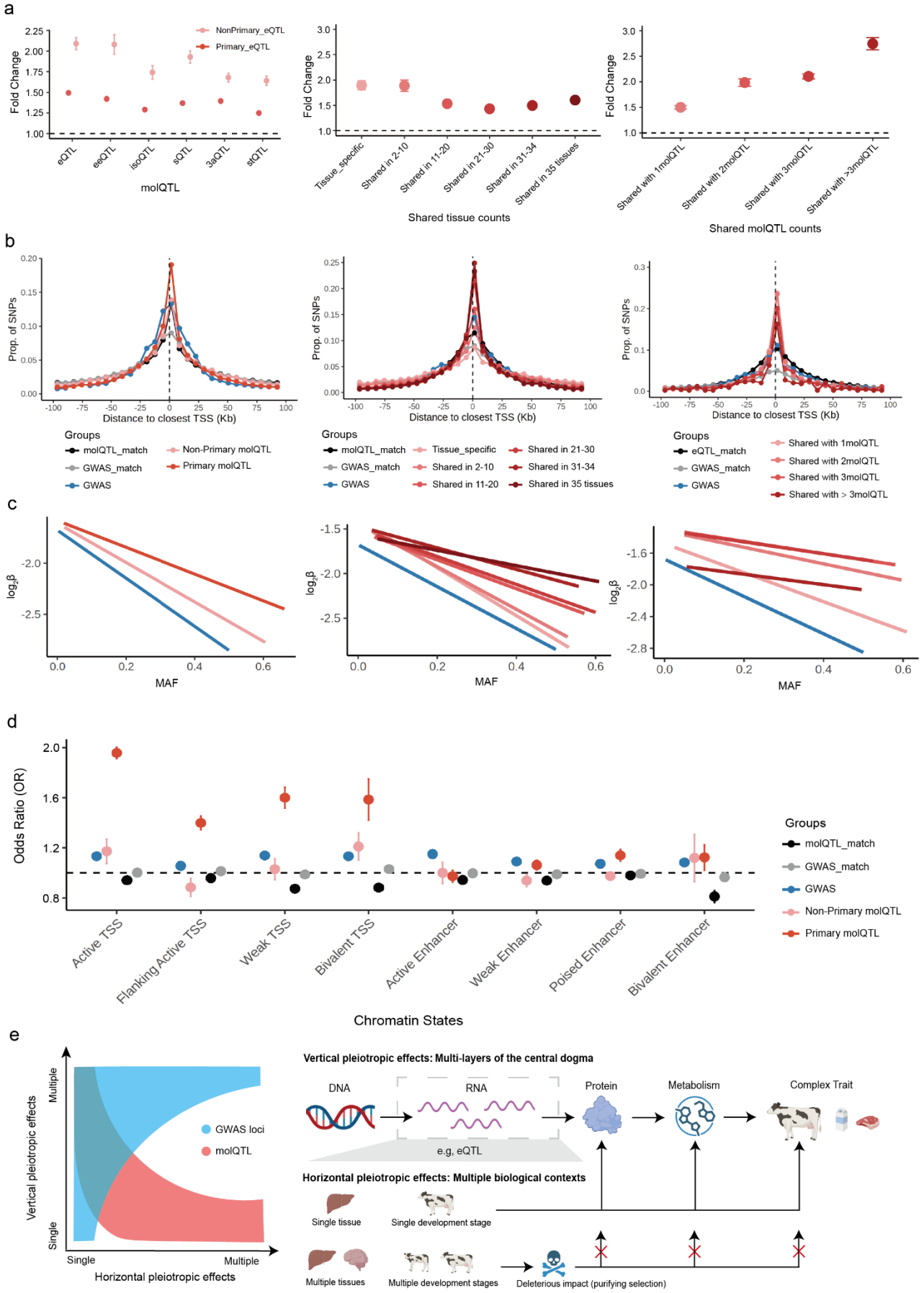

a
Fold Change
NonPrimary_eQTL
Primary_eQTL
eQTL
eeQTL
isoQTL
sQTL
3aQTL
stQTL
molQTL
Tissue_specific
Shared in 2-10
Shared in 11-20
Shared in 21-30
Shared in 31-34
Shared in 35 tissues
Shared tissue counts
Shared with 1molQTL
Shared with 2molQTL
Shared with 3molQTL
Shared with >3molQTL
Shared molQTL counts
b
Prop. of SNPs
Distance to closest TSS (Kb)
Groups
molQTL_match
GWAS_match
GWAS
Non-Primary molQTL
Primary molQTL
Tissue_specific
Shared in 2-10
Shared in 11-20
Shared in 21-30
Shared in 31-34
Shared in 35 tissues
eQTL_match
Shared with 1molQTL
Shared with 2molQTL
Shared with 3molQTL
Shared with > 3molQTL
c
log2β
MAF
d
Odds Ratio (OR)
Active TSS
Flanking Active TSS
Weak TSS
Bivalent TSS
Active Enhancer
Weak Enhancer
Poised Enhancer
Bivalent Enhancer
Chromatin States
Groups
molQTL_match
GWAS_match
GWAS
Non-Primary molQTL
Primary molQTL
e
Vertical pleiotropic effects
Multiple
Single
GWAS loci
molQTL
Horizontal pleiotropic effects
Vertical pleiotropic effects: Multi-layers of the central dogma
DNA
RNA
Protein
Metabolism
Complex Trait
e.g, eQTL
Horizontal pleiotropic effects: Multiple biological contexts
Single tissue
Single development stage
Multiple tissues
Multiple development stages
Deleterious impact (purifying selection)

Fig. 6 | Enrichment between three classes of molQTL and GWAS loci from 44 complex traits.

a, Enrichment (odds ratio) of three classes of molQTL with genome-wide associations (GWAS) of 44 cattle traits using QTLenrich (v.2.0), including primary and non-primary molQTL (**left**), tissue-sharing eQTL (**mid**) and pleiotropic eQTL (**right**). **b**, Distance to nearest TSS regions of three classes of molQTL, GWAS loci, MAF-matched GWAS loci and MAF-matched molQTL, including primary and non-primary molQTL (**left**), tissue-sharing molQTL (**mid**) and pleiotropic eQTL (**right**). **c,** Selection pressure (measured by MAF-effect size relationship) of three classes of molQTL and GWAS loci, including primary and non-primary molQTL (**left**), tissue-sharing molQTL (**mid**) and pleiotropic eQTL (**right**). **d,** Functional enrichment between primary/non-priamry molQTL and regulatory elements. **e,** A genetic regulatory model of complex traits. The left plot represents the distribution of molQTL (red) and GWAS loci (blue) across different levels of biological contexts (x-axis) and layers of the central dogma (y-axis). The right plot represents the process of how regulatory genetic variants under multiple layers of the central dogma and specific biological contexts shapes GWAS signals.

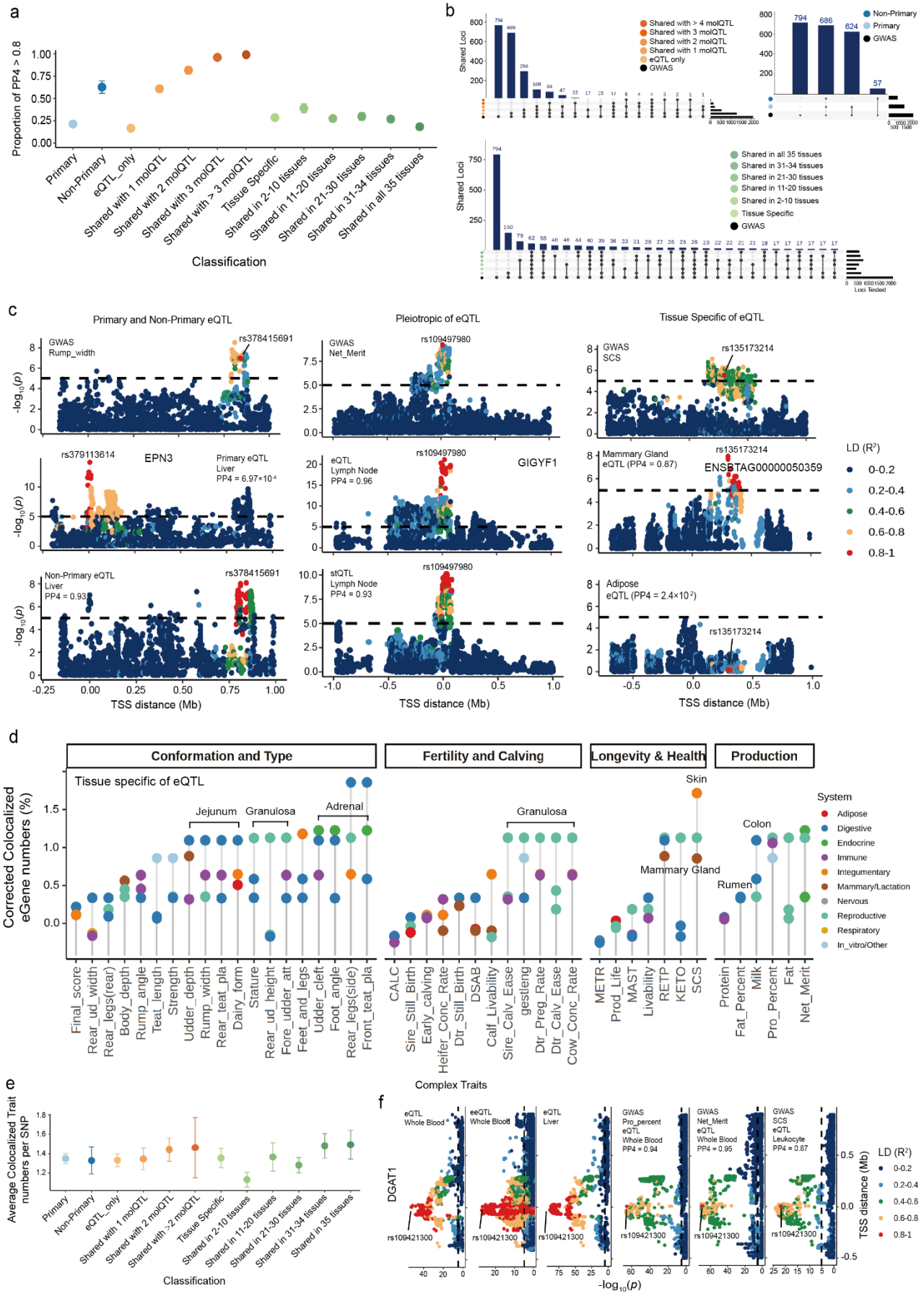

a
Proportion of PP4 > 0.8
Primary
Non-Primary
eQTL_only
Shared with 1 molQTL
Shared with 2 molQTL
Shared with 3 molQTL
Shared with > 3 molQTL
Tissue Specific
Shared in 2-10 tissues
Shared in 11-20 tissues
Shared in 21-30 tissues
Shared in 31-34 tissues
Shared in all 35 tissues
Classification
b
Shared Loci
Shared with > 4 molQTL
Shared with 3 molQTL
Shared with 2 molQTL
Shared with 1 molQTL
eQTL only
GWAS
Non-Primary
Primary
GWAS
Shared in all 35 tissues
Shared in 31-34 tissues
Shared in 21-30 tissues
Shared in 11-20 tissues
Shared in 2-10 tissues
Tissue Specific
GWAS
Loci Tested
c
Primary and Non-Primary eQTL
Pleiotropic of eQTL
Tissue Specific of eQTL
GWAS Rump_width
rs378415691
GWAS Net_Merit
rs109497980
GWAS SCS
rs135173214
rs379113614
EPN3
Primary eQTL Liver PP4 = 6.97×10⁻⁴
eQTL Lymph Node PP4 = 0.96
GIGYF1
Mammary Gland eQTL (PP4 = 0.87)
ENSBTAG00000050359
Non-Primary eQTL Liver PP4 = 0.93
stQTL Lymph Node PP4 = 0.93
Adipose eQTL (PP4 = 2.4×10⁻²)
-log10(p)
TSS distance (Mb)
LD (R²)
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1
d
Conformation and Type
Fertility and Calving
Longevity & Health
Production
Tissue specific of eQTL
Corrected Colocalized eGene numbers (%)
Jejunum
Granulosa
Adrenal
Skin
Colon
Mammary Gland
Rumen
System
Adipose
Digestive
Endocrine
Immune
Integumentary
Mammary/Lactation
Nervous
Reproductive
Respiratory
In_vitro/Other
Final_score
Rear_ud_width
Rear_legs(rear)
Body_depth
Rump_angle
Teat_length
Strength
Udder_depth
Rump_width
Rear_teat_pla
Dairy_form
Stature
Rear_ud_height
Fore_udder_att
Feet_and_legs
Udder_cleft
Foot_angle
Rear_legs(side)
Front_teat_pla
CALC
Sire_Still_Birth
Early_calving
Heifer_Conc_Rate
Dtr_Still_Birth
DSAB
Calf_Livability
Sire_Calv_Ease
gestleng
Dtr_Preg_Rate
Dtr_Calv_Ease
Cow_Conc_Rate
METR
Prod_Life
MAST
Livability
RETP
KETO
SCS
Protein
Fat_Percent
Milk
Pro_Percent
Fat
Net_Merit
Complex Traits
e
Average Colocalized Trait numbers per SNP
Classification
f
DGAT1
eQTL Whole Blood
eeQTL Whole Blood
eQTL Liver
GWAS Pro_percent eQTL Whole Blood PP4 = 0.94
GWAS Net_Merit eQTL Whole Blood PP4 = 0.95
GWAS SCS eQTL Leukocyte PP4 = 0.87
rs109421300
-log10(p)
TSS distance (Mb)
LD (R²)

Fig. 7 | Colocalization between three classes of molQTL and GWAS loci from 44 complex traits.

a, Proportion of colocalized molQTL (PP.H4 > 0.8) in three classes molQTL, including primary and non-primary molQT, tissue-shared molQTL and pleiotropic eQTL. **b**, Numbers of colocalized GWAS loci by each class of molQTL, including primary and non-primary molQTL (**right and top**), tissue-shared molQTL (**bottom**) and pleiotropic eQTL (**left and top**). Each blue bar represents a combination of different molQTL in each class. Each black bar (right) represents the number of GWAS loci colocalized in the respective combination. **c,** Manhattan plots showing the -$\log_{10}$(p) distribution of specific associated signals between GWAS loci and three classes of molQTL, including rump width (GWAS) and non-primary eQTL in liver across genetic variants within the locus of *EPN3*, net merit (GWAS) and pleiotropic eQTL (with stQTL) in lymph nodes across genetic variants within the locus of *GIGYF1*, and SCS (GWAS) and tissue-specific eQTL in mammary gland across genetic variants within the locus of *ENSBTAG00000050359*. Variants are colored according to their linkage disequilibrium (LD, r^2) with the lead SNP. The two-sided P values of SNPs were calculated by the linear mixed regression model. **d**, Corrected colocalized eGene counts (PP.H4 > 0.8) between each tissue and each complex trait in eQTL results. Colocalized eGene counts were corrected by detected eQTL counts. **e,** Average colocalized trait counts per SNP (PP.H4 > 0.8) in three classes of molQTL, including primary and non-primary molQT, tissue-shared molQTL and pleiotropic eQTL **f,** Manhattan plots showing the -$\log_{10}$(p) distribution of the association signal between pleiotropic GWAS loci and molQTL for protein percent, SCS and net merit (GWAS) with eQTL and eeQTL in whole blood and liver across genetic variants within the locus of *DGAT1*. Variants are colored according to their linkage disequilibrium (LD, r^2) with the lead SNP rs109421300. The two-sided P values of lead SNPs were calculated by the linear mixed regression model.

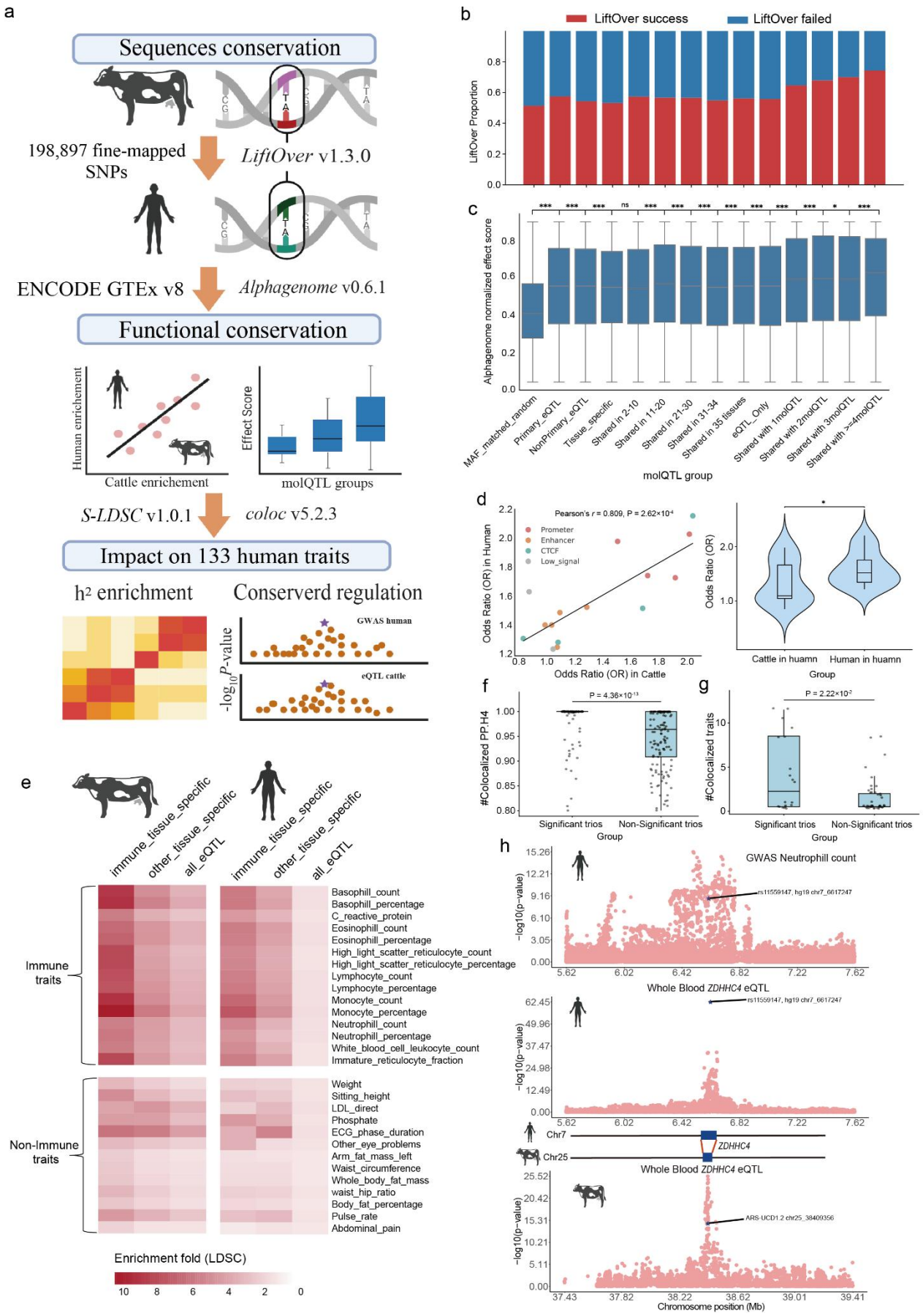
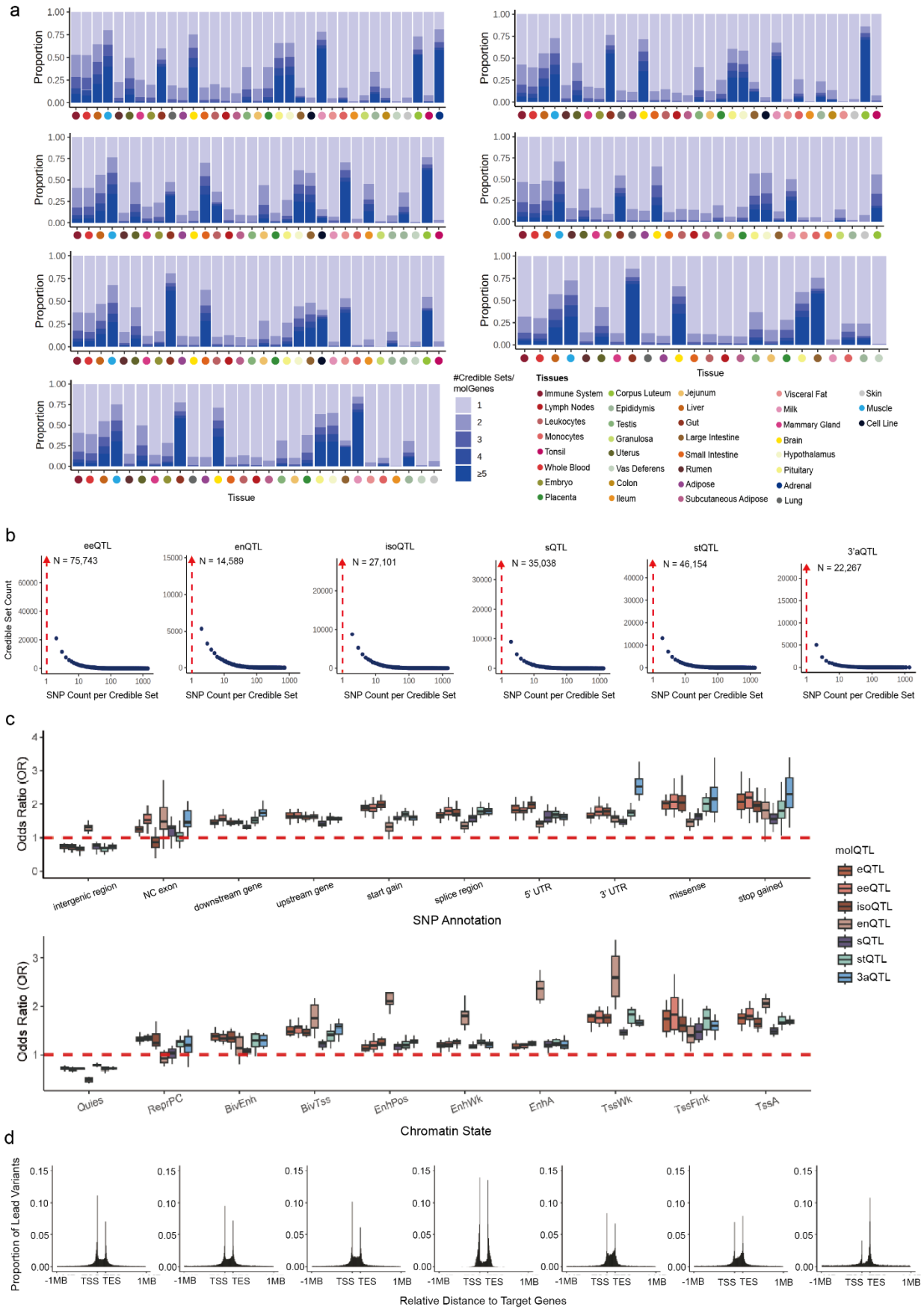
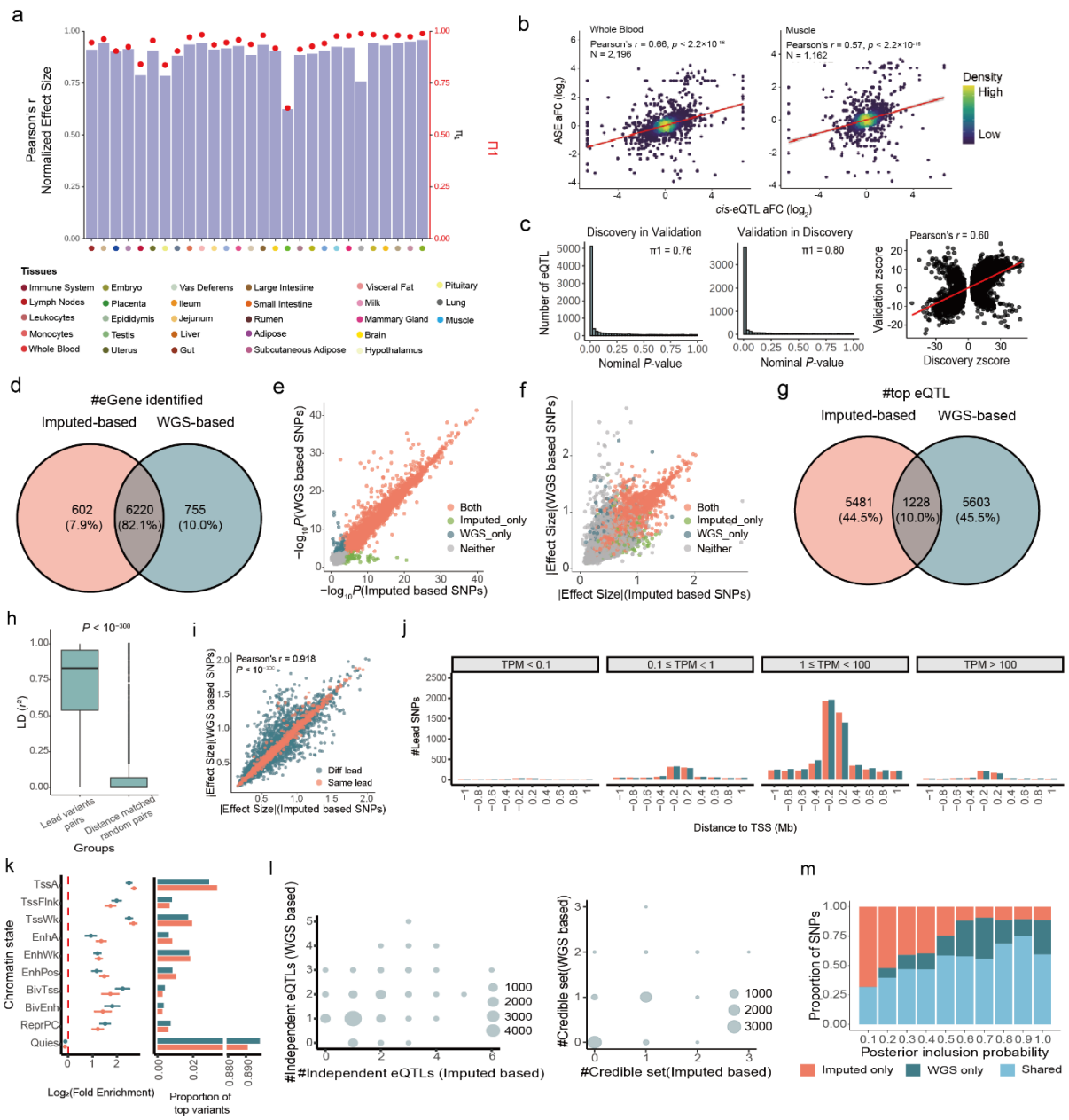

a
Sequences conservation
198,897 fine-mapped SNPs
LiftOver v1.3.0
ENCODE GTEx v8
Alphagenome v0.6.1
Functional conservation
Human enrichement
Cattle enrichement
Effect Score
molQTL groups
S-LDSC v1.0.1
coloc v5.2.3
Impact on 133 human traits
h² enrichment
Conserverd regulation
GWAS human
eQTL cattle
-log10 P-value
b
LiftOver success
LiftOver failed
LiftOver Proportion
c
Alphagenome normalized effect score
MAF_matched_random
Primary_eQTL
NonPrimary_eQTL
Tissue_specific
Shared in 2-10
Shared in 11-20
Shared in 21-30
Shared in 31-34
Shared in 35 tissues
eQTL_Only
Shared with 1molQTL
Shared with 2molQTL
Shared with 3molQTL
Shared with >=4molQTL
molQTL group
d
Pearson's r = 0.809, P = 2.62×10⁻⁴
Prometer
Enhancer
CTCF
Low_signal
Odds Ratio (OR) in Human
Odds Ratio (OR) in Cattle
Odds Ratio (OR)
Cattle in huamn
Human in huamn
Group
f
P = 4.36×10⁻¹³
#Colocalized PP.H4
Significant trios
Non-Significant trios
Group
g
P = 2.22×10⁻³
#Colocalized traits
Significant trios
Non-Significant trios
Group
e
immune_tissue_specific
other_tissue_specific
all_eQTL
immune_tissue_specific
other_tissue_specific
all_eQTL
Immune traits
Basophill_count
Basophill_percentage
C_reactive_protein
Eosinophill_count
Eosinophill_percentage
High_light_scatter_reticulocyte_count
High_light_scatter_reticulocyte_percentage
Lymphocyte_count
Lymphocyte_percentage
Monocyte_count
Monocyte_percentage
Neutrophill_count
Neutrophill_percentage
White_blood_cell_leukocyte_count
Immature_reticulocyte_fraction
Non-Immune traits
Weight
Sitting_height
LDL_direct
Phosphate
ECG_phase_duration
Other_eye_problems
Arm_fat_mass_left
Waist_circumference
Whole_body_fat_mass
waist_hip_ratio
Body_fat_percentage
Pulse_rate
Abdominal_pain
Enrichment fold (LDSC)
h
GWAS Neutrophill count
rs11559147; hg19 chr7_6617247
-log10(p-value)
Whole Blood ZDHHC4 eQTL
Chr7
Chr25
ZDHHC4
ARS-UCD1.2 chr25_38409356
Chromosome position (Mb)

Fig. 8 | Cross-species analysis identifies conserved regulatory effects between cattle and humans.

a, Overview of the cross-species analysis workflow. At the sequence-conservation level, fine-mapped cattle expression quantitative trait loci (eQTL) were projected onto the human genome using LiftOver. Functional conservation was assessed using ENCODE regulatory annotations and effect-score predictions from the AlphaGenome model. Conserved regulatory activity and its contribution to human complex traits were evaluated using 133 genome-wide association study (GWAS) summary statistics from UK Biobank and human GTEx expression data. **b,** Stacked bar plot showing the proportion of cattle eQTL from different groups, and a minor-allele-frequency (MAF)-matched random set, that could be successfully lifted over to the human genome. **c,** Box plots of standardized AlphaGenome-predicted effect scores for lifted-over variants from different cattle QTL groups; only tissue-level, gene-expression–related tracks were included. P values were calculated using two-sided unpaired t tests. **d,** Left, correlation between enrichment of lifted-over cattle eQTL in human ENCODE regulatory elements and enrichment of human GTEx eQTL in the same elements. Right, violin plots comparing the magnitude and distribution of enrichment estimates between lifted-over cattle eQTL and human GTEx eQTL. **e,** Stratified LD score regression (S-LDSC) heritability enrichment of tissue-specific eQTL annotations from cattle GTEx and human GTEx across complex traits; the heat map shows enrichment fold change. **f,** For 344 cattle–human orthologous gene–SNP–trait trios, comparison of the distribution of colocalization posterior probability (PPH4) with human complex traits between trios that were significant versus not significant in cattle GTEx. P values were calculated using two-sided unpaired t tests. **g,** For the same 344 trios, comparison of the number of human complex traits showing evidence of colocalization between cattle GTEx–significant versus non-significant trios. P values were calculated using two-sided unpaired t tests. **h,** Example of a conserved whole-blood eQTL at *ZDHHC4* in cattle and humans that also influences human neutrophil count; human rs11559147 is homologous to the cattle ARS-UCD1.2 chr25_38409356 variant.

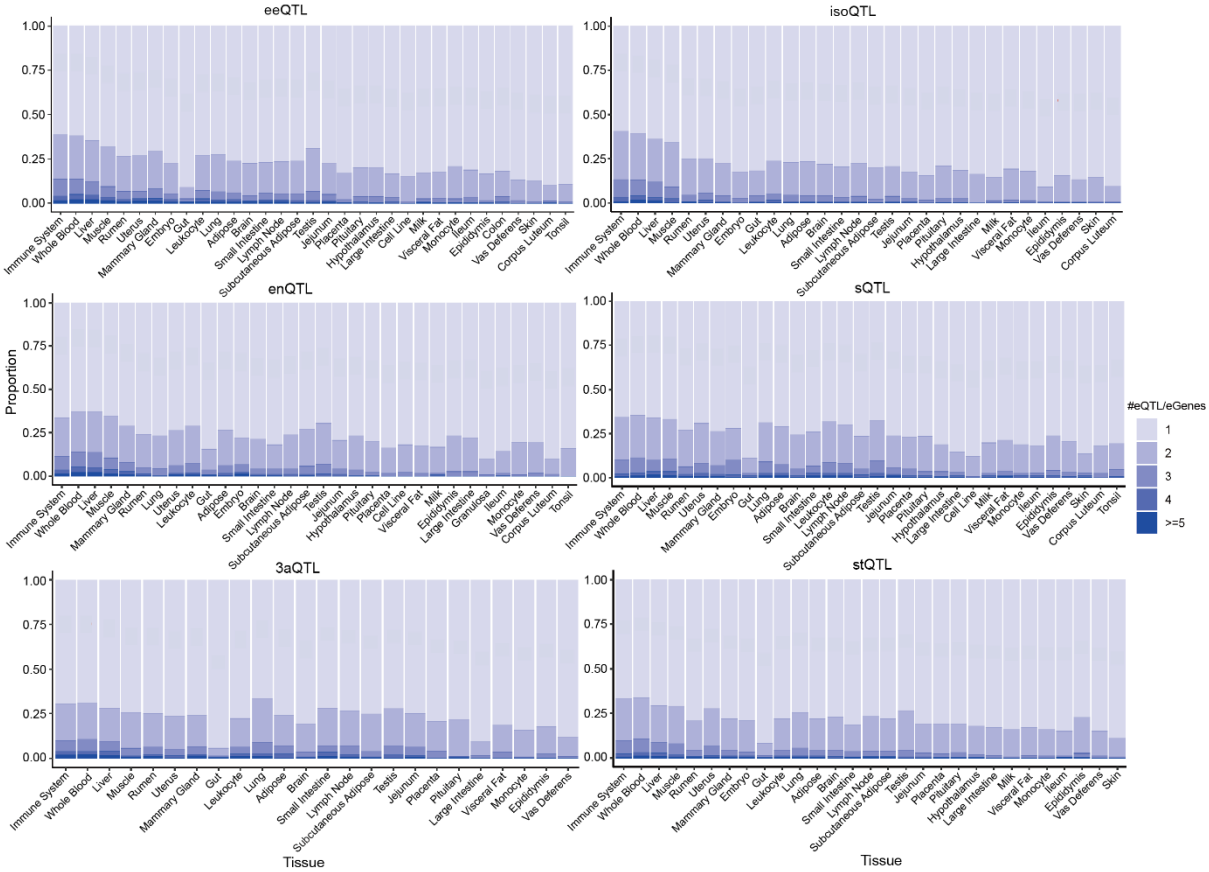

a
Proportion
Tissue
#Credible Sets/ molGenes
1
2
3
4
≥5
Tissues
Immune System
Lymph Nodes
Leukocytes
Monocytes
Tonsil
Whole Blood
Embryo
Placenta
Corpus Luteum
Epididymis
Testis
Granulosa
Uterus
Vas Deferens
Colon
Ileum
Jejunum
Liver
Gut
Large Intestine
Small Intestine
Rumen
Adipose
Subcutaneous Adipose
Visceral Fat
Milk
Mammary Gland
Brain
Hypothalamus
Pituitary
Adrenal
Lung
Skin
Muscle
Cell Line
b
eeQTL
N = 75,743
enQTL
N = 14,589
isoQTL
N = 27,101
sQTL
N = 35,038
stQTL
N = 46,154
3'aQTL
N = 22,267
Credible Set Count
SNP Count per Credible Set
c
Odds Ratio (OR)
intergenic region
NC exon
downstream gene
upstream gene
start gain
splice region
5' UTR
3' UTR
missense
stop gained
SNP Annotation
molQTL
eQTL
eeQTL
isoQTL
enQTL
sQTL
stQTL
3aQTL
Quies
ReprPC
BivEnh
BivTss
EnhPos
EnhWk
EnhA
TssWk
TssFlnk
TssA
Chromatin State
d
Proportion of Lead Variants
-1MB
TSS TES
1MB
Relative Distance to Target Genes

Extended Data Fig. 1 | a, The distribution of the number of fine-mapped *cis*-molQTL per gene across seven molQTL is shown as blue stacked bars across 35 tissues. Tissues are ordered by decreasing sample size. **b,** The size distribution of fine-mapped credible sets for all identified *cis*-molQTL across 35 tissues. The red arrow indicates count of credible sets containing only one single candidate variant. **c,** Functional enrichment of lead SNPs in seven molQTL with regulatory elements (top: SNP classification, bottom: chromatin states). **d,** Distribution of the distance to TSS regions across lead SNPs in seven molQTL.

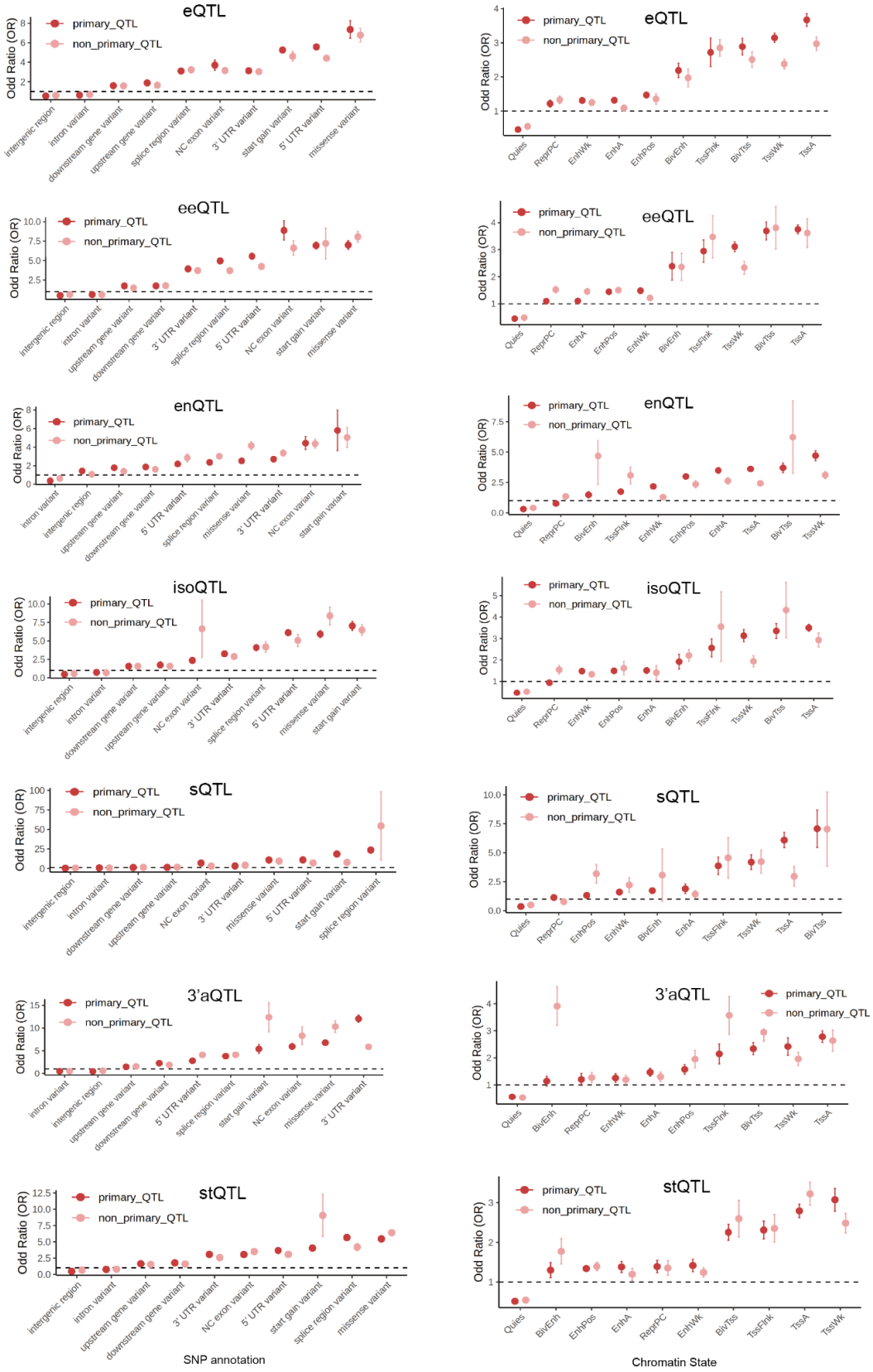


Extended Data Fig. 2 | a, Internal validation of *cis*-eQTL. We selected 28 tissues with sample sizes > 100 and randomly split samples into equally sized discovery and validation cohorts. eQTL mapping was performed independently in each cohort. Bars show Pearson's correlation of normalized effect sizes between the two groups, and red dots represent Storey's π_1 statistics, indicating the replication rate of *cis*-eQTL. **b,** Spearman's correlation between *cis*-eQTL effect sizes (allelic fold change, $\log_2$ scale) and corresponding effect sizes derived from allele-specific expression (ASE) analysis. **c,** External validation of *cis*-eQTL using an independent Holstein population ($n = 99$). Storey's π_1 statistics and Spearman's correlation were used to assess replication between external population and CattleGTEx Phase1 population in blood. **d,** Robustness of our approach by comparing molQTL mapping

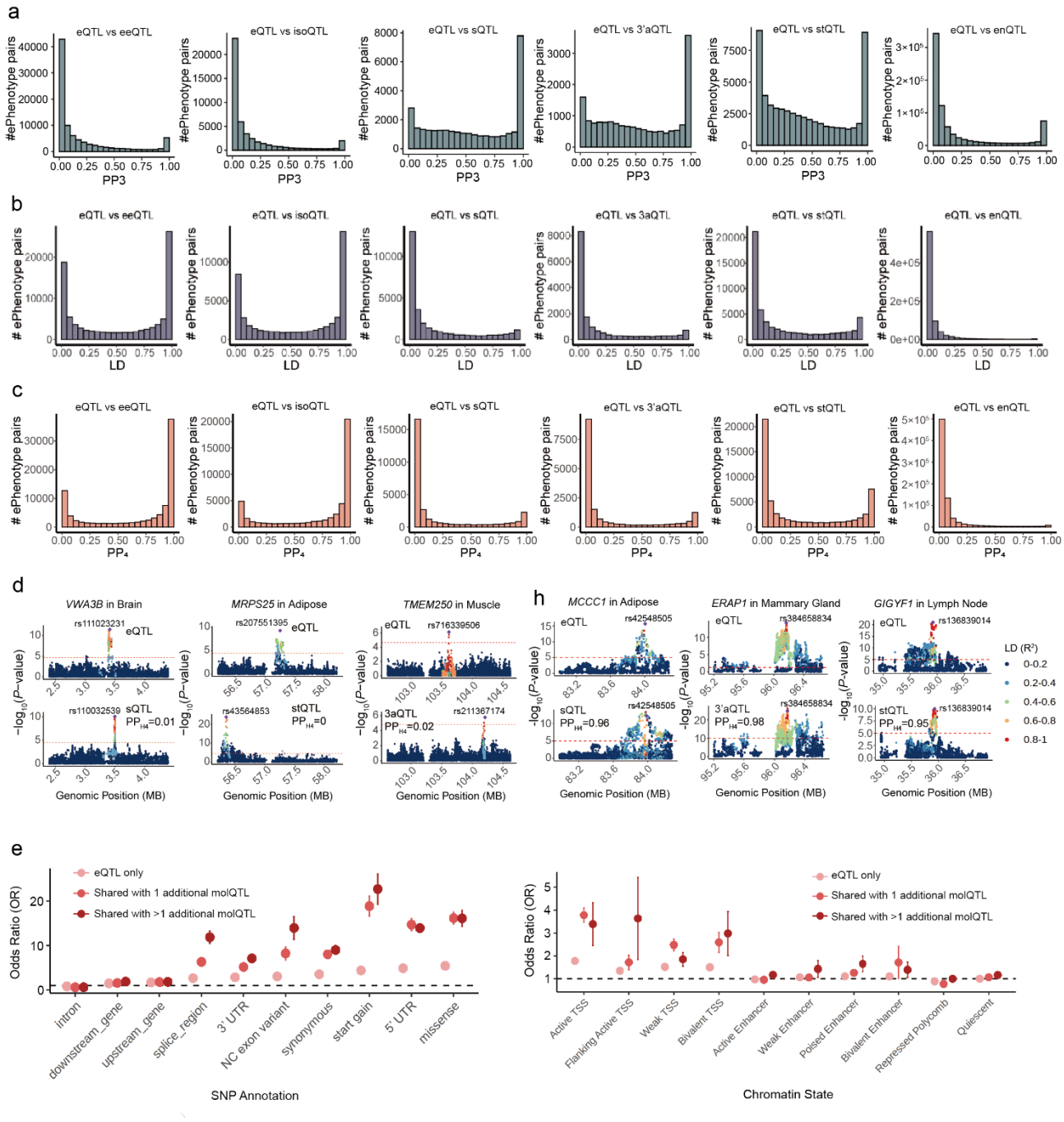
using RNA-seq imputed versus whole-genome sequencing (WGS)-called genotypes in an independent blood-tissue cohort (n = 99). Venne plot represents the overlapping of eGenes detected using imputed versus WGS–called genotypes. **e,** Comparison of significance (-log10 P) across different gene categories. Each point represents a gene. Genes are categorized as Both (n = 5,798): eGene detected by both WGS–called and imputed genotypes; WGS_based_only (n = 552): eGene detected only by WGS–called genotypes; Imputed_based _only (n = 1,792): eGene detected only by imputed genotypes; Neither (n = 7,935): Non-eGene in either dataset. **f,** Comparison of effect sizes (beta) of top variants across different gene categories. Each point represents a gene. **g,** Overlap of top variants using imputed versus WGS–called genotypes. **h,** Linkage disequilibrium (LD, r2) among lead variants of the same eGenes (n = 5,142) identified using imputed- and WGS-called. The "Distance-matched random set" contains SNP pairs with similar genomic distances as the lead variant pairs. The boxplot shows the median as the central line, the 25th–75th percentiles as the box limits, and whiskers extending 1.5× the interquartile range. Statistical significance was determined using a two-sided Student's t-test. **i,** Correlation of effect sizes for lead variants of eGenes identified using imputed- and WGS-based genotypes. Same lead indicates that the shared eGenes have the same lead variant, while Diff lead indicates that the shared eGenes have distinct lead variants. P-values were calculated using the asymptotic t approximation. **j,** Number of top variants detected using imputed and WGS-called genotypes, plotted by distance to the TSS and stratified by median gene expression across samples. **k,** Functional enrichment (log2 Fold change, mean ± s.d., left) and proportion (right) of eQTL detected using WGS-called versus imputed genotypes across chromatin states. **l,** The left panel represents numbers of independent eQTL per gene detected using WGS-called and imputed genotypes. Point size reflects the number of eGenes with the corresponding number of independent eQTL. The right panel represents numbers of credible sets per gene detected in fine-mapping using WGS-called and imputed genotypes. Point size reflects the number of eGenes with the corresponding number of credible sets. **m,** Percentage of fine-mapped variants detected using WGS-called and imputed genotypes, shown across different posterior inclusion probability (PIP) thresholds



Extended Data Fig. 3 | The distribution of the number of conditionally independent *cis*-molQTL per gene across six molQTL is shown as blue stacked bars across 35 tissues. Tissues are ordered by decreasing sample size.

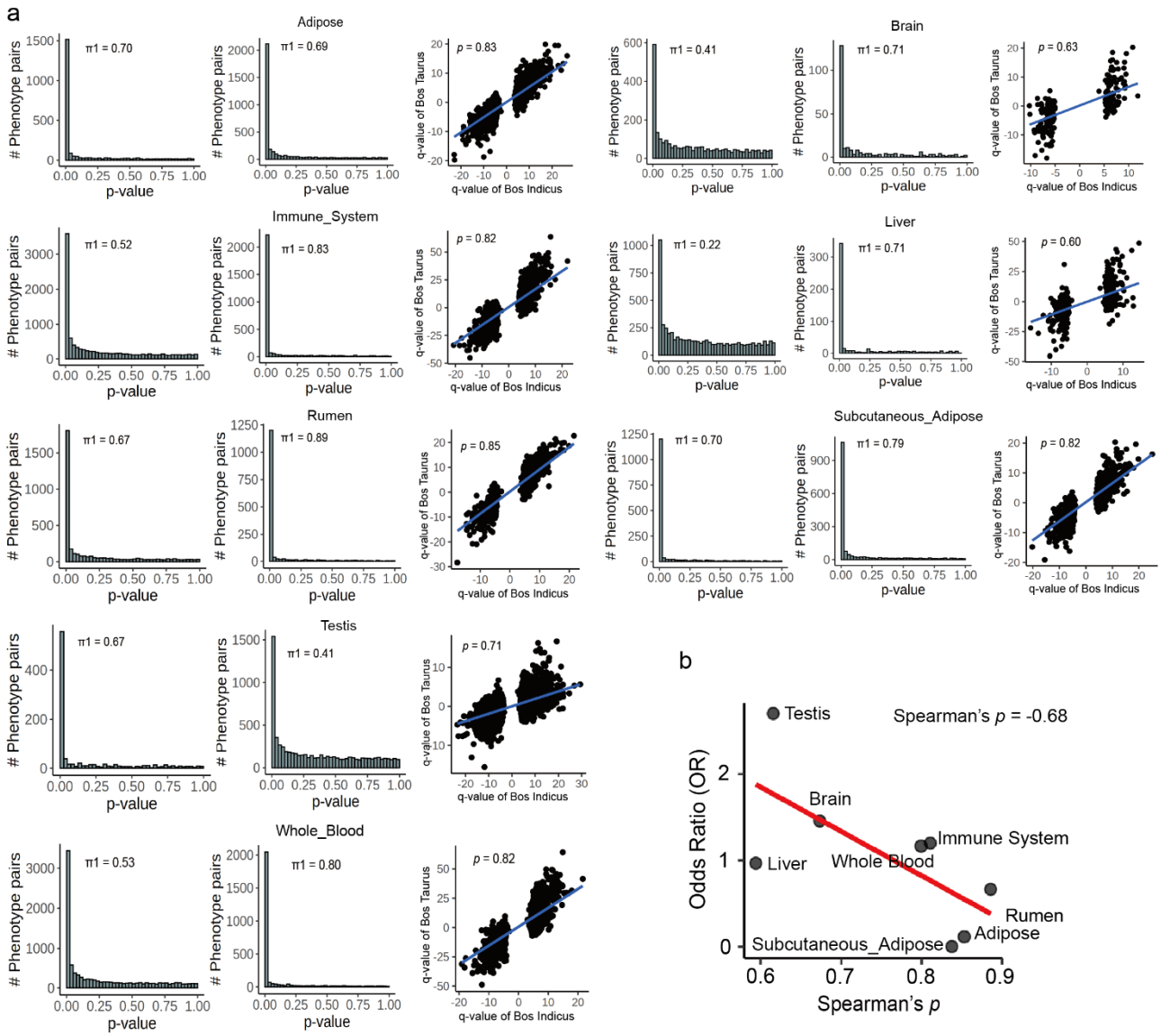


Extended Data Fig. 4 | Functional enrichment of primary and non-primary effects in six molQTL with regulatory elements (left: SNP classification, right: chromatin states).

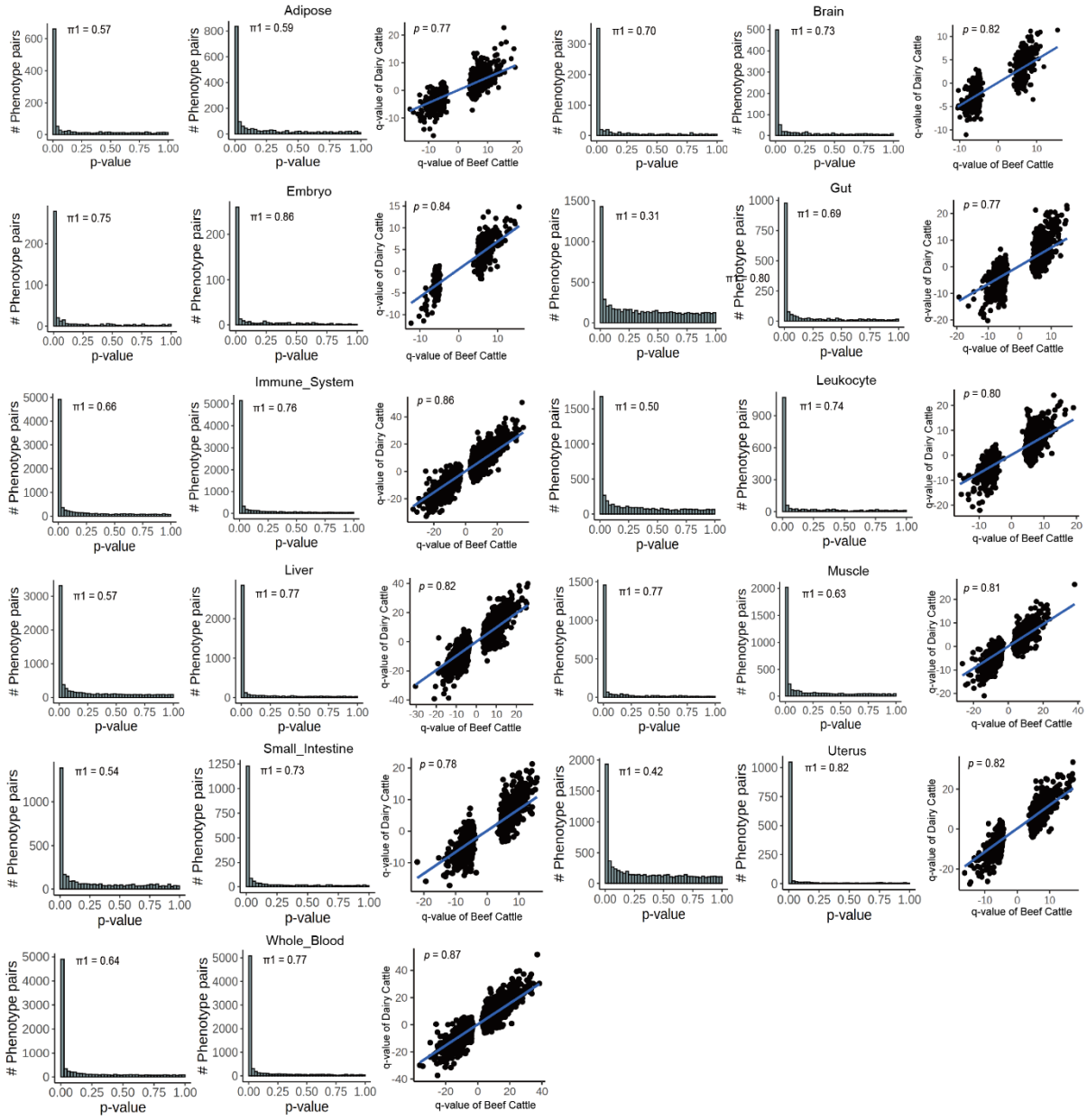


Extended Data Fig. 5 | a-c, Colocalization analyses eQTL across other six molecular phenotypes of the same molGene. PP_3: the posterior probability that the two molQTL are driven by distinct causal variants (**a**). PP_4: the posterior probability that the two molQTL share a single causal variant (**c**). LD (r^2): the linkage disequilibrium of lead SNPs of two molecular phenotypes derived from the molGene (**b**). **d,** Examples of distinct signals between eQTL and other six molQTL within the same molGene.

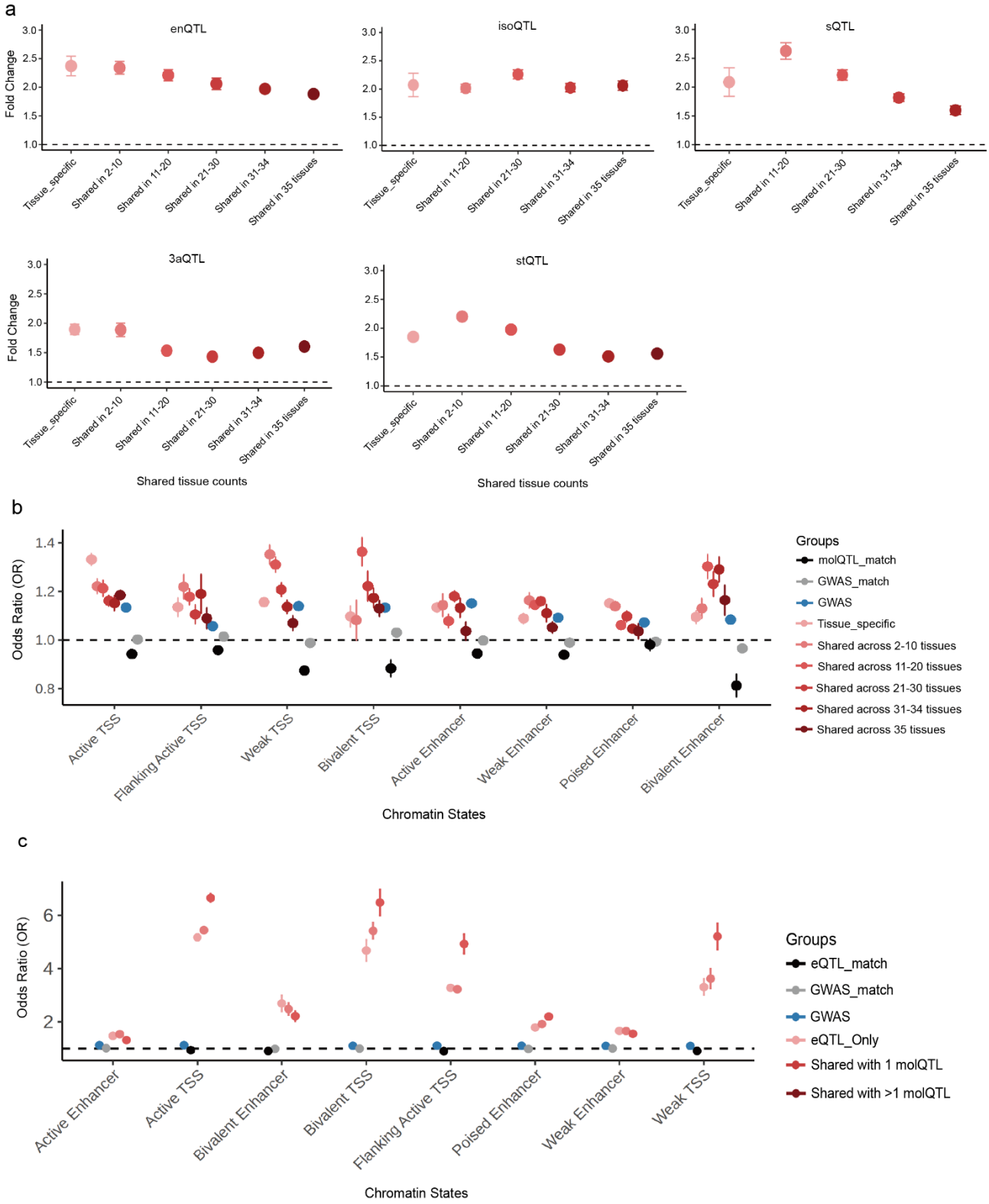
The two-sided P values of lead SNPs were calculated by the linear mixed regression model. **e,** Functional enrichment of pleiotropic eQTL with regulatory elements (left: SNP classification, right: chromatin states).



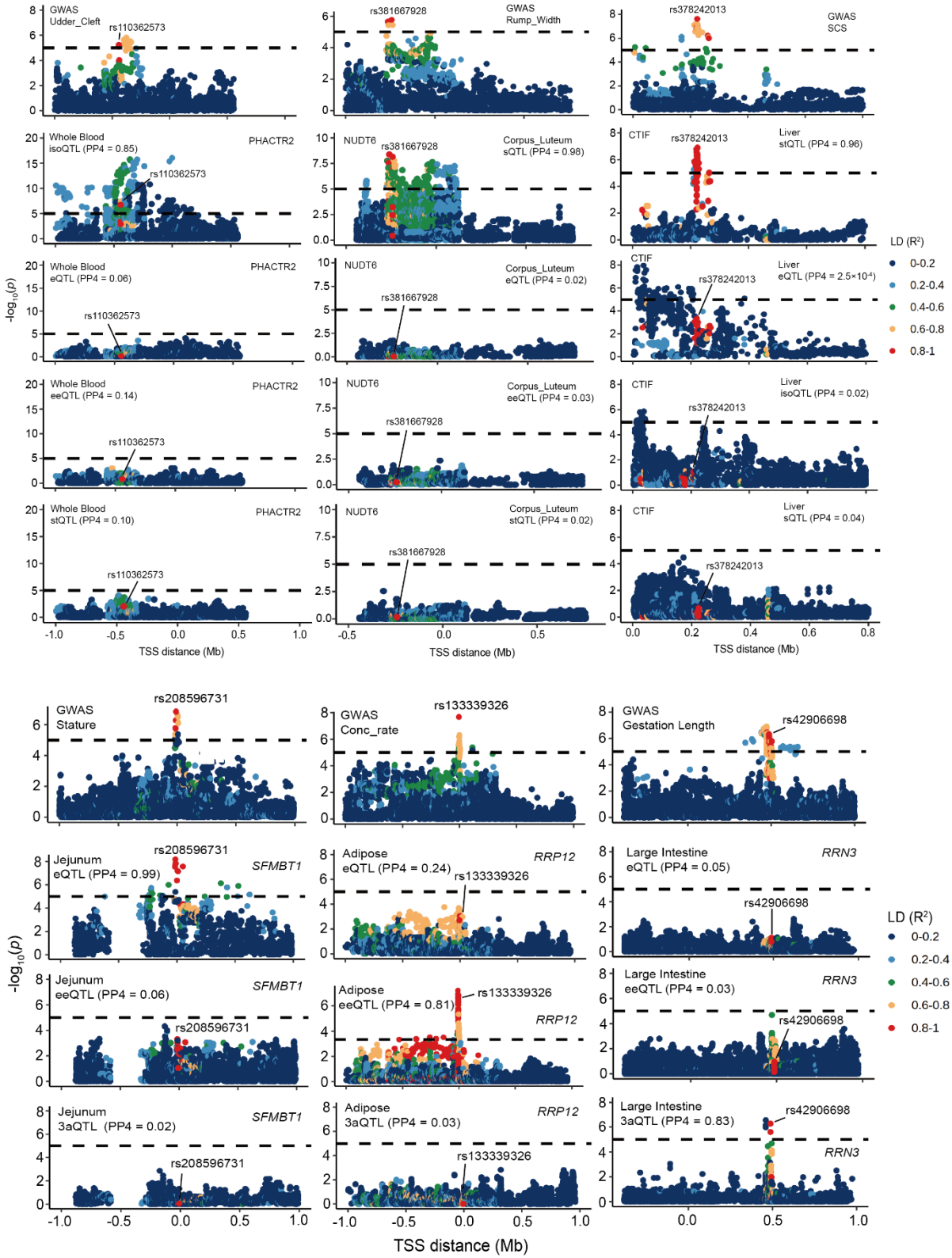
Extended Data Fig. 6 | a, Storey's π_1 statistics implemented in the qvalue package (v2.32.0)[119] and Spearman's correlation were used to assess replication of effect sizes between *B. taurus* and *B. indicus* across eight tissues. In each tissue, the p-value of the eQTL of *B. taurus* was used as observation (**left**) and validation data (**mid**), respectively. **b,** Spearman's correlation between odds ratio (OR) of enrichment between tissue-specific eQTL and strongly selected regions and spearman's correlation of effect sizes between *B. taurus* and *B. indicus*.



Extended Data Fig. 7 | Storey's π_1 statistics implemented in the qvalue package (v2.32.0)[119] and Spearman's correlation were used to assess replication of effect sizes between dairy cattle and beef cattle across 11 tissues**.** In each tissue, the p-value of the eQTL of dairy cattle was used as observation (**left**) and validation data (**mid**), respectively.

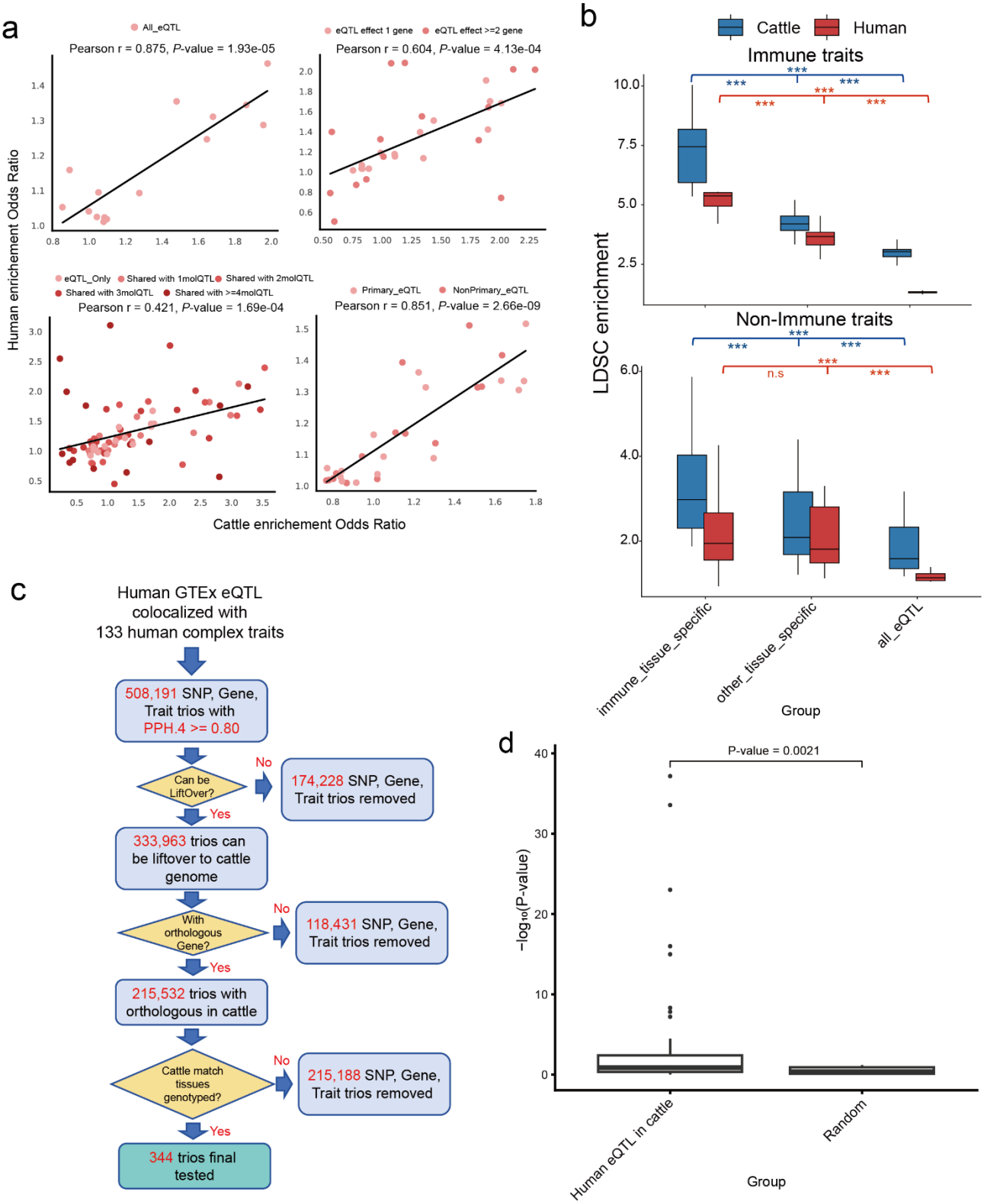


Extended Data Fig. 8 | a, Enrichment (odds ratio) of tissue specific/shared molQTL with genome-wide associations (GWAS) of 44 cattle traits using QTLenrich (v.2.0)[54]. **b,** Functional enrichment between tissue specific/shared molQTL and regulatory elements. **c,** Functional enrichment between pleiotropic eQTL and regulatory elements.



Extended Data Fig. 9 | Manhattan plots showing the -log$_{10}$(p) distribution of distinct colocalization between GWAS loci and each of seven molQTL, including udder cleft (GWAS) and isoQTL in whole blood across genetic variants within the locus of *PHACTR2*, rump width (GWAS) and sQTL in corpus leteum across genetic variants within the locus of *NUDT6*, SCS (GWAS) and stQTL in liver

across genetic variants within the locus of *CTIF,* stature (GWAS) and eQTL in jejunum across genetic variants within the locus of *SFMBT1,* conception rates (GWAS) and eeQTL in adipose across genetic variants within the locus of *RRP12* and gestation length (GWAS) and 3'aQTL in large intestine across genetic variants within the locus of *RRN3*. Variants are colored according to their linkage disequilibrium (LD, r^2) with the lead SNP. The two-sided P values of SNPs were calculated by the linear mixed regression model.



Extended Data Fig. 10 | a, Correlation between enrichment odds ratios (ORs) for different cattle eQTL groups in cattle FAANG annotations and the corresponding enrichment ORs after liftOver to the human genome in ENCODE annotations. **b,** Comparison of stratified LD score regression (S-LDSC) heritability enrichment of tissue-specific eQTL annotations from cattle and humans across

immune versus non-immune traits. P values were calculated using two-sided unpaired t tests; $P < 0.05$ (), $P < 0.01$ (), and $P < 0.001$ (). **c,** Schematic of filtering steps and counts used to derive cattle orthologous, segregating trios from 508,191 gene–variant–trait trios obtained by colocalization between human GTEx eQTL and 133 human complex traits. Filters were applied sequentially: (1) the human variant could be lifted over from the human genome to the cattle genome; (2) the corresponding eGene had an ortholog in cattle; (3) the corresponding tissue was available in the cattle dataset; and (4) the lifted-over cattle locus harbored a segregating SNP. This yielded 344 retained trios. **d,** Comparison of association significance in cattle for the 344 retained trios tested using a single-variant linear mixed model versus a random set of matched orthologous loci. P values were calculated using two-sided unpaired t tests.