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[40,56]*, 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

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

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

204 **Results**

205 **Data summary and molecular phenotyping**

206 In total, we collected 27,681 RNA-seq samples spanning 103 bovine tissues (**Supplemen-**
207 **tary Table** 1-2). After quality control (**Fig.** 2a**; Supplementary Fig.** 1**; Supplementary Note**),
208 19,789 high-quality RNA-seq samples of 88 tissues were retained for further analysis, and their gen-
209 otypes were imputed using the population structure matched reference panel described above[34]. Fol-
210 lowing extensive quality control and independent validation of genotype imputation accuracy (**Sup-**
211 **plementary Notes**), we obtained 4,317,531 common variants (minor allele frequency, MAF > 0.05),
212 with a median concordance rate of 0.97 and genotype Pearson's correlation-r^2 of 0.91 **(Fig.** 2b**)**. On
213 average, 4,576 variants were available within the *cis*-window (±1 Mb from the transcription start site,
214 TSS) of expressed genes (Transcript per million, TPM > 0.1 and read count > 6) **(Supplementary Fig.**

215 2a-d**)**. Compared with all 28,958,211 variants in the full reference panel, RNA-seq imputed variants
216 were more enriched in gene bodies and *cis*-regulatory elements (e.g., promoter and enhancer) but
217 depleted in intergenic regions **(Fig.** 2c**; Supplementary Fig.** 2e**)**. After removing duplicated samples
218 within each tissue based on identity-by-state (IBS ≥ 0.9), the final dataset comprised 12,422 RNA-
219 seq samples from 8,956 individuals across 43 tissues, each represented by more than 40 samples
220 **(Supplementary Fig.** 2f-g**)**. For 5,170 samples lacking breed information, their breed identity was
221 inferred using imputed genotypes by using a linear support vector machine (Linear SVM)[35], achieving
222 an overall prediction accuracy of 97.7% with 10-fold stratified cross validation **(Supplementary Fig.**
223 3; **Supplementary Table** 5-6**)**.

224 Using these RNA-seq datasets, we defined and quantified seven classes of molecular phenotypes,
225 including four abundance-based phenotypes (expression levels for 20,875 protein-coding and
226 lncRNA genes, 102,033 exons, 20,690 isoforms, and 27,653 enhancers), and three structure-based
227 phenotypes (68,258 alternative splicing, 17,995 RNA stability, and 18,387 3'UTR alternative poly-
228 adenylation-3'UTR APA) **(Fig.** 2a**)**. Across all molecular phenotypes, samples clustered primarily by
229 tissue type, and then reflected developmental origin, with samples derived from the same embryonic
230 germ layer tending to cluster together **(Fig.** 2d, **Supplementary Fig.** 4**)**. Genes exhibiting high inter-
231 tissue variability in expression were significantly enriched in tissue-specific biological processes,
232 such as nervous system processes (FDR = 8.81×10^{-30}) and visual perception (FDR = 6.82×10^{-6}),
233 whereas genes with low inter-tissue variability were enriched in fundamental cellular functions such
234 as protein localization (FDR = 3.15×10^{-32}) and translation (FDR = 2.09×10^{-27}) **(Supplementary**
235 **Fig.** 5a, c; **Supplementary Table** 7**)**. As expected, lncRNA genes exhibited higher tissue-specificity
236 compared to protein-coding genes[22] **(Supplementary Fig.** 5b**)**. Functional enrichment analysis of
237 tissue-specific genes (defined by FDR < 0.05 and $|\log_2$ fold change$|$ > 2 using the Wilcoxon rank-sum
238 test[36]) recapitulated well-established tissue biology **(Fig.** 2d**)**. For example, genes upregulated in im-
239 mune tissues were enriched in regulation of immune system processes (FDR = 1.77×10^{-70}), while
240 genes upregulated in the digestive system were enriched in monocarboxylic acid transport (FDR =
241 4.71×10^{-8}) **(Fig.** 2d; **Supplementary Table** 8**)**. Altogether, these results demonstrated that the de-
242 rived molecular phenotypes from publicly available RNA-seq data can robustly capture tissue-spe-
243 cific biological programs, providing a foundation for dissecting the regulatory architecture underlying
244 tissue-specific gene regulation.

245 **Fine-mapping and validation of regulatory variants**

246 Following rigorous quality control and the mitigation of confounding factors via a linear mixed
247 model implemented in OmiGA[37] (**Methods**; **Supplementary Fig.** 6), we estimated the *cis*-heritabil-
248 ity (*cis*-h^2) of seven molecular phenotypes (**Supplementary Fig.** 7), and subsequently mapped mo-
249 lecular quantitative trait loci (molQTL) in the *cis*-windows across 43 tissues. These molQTL in-
250 cluded four types of abundance-based signals: eQTL for gene expression, eeQTL for exon expres-
251 sion, enQTL for enhancer expression, isoQTL for isoform expression; and three types of structure-
252 based signals: sQTL for alternative splicing, 3'aQTL for 3'UTR APA, and stQTL for RNA stability
253 (**Fig.** 3a). We detected 19,180 (91.9% of tested genes) eGenes (with at least one significant variant)
254 for gene expression, 16,358 (78.4%) eeGenes for exon expression, 7,529 (36.1%) isoGenes for iso-
255 form expression, 5,903 (28.3%) enGenes for enhancer expression, 8,454 (40.5%) sGenes for alterna-
256 tive splicing, 15,205 (72.9%) stGenes for RNA stability, and 7,851 (37.6%) 3'aGenes for 3'UTR
257 APA in 35 tissues, collectively termed molGenes (**Fig.** 3a; **Supplementary Table** 9). Of them, there
258 were 18,192 (82.9% of tested genes) genes associated with at least two types of molQTL, and 747
259 (3.4%) genes that were simultaneously associated with all seven molQTL. In addition, 2,203 (10%)
260 genes were associated with a single molQTL only, including 865 eGenes, 412 eeGenes, 298 sGenes,
261 226 isoGenes, 183 enGenes, 147 stGenes, and 72 3'aGenes. Further statistical fine-mapping analyses
262 for molGenes identified 771,007 credible sets (variants with cumulative PIP ≥ 95%), among which
263 a substantial fraction contained only a single variant, such as 37.7% (85,604) of eQTL, 41.8% (75,769)
264 of eeQTL, 29.2% (14,547) of enQTL, 35.2% (26,774) of isoQTL, 44.0% (35,710) of sQTL, 40.8%
265 (46,077) of stQTL, and 52.3% (21,679) of 3'aQTL (**Fig**. 3b**; Extended Data Fig.** 1a-b).

266 To assess the reliability of molQTL identified above, we conducted internal and external vali-
267 dations, allele-specific expression (ASE) analysis at individual level, and functional enrichment (**Ex-**
268 **tended Data Fig.** 2). Internal validation across 28 tissues with sample size over 100 yielded a high
269 replication rate (π_1, mean = 0.93) and strong effect-size consistency (mean Pearson's r = 0.89) (**Ex-**
270 **tended Data Fig**. 2a). External validation using blood eQTL in an independent population of 99
271 Holstein cattle achieved a decreased but remain robust replication rate of 0.76 and effect size con-
272 cordance of r = 0.6 (**Extended Data Fig.** 2c). At the individual level, lead eQTL effect sizes were
273 significantly correlated with ASE in whole blood (r = 0.66) and muscle (r = 0.57) (**Extended Data**
274 **Fig.** 2b). Beyond statistical validation, the identified molQTL recapitulated known regulatory logic
275 for corresponding molecular phenotypes: abundance-based signals (e.g., eQTL and isoQTL) clus-
276 tered 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.

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

314 **Pleiotropic effects of genomic variants on molecular phenotypes**

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

326 Despite the prevalence of layer-specific regulation, 40.0%, 20.7%, 6.2%, 5.1%, 16.6% and 10.7 %
327 of eGenes were colocalized (PP.H4 > 0.8) with eeGenes, isoGenes, sGenes, 3aGenes, stGenes and
328 enGenes, respectively (**Extended Data Fig.** 5c). We identified 41,425 pleiotropic eQTL (including
329 31,044 primary and 10,381 non-primary eQTL) that modulated at least one additional molecular phe-
330 notype, with some variants (2.4%) acting as master regulators for up to six molecular phenotypes
331 (**Fig**. 3g). Crucially, non-primary eQTL contributed a disproportionately higher fraction of pleiotropic
332 effects when coordinating gene expression with enhancer activity (enQTL) or structural phenotypes,
333 whereas primary eQTL dominated the regulation of abundance-based phenotypes such as exon and
334 isoform expression (**Fig**. 3g). Furthermore, the contribution of non-primary eQTL increased progres-
335 sively with the degree of pleiotropy (**Fig**. 3h). This suggest that while primary eQTL drive the ma-

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

343 **Tissue-sharing patterns of regulatory effects**

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

357 To systematically characterize tissue specificity of eQTL between any two tissues, we first iden-
358 tified three classes of tissue-specific eGenes (**Fig**. 4e) : **1) Specific-expression**, defined as genes ex-
359 pressed in only one of each of the tissue pairs (e.g., *INMT* between liver and brain, **Fig**. 4f); **2) Spe-**
360 **cific-eGenes**, defined as detected eGenes in only one tissue despite being expressed in both tissues
361 (e.g., *C21of140* between testis and muscle, **Fig**. 4g); **3) Specific-eQTL**, defined as lead variants of
362 shared eGenes were weakly linked (LD < 0.2), such as *OTOR* between muscle and testis (LD = 0.02;
363 **Fig**. 4h). We defined shared-eQTL when lead variants of same eGenes were highly correlated (LD ≥
364 0.8) (**Fig**. 4e). These shared-eQTL were further divided into three classes according to their direction
365 and magnitude of effect sizes between tissues: **1) Opposite**, eQTL exhibiting opposite effect direc-
366 tions between tissues (e.g., *GPR39* between mammary gland and adipose, **Fig**. 4i); **2) Heterogeneous**,

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

371 Among all eGenes, **Specific-eGenes** constituted the largest proportion (64%) across tissues, fol-
372 lowed by **Specific-expression** (7%)**,** and **Specific-eQTL** (2%, **Fig**. 4l). Among shared-eQTL, **Ho-**
373 **mogeneous** constituted the largest proportion across tissues (63%), followed by **Heterogeneous**
374 (33%) and **Opposite** (4%)**.** These shared-eQTL showed shorter distances to TSS, higher MAF, larger
375 effect sizes, and stronger enrichment in active regulatory elements compared with other eQTL (**Fig**.
376 4m; **Supplementary Fig**. 13). Notably, among shared-eQTL, **opposite** showed a strongest enrich-
377 ment in enhancers and flanking TTS regions (**Fig**. 4m), indicating that these variants may participate
378 in context-dependent or antagonistic regulatory mechanisms that fine-tune gene expression in a tis-
379 sue-specific manner.

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

387 **Deciphering the regulatory architecture underlying population selection signatures**

388 To explore the regulatory conservation between *B. taurus* and *B. indicus*, we conducted pop-
389 ulation-stratified eQTL mapping across eight tissues, each with at least 40 individuals for both two
390 subspecies (**Methods**; **Supplementary Fig**. 15a-b). In general, effect sizes of population-stratified
391 eQTL clustered primarily by tissue types and then by subspecies (**Fig**. 5a). The eQTL showed higher
392 replication rates (mean π_1 = 0.73) and effect-size concordance (mean Spearman's ρ = 0.76) between
393 subspecies than between tissues (**Fig**. 5a; **Extended Data Fig.** 6a). Even though they diverged at least
394 300,000 years ago, these results highlighted the conservation of regulatory effects between the two
395 subspecies[39-41].

396 To further explore the regulatory adaptation of *B. indicus* to heat, humidity and parasites chal-
397 lenges in tropical environments, we scanned the genome-wide selection sweeps between 200 *B. tau-*
398 *rus* and 142 *B. indicus* with WGS data (> 10×) via an F_{ST}-based approach[34] (**Supplementary Table**
399 11; **Methods**). The 248,679 genomic regions tested were evenly divided into 10 deciles (F_{ST}1-10)
400 based on their F_{ST} values from the smallest (no selection) to the largest (strongest selection). For
401 example, the F_{ST}10 genomic region contained *DNAJC18* (F_{ST} = 0.86), *HOXC4* (F_{ST} = 0.77), and
402 *FAAP20* (F_{ST} = 0.77), which were previously reported to be implicated in environmental adaptation
403 between cattle populations[42-44] (**Supplementary Fig**. 16a). The enrichment of fine-mapped molQTL
404 in each of 10 deciles was then tested. Notably, tissue-specific molQTL showed a significantly higher
405 enrichment in regions with larger F_{ST} values (**Fig**. 5b; **Supplementary Fig**. 16b). As expected, tissues
406 with higher enrichment of eQTL in strongly selected regions showed lower correlation of effect-size
407 between *B. taurus* and *B. indicus* in the population-stratified analysis (**Extended Data Fig.** 6b).
408 Among all tissues, molQTL specific in skin, adipose, immune and digestive tissues displayed the
409 highest enrichment in the F_{ST}10 group (**Fig**. 5c; **Supplementary Fig**. 16c). The eGenes of these tis-
410 sues located in strongly selected regions were significantly involved in immune-related functions
411 such as choline metabolism (FDR = 0.02) and bacterial invasion (FDR = 0.01; **Fig**. 5c). For example,
412 the lead eQTL (rs135226326) of *FUT2*, a skin-specific eGene, resided in a strong selection region
413 (F_{ST} = 0.54; **Fig.** 5d). The G allele of rs135226326 was associated with increased *FUT2* expression
414 and occurred at a higher frequency in *B. indicus*, potentially enhancing immune responsiveness in
415 extreme environments and thereby contributing to adaptive differentiation between *B. taurus* and *B.*
416 *indicus* due to the role of *FUT2* in mediating antiviral immune defence[45]. These results suggest that
417 regulatory variants in immune and metabolic tissues have played important roles in the divergence of
418 *B. taurus* and *B. indicus*, likely reflecting selection pressures related to environmental, metabolism,
419 and immune defence adaption[2].

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

428 between dairy and beef were brain-related tissues, particularly hypothalamus, rather than skin and
429 immune tissues (**Fig**. 5g; **Supplementary Fig**. 17b). Brain-specific eGenes located in the strongly
430 selected regions (F_{ST}10) were significantly enriched in pathways related to growth and developmental
431 regulation, including glutathione metabolism (FDR = 0.03) and the GnRH signalling pathway (FDR
432 = 0.03; **Fig**. 5g). This indicated that artificial selection between dairy and beef cattle has preferentially
433 targeted neuroendocrine and growth-regulatory mechanisms. For example, the lead eQTL
434 rs383563305 for *HTR2B*, a brain-specific eGene, resided in a region under strong selection (F_{ST} =
435 0.58; **Fig**. 5h). The allele G in rs383563305 was associated with upregulated expression of *HTR2B*
436 and showed a higher frequency in beef compared to dairy cattle (**Fig**. 5i). As *HTR2B* mediates sero-
437 tonergic regulation of hypothalamic appetite control, selection on its eQTL likely reflects divergent
438 breeding pressures on feed intake, energy allocation and growth-related traits between beef and dairy
439 cattle[46]. Altogether, these results suggest that both long-term adaptive evolution and intensive artifi-
440 cial breeding tended to target on variants with tissue-specific effects rather than those with ubiquitous
441 effects.

442 **The genetic and molecular architecture underlying bovine complex traits**

443 To investigate the regulatory mechanisms underlying complex traits in cattle, we examined
444 2,070 GWAS loci of 44 complex traits of economic importance, encompassing 22 body conformation,
445 6 milk production, 8 reproduction, and 8 health traits (**Supplementary** Table 13). We systematically
446 explored enrichment patterns of these GWAS loci against all seven types of molQTL and different
447 categories of regulatory variants based on effect size, tissue-sharing pattern, and pleiotropy. Overall,
448 eQTL showed the strongest enrichment with GWAS loci, followed by eeQTL, 3'aQTL, sQTL,
449 isoQTL, and stQTL (**Fig**. 6a). Notably, compared to primary effects, non-primary effects exhibited
450 higher enrichment in GWAS loci, consistent across all types of molQTL (**Fig**. 6a). Compared to tis-
451 sue-shared effects, tissue-specific effects were more enriched in GWAS loci (**Fig**. 6a; **Extended Data**
452 **Fig.** 8a). The eQTL with stronger pleiotropy showed higher enrichment in GWAS loci (**Fig**. 6a).
453 These observations were further supported by their similarity to GWAS loci in terms of distance to
454 TSS, population selection pressure (measured by MAF-effect size relationship), and enrichment in
455 regulatory elements (**Fig**. 6b-d; **Extended Data Fig.** 8b-c). We thereby propose a genetic regulatory
456 model in which genetic effects spanning multiple layers of the central dogma (vertical pleiotropic
457 effects) are more likely to propagate to complex traits, whereas variants acting across multiple bio-
458 logical contexts of a single biological layer (horizontal pleiotropic effects) may incur greater fitness
459 costs and are therefore less likely to translate their effects to higher-order phenotypes (**Fig**. 6e).

460 To pinpoint putative causal variants and genes governing complex traits, we performed colo-
461 calization analyses of molQTL and GWAS loci. Overall, 75% (1,545) of GWAS loci, ranging from
462 33% (5 loci) for calving-to-conception interval to 96% (96 loci) for milk protein percentage, were
463 colocalized (PP4 > 0.8) with at least one molQTL (**Supplementary Fig.** 18a). Each molQTL type
464 also made unique contributions to GWAS colocalization. Among them, eeQTL accounted for the
465 largest number of unique colocalizations (n = 64), followed by eQTL (29), stQTL (24), sQTL (21),
466 isoQTL (17), enQTL (15), and 3′aQTL (12; **Supplementary Fig.** 18b). For instance, an isoQTL of
467 *PHACTR2* in whole blood was exclusively associated with the udder cleft trait, while an sQTL of
468 *NUDT6* in the corpus luteum was exclusively associated with rump width trait (**Extended Data Fig.**
469 9). Together, these findings underscored the importance of integrating multiple molecular phenotypes
470 to elucidating the molecular basis of complex traits. Consistent with enrichment analyses, non-pri-
471 mary, tissue-specific, and pleiotropic eQTL showed higher colocalization proportions (PP4 > 0.8)
472 with GWAS loci compared to their counterparts (**Fig**. 7a). Each category also made distinct contri-
473 butions to explaining GWAS signals: non-primary eQTL contributed 57 additional colocalizations
474 beyond primary eQTL, tissue-specific eQTL contributed 79 unique colocalizations, and pleiotropic
475 eQTL accounted for 34 unique colocalizations (**Fig**. 7b). For example, a colocalization (PP4 = 0.93)
476 between rump width and *EPN3* in the liver was detected for a non-primary eQTL (rs378415691) after
477 conditioning on the primary eQTL (rs379113614). As an endocytic adaptor protein, *EPN3* regulates
478 lipid transport and signal transduction, potentially influencing growth and development in cattle[47].
479 For pleiotropic eQTL, *GIGYF1*, which is related to adverse metabolic health related to higher glyce-
480 mic levels and fat mass[48], showed high colocalization with the net merit trait through both eQTL (PP4
481 = 0.96) and stQTL (PP4 = 0.93) effects driven by the same lead variant (rs109407080). In addition,
482 the lead eQTL rs135173214 of *ENSBTAG00000050359* exhibited strong colocalization with somatic
483 cell score (SCS) specifically in the mammary gland (PP4 = 0.87) but not in other tissues (e.g., adipose
484 PP4 = 2.4×10^{-2}), highlighting tissue-specific regulatory control of this gene in mastitis-related traits
485 (**Fig**. 7c).

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

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

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

508 **Evolutionary constraint of regulatory effects between cattle and humans**

509 To explore the evolutionary constraint of regulatory effects, we aligned 198,897 fine-mapped
510 cattle eQTL to the human genome *via* LiftOver (referred to as orthologous variants) and then system-
511 atically evaluated the impact of these orthologous variants on genome function and complex traits in
512 humans **(Fig.** 8a; See **Methods)**. Compared to MAF-matched random variants (51.27%) in cattle,
513 fine-mapped eQTL exhibited a higher mapping rate to the human genome, with 53.25~74.45% across
514 different groups of eQTL, indicating cross-species conservation of regulatory variants at the DNA
515 sequence level. Primary eQTL (57.53%) were more evolutionarily constrained than non-primary ones
516 (54.34%). Notably, variants with higher pleiotropic effect were more evolutionarily constrained (**Fig.**
517 8b). We then used *AlphaGenome* to explore the functional impact of these orthologous variants in the
518 human genome[53]. Compared with MAF-matched random variants, orthologous variants of cattle
519 eQTL exhibited significantly higher predicted quantile scores, indicative of the functional conserva-
520 tion of regulatory variants between cattle and humans. Orthologous variants of tissue-shared eQTL
521 showed higher *AlphaGenome*-predicted effects than those of tissue-specific variants, and variants
522 with higher pleiotropic effect on molecular phenotypes also displayed higher *AlphaGenome*-predicted

523 effects (**Fig.** 8c). To test whether orthologous variants of cattle eQTL tended to overlap the same
524 human regulatory elements as eQTL detected in human GTEx[54], we assessed their enrichment across
525 15 human chromatin states predicted from ENCODE data[55]. Notably, human-orthologous variants of
526 cattle eQTL exhibited an enrichment profile across these chromatin states that closely resembled that
527 of human GTEx eQTL (Pearson's r = 0.809, P-value = 2.62e-4) (**Fig. 8d; Supplementary Fig.** 22a-
528 d). Notably, variants with higher pleiotropic effect exhibited a higher enrichment in CTCF and en-
529 hancers (**Supplementary Fig.** 21d).

530 To further investigate whether cattle eQTL can provide evolutionary evidence for dissecting
531 complex traits in humans, we employed S-LDSC to partition heritability of 133 human complex traits
532 using orthologous variants of cattle eQTL[56] (**Supplementary Table** 15). These orthologous variants
533 were significantly enriched for heritability across all complex traits (**Supplementary Table** 16-17).
534 Orthologous variants of tissue-specific cattle eQTL showed significantly higher enrichment than
535 those of the remaining cattle eQTL, consistent with reports from human studies[57] (**Fig.** 8e). For in-
536 stance, compared to immune-specific eQTL detected in human GTEx, orthologous variants of cattle
537 immune-specific eQTL displayed significantly higher enrichment in human immune-related traits
538 (**Fig 8e; Extended Data Fig.** 10), supporting the idea that immune-related regulatory programs
539 showed substantial conservation between cattle and humans[5,6]. Furthermore, we performed colocali-
540 zation between human GTEx eQTL and 133 human GWAS traits, yielding 508,191 SNP–gene–tissue
541 trios (PP.H4 > 0.8). Among them, 344 SNP–gene–tissue trios were tested in the CattleGTEx (**Ex-**
542 **tended Data Fig.** 10c**; Supplementary Table** 18), and 158 of them were also eQTL in cattle (FDR
543 < 0.05), which were significantly more than that of MAF-matched random variants (P-value = 0.0021)
544 (**Extended Data Fig.** 10d). Moreover, trios that were significant in the CattleGTEx exhibited stronger
545 colocalization (higher PP.H4 values) with human complex traits than non-significant trios (P-value
546 = 4.36×10^{-13}; **Fig. 8f**) and were associated with more complex traits (**Fig. 8g**, P-value = $2.22\times10^{-}$
547 ^{2}). For instance, a variant (rs11559147 in humans and rs210096698 in cattle) showed conserved reg-
548 ulatory effects on *ZDHHC4* expression in whole blood in both humans and cattle, and it was signifi-
549 cantly colocalized with a GWAS locus of neutrophil count in humans (PP.H4 = 0.99; **Fig. 8h**). The
550 *ZDHHC* family encodes palmitoyltransferases that catalyze reversible protein S-palmitoylation[58],
551 thereby modulating membrane localization and signalling complex assembly of immune regulators
552 and fine-tuning key immune signalling cascades[59,60]. Altogether, these results suggest the presence of
553 evolutionary constraint regulatory variants between humans and cattle, which can facilitate the prior-
554 itization 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

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

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

618 gene regulatory networks[69-72], will be essential to capture downstream and complex regulatory mech-
619 anisms that cannot be resolved at the transcriptomic level alone. **4**) We only conducted molQTL
620 mapping at bulk-tissue resolution and therefore did not fully capture regulatory heterogeneity across
621 cell types and states[73-75]. Although we explored cell-type-dependent regulation using computational
622 deconvolution approaches in a companion study[21], population-scale single-cell molQTL mapping[73,76],
623 together with spatially resolved transcriptomic data[77,78], will be required to dissect regulatory mech-
624 anisms operating at finer cellular and spatiotemporal scales. **5**) This study considered a limited set of
625 biological contexts, primarily focusing on different tissue types and breeds. Future development will
626 incorporate additional contexts, including sex, developmental stages, and diverse environmental con-
627 ditions[79-81], which will require the establishment of a FarmGTEx tissue biobank with standardized
628 sampling protocols, controlled environmental conditions, and comprehensive metadata collection, to
629 characterize context-dependent regulatory mechanisms in cattle and further advance our understand-
630 ing of the genetic architecture of complex traits. **6**) The regulatory variants prioritized through com-
631 putational analyses in this study remain largely inferential. Experimental validation, including *in vitro*
632 assays and *in vivo* functional studies[82,83], will be crucial to confirm causal regulatory effects and to
633 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.

654 **Online Methods**

655 **Ethics and sample collection**

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

661 **Bulk RNA-seq data processing**

662 We downloaded 24,975 RNA-seq datasets from the Sequence Read Archive (SRA,
663 https://www.ncbi.nlm.nih.gov/sra) and 1,385 datasets from the Genome Sequence Archive (GSA,
664 https://ngdc.cncb.ac.cn/gsa) using the Aspera command-line tool (**Supplementary Table** 1-2). In ad-
665 dition, we newly generated 1,268 RNA-seq datasets from 3 tissues of 533 Holstein cattle, including
666 514 samples in lung, 459 sample in rumen and 295 samples in lymph node (See **Supplementary**
667 **Notes**). After raw RNA sequencing data were obtained, the adapters were removed along with low
668 quality reads using Fastp (v0.23.4) with parameters: --trim_front 3, --trim_tail 3, --length_required
669 36, --cut_right, --cut_window_size 4 and --cut_mean_quality 15[84]. Clean reads were then aligned to
670 the B. taurus reference genome (ARS-UCD1.2, Ensembl v108[85]) using STAR (v 2.7.10b) with pa-
671 rameters: --chimSegmentMin 10, --twopassMode Basic, --outFilterMismatchNmax 10, --
672 chimOutJunctionFormat 1, --outFilterType BySJout, --alignSJoverhangMin 8 and -- alignSJDBover-
673 hangMin 1[86]. In total, 19,790 samples with >10M mapped reads, >40% uniquely mapped reads,
674 and >40% expressed genes were retained for downstream analyses. The detailed mapping quality of
675 all the RNA-seq samples is summarized in **Supplementary Table** 4.

676 **Genotype calling and imputation of RNA-seq samples**

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

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

694 **Quantification of molecular phenotypes**

695 After mapping bias removal, the corrected bam file was used for molecular phenotype definition and
696 quantification. Normalized expression (TPM) of 23,103 protein-coding and lncRNA genes (Bos_tau-
697 rus. ARS-UCD1.2, Ensembl v108) were generated using Stringtie (v.2.2.1)[91]. Raw read counts for all
698 20,875 genes and 102,033 exons were obtained with featureCounts (v2.0.3)[92]. Transcribed enhancer
699 regions were annotated, and a custom SAF file was prepared after excluding those overlapping with
700 gene bodies[93], then the raw counts were also quantified using featureCounts (v.2.0.3)[92]. For transcript-
701 level expression, Salmon (v1.10.1)[94] was used in mapping-based mode with clean reads as input to
702 quantify transcript expression, generating both TPM and raw count outputs. The index for Salmon
703 was built with the decoy sequences generated using SalmonTools[94]. For alternative splicing, Leafcut-
704 ter (v.0.2.9)[95] was used to quantify the splicing events as following: 1) mapping bias–corrected BAM
705 files from the previous step were converted into junction files using the bam2junc.sh script provided
706 by Leafcutter; (2) intron clusters were identified with leafcutter_cluster.py and annotated to genes
707 using map_clusters_to_genes.R; (3) introns with zero read counts in more than 50% of samples or
708 with fewer than max(10 or 0.1n) unique values in each tissue (where n is the sample size) were re-
709 moved; (4) low-complexity introns were filtered out based on the z score of cluster read fractions
710 across individuals; (5) intron excision ratios were calculated, and the filtered counts were converted
711 to BED format using prepare_phenotype_table.py. Finally, excision levels of introns were normalized
712 as percent spliced-in (PSI) values. For the quantification of 3'UTR alternative polyadenylation
713 (3'UTR APA), Dapars (v2.1) was used with the following steps[96]: 1) distal polyadenylation sites from
714 the Ensembl annotation (v108) were extracted using the script DaPars_Extract_Anno.py provided by
715 Dapars (v2.1) package, 2) genomecov function in BEDTools (v2.31.0) was utilized to obtain wiggle

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

728 **molQTL mapping**

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

744 **Conditional and fine-mapping analysis of molQTL**

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

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

754 **Colocalization analysis between molQTL**

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

766 **Tissue-sharing/specificity analysis of molQTL**

767 We used multivariate adaptive shrinkage (MashR, v0.2.57)[108] to characterize cross-tissue sharing pat-
768 terns of molQTL. For the overall estimation, we extracted the z scores (slope / slope_se) of the top
769 molQTL for each molGene across all tissues in which it was detected and used these top-effect
770 molQTL–gene pairs as input to MashR. In addition, one million molQTL-gene pairs were randomly
771 sampled from all nominal associations tested by OmiGA across tissues[37], and their corresponding z-
772 scores were obtained. Pairs with missing values in more than five tissues were removed to avoid
773 inflating tissue-specific signals driven by sparse observations rather than true biological effects; for
774 the remaining pairs, missing z scores were assigned a standard error of 1×10^{6}. MashR was then ap-
775 plied to estimate the local false sign rate (LFSR) for each tissue, and molQTL with LFSR < 0.05 were
776 considered active in that tissue. Tissue similarity was assessed by computing pairwise Spearman cor-
777 relations 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[55]. 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[110], 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)[111], 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[112]. Significant enrichment results (FDR < 0.05) were used for statistical plot.

cis-molQTL-GWAS colocalization

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

817 **Function enrichment analysis of gene lists**

818 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were per-
819 formed using the clusterProfiler (v.4.0)[113]. The GO terms and KEGG pathways of selected genes were
820 enriched in the org.Bt.eg.db and KEGG-bta databases, using the erichGO and enrichKEGG functions,
821 respectively, with a threshold parameter of "pvalueCutoff = 0.05". We considered a GO term or a
822 KEGG pathway significant if it had FDR < 0.05.

823 **Population-stratified molQTL analysis**

824 To investigate the difference in *cis*-eQTL between breeds, global ancestry of all samples within each
825 tissue was estimated by ADMIXTURE (v1.3.0) with K=2[114]. To ensure clear genetic assignments of
826 samples to four breed groups (i.e., *B. taurus, B. indicus,* dairy, and beef cattle), we matched the pop-
827 ulation labels with ancestral lineage composition, and considered the label correct if the lineage purity
828 was greater than 80%. Subsequently, other samples with ancestral purity greater than 80% (even if
829 the original label was not) were defined as belonging to this breed. We then conducted *cis*-eQTL
830 mapping separately in *B. taurus*, *B. indicus*, dairy and beef cattle populations using the same proce-
831 dures as described in **molQTL mapping** above**.** Replications of *cis*-eQTL between populations
832 within a tissue and tissues within a population were quantified using Storey's π_1 statistic and Pear-
833 son's correlation of normalized eQTL effect sizes and compared to replication between tissues within
834 a population and populations within a tissue[22]. To further characterize sharing patterns across popu-
835 lations, we conducted a meta-analysis of *cis*-eQTL across tissues and population groups using the
836 MashR package (v0.2.57)[108], following the same pipeline used for **Tissue-sharing/specificity anal-**
837 **ysis of molQTL**.

838 **Selective sweep and molQTL enrichment analysis**

839 To detect selection signatures between *B. taurus* and *B. indicus,* we selected 200 *B. taurus* (50 Hol-
840 stein, 50 Jersey, 50 Simental and 50 Angus) and 142 *B. indicus* (20 Leiqiong, 20 Butana, 11 Horro,
841 11 Boran, 10 Arsi, 10 Goffa, 10 Mursi, 9 Barka, 9 Fogera, 9 Afar, 9 Sheko, 7 Baoule and 7 Nelore)
842 with WGS data (>10x) from the genotype imputation panel[31] (**Supplementary Table 3**) with
843 28,958,211 SNPs. To detect selection signatures between beef and dairy cattle, we selected 651 dairy
844 (Holstein) and 552 beef (326 Simmental, 64 Angus, 60 Charolais, 58 Limousin and 44 Hereford)
845 cattle from the genotype imputation panel[31] (**Supplementary Table 4**) with 28,958,211 SNPs. We
846 analyzed genetic differentiation (F_{ST}) using the vcftools (v0.1.16) tool[115], employing a sliding win-
847 dow of 30 Kb with a step size of 10 Kb. The windows were divided into deciles based on their F_{ST}
848 values, forming ten groups ($F_{ST}1$–$F_{ST}10$) with increasing selection intensity. Enrichment analysis was
849 then performed between each group and the molQTL from credible sets derived through fine mapping.
850 Odds ratio (OR) was calculated to evaluate enrichment levels between selected regions and molQTL
851 with the following formula[21]:

852 $$OR = \frac{a/b}{c/d}$$

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

857 **Cross-species conservation analysis between cattle and humans**

858 To evaluate cross-species conservation between cattle and humans, we mapped fine-mapped cattle
859 eQTL to the human reference genome (hg38) using the UCSC liftOver tool[116]. SNPs that were suc-
860 cessfully converted were considered as orthologous (sequence-level conserved) variants between cat-
861 tle and humans. To assess conservation at the functional level, we tested whether enrichment of eQTL
862 within regulatory elements was preserved between species. Specifically, we performed: (i) enrich-
863 ment analysis of cattle eQTL in cattle regulatory elements[55]; (ii) enrichment analysis of liftOvered
864 cattle eQTL in human regulatory elements[117]; and (iii) enrichment analysis of human eQTL from
865 GTEx v8 in human regulatory elements. We then quantified concordance in enrichment patterns
866 across the three analyses using Pearson's correlation to assess cross-species functional conservation.

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

871 We next collected GWAS summary statistics for 133 human complex traits (**Supplementary Table**
872 **7**) from the Pan-UK Biobank[118] and applied LD score regression (*LDSC* v1.0.1)[56] to estimate the
873 proportion of SNP-heritability attributable to each predefined eQTL group. We conducted
874 colocalization analysis between human GTEx v8 eQTL and GWAS of human complex traits using
875 coloc v5.2.3[91], and considered a posterior probability threshold of PP.H4 > 0.8 as high-confidence
876 shared signals. To ensure biologically interpretable cross-species comparisons, we retained only
877 tissue-gene-SNP trios that satisfied all of the following criteria: (i) genes have identifiable human–
878 cattle orthologs; (ii) the tissue is matched between humans and cattle; and (iii) the orthologous locus
879 is segregating in both the human GTEx and cattle GTEx datasets. As a control, we randomly selected
880 an equal number of trios. We then tested whether SNPs from the retained trios were also associated
881 with gene expression in CattleGTEx using a single-SNP linear mixed model, in which the genomic
882 relationship matrix (GRM) was included as a random effect to account for population structure and
883 relatedness; this analysis was performed using GCTA (v1.93.2)[111]. Fixed-effect covariates were kept
884 identical to those used in the cattle eQTL mapping. SNPs with FDR < 0.05 were considered
885 significant.

886 **Statistics and reproducibility**

887 No statistical method was used to predetermine the sample size. The details of data exclusions for
888 each specific analysis are available in the Methods section. For all the boxplots, the horizontal lines
889 inside the boxes show the medians. Box bounds show the lower quartile (Q1, the 25th percentile) and
890 the upper quartile (Q3, the 75th percentile). Whiskers are minima (Q1 − 1.5× IQR) and maxima (Q3
891 + 1.5× IQR), where IQR is the interquartile range (Q3–Q1). Outliers are shown in the boxplots unless
892 otherwise stated. The experiments were not randomized, as all the datasets are publicly available from
893 observational studies. The investigators were not blinded to allocation during experiments and
894 outcome assessment, as the data were not from controlled randomized studies.

895 **Data availability**

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

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

903 **Code availability**

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

909 **Acknowledgments**

910 L. F. was supported by Agriculture and Food Research Initiative Competitive grants nos. 2022-
911 67015-36215 from the USDA National Institute of Food and Agriculture, seed-funding from
912 CellFood Hub (AUFF), Danmarks Frie Forskningsfond nos. 5254-00069B. B. L. acknowledges
913 funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC) with grant
914 BB/X009505/1.

915

916

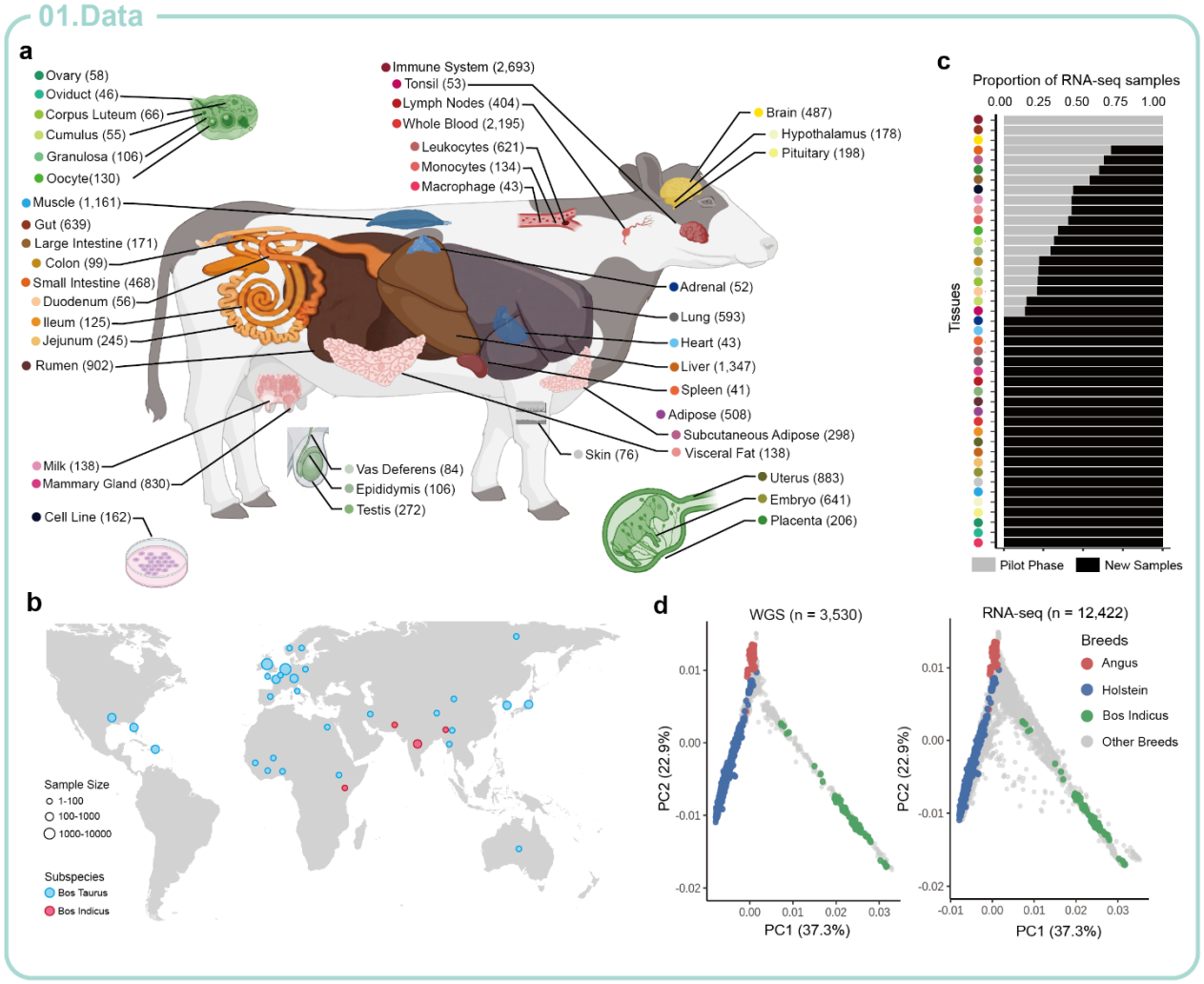
1. Frantz, L.A.F., Bradley, D.G., Larson, G. & Orlando, L. Animal domestication in the era of ancient genomics. *Nature Reviews Genetics* **21**, 449-460 (2020).
2. Utsunomiya, Y.T. *et al.* Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics* **50**, 557-568 (2019).
3. 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 Genetics* **57**, 2578-2588 (2025).

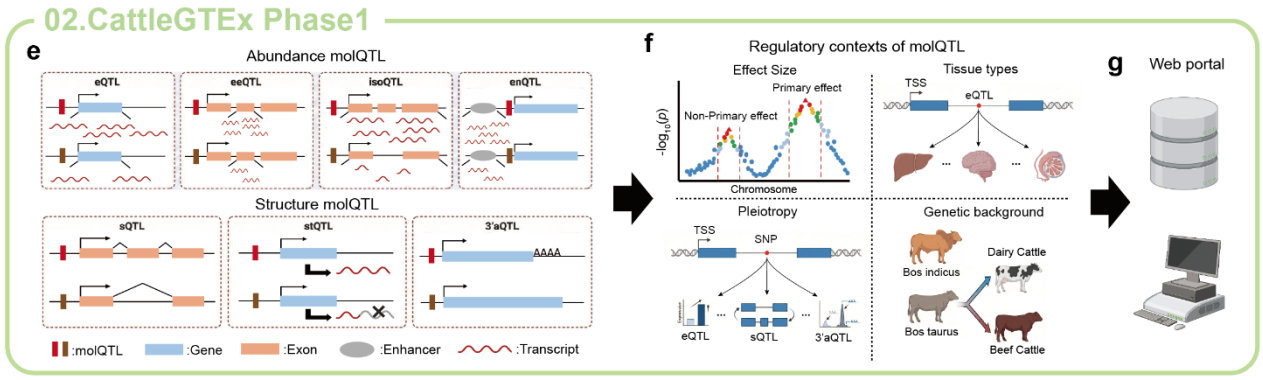
30. Lee, S. & Mostafavi, H. Missing regulatory effects on complex traits: Contribution of distal variants. *Cell Genomics* **5**, 100809 (2025).
31. Yao, D.W., O'Connor, L.J., Price, A.L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics* **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. 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).
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. 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).
111. 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).
112. 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).
113. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
114. 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).

115. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
116. Perez, G. *et al*. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53**, D1243-d1249 (2025).
117. 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).
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

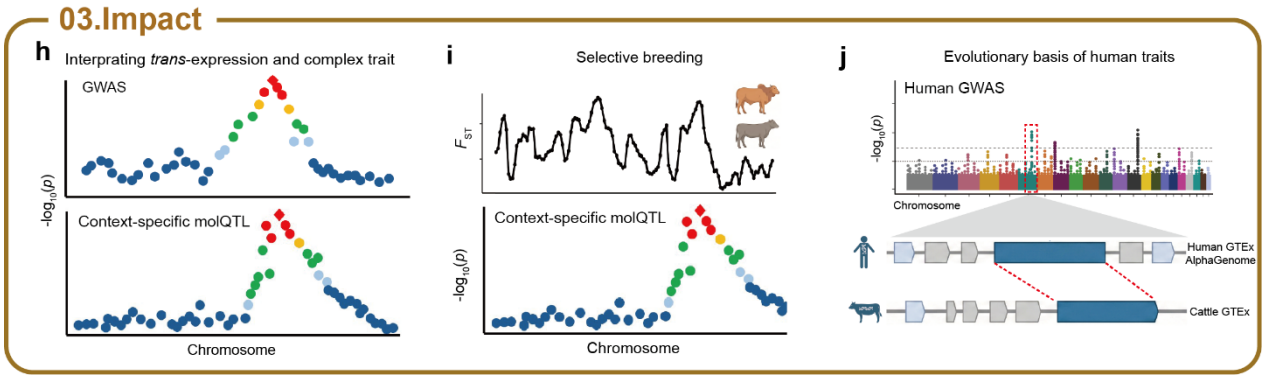
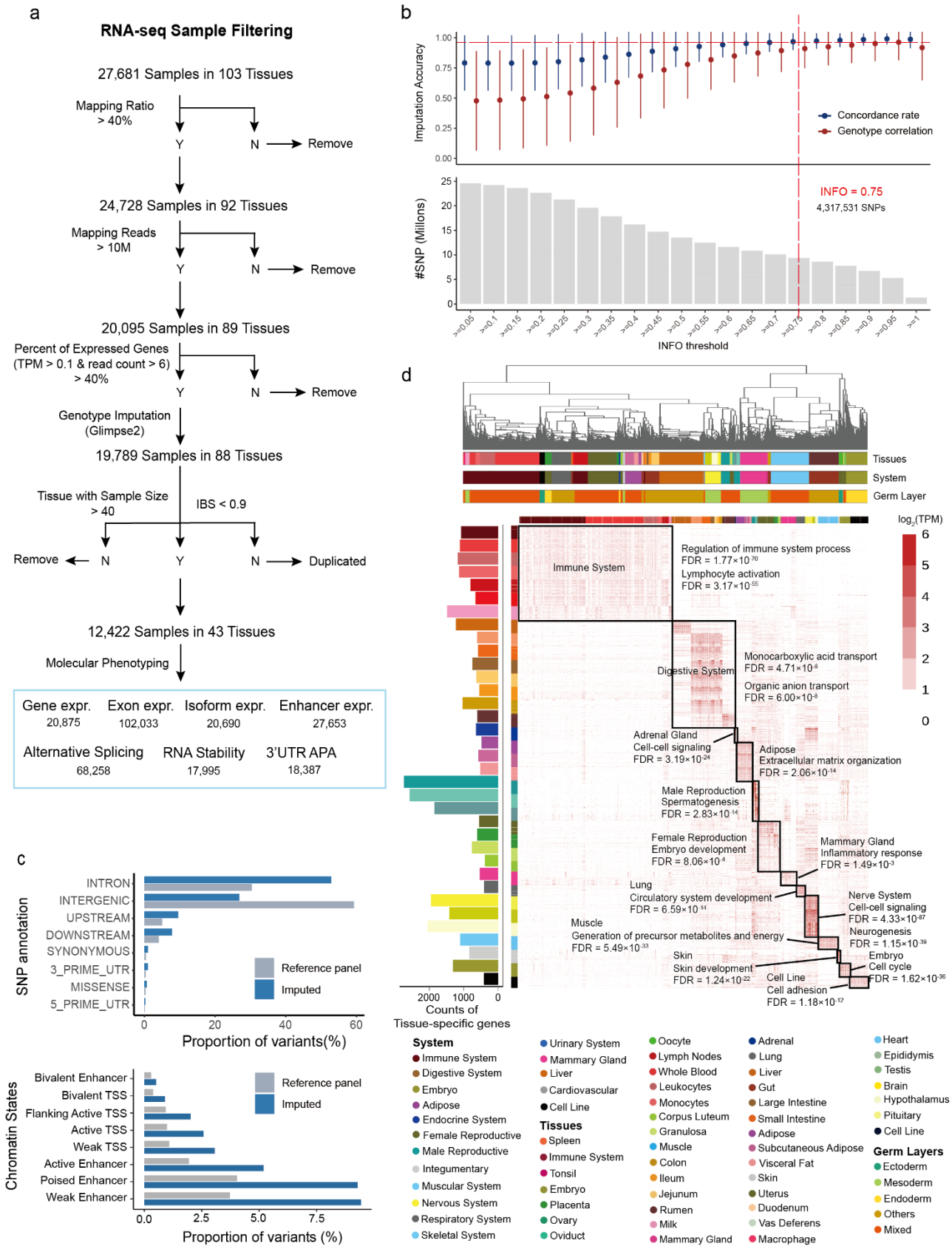
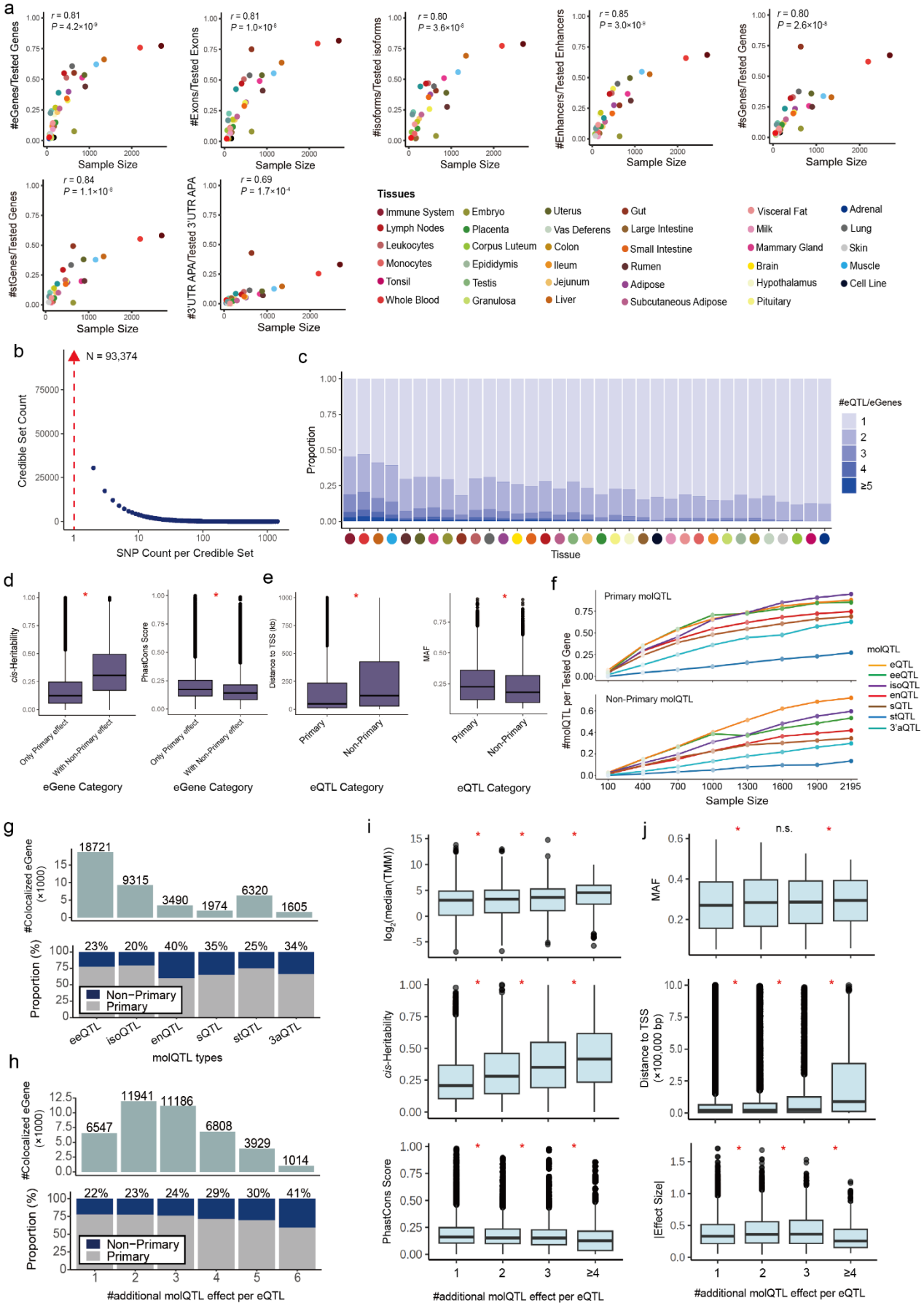

03.Impact
h
Interpreting trans-expression and complex trait
GWAS
Context-specific molQTL
Chromosome
i
Selective breeding
Context-specific molQTL
Chromosome
j
Evolutionary basis of human traits
Human GWAS
Chromosome
Human GTEx
AlphaGenome
Cattle GTEx

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-g**, 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/. **h-j**, Applications of the CattleGTEx Phase1 resource. Analyses include 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
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

1172

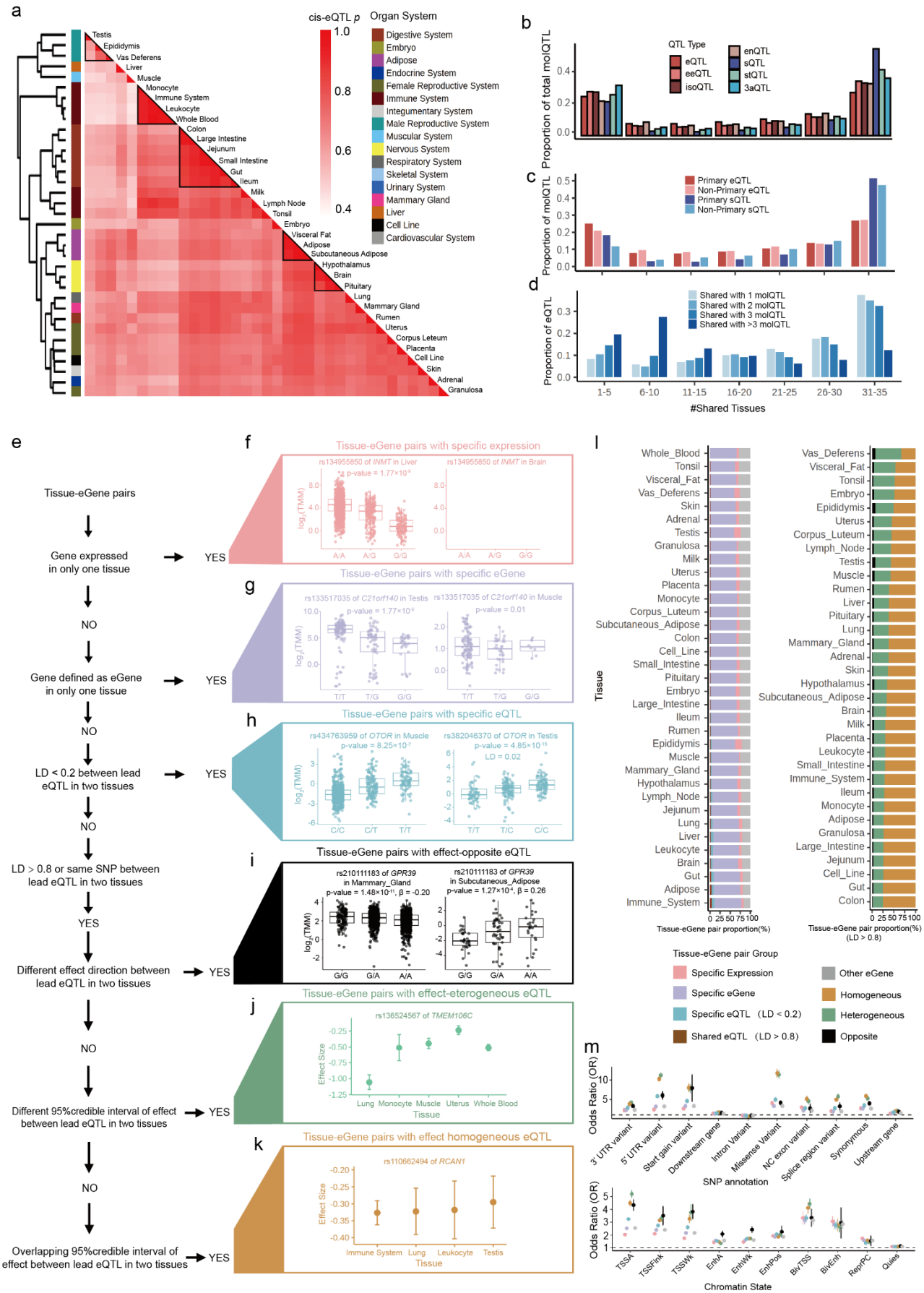
1173 **Fig. 2 | Genotype imputation and molecular phenotyping. a**, Workflow for sample quality control
1174 (QC), duplicate removal, and molecular phenotyping. Y: Sample passed the filtering, N: Sample
1175 didn't pass the filtering, TPM: Transcript per million, IBS: Identity-by-state, expr.: expression, APA:
1176 Alternative polyadenylation. **b**, Imputation accuracy assessed by concordance rate and genotype cor-
1177 relation using an external dataset, which include 15,952 samples with paired RNA-seq and whole-
1178 genome sequencing data (WGS, 30×). The top panel shows the distribution of concordance rates and
1179 genotype correlations across various INFO score cutoffs (reported by Glimpse2); the bottom panel
1180 displays the number of SNPs retained at each cutoff. **c**, Distribution of 4,317,531 imputed SNPs (from
1181 RNA-seq) and 28,958,211 reference panel SNPs across eight genomic features and eight chromatin
1182 states. TSS: Transcription start site. **d**, **Top:** Hierarchical clustering of 43 tissues based on tissue-
1183 specific gene expression, annotated by tissue type, organ system, and germ layer origin. **Bottom:**
1184 Number of tissue-specific genes (left), corresponding expression heatmap across the 43 tissues (cen-
1185 tre), 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

1187 **Fig. 3 | Regulatory architecture of seven molecular phenotypes across 35 tissues.**

1188 **a,** Pearson's correlation (*r*) between the proportion of molGenes (i.e., genes with significant molecu-
1189 lar quantitative trait loci (molQTL) / all tested genes) and sample size across 35 tissues. *P* values were
1190 calculated using a two-sided Student's *t*-test. eQTL, expression QTL; eeQTL, exon expression QTL;
1191 isoQTL, isoform expression QTL; enQTL, enhancer expression QTL; sQTL, splicing QTL; stQTL,
1192 QTL for RNA stability of genes; 3'aQTL, QTL for 3' UTR alternative polyadenylation (APA). **b,**
1193 The size distribution of fine-mapped credible sets for all identified *cis*-eQTL across 35 tissues. The
1194 red arrow indicates 93,374 credible sets containing only one single candidate variant. **c,** The distribu-
1195 tion of the number of conditionally independent *cis*-eQTL per gene is shown as blue stacked bars
1196 across 28 tissues. Tissues are ordered by decreasing sample size. **d,** Characteristics of genes with and
1197 without non-primary *cis*-eQTL. (Left) cis-heritability of mRNA and lncRNA expression; and (Right)
1198 sequence constraint (PhastCons score). Statistical significance was determined using a two-sided Stu-
1199 dent's t-test. *: $P < 0.05$. **e,** Characteristics of primary and non-primary *cis*-eQTL. (Left) distance to
1200 the target gene; and (Right) Minor Allele Frequency (MAF). Statistical significance was determined
1201 using a two-sided Student's t-test. *: $P < 0.05$. **f,** Numbers of detected primary eQTL and non-primary
1202 eQTL per tested gene across different sample size in whole blood. **g,** eQTL colocalized with other
1203 molQTL (top) and the proportion of primary versus non-primary eQTL among them (bottom). Exact
1204 numbers and proportions are shown above each bar. **h,** Pleiotropic eQTL regulating varying numbers
1205 of molecular phenotypes (top) and the proportion of primary versus non-primary eQTL among them
1206 (bottom). Exact numbers and proportions are shown above each bar. **i,** Characterization of genes
1207 harboring pleiotropic cis-eQTL. The box plots show the relationship between the number of addi-
1208 tional molQTL regulated per lead cis-eQTL (X-axis, excluding the eQTL itself) and three properties:
1209 (Top) mRNA and lncRNA expression level, measured as $\log_2$(median TMM) (bp); (Middle) cis-her-
1210 itability of mRNA and lncRNA expression; and (Down) sequence constraint, quantified by PhastCons
1211 scores. In each box plot, the median is shown as the central line, the 25th – 75th percentiles define
1212 the box limits, and whiskers extend 1.5×the interquartile range. Statistical significance was deter-
1213 mined using a two-sided Student's t-test. *: $P < 0.05$; n.s.: not significant. **j,** Characterization of plei-
1214 otropic cis-eQTL. The box plots show the relationship between the number of additional molQTL
1215 regulated per lead cis-eQTL (X-axis, excluding the eQTL itself) and three properties: (Top) Minor
1216 Allele Frequency (MAF); (Middle) distance to the target gene (bp); and (Down) absolute effect size
1217 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
Organ 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
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
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
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

1221 **Fig. 4 | Tissue-specific regulatory effects.**

1222 **a,** Heatmap showing pairwise Spearman's correlation (ρ) of *cis*-eQTL effect sizes across 51 tissues.
1223 Tissues were hierarchically clustered using the complete linkage method based on the maximum dis-
1224 tance of ρ. **b,** Proportion of molQTL active across tissues, measured by the number of tissues with
1225 a *mashr* local false sign rate (LFSR, equivalent to FDR) < 0.05. **c,** Proportion of primary and non-
1226 primary *cis*-eQTL and *cis*-sQTL active across different numbers of tissues. **d,** Proportion of *cis*-eQTL
1227 shared with different numbers of other molQTL active across different numbers of tissues. **e,** Pipeline
1228 for classifying different tissue-eGene pairs in each tissue. LD means linkage disequilibrium. **f,** Ex-
1229 ample of expression-specific tissue-eGene pairs, including gene name, lead eQTL, p-value and three
1230 genotypes in gene-expressed tissue (left) and non-expressed tissue (right). The two-sided P value of
1231 SNPs was calculated by the linear mixed regression model. **g,** Example of eGene-specific tissue-
1232 eGene pairs, including gene name, lead eQTL, p-value and three genotypes in eGene-defined tissue
1233 (left) and non-eGene tissue (right). The two-sided P value of SNPs was calculated by the linear mixed
1234 regression model. **h,** Example of eQTL-specific tissue-eGene pairs, including gene name, lead eQTL,
1235 p-value, three genotypes and LD between lead eQTL from two tissues. The two-sided P value of
1236 SNPs was calculated by the linear mixed regression model. **i,** Example of effect-opposite tissue-
1237 eGene pairs, including gene name, lead eQTL, p-value, three genotypes and effect sizes from lead
1238 eQTL in two tissues. The two-sided P value of SNPs was calculated by the linear mixed regression
1239 model. **j,** Example of effect-heterogeneous tissue-eGene pairs. The title represents gene name and
1240 lead eQTL. Y axis represents the 95% credible interval of effect size in lead eQTL. The first tissue in
1241 X axis represents the tested tissue and others represent tissues with non-overlapped interval with the
1242 tested tissue. **k,** Example of effect-homogeneous tissue-eGene pairs, The title represents gene name
1243 and lead eQTL. Y axis represents the 95% credible interval of effect size in lead eQTL. The first
1244 tissue in X axis represents the tested tissue and others represent tissues with overlapped interval with
1245 the tested tissue. **l,** Proportion of each tissue-eGene classification. Bar plot in the left represents pro-
1246 portion of expression-specific, eGene-specific, eQTL-specific (LD < 0.2), eQTL-shared (LD > 0.8)
1247 and other tissue-eGene pairs, whereas the right one represents proportion of eQTL-shared (LD > 0.8)
1248 tissue-eGene pairs, including effect-opposite, effect-heterogeneous and effect-homogeneous. **m,**
1249 Functional enrichment between each group of tissue-eGene and regulatory elements (top: SNP clas-
1250 sification, bottom: chromatin states).

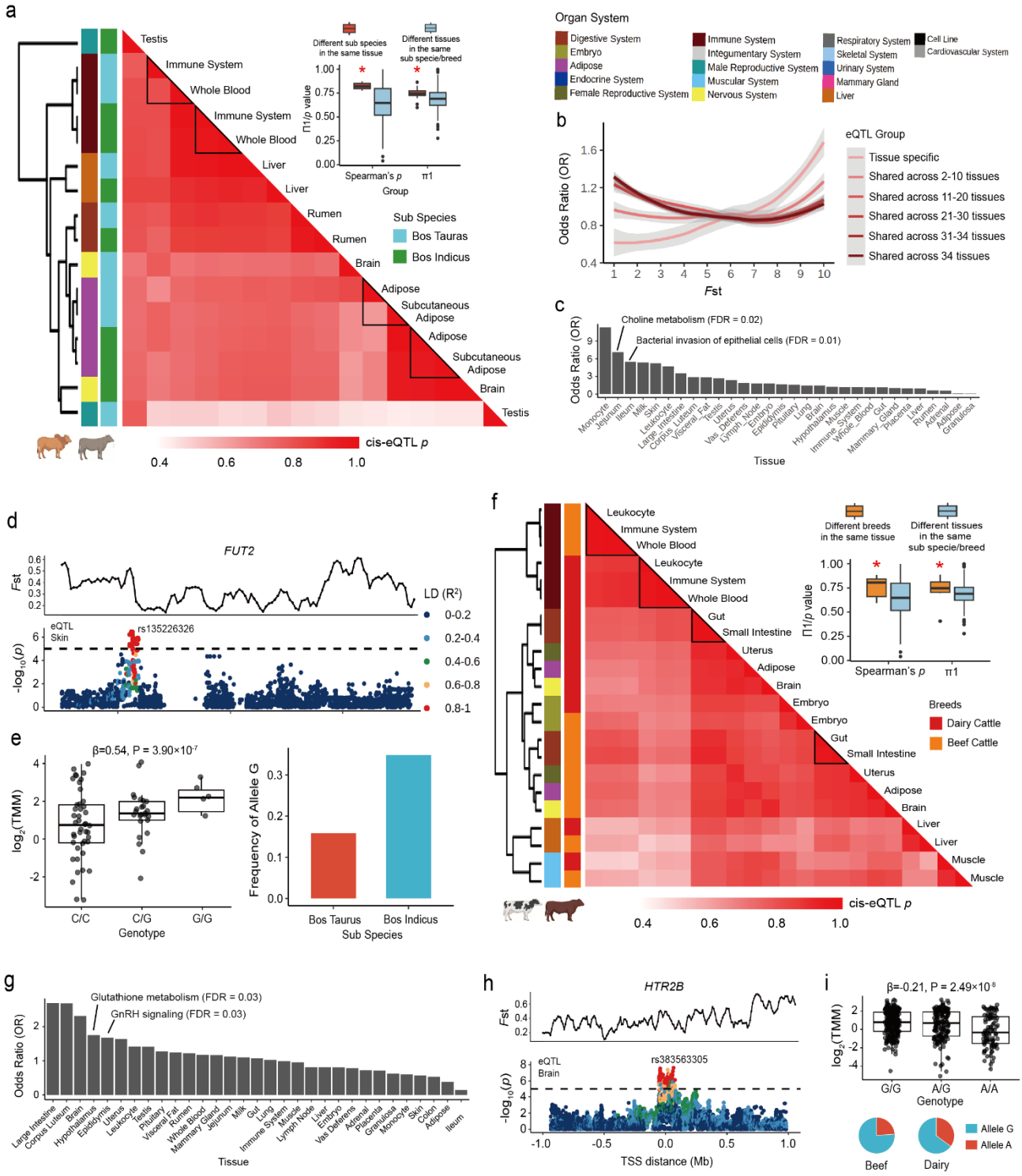
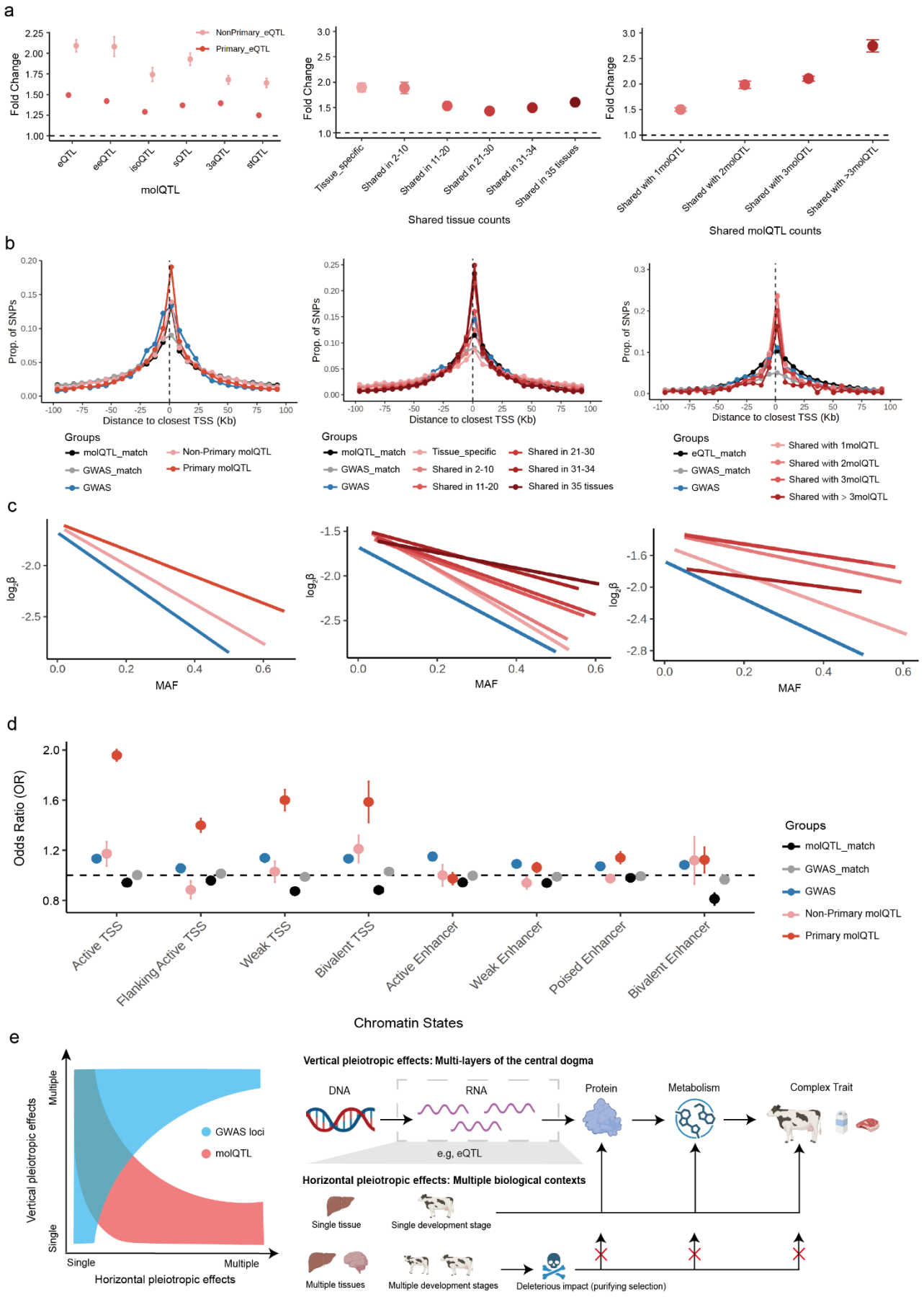


Fig. 5 | Liking molQTL to selection signatures between subspecies and breeds.

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

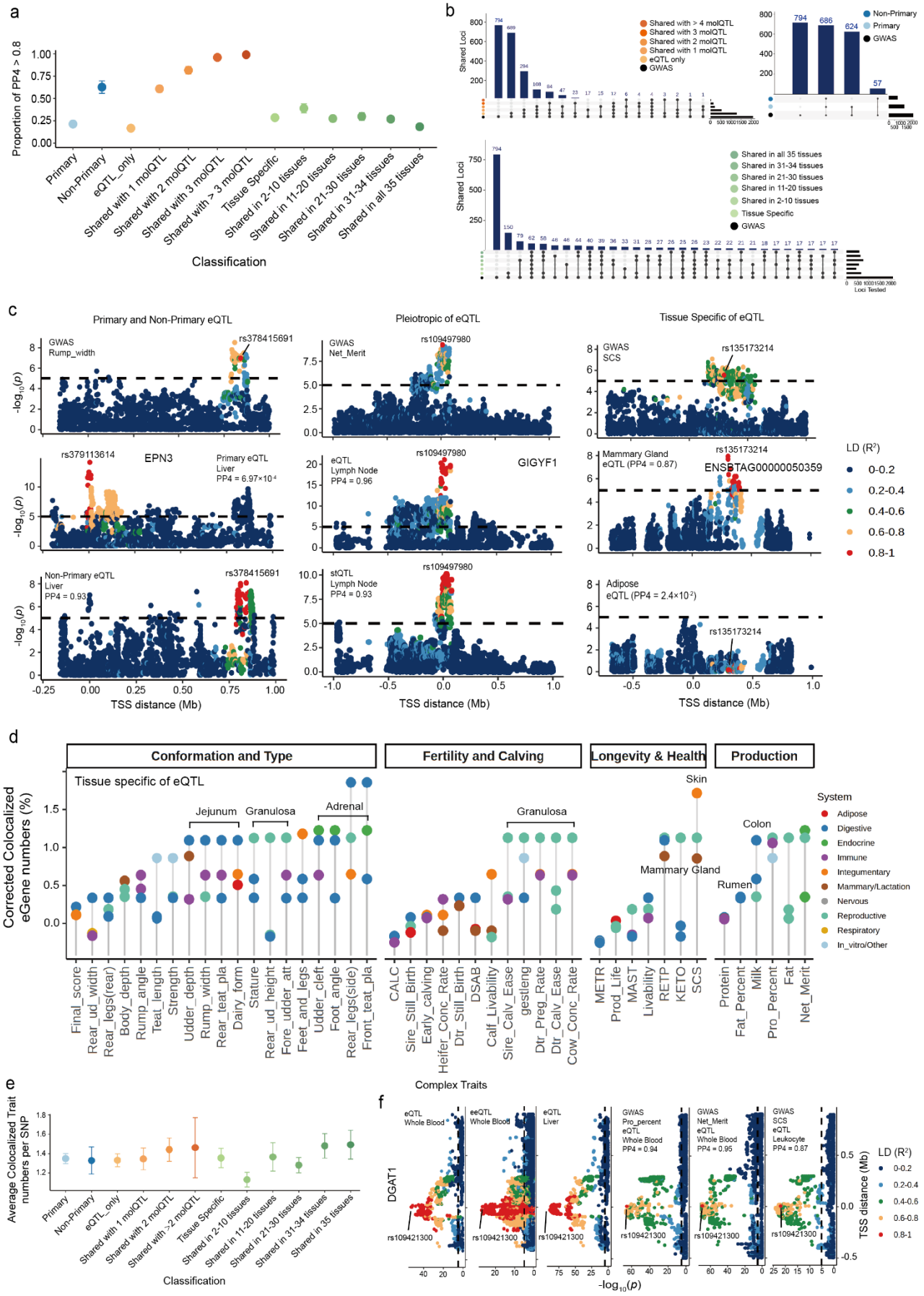
p-value between two groups were determined using a two-sided Student's t-test. *: P < 0.05. **b,** Enrichment (odds ratio) of shared and tissue-specific molQTL across 10 F_{ST} deciles. The lines were fitted by using the geom_smooth function from ggplot2 (v3.3.2) in R (v4.0.2). **c,** Enrichment (odds 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)

1279

1280 **Fig. 6 | Enrichment between three classes of molQTL and GWAS loci from 44 complex traits.**

1281 **a,** Enrichment (odds ratio) of three classes of molQTL with genome-wide associations (GWAS) of
1282 44 cattle traits using QTLenrich (v.2.0), including primary and non-primary molQTL (**left**), tissue-
1283 sharing eQTL (**mid**) and pleiotropic eQTL (**right**). **b**, Distance to nearest TSS regions of three classes
1284 of molQTL, GWAS loci, MAF-matched GWAS loci and MAF-matched molQTL, including primary
1285 and non-primary molQTL (**left**), tissue-sharing molQTL (**mid**) and pleiotropic eQTL (**right**). **c,** Se-
1286 lection pressure (measured by MAF-effect size relationship) of three classes of molQTL and GWAS
1287 loci, including primary and non-primary molQTL (**left**), tissue-sharing molQTL (**mid**) and plei-
1288 otropic eQTL (**right**). **d,** Functional enrichment between primary/non-priamry molQTL and regula-
1289 tory elements. **e,** A genetic regulatory model of complex traits. The left plot represents the distribution
1290 of molQTL (red) and GWAS loci (blue) across different levels of biological contexts (x-axis) and
1291 layers of the central dogma (y-axis). The right plot represents the process of how regulatory genetic
1292 variants under multiple layers of the central dogma and specific biological contexts shapes GWAS
1293 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)

1295 **Fig. 7 | Colocalization between three classes of molQTL and GWAS loci from 44 complex traits.**

1296 **a,** Proportion of colocalized molQTL (PP.H4 > 0.8) in three classes molQTL, including primary and
1297 non-primary molQT, tissue-shared molQTL and pleiotropic eQTL. **b**, Numbers of colocalized GWAS
1298 loci by each class of molQTL, including primary and non-primary molQTL (**right and top**), tissue-
1299 shared molQTL (**bottom**) and pleiotropic eQTL (**left and top**). Each blue bar represents a combina-
1300 tion of different molQTL in each class. Each black bar (right) represents the number of GWAS loci
1301 colocalized in the respective combination. **c,** Manhattan plots showing the -$\log_{10}$(p) distribution of
1302 specific associated signals between GWAS loci and three classes of molQTL, including rump width
1303 (GWAS) and non-primary eQTL in liver across genetic variants within the locus of *EPN3*, net merit
1304 (GWAS) and pleiotropic eQTL (with stQTL) in lymph nodes across genetic variants within the locus
1305 of *GIGYF1*, and SCS (GWAS) and tissue-specific eQTL in mammary gland across genetic variants
1306 within the locus of *ENSBTAG00000050359*. Variants are colored according to their linkage disequi-
1307 librium (LD, r^2) with the lead SNP. The two-sided P values of SNPs were calculated by the linear
1308 mixed regression model. **d**, Corrected colocalized eGene counts (PP.H4 > 0.8) between each tissue
1309 and each complex trait in eQTL results. Colocalized eGene counts were corrected by detected eQTL
1310 counts. **e,** Average colocalized trait counts per SNP (PP.H4 > 0.8) in three classes of molQTL, in-
1311 cluding primary and non-primary molQT, tissue-shared molQTL and pleiotropic eQTL **f,** Manhattan
1312 plots showing the -$\log_{10}$(p) distribution of the association signal between pleiotropic GWAS loci and
1313 molQTL for protein percent, SCS and net merit (GWAS) with eQTL and eeQTL in whole blood and
1314 liver across genetic variants within the locus of *DGAT1*. Variants are colored according to their link-
1315 age disequilibrium (LD, r^2) with the lead SNP rs109421300. The two-sided P values of lead SNPs
1316 were calculated by the linear mixed regression model.

1317

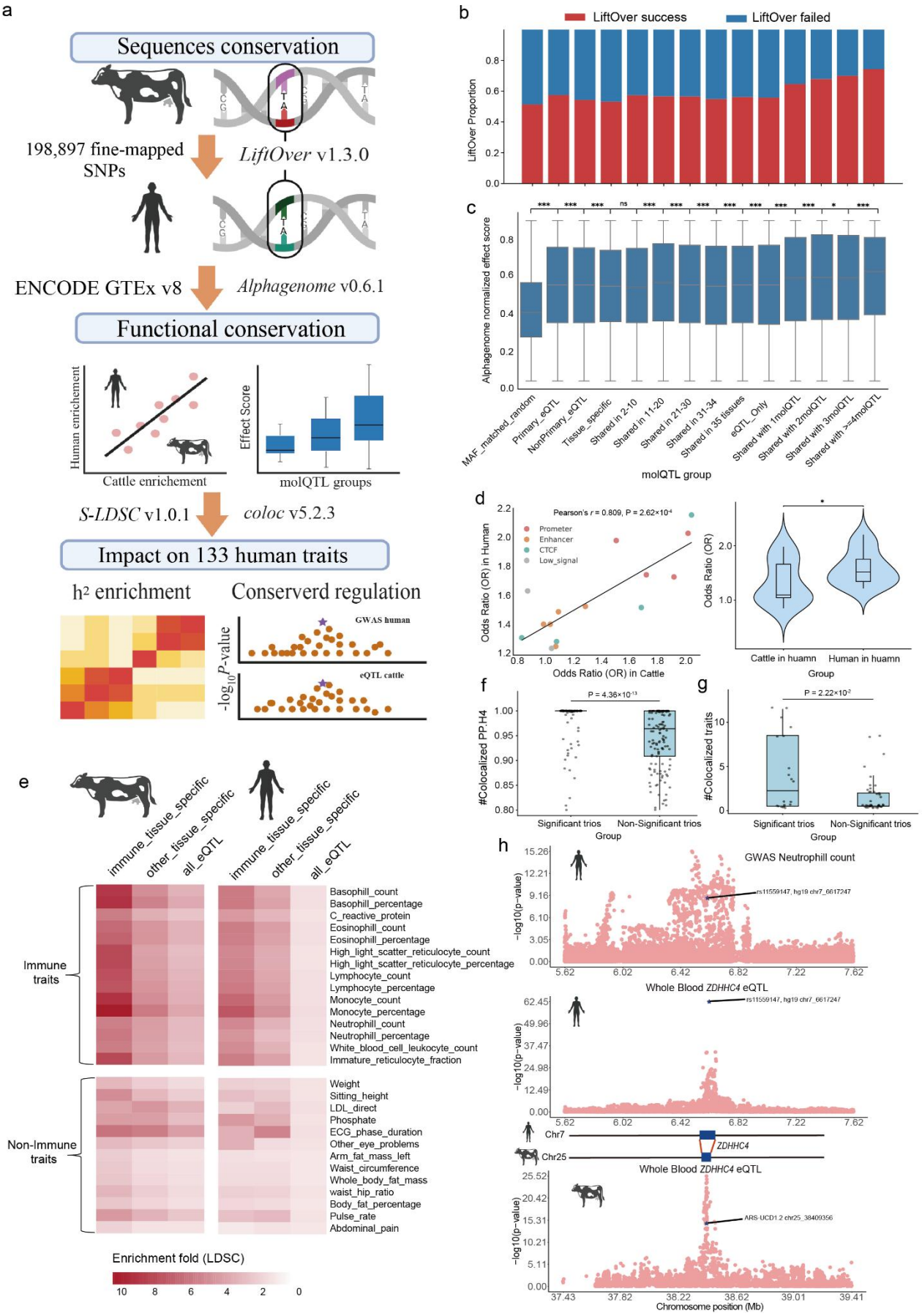
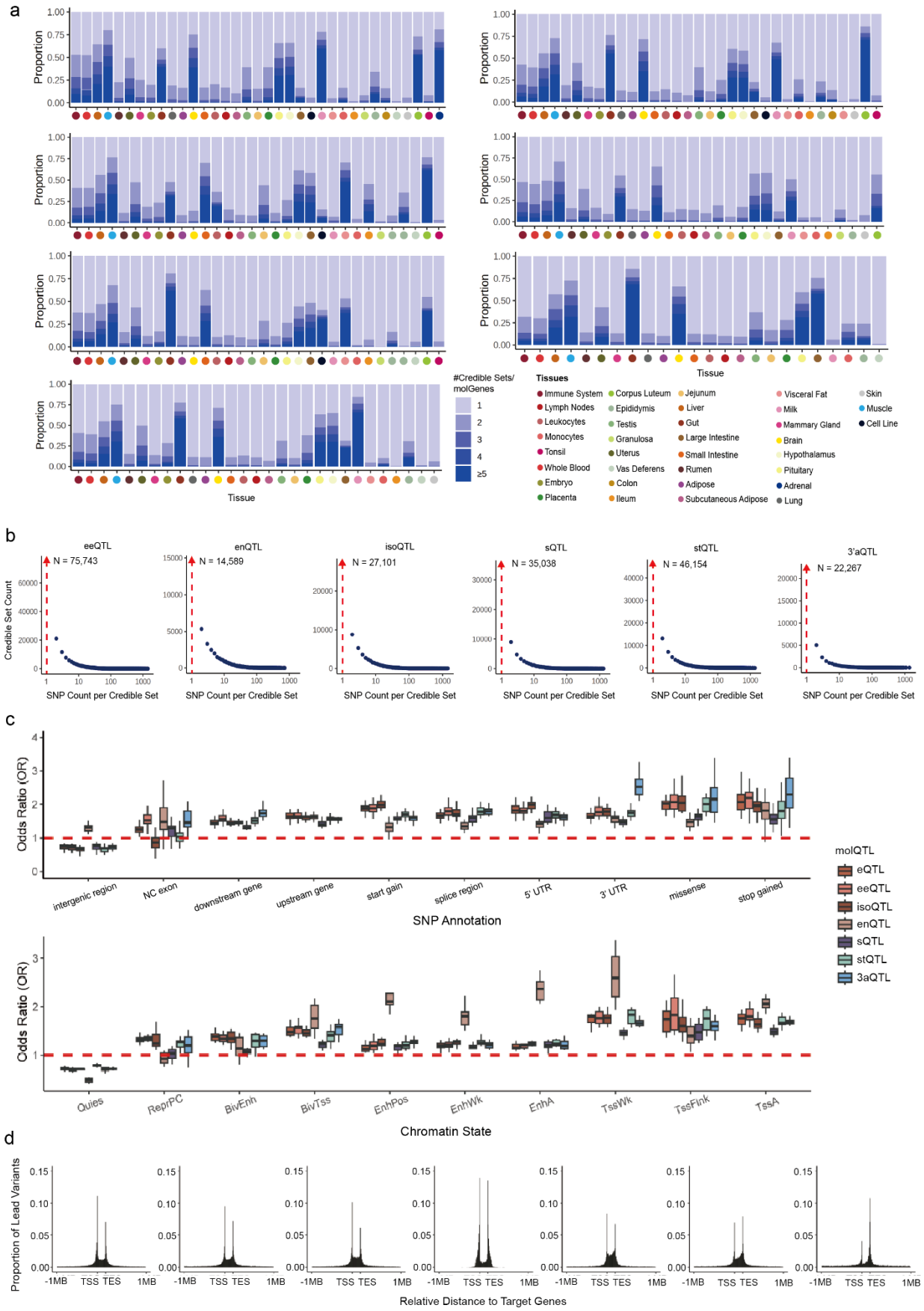
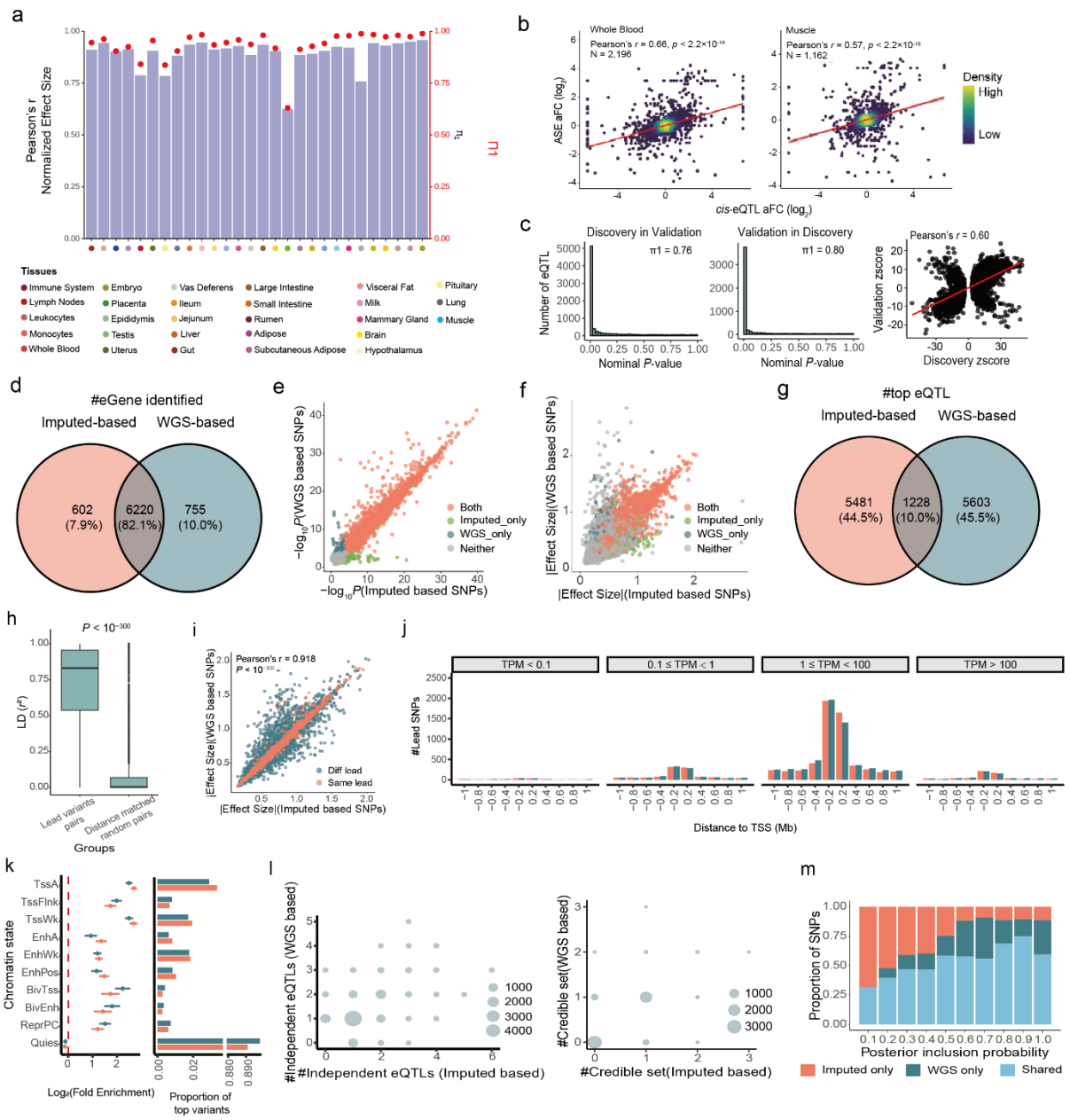

a
Sequences conservation
198,897 fine-mapped SNPs
LiftOver v1.3.0
ENCODE GTEx v8
Alphagenome v0.6.1
Functional conservation
Human enrichement
Cattle enrichement
Effect Score
molQTL groups
S-LDSC v1.0.1
coloc v5.2.3
Impact on 133 human traits
h2 enrichment
Conserverd regulation
GWAS human
eQTL cattle
-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 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.

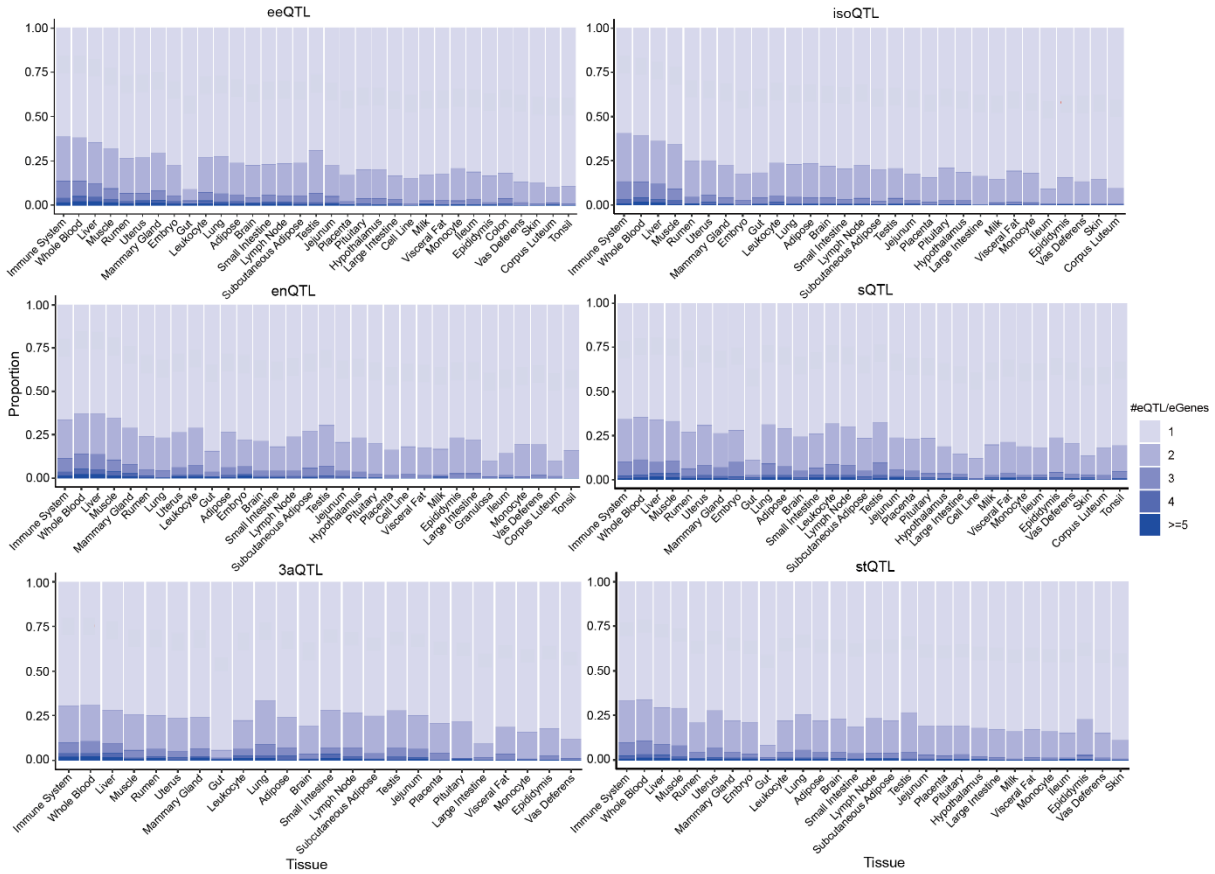

a
Proportion
Tissue
#Credible Sets/ molGenes
1
2
3
4
≥5
Tissues
Immune System
Lymph Nodes
Leukocytes
Monocytes
Tonsil
Whole Blood
Embryo
Placenta
Corpus Luteum
Epididymis
Testis
Granulosa
Uterus
Vas Deferens
Colon
Ileum
Jejunum
Liver
Gut
Large Intestine
Small Intestine
Rumen
Adipose
Subcutaneous Adipose
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
RepriPC
BivEnh
BivTss
EnhPos
EnhWk
EnhA
TssWk
TssFlnk
TssA
Chromatin State
d
Proportion of Lead Variants
-1MB
TSS TES
1MB
Relative Distance to Target Genes

1344 **Extended Data Fig. 1 | a,** The distribution of the number of fine-mapped *cis*-molQTL per gene across
1345 seven molQTL is shown as blue stacked bars across 35 tissues. Tissues are ordered by decreasing
1346 sample size. **b,** The size distribution of fine-mapped credible sets for all identified *cis*-molQTL across
1347 35 tissues. The red arrow indicates count of credible sets containing only one single candidate variant.
1348 **c,** Functional enrichment of lead SNPs in seven molQTL with regulatory elements (top: SNP classi-
1349 fication, bottom: chromatin states). **d,** Distribution of the distance to TSS regions across lead SNPs
1350 in seven molQTL.

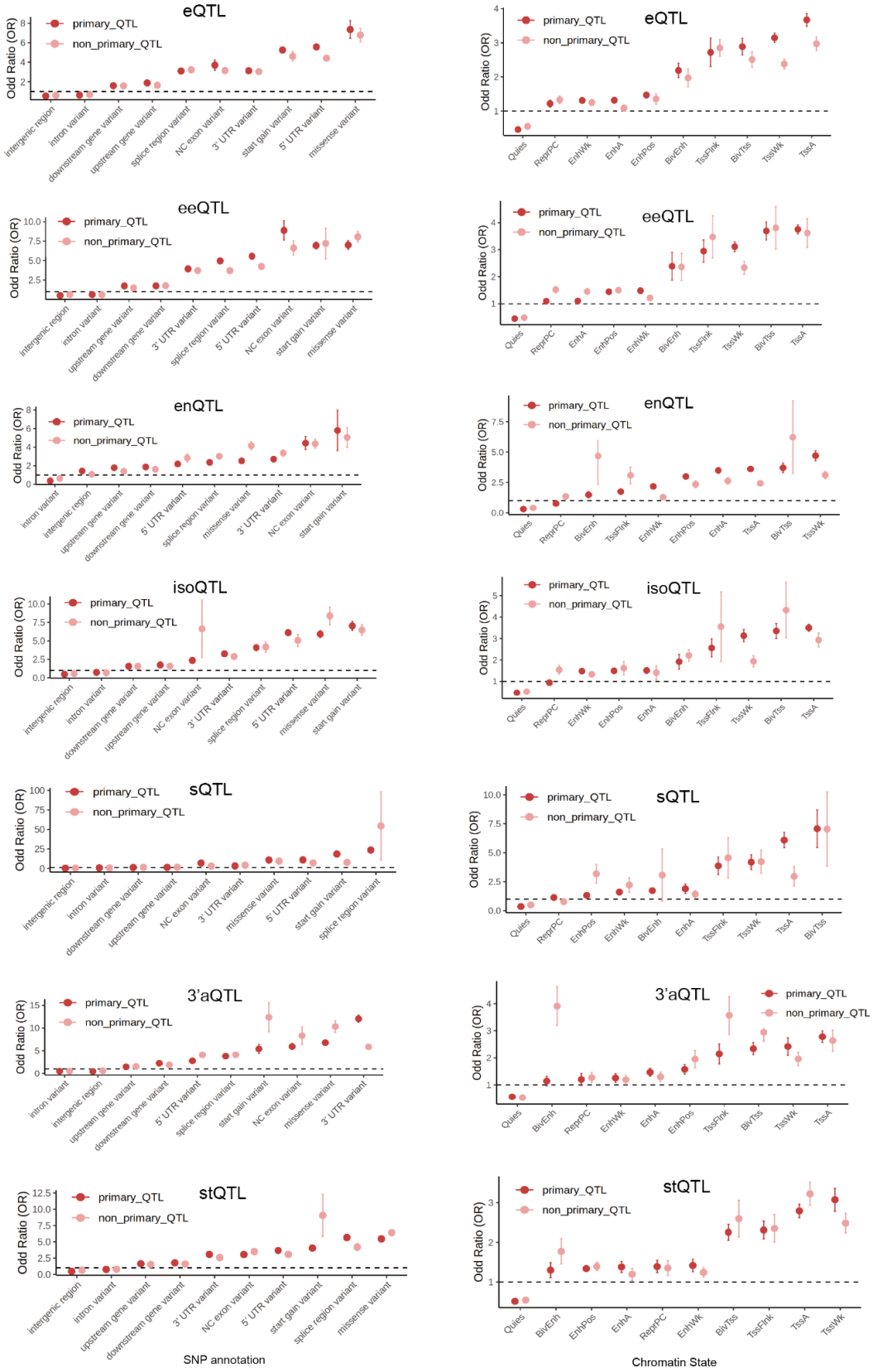


Extended Data Fig. 2 | a, Internal validation of *cis*-eQTL. We selected 28 tissues with sample sizes > 100 and randomly split samples into equally sized discovery and validation cohorts. eQTL mapping was performed independently in each cohort. Bars show Pearson's correlation of normalized effect sizes 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

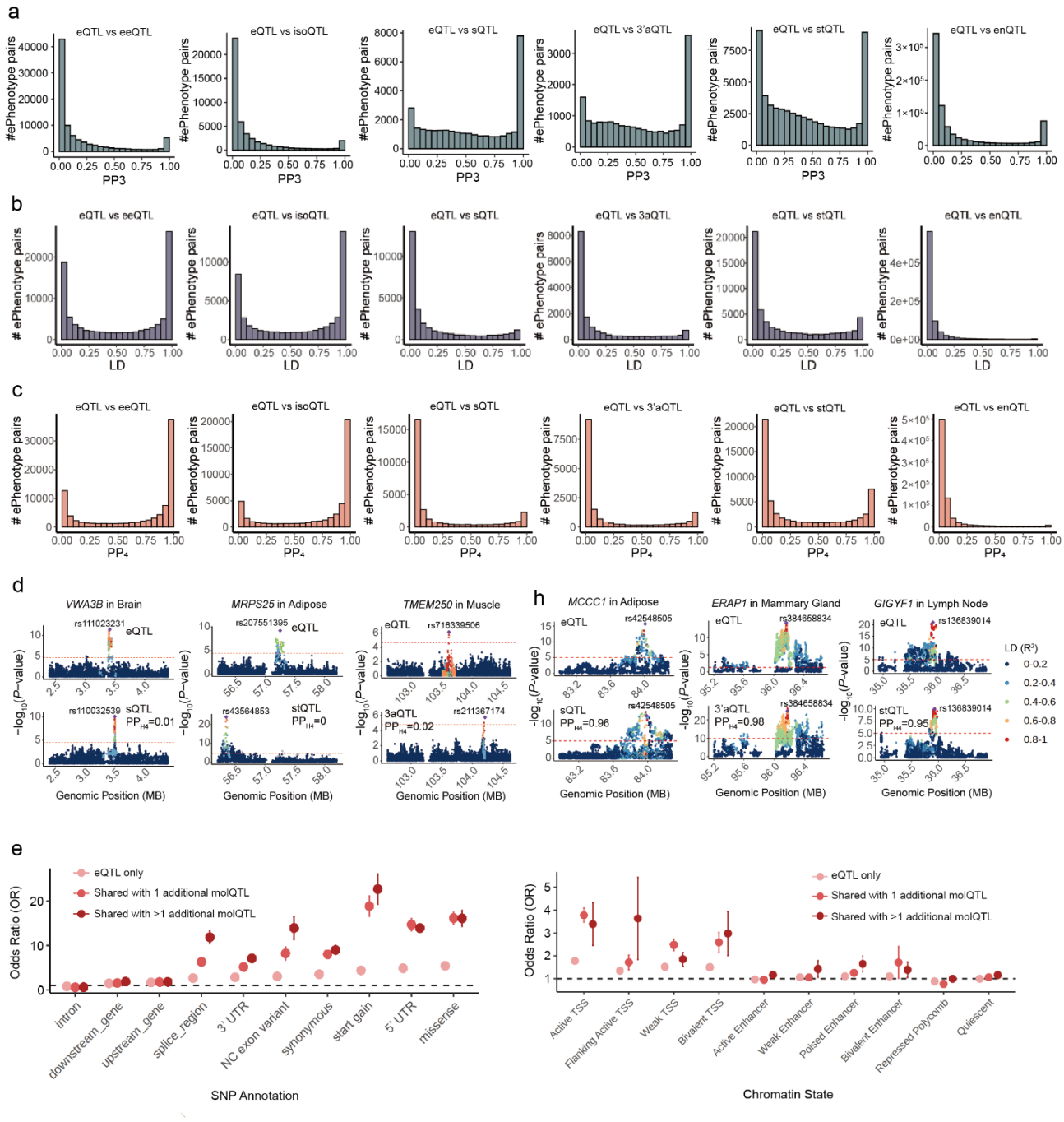
using RNA-seq imputed versus whole-genome sequencing (WGS)-called genotypes in an independent blood-tissue cohort (n = 99). Venne plot represents the overlapping of eGenes detected using imputed versus WGS–called genotypes. **e,** Comparison of significance (-log10 P) across different gene categories. 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.

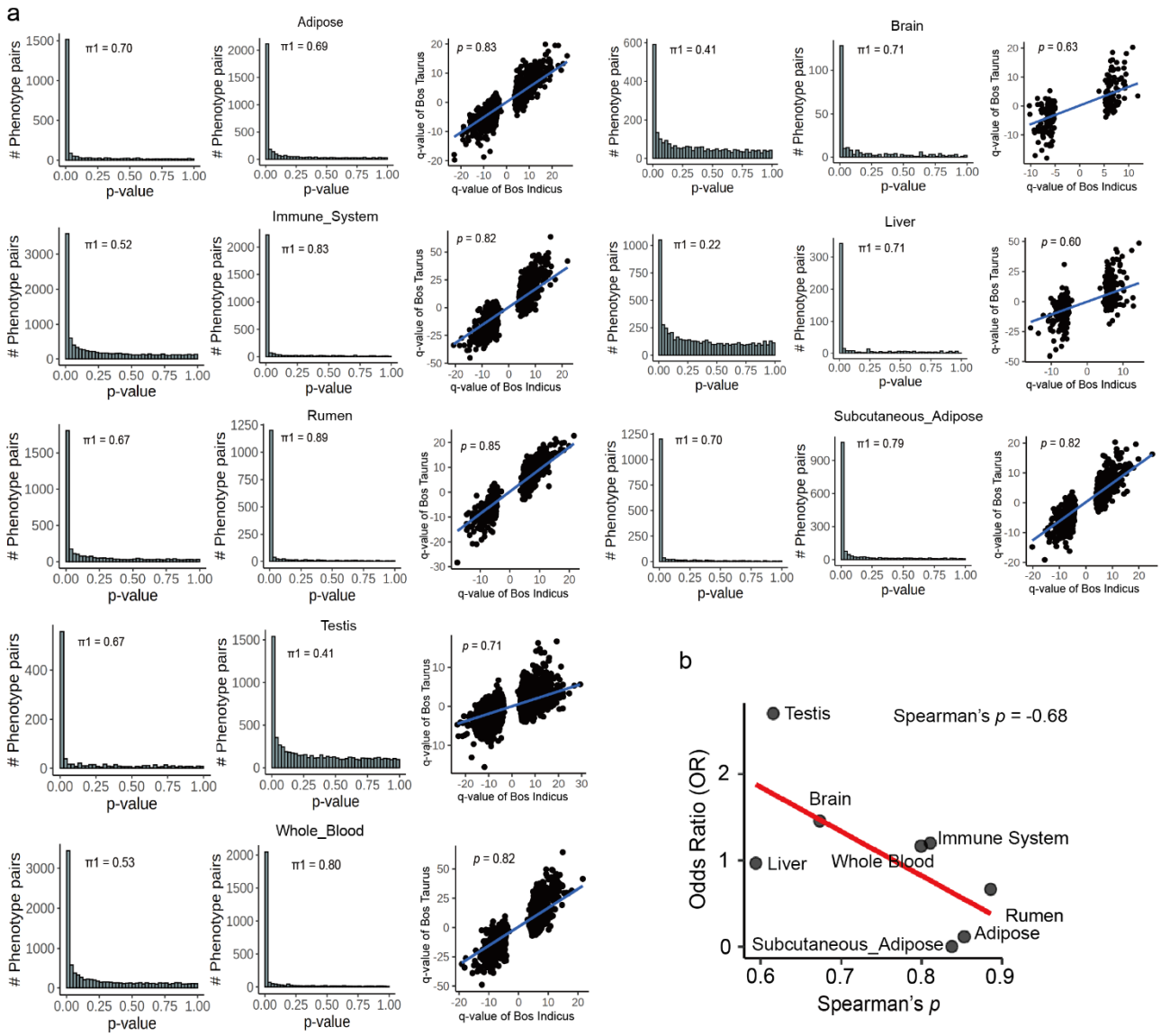


Extended Data Fig. 4 | Functional enrichment of primary and non-primary effects in six molQTL with regulatory elements (left: SNP classification, right: chromatin states).

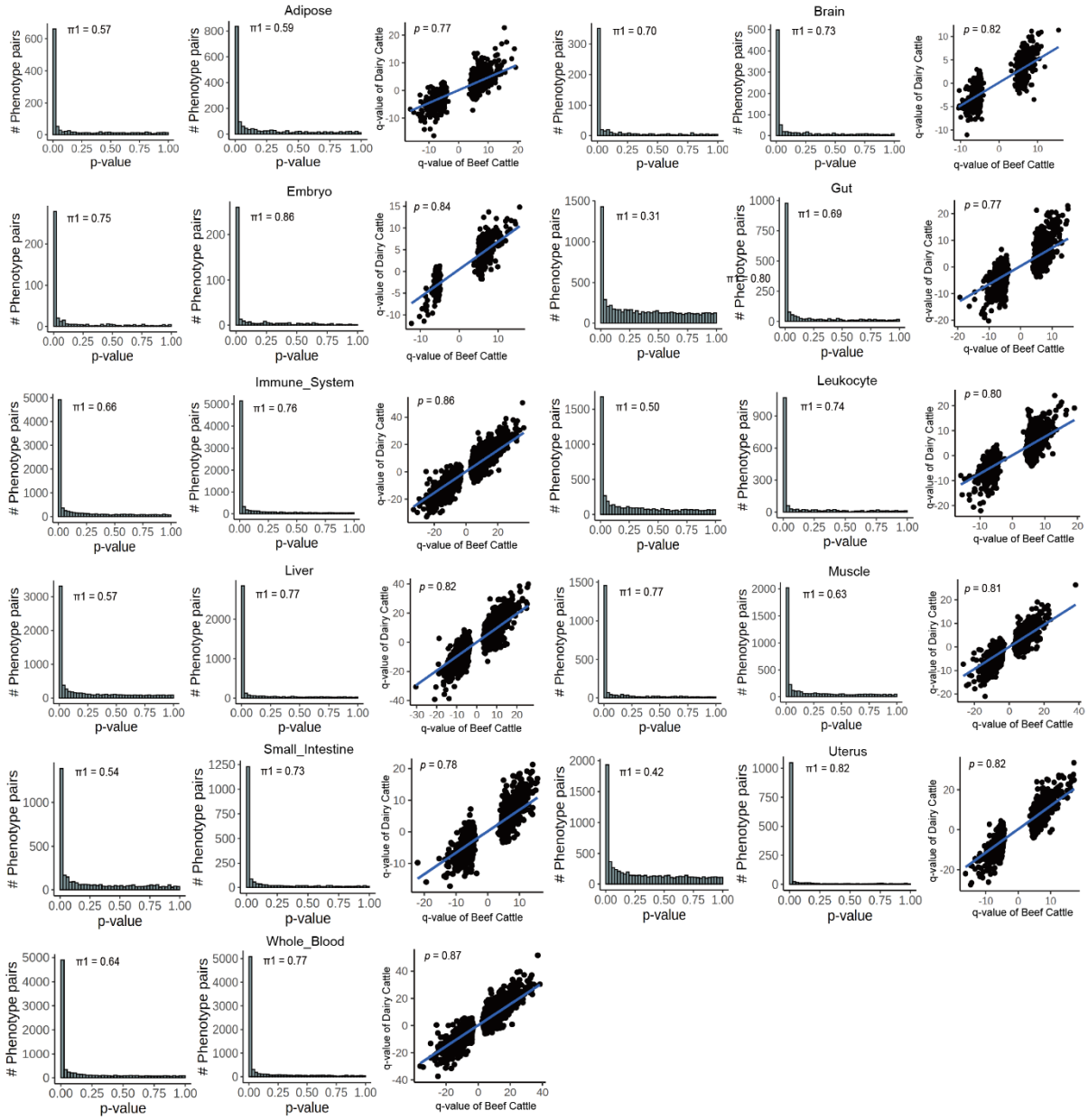


Extended Data Fig. 5 | a-c, Colocalization analyses eQTL across other six molecular phenotypes of the same molGene. PP_3: the posterior probability that the two molQTL are driven by distinct causal variants (**a**). PP_4: the posterior probability that the two molQTL share a single causal variant (**c**). 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.

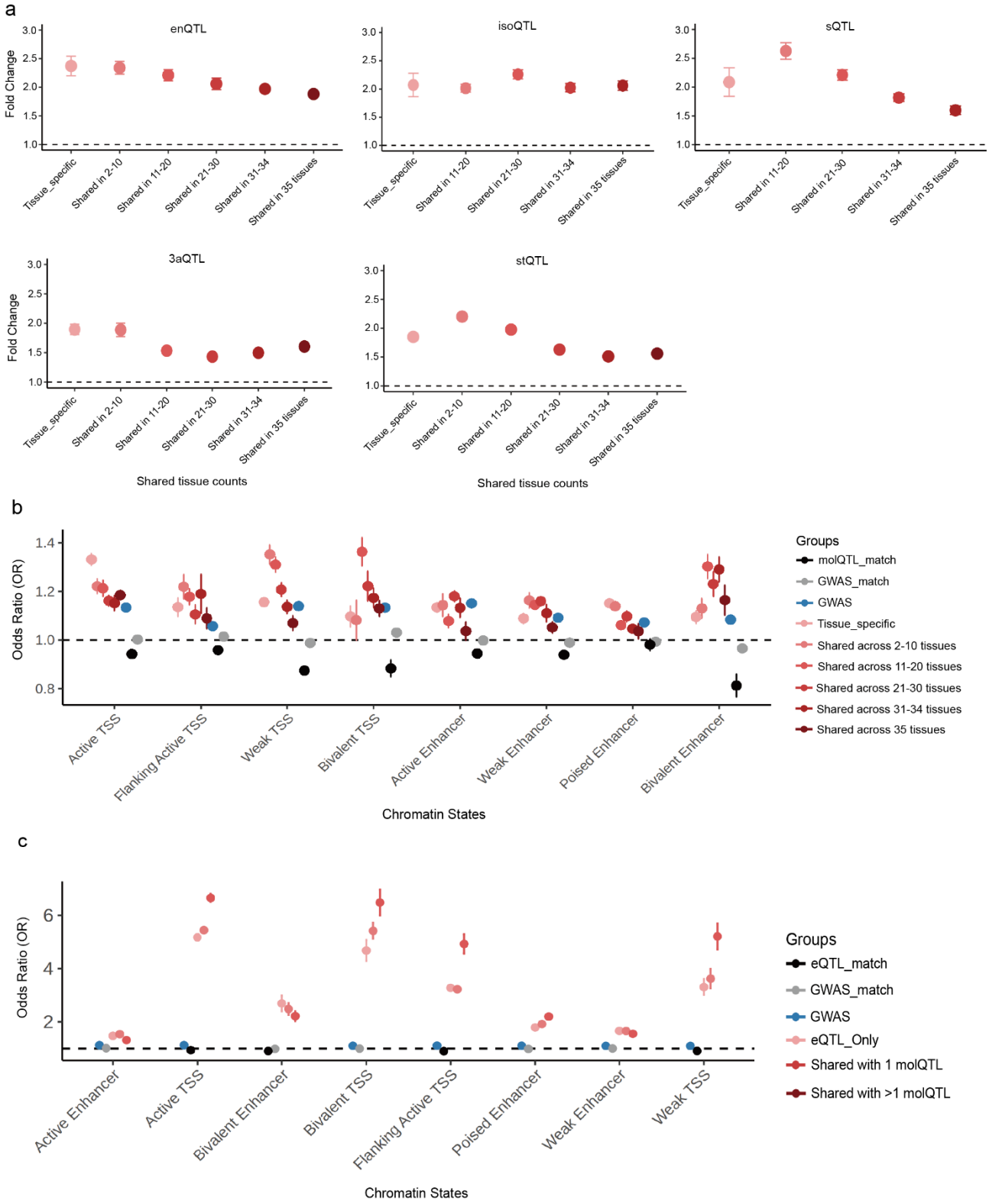
1401 The two-sided P values of lead SNPs were calculated by the linear mixed regression model. **e,** Func-
1402 tional enrichment of pleiotropic eQTL with regulatory elements (left: SNP classification, right: chro-
1403 matin states).



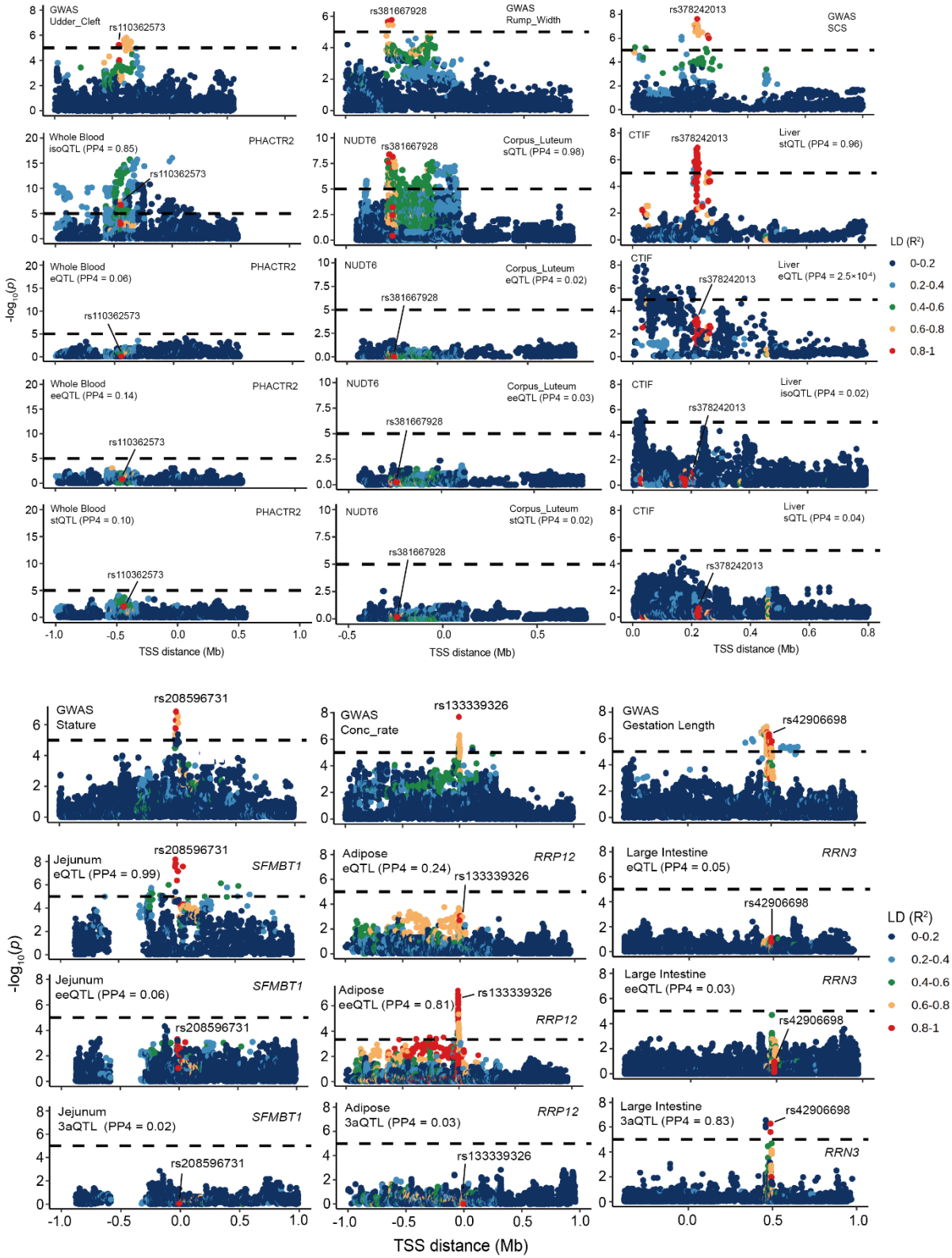
Extended Data Fig. 6 | a, Storey's π_1 statistics implemented in the qvalue package (v2.32.0)[119] and Spearman's correlation were used to assess replication of effect sizes between *B. taurus* and *B. indicus* across eight tissues. In each tissue, the p-value of the eQTL of *B. taurus* was used as observation (**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.

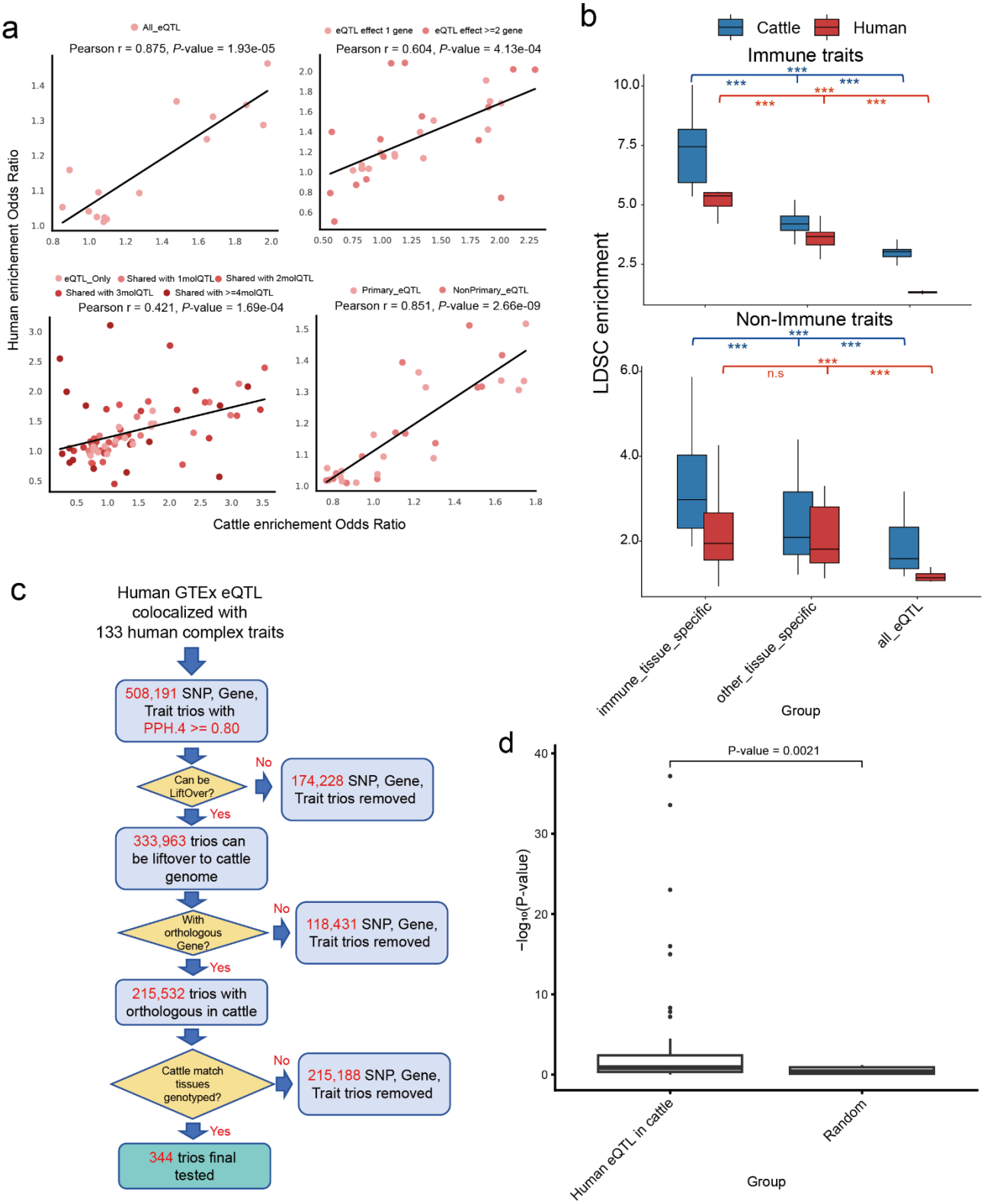


Extended Data Fig. 8 | a, Enrichment (odds ratio) of tissue specific/shared molQTL with genome-wide associations (GWAS) of 44 cattle traits using QTLenrich (v.2.0)[54]. **b,** Functional enrichment between tissue specific/shared molQTL and regulatory elements. **c,** Functional enrichment between pleiotropic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

1428 across genetic variants within the locus of *CTIF,* stature (GWAS) and eQTL in jejunum across genetic
1429 variants within the locus of *SFMBT1,* conception rates (GWAS) and eeQTL in adipose across genetic
1430 variants within the locus of *RRP12* and gestation length (GWAS) and 3'aQTL in large intestine across
1431 genetic variants within the locus of *RRN3*. Variants are colored according to their linkage disequilib-
1432 rium (LD, r^2) with the lead SNP. The two-sided P values of SNPs were calculated by the linear mixed
1433 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.