Comparative immune-relevant transcriptome reveals the evolutionary basis of complex traits

386 RNA-seq samples
10 immune tissues

Gene expression conservation

GWAS loci of complex traits

Transcriptome-based phylogeny

Wenjing Yang, Jian Yu, Yuelin Yao, ..., Lei Zhou, Lingzhao Fang, Jianfeng Liu

lingzhao.fang@qgg.au.dk (L.F.)
liujf@cau.edu.cn (J.L.)

Highlights

Comparative transcriptome of 10 immune tissues from 386 samples in six species

Gene expression of orthologous genes was generally conserved across species

Transcriptionally conserved genes were enriched for trait heritability

Yang et al., iScience 25, 105572
December 22, 2022 © 2022 The Author(s)
https://doi.org/10.1016/j.isci.2022.105572

OPEN ACCESS
Comparative immune-relevant transcriptome reveals the evolutionary basis of complex traits

Wenjing Yang,1 Jian Yu,1 Yuelin Yao,2 Siqian Chen,1 Bingru Zhao,1 Shuli Liu,3,4 Lei Zhou,1 Lingzhao Fang,2,5,* and Jianfeng Liu1,6,7,*

SUMMARY
Comparing transcriptome can help us reveal the genetic and evolutionary architecture underlying complex phenotypes within and between species. Here, by analyzing 386 publicly available RNA sequencing samples using a uniform bioinformatics pipeline, we systematically compared expression profiles of 10 immune-relevant tissues across humans, mice, pigs, cattle, sheep, and chickens. In general, we demonstrated that gene expression of orthologous genes was conserved within tissues across species. By integrating our findings with results of genome-wide association studies (GWAS) from 17 health-relevant traits in humans and 16,539 health-relevant quantitative trait loci (QTLs) in animals, we found that transcriptionally conserved genes were significantly enriched for more heritability of complex traits, compared to species-specific genes. In conclusion, our results advanced the knowledge of the transcriptome evolution of immune tissues and demonstrated that multi-species transcriptome comparison is highly informative for understanding the genetics of complex traits/disease.

INTRODUCTION
Variation in gene expression often underlies differences in complex phenotypes within and between species. Comparing gene expression patterns in multiple tissues and cell types across species could help understand the molecular and evolutionary processes shaping phenotypic diversity. For instance, comparative transcriptome between humans and mice or other primates has been employed to interpret the molecular mechanisms underlying brain-related traits and diseases in humans (Nowick et al., 2009; Xu et al., 2010; Zheng-Bradley et al., 2010).

The immune system is an interactive network of lymphoid organs, cells, humoral factors, and cytokines, which plays a crucial role in the host defense against pathogens and diseases (Parkin and Cohen, 2001). Specialized tissues for adaptive immunity in vertebrate include primary (i.e., bone marrow and thymus, responsible for hematopoiesis and lymphocyte development) and secondary (i.e., spleen, lymph nodes, and gut-associated lymphoid tissue, mainly for immune responses) lymphoid tissues. Among the secondary lymphoid tissues, the spleen or a spleen-like organ is most conserved, being present in all jawed vertebrates, whereas lymph nodes are more recently derived tissues, being restricted to mammals (Hirano, 2015). Due to the coevolution of the immune system and pathogens (Danilova, 2012), there is a constant pressure from pathogens on the host immune system. A well-known example of host-pathogen coevolution is the coevolution of humans and Staphylococcus aureus (Abi-Rached et al., 2007).

Cattle, sheep, pig, and chicken are four of the most important domestic animal species that feed the global population with nutritionally valuable animal protein (Adesogan et al., 2020). These species are also important as biomedical models to advance our understanding of evolutionary biology, human developmental biology, and disease. Compared to mouse and other model organisms which are often highly inbred and raised in the laboratory rather than in nature, these livestock species live close to humans, and thereby they may share similar pathogens and host-pathogen co-evolution processes. It is estimated that 60% of human pathogens reported have originated in animals and the risk of zoonoses is predicted to continue to increase (King et al., 2006; Klous et al., 2016). In addition, with the increase in data availability, there are a large amount of immune-related data in the public domain for livestock species (e.g., 7,922, 6,761, 896, 960 health-relevant QTLs in AnimalQTLdb database for cattle, pig, sheep, and chicken, respectively) (Hu et al., 2018). It is therefore of interest to compare gene expression and functional features of lymphoid...
tissues across farm species and humans to explore the development and evolution of the immune system and their impacts on health-relevant traits.

From an evolutionary viewpoint, although it has been reported that the heritability of complex traits was enriched in genomic regions with constrained DNA sequence across species (Hujoel et al., 2019), the evolution of anatomy and complex phenotypes is thought to primarily depend on the evolution of gene expression and regulation across tissues (Britten and Davidson, 1971; King and Wilson, 1975). By examining the heritability enrichments of complex diseases/traits among diverse molecular signatures (e.g., genes and regulatory elements) that show specific activities in tissues and cell types, recent studies found that immune/blood system was specifically and significantly associated with multiple immune, health and reproductive traits in both humans and cattle (Fang et al., 2020; Finucane et al., 2018). Therefore, broad questions still remain as to what extent the transcriptome is conserved in immune-relevant tissues across species, and how this knowledge can provide novel insights into the evolutionary and molecular processes shaping immune/health traits in humans. Furthermore, cross-species comparisons at different biological layers (e.g., genome, transcriptome, and epigenome) can take full advantage of biological information across species, providing additional knowledge for the ongoing genetic improvement program in livestock (Fang et al., 2020; Liu et al., 2020; Raymond et al., 2020).

Here, we collected 386 publically available RNA-seq samples with high quality from 10 immune-relevant tissue/cell types (i.e., blood, bone marrow, spleen, thymus, lymph node, small intestine, large intestine, CD4+ T cells, CD8+ T cells, and liver) and testis (testis is one of the most well-studied tissues in the comparative transcriptome studies, which we used here to compare with immune-relevant tissues) across human, mouse and four livestock species, including pig, cattle, sheep and chicken (Figure 1A and Table S1). We
hereafter referred to these tissue/cell types as the “tissues” in the following analysis. We systematically examined the similarities and differences in the transcriptome of these immune tissues across all these six species. We then integrated our findings with large-scale GWAS results from 17 health-relevant traits in humans and 16,539 QTLs of 10 health-relevant traits in animals (https://www.animalgenome.org/cgi-bin/QTLdb/index) to explore the evolutionary and genetic basis of such traits in humans and livestock. We found that in general, transcriptionally conserved genes were significantly enriched more heritability of complex traits than species specifically expressed genes. Based on these findings, we further proposed several candidate genes for health-relevant traits/disease in humans. Our results advanced the knowledge of organ evolution, which is a cornerstone in understanding the immune-transcriptome evolution in humans and livestock and the genetic and evolutionary architecture underlying complex traits/diseases.

RESULTS
Sample clustering and gene expression phylogenies
In total, using a uniform bioinformatics pipeline, we analyzed 386 publicly available RNA-seq samples from 10 immune-relevant tissues and testis across humans, mice, and four farm animals, yielding 15,500,306,304 clean reads with an averaged uniquely mapping rate of 87.38% (Figure 1A). The details of mapping statistics of all samples were summarized in Table S1. A total of 8,324 one-to-one orthologous genes were used for the subsequent analysis.

Based on the profiles of gene expression, the cluster of all individual samples is shown in Figure 1B, and the hierarchical clustering of tissues and species based on median expression of orthologous genes was showed in Figure 2A. Liver and testis formed two dominant clusters which showed “tissue-dominated clustering” pattern, consistent with the previous report (Brawand et al., 2011), indicating that the technical bias is limited after analyzing data in a uniform way. All the intestine samples from different species were clustered together, similar to blood/immune tissues (i.e., blood, bone marrow, spleen, thymus, lymph node, CD4+ T cells, and CD8+ T cells). Within these intestine and blood/immune systems, tissues tended to cluster first by species. This exhibiting the “species signal” phenomenon observed previously (Musser and Wagner, 2015), that is, gene expression is more similar among these tissues within a particular species than matched tissues between species. It is likely to reflect the concerted gene expression evolution of tissues within the immune system due to their shared host-pathogen co-evaluation processes within a species. For instance, within the immune system, chicken immune tissues clustered together rather than with their mammalian counterparts, corresponding to the split between birds and mammals at a phylogenetic distance of ~310 million year (Brawand et al., 2011). In addition, we noticed that cattle and sheep always clustered together, probably reflecting the underlying biological similarities in ruminants. A previous study proposed that tissues with more tissue-specific genes tended to be more transcriptionally conserved between humans and mouse (Lin et al., 2014). We thus identified tissue-specific genes within each species and calculated the percentage of shared tissue-specific genes among six species (Figure 2B). As expected, liver and testis shared more tissue-specific genes than other immune-relevant tissues. A linear mixed model was further used to estimate the contribution of tissue and individual to gene expression variation. We found that the variance in gene expression among tissues (average 30% of the total variance) was greater than that among species (average 25% of the total variance, Figure S1 and Table S2).

To obtain a global view of rates of gene-expression evolution, we reconstructed expression trees for the transcriptome in 10 tissues (CD8+ T cells were not included because of missing values, Figures 2C and S2). Trees clearly recapitulated the known mammalian phylogeny, roughly consistent with previous reports (Brawand et al., 2011): take bone marrow as an example, the tree separated chicken and mammals first, then grouped human and mouse together whereas pig, cattle, and sheep clustered together (Figure 2C). Of specific note, pigs rather than mice clustered together with humans in the lymph node phylogeny, which might be because domestic pigs are closely related to humans in terms of anatomy, genetics, and physiology, which are potential better animal models than mice to study human diseases (Meurens et al., 2012). The total branch lengths of trees vary widely among tissues (Figure 2D). Compared to other tissues, testis has the longest branches, reflecting the rapid evolution of gene expression in this organ, which was in line with the previous findings (Brawand et al., 2011; Fang et al., 2020). Within the immune tissues, lymph node evolves slowest (permutation test, p < 0.0001), suggesting that it may have experienced stronger purifying selection and/or less positive selection than other tissues during mammalian evolution, while blood and small intestine evolves fastest (permutation test, p < 0.0001).
We considered a gene to be expressed if it had the normalized expression value (TMM-normalized CPM) greater than 0.1. We then observed a significant correlation (Pearson's correlation values ranging from 0.78 to 0.92, p < 0.004) of the number of expressed genes in each tissue between human and mouse, cattle, sheep, pig or chicken (Figures 3A and S3A–S3D; Table S4). In general, blood expressed the least number of orthologous genes (average n = 7,152) while testis expressed the greatest (average n = 8,037) among species (Table S4). We found that the majority of orthologous genes (>78%) were expressed ubiquitously across all these immune-relevant tissues, which was consistent across species (Figures 3B and S3E–S3I). In addition, the number of tissues in which each gene was expressed was positively correlated between human and mouse, cattle, sheep, pig, or chicken (Pearson's correlation values ranging from 0.61 to 0.73, p < 2.2e-16, Figure 3D). However, an average of 69 out of 8,384 orthologous genes were not expressed in any sample across species, which were significantly enriched in non-immune associated functions such as neuropeptide signaling pathway, visual perception, and inner ear development (Figure S4A and Table S5).

Global conservation of gene expression across species

To explore the complexity of transcriptome in each tissue and its conservation across species, we calculated the fraction of total RNAs accounted for by the top 100 expressed genes in each tissue per species (Melé et al., 2015). Immune-relevant tissues showed relatively high transcriptional complexity (15.3-53.6%...
of total transcriptional output explained by top 100 expressed genes), whereas liver was the least complex tissue, where >55% of total transcriptional output explained by top 100 expressed genes in all species (Figures 3Ca and S3J–S3N). The transcriptional complexity was significantly correlated between human and mouse, cattle, sheep, pig, or chicken (Pearson's r = 0.64–0.92, p < 0.05, Figure 3D). In addition, we sorted all orthologous genes according to their median level of expression in each tissue and observed that all six species shared most in the top (highest expression) and bottom (lowest expression) 10% genes (Figure 3E). The functional annotation of the top 10% shared genes showed the known tissue-relevant biology and the bottom 10% of shared genes generally participated in basic cellular functions such as cell fate commitment and specification (Figures S4Ba and S4C; Table S5). Generally, these results demonstrate that gene expression profiles of orthologous genes are conserved at a certain degree within corresponding tissues among species.

**Comparison of gene expression between humans and other species**

Expression levels of pairs of orthologous genes were positively correlated across tissues, with median Pearson's correlation values ranging from 0.32 to 0.54 across species (p < 2.2 × 10^{-16}) (Figure 4A). We then calculated the Pearson’s correlation between the matched tissues of humans and other species, and it showed a significant correlation of 0.66 to 0.87 (permutation test p < 0.0001) (Figure S6).

Pairwise differential expression analysis of genes between human and mouse, cattle, sheep, pig, or chicken in each tissue revealed that, on average, 49.5% of expressed genes were significantly differentially
Figure 4. Differential gene expression analysis

(A) Cumulative distribution of Pearson’s correlation $r$ across 8 shared tissues in gene expression between human and mouse, cattle, pig, sheep, or chicken. 8,309 expressed orthologous genes were included in the distribution, and the estimated median value of $r$ is indicated on the horizontal axis of each graph. The background distribution of $r$ obtained by randomizing genes (solid curve) or randomizing samples (dashed curve) were shown in gray. The statistical significance was calculated using the Mann-Whitney U test comparing Pearson’s correlation values for orthologs to the background distribution of $r$ for randomly paired genes between human and mouse, cattle, pig, sheep, or chicken.

(B) Differential gene expression analysis of orthologous genes in human compared to mouse, cattle, pig, sheep, and chicken. The orange and blue bars correspond to the percentage of expressed orthologous genes with significantly (Benjamini-Hochberg corrected $p < 0.05$) higher and lower expression, respectively, in human compared to mouse, cattle, pig, sheep, or chicken.

(C) UpSet diagram analysis and functions of human-species conserved genes in liver. Human, human-specific genes; Mouse, genes only conserved between human and mouse; Pig, genes only conserved between human and pig; Cattle, genes only conserved between human and cattle; Sheep, genes only conserved between human and sheep; Chicken, genes only conserved between human and chicken. The functions displayed on the bar chart correspond to columns of the same color.
expressed (\(|\log_{2} FC| > 1, \text{FDR} < 0.05\)) between species (Figure 4B). We further investigated the function of genes (average \(n = 4,112\)) that are consistently differentially expressed in humans compared to other five species, and found that genes with higher expression in humans are significantly enriched in neurons and synapses such as the modulation of chemical synaptic transmission and neuron projection guidance. Immune molecules play integral roles in the CNS throughout neural development, including affecting neurogenesis, neuronal migration, axon guidance, synapse formation, activity-dependent refinement of circuits, and synaptic plasticity (Garay and McAllister, 2010). A better understanding of the shared mechanisms between immunological and neural synapses could aid in the development of new therapeutic modalities for immunological, neurological, and neuroimmunological disorders alike (Dustin, 2012). However, future works are required to fully understand functions of these neuronal genes in immune tissues.

While those with lower expression in humans are enriched in cell cycle and antigen processing (Figure S7; Table S6). We then investigated the function of genes that were conservatively expressed in humans compared to other species and found that genes from different comparisons revealed distinct function (Table S7). For instance, those conserved genes only between humans and cattle in liver were enriched in the regulation of phospholipid metabolic process, while those conserved only between humans and chicken were enriched in macroautophagy (Figure 4C).

Impact of transcriptional evolution on the genetics of complex traits

To better understand the genetic architecture underlying complex traits from an evolutionary point of view, we tested whether transcriptionally conserved genes explained more phenotypic variation than diverged genes and sequence-conserved genes using stratified linkage disequilibrium score regression (LDSC) in humans. We analyzed GWAS summary statistics of 17 human complex traits and disease with an average sample size of 188,921 (Table S8). We ranked (from the largest to smallest) genes according to their degree of differential expression (Figure 4B, measured by p-value), considering the top 1,000 as conserved genes between human and mouse, cattle, pig, sheep, or chicken in each tissue. We considered genes (average \(n = 792\)) with \(|\log_{2} FC| > 1\) and FDR < 0.05 and shared between humans and the other five species as human-specific genes. In addition, we selected the top 1,000 sequence-conserved orthologous genes in humans and other species. The details of sequence-conserved genes, transcriptionally conserved, and human-specific genes in each tissue were shown in Table S9. We found that genes with conserved expression had significantly higher heritability than human-specific ones (Student t-test, \(p < 0.0001\)) and similar heritability to sequence-conserved genes (Student t-test, \(p > 0.05\)) across tissues and traits (Figures 5A, S8, and S9; Table S10). We then integrated the AnimalQTL database with those transcriptionally conserved genes and found that human-cattle transcriptionally conserved genes significantly overlapped health QTLs in cattle (Figure 5B and Table S11, Fisher exact test, FDR<0.05). This may be because cattle, as one of the most well-studied domestic animals, have the most powerful QTLs compared to the other three farm animals. Next, we divided cattle health QTLs into 8 categories based on trait classes and found that the human and cattle conservatively expressed genes in CD4 + T cells and CD8 + T cells were significantly overlapping with disease QTLs. Whereas, the human and cattle conservatively expressed genes in bone marrow were significantly overlapped with general health parameters (Figure S10 and Table S11, Fisher exact test, FDR<0.05).

We further explored the contribution of conserved genes to different complex traits in humans. For instance, liver has been reported to be significantly associated with levels of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, and total cholesterol (TC) (Kundaje et al., 2015). By integrating GWAS summary statistics of LDL, HDL, TC, and triglycerides, we observed that genes with conserved expression in liver between humans and sheep showed higher heritability enrichment of these traits than genes with conserved expression between human and other species (Figure 5C). We found that two genes, TRPS1 (the top SNP \(p = 4.271E-17\) in GWAS of HDL) and RAF1 (the top SNP \(p = 8.406E-17\) in GWAS of TC), only showed expression conservation in liver between human and sheep (Figures 5D, SE, S11A, and S11B). TRPS1 encodes an atypical member of the GATA-type family of transcription factors, which was associated with HDL in a meta-analysis of 46 GWASs of lipids (Teslovich et al., 2010). RAF1 is the cellular homolog of viral raf gene (v-raf). The encoded protein MAP3K functions downstream of the Ras family of membrane-associated GTPases to which it binds directly. RAF1 was significantly associated with TC in previous GWAS studies (Willer et al., 2013).

Crohn’s disease (CD), ulcerative colitis (UC), and inflammatory bowel disease (IBD) are three major intestinal diseases. In general, we found that genes with conserved expression between humans and pigs had the
The highest heritability enrichment of these three diseases compared to those conserved between humans and other species (Figures 6A and 6B). We took TNFSF15 in the large intestine and TYK2 in the small intestine, which showed conserved expression only between humans and pigs, as examples later in discussion (Figures 6C, 6D, and S11C–S11E). TNFSF15 is a unique cytokine that functions in the modulation of vascular homeostasis and inflammation, which has been previously identified as a susceptibility gene for CD in Japanese, Koreans, and UK cohorts (Tremelling et al., 2008; Yamazaki et al., 2005; Yang et al., 2008). TYK2, as a therapeutic target in the treatment of autoimmune and inflammatory diseases, which plays a key role in immunity and apoptosis pathways, is also identified as a CD susceptibility gene by the meta-GWAS analysis (Gonciarz et al., 2021; Wang et al., 2009).

A previous study has shown that primary biliary cirrhosis (PBC) and systemic lupus erythematosus (SLE) are significantly associated with lymph nodes (Kundaje et al., 2015). We found that genes with conserved expression in lymph nodes across six species had the highest heritability enrichment of these two traits.
We identified TNPO3 (the top SNP p = 5.07975E-23 in GWAS of PBC and 1.8609E-45 in GWAS of SLE) with conserved expression in lymph nodes across six species (Figures S12B–S12D). TNPO3 is a nuclear importer, which is required for the infection of several lentiviruses including HIV (Larue et al., 2012). IRF5-TNPO3 locus contains SLE/PBC-associated variants (Hirschfield et al., 2010; Kottyan et al., 2014), where IRF5 is activated by pattern recognition receptors such as Toll-like receptor 7 and 9 (TLR7 and TLR9) and is a critical regulator of the immune response to infection (Cham et al., 2012).

**DISCUSSION**

In this study, we compared the transcriptomes of 10 immune-relevant tissues in human, mouse, cattle, sheep, pig, and chicken. Despite the significant differences in experiment conditions and sample characteristics, we observed that gene expression profiles in these tissues were highly conserved across species. This is in line with previous findings of gene expression between mouse and human (Zheng-Bradley et al., 2010).

The interpretation of the molecular mechanisms underlying complex traits has always been the research focus of genetics. By integrating large-scale GWAS data with tissue-specific genes, previous studies linked complex traits with their specifically relevant tissues, offering valuable information for fine-mapping causal genes/variants and for functionally validating GWAS hits through selecting the “right” tissues and cell types. Our analysis demonstrates that orthologous genes conservatively expressed across species...
contribute more to the heritability of complex traits than species-specific genes and similar heritability to sequence-conserved genes from an evolutionary perspective. Our results will serve as a valuable source for the human and livestock science community to interpret GWAS findings, to design follow-up validation experiments by choosing the most appropriate test animals as we know that not all species can be used as human disease models, as well as to help gain more novel insights into the genetic and biological mechanisms underpinning complex traits.

Limitations of the study
We noticed some limitations in our current study. We are limited by the availability of transcriptomic data. Though we tried our utmost to select healthy and adult samples for our analysis, there were differences in experimental conditions and sample characteristics that were beyond our control, which might affect gene expression. Also, due to the limited sample size, we, therefore, focused on cis-regulators of conserved genes by extending 50 kb around these genes. In order to study trans-eQTLs, we need a large amount of samples for each tissue due to their relatively small effects (Battle et al., 2017). Additionally, the cell-type composition of tissues could confound our interpretation of results. As we showed in Figure S5B and Figure S9, CD4+ T cells and CD8+ T cells groups had distinct enrichments across human complex traits and animal QTLs. Therefore, single-cell expression data may help further uncover the genetic basis of complex traits. However, this is not the only explanation, as considerable expression level differences were also observed between matched primary cell types (Alam et al., 2020). Phenotypic differences between species may be due to differences in multiple biological layers, e.g., regulatory elements, chromatin conformation, protein, metabolite, and microbiome (Alam et al., 2020; Arendt et al., 2016).

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Data collection
- METHOD DETAILS
  - Gene expression analysis
  - Sample clustering and tissue-specific gene expression analysis
  - Estimation of tissue and individual contribution to gene expression variation
  - Gene expression phylogenies
  - Stratified LD score regression for human complex traits
  - Gene Ontology enrichment analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105572.

ACKNOWLEDGMENTS
This work was supported by National Key Research and Development Program of China (2021YFD1200801), National Natural Science Foundation of China (31972563, 32272844) and the Earmarked Fund for China Agriculture Research System (No. CARS-pig-35).

AUTHOR CONTRIBUTIONS
L.F. and J.L. conceived and designed the project. W.Y., J.Y., Y.Y., S.C., B.Z., S.L., and L.Z., performed bioinformatics analyses. W.Y., L.F., and J.L. drafted and revised the article. All authors read, edited, and approved the final article.

DECLARATION OF INTERESTS
The authors declare no competing interests.
REFERENCES

Abi-Rached, L., Dorighi, K., Norman, P.J., Yawata, M., and Parham, P. (2007). Episodic nature of mature selection shaped the interactions of IgA-fc with FcαRI and bacterial decay proteins. J. Immunol. 178, 7943–7954. https://doi.org/10.4049/jimmunol.178.12.7943.

Adesogan, A.T., Havelaar, A.H., McKune, S.L., Eilitta, M., and Dahl, G.E. (2020). Animal source foods: sustainability problem or malnutrition and sustainability solution? Perspective matters. Global Food Secur. 25, 100325. https://doi.org/10.1016/j.gfs.2019.100325.

Alam, T., Agrawal, S., Severin, J., Young, R.S., Andersson, R., Arner, E., Hasegawa, A., Luzio, M., Ramlowski, J.A., Abugessaisa, I., et al. (2020). Comparative transcriptomics of primary cells in vertebrates. Genome Res. 30, 951–961. https://doi.org/10.1101/gr.255679.119.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., and Wagner, G.P. (2016). The origin and evolution of cell types. Nat. Rev. Genet. 17, 744–757.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using ‘lme4’. J. Stat. Softw. 67. https://doi.org/10.18637/jss.v067.i01.

GTEx Consortium; Laboratory Data Analysis &Coordinating Center LDACC—Analysis Working Group, Statistical Methods group—Analysis Working Group, Enhancing GTEx eGTEx groups; NIH Common Fund; NIH/NCI; NIH/NIHGR; NIH/NIH; NH/NIH; Biospecimen Collection Source Site—NDBI (2017). Genetic effects on gene expression across human tissues. Nature 550, 204–213. https://doi.org/10.1038/nature24277.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B., Scheipl, F., and Grothendieck, G. (2011). Package ‘lme4’. Linear mixed-effects models using S4 classes. R package version 1. https://cran.r-project.org/web/packages/lme4/index.html.

Bedford, T., and Hult, D.L. (2009). Optimization of gene expression by natural selection. Proc. Natl. Acad. Sci. USA 106, 1133–1138.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequencing data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al. (2011). The evolution of gene expression levels in mammalian organs. Nature 478, 343–348. https://doi.org/10.1038/nature10532.

Britten, R.J., and Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. 46, 111–138. https://doi.org/10.1080/00335509.1971.1068630.

Cham, C.M., Ko, K., and Niewold, T.B. (2012). Interferon Regulatory Factor 5 in the Pathogenesis of 723 Transcripts Erythematous. Clin. Dev. Immunol.

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S.J., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393.

Danilova, N. (2012). Evolution of adaptive immunity. Adv. Exp. Med. Biol. 738, 218–235. https://doi.org/10.1007/978-1-4614-1680-7_13.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq reads with STAR. Curr. Protoc. Bioinformatics 51, 11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

Dustin, M.L. (2012). Signaling at neuro/immune synapses. J. Clin. Invest. 122, 1149–1155.

Fang, L., Cai, W., Liu, S., Canela-Xandri, O., Gao, Y., Jiang, J., Rawlik, K., Li, B., Schroeder, S.G., Rosen, B.D., et al. (2020). Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. Genome Res. 30, 790–801. https://doi.org/10.1101/gr.250704.119.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Fara, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235.

Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., L-rateau, C., Shkriabai, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. 50, 621–629. https://doi.org/10.1038/s41588-018-0081-4.

Garay, P.A., and McAllister, A.K. (2010). Novel developments of Animal QTLdb and CorrDB. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB.

Hirano, M. (2015). Evolution of vertebrate adaptive immunity: immune cells and tissues, and AID/APOBEC cytidine deaminases. Bioessays 37, 877–887. https://doi.org/10.1002/bies.201400178.

Hirschfield, G.M., Liu, X., Han, Y., Gorlov, I.P., Lu, Y., Xu, C., Lu, Y., Chen, W., Juran, B.D., Collescu, C., et al. (2010). Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis. Nat. Genet. 42, 655–657. https://doi.org/10.1038/ng.631.

Hu, Z.-L., Park, C.A., and Reecy, J.M. (2018). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB.

Hujoel, M.L.A., Gazal, S., Hormozdari, F., van de Geijn, B., and Price, A.L. (2019). Disease heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function across species. Am. J. Hum. Genet. 104, 611–624. https://doi.org/10.1016/j.ajhg.2019.02.008.

King, D.A., Peckham, C., Waage, J.K., Brownlie, J., and Woolhouse, M.E.J. (2006). Infectious diseases: preparing for the future. Science 313, 1392–1393. https://doi.org/10.1126/science.1129134.

King, M.-C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science 188, 107–116.

Klous, G., Huss, A., Heederick, D.J.J., and Coutinho, R.A. (2016). Human–livestock contacts and their relationship to transmission of zoonotic pathogens, a systematic review of literature. One Health 2, 65–76. https://doi.org/10.1016/j.ohene.2016.03.001.

Kolde, R. (2012). Pheatmap: Pretty Heatmaps, 1. p. 726. https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf.

Kottyan, L.C., Zoller, E.E., Bene, J., Lu, X., Kelly, J.A., Rupert, A.M., Lessard, C.J., Vaughn, S.E., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes varies share. Hum. Mol. Genet. 24, 582–596. https://doi.org/10.1093/hmg/ddu435.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. https://doi.org/10.1038/nature14248.

Larue, R., Gupta, K., Wuenisch, C., Shkriabai, N., Kessl, J.J., Danhart, E., Feng, L., Taltyovn, O., Christ, F., Van Duyne, G.D., et al. (2012). Interaction of the HIV-1 inverse with transportin 3 protein (TNPO3 or TRN-SR2). J. Biol. Chem. 287, 34044–34058.

iScience Article

Received: November 10, 2021
Revised: August 15, 2022
Accepted: November 9, 2022
Published: November 14, 2022
Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Lin, S., Lin, Y., Nery, J.R., Urich, M.A., Breschi, A., Davis, C.A., Dobin, A., Zaleski, C., Beer, M.A., Chapman, W.C., et al. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. Proc. Natl. Acad. Sci. USA 111, 17224–17229. https://doi.org/10.1073/pnas.1413624111.

Liu, S., Yu, Y., Zhang, S., Cole, J.B., Tenesa, A., Wang, T., McDanel, T.G., Ma, L., Liu, G.E., and Fang, L. (2020). Epigenomics and genotypen-phenotype association analyses reveal conserved genetic architecture of complex traits in cattle and human. BMC Biol. 18, 80. https://doi.org/10.1186/s12915-020-00792-6.

Melé, M., Ferreira, P.G., Reverter, F., Deluca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al. (2015). Human genomics. The human transcriptome across tissues and individuals. Science 348, 660–665. https://doi.org/10.1126/science.aaa0355.

Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., and Gerdts, V. (2012). The pig: a model for human infectious diseases. Trends Microbiol. 20, 50–57. https://doi.org/10.1016/j.tim.2011.11.002.

Musser, J.M., and Wagner, G.P. (2015). Character trees from transcriptome data: origin and individuation of morphological characters and the so-called “species signal”. J. Exp. Zool. B Mol. Dev. Evol. 324, 588–604. https://doi.org/10.1002/jez.b.22636.

Nowick, K., Gernat, T., Almaas, E., andStubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proc. Natl. Acad. Sci. USA 106, 22358–22363. https://doi.org/10.1073/pnas.0911376106.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Parkin, J., and Cohen, B. (2001). An overview of the immune system. Lancet 357, 1777–1789. https://doi.org/10.1016/S0140-6736(00)04904-7.

Raymond, B., Yengo, L., Costilla, R., Schrooten, C., Bouwman, A.C., Hayes, B.J., Veerkamp, R.F., and Visscher, P.M. (2020). Using prior information from humans to prioritize genes and gene-associated variants for complex traits in livestock. PLoS Genet. 16, e1008780. https://doi.org/10.1371/journal.pgen.1008780.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp16.1.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713.

Tremelling, M., Berzuni, C., Massey, D., Bredin, F., Price, C., Dawson, C., Bingham, S.A., and Parkes, M. (2008). Contribution of TNFSF15 gene variants to Crohn’s disease susceptibility confirmed in UK population. Inflamm. Bowel Dis. 14, 733–737. https://doi.org/10.1002/ibd.20399.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Mach. Learn. 87, 33–54.

Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J.P., Russell, R.K., Sleiman, P.M.A., Imielinski, M., Glessner, J., Hou, C., et al. (2009). Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. Am. J. Hum. Genet. 84, 399–405. https://doi.org/10.1016/j.ajhg.2009.01.062.

Wang, Z.-Y., Leushkin, E., Lieichi, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Gritzner, F., Cardoso-Moreira, M., Janich, P., et al. (2020). Transcriptome and translomate co-evolution in mammals. Nature 588, 642–647.

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283. https://doi.org/10.1038/ng.2797.

Xu, A.G., He, L., Li, Z., Xu, Y., Li, M., Fu, X., Yan, Z., Yuan, Y., Menzel, C., Li, N., et al. (2010). Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. PLoS Comput. Biol. 6, e1000843. https://doi.org/10.1371/journal.pcbi.1000843.

Yamazaki, K., McGovern, D., Ragoussis, J., Paolucci, M., Butler, H., Jewell, D., Cardon, L., Takao, M., Tanaka, T., Ichinomi, T., et al. (2005). Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn’s disease. Hum. Mol. Genet. 14, 3499–3506. https://doi.org/10.1093/hmg/ddi379.

Yang, S.K., Lim, J., Chang, H.S., Lee, I., Li, Y., Liu, J., and Song, K. (2008). Association of TNFSF15 with Crohn’s disease in Koreans. Am. J. Gastroenterol. 103, 1437–1442. https://doi.org/10.1111/j.1572-0241.2007.01752.x.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–297.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Grón, C.G., et al. (2018). Nucleic Acids Res. 46, D754-D761. https://doi.org/10.1093/nar/gky1098.

Zheng-Bradley, X., Rung, J., Parkinson, H., and Brahma, A. (2010). Large scale comparison of global gene expression patterns in human and mouse. Genome Biol. 11, R124. https://doi.org/10.1186/gb-2010-11-12-r124.
STAR METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited Data      |        |            |
| 386 publicly available RNA-seq data | NCBI SRA database | https://www.ncbi.nlm.nih.gov/sra/ PRJNA516565, PRJNA646376, PRJEB22535, PRJNA263600, PRJNA294306, PRJEB27455, PRJNA526257, PRJNA379574, PRJNA573919, PRJNA251439, PRJEB25677, PRJNA329305, PRJNA177791, PRJEB28745, PRJNA279487, PRJNA412404, PRJEB38599, PRJNA503784, PRJNA331082, PRJNA579994, PRJNA204941, PRJEB12891, PRJNA492316, PRJNA548207, PRJNA494560, PRJEB4337, PRJNA552599, PRJNA52875, PRJNA450828, PRJEB6971, PRJNA633865, PRJNA280600, PRJNA386625, PRJEB24867, PRJNA375882, PRJNA490481, PRJNA641931, PRJNA631158, PRJNA286225, PRJNA454272, PRJEB22693, PRJNA360164, PRJNA143627, PRJNA471866, PRJNA510331, PRJEB26391, PRJEB19386, PRJEB37735, PRJNA529662, PRJNA392949, PRJNA362606, PRJNA308292, PRJEB19199, PRJEB6169 |

Software and algorithms

- Trimmomatic (version 0.39) Bolger et al. (2014) https://github.com/timflutre/trimmomatic
- STAR (version 2.7.0e) Dobin and Gingeras (2015) https://github.com/alexdobin/STAR
- featureCounts (version 2.0.0) Liao et al. (2014) http://subread.sourceforge.net/
- EdgeR (v. 3.28.1) Robinson et al. (2010) https://bioconductor.org/packages/edgeR/
- Rsne Van der Maaten and Hinton (2008) https://cran.r-project.org/web/packages/Rsne/index.html
- pheatmap Kolde (2012) https://CRAN.R-project.org/package=pheatmap
- lme4 Bates et al. (2011) https://cran.r-project.org/web/packages/lme4/index.html
- ape Paradis et al. (2004) https://cran.r-project.org/web/packages/ape/index.html
- LDSC Finucane et al. (2015) https://github.com/bulik/LDSC
- clusterProfiler v3.14.3 Yu et al. (2012) https://github.com/YuLab-SMU/clusterProfiler

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jianfeng Liu (liujf@cau.edu.cn).

Materials availability
This study did not generate new unique reagents or material.

Data and code availability
This paper analyzes existing, publicly available data. The details for the datasets are listed in the Table S1. This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.
EXPERIMENTAL MODEL AND SUBJECT DETAILS

Data collection
Eleven tissues (blood, bone marrow, spleen, thymus, lymph node, small intestine, large intestine, CD4+ T cells, CD8+ T cells, liver, testis) of at least 3 healthy adult individuals (except for 26 samples of chicken, which were all healthy individuals but did not meet adult standards) for each of six vertebrates (human, mouse, cattle, sheep, pig and chicken) were collected. A total of 386 RNA-seq samples were download from 54 studies of SRA database. We provided references where RNA-seq data were retrieved and summarized details of all RNA-seq samples in Table S1. The number of expressed genes (TMM-normalized CPM >0.1) and the number of mapping reads in each sample showed no significant correlation (Figure S1).

METHOD DETAILS

Gene expression analysis
We used genome assemblies GRCh38 (human), GRCm38 (mouse), UCD1.2 (cattle), Oar_v1.0 (sheep), Sscrofa11.1 (pig) and GRCg6a (chicken) for our analysis. Gene annotations from Ensembl (Zerbino et al., 2018) release 101 were used for the six species. We retrieved the 1:1 orthologous for the species in our set from the Ensembl database release 101 and extracted 8,324 orthologous that have 1:1 orthology relationships among all the species in our study.

Then we analyzed all RNA-seq data uniformly using the following bioinformatics pipeline. Firstly, we removed contaminating adapter molecules, reads containing poly(N), and low-quality reads using Trimmomatic (version 0.39) (Bolger et al., 2014), obtaining a total of 15,500,306,304 clean reads. We then mapped the reads from each library against the species reference genome using STAR (version 2.7.0e) (Dobin and Gingeras, 2015), resulting in an averaged uniquely mapping rate of 87.38% (Table S1). We used feature-Counts (version 2.0.0) (Liao et al., 2014) to generate read counts for the set of genes. We normalized the count data using the method trimmed mean of M-values (TMM) as implemented in the package EdgeR (v. 3.28.1) (Robinson et al., 2010). Log2-transformed values of (TMM-normalized CPM +1) for genes with >0.1 TMM-normalized CPM in more than 80% of the samples were used for subsequent analyses. After filtering, 8,309 genes remained.

Sample clustering and tissue-specific gene expression analysis
T-Distributed Stochastic Neighbor Embedding (t-SNE) (Figure 1B) and heatmap (Figure 2A) were applied to visualize the samples. We performed t-Distributed Stochastic Neighbor Embedding (t-SNE) using Rtsne package (Van der Maaten and Hinton, 2008), projecting samples to a two-dimensional space based on TMM-normalized CPM expression values of orthologous genes. We calculated the normalized log2 (TMM-normalized CPM +1)-transformed median gene expression in each tissue of each species separately and then performed hierarchical clustering using Ward.D2 method in R package pheatmap (Kolde, 2012), to explore the relationship of tissues among species based on the median gene expression.

We divided tissues into four categories according to the results of clustering, including liver; testis; large intestine and small intestine; blood, bone marrow, spleen, thymus, lymph node, CD4+ T cells, and CD8+ T cells. We detected genes with tissue-specific expression using EdgeR package which compares gene expression of samples in a given tissue to those in the remaining tissues in each species, by excluding tissues in the same category. We adjusted p-values for multiple testing using Benjamini and Hochberg methods (FDR). We considered genes with a log2-transformed fold-change >1 and FDR <0.05 as tissue-specific genes. Then we calculated the percentage of tissue specific-genes shared by six species per tissue.

We also employed EdgeR package to detect differential expressed genes (DEGs) in each tissue between humans and mouse, cattle, sheep, pig, chicken. The analyses were performed in a similar manner as the tissue-specific genes. Genes with absolute log2-transformed fold-change values > 1 and FDR <0.05 were detected as DEGs.

Estimation of tissue and individual contribution to gene expression variation
We used a linear mixed model (LMM) implemented in R package lme4 (Bates et al., 2015) to assess the contribution of tissues and species to variation in gene expression. Gene expression is modeled as a function of tissues and species (considered as random factors). Gene expression were log2-normalized. To
obtain the variance components, we divide the limited maximum likelihood estimators of random effects of tissue, species, and residual variances by their sum (Table S2).

Gene expression phylogenies

We used the variance-based approach to estimate expression divergence (Wang et al., 2020), which was basis on the assumption that gene-expression evolution represents the succession of independent changes in gene-expression levels, consistent with Brownian-motion-based models of gene-expression evolution (Bedford and Hartl, 2009). Therefore, for example, differences in population-average expression between human (\(e_h\)) and cattle (\(e_c\)) can be quantified as,

\[
\text{var}_g \left( \hat{e}_{g,H} - \hat{e}_{g,C} \right) - \frac{\hat{\theta}_H}{n_H} - \frac{\hat{\theta}_C}{n_C}
\]

where \(\hat{e}_{g,H}\) corresponds to the variance of the log2(TMM-normalized CPM +1)-transformed median expression values across biological replicates in a particular tissue of human across 8309 1:1 orthologous genes (g). The obtained values were corrected for the sampling variance (\(\hat{\theta}_j\)), stemming from variation in expression between individuals within species and measurement error. \(n\) is the number of replicates for a particular species-tissue combination, and \(\hat{\theta}\) is calculated as the average variance of expression levels of a gene (\(e_g\)) across biological replicates (r):

\[
\hat{\theta} = \text{mean}_{r \in \{\text{replicates}\}} \text{var}_{g} \left( \hat{e}_{g,r} \right)
\]

To scale the expression change and to understand how much it contributes to expression variation, we divide the metrics obtained in Equation 1 by the variance of expression levels across genes, averaged across all six species. We use the normalized metric as an estimate of the expression divergence between species (in this example between human and cattle)

\[
d_{e_{H-C}} = \frac{\text{var}_{g} \left( \hat{e}_{g,H} - \hat{e}_{g,C} \right) - \frac{\hat{\theta}_H}{n_H} - \frac{\hat{\theta}_C}{n_C}}{\left( \sum_{s \in \{\text{species}\}} \text{var}_{g} \hat{e}_{g,s} - \frac{\hat{\theta}_s}{n_s} \right)^{\frac{1}{2}}}
\]

On the basis of pairwise distance matrices between species, for each tissue, we constructed gene-expression trees using the neighbour-joining (NJ) approach. The NJ trees and the total tree length were constructed using functions in the ape package (Paradis et al., 2004) in R. The reliability of branching patterns was assessed with bootstrap analyses (1:1 orthologous genes were randomly sampled with replacement 1000 times). The bootstrap values are the proportions of replicate trees that share the branching pattern of the majority-rule consensus tree shown in the figures.

The previously used spearman-rank-correlation-based approach gives very similar estimates across tissues (Table S3) to what we observe in our variance-based approach (Pearson’s \(r = 0.92\), \(p = 5.71E-05\)).

Stratified LD score regression for human complex traits

Then we applied stratified LD score regression (Finucane et al., 2015) to partition heritability of human complex traits and test which group of human-species’ transcriptionally conserved genes contribute the more genetic variance of complex traits, which relies on the fact that the \(\chi^2\) association statistic for SNP \(j\) includes the effects of all SNPs that in LD with SNP \(j\). That is, the expected \(\chi^2\) statistic of SNP \(j\) is

\[
E \left[ \chi^2_j \right] = N \sum_{C} \rho_{C} f(j, C) + nA + 1
\]

where \(N\) is sample size, \(C\) indexes categories, \(f(j, C)\) is the LD score of SNP \(j\) with respect to category \(C\), defined as \(f(j, C) = \sum_{k \neq j} r_{jk}^2\), where \(r_{jk}\) is the squared correlation between SNPs \(j\) and \(k\) in the population. \(a\) is a constant that reflects population structure and other sources of confounding. The enrichment of a category was defined to be the proportion of SNP-heritability in the category divided by the proportion of SNPs. Specifically speaking, we estimate the regression coefficient \(\tau_C\) quantifies the importance of annotation \(C\). \(\tau_C\) will equal 0 if \(C\) is not enriched, will be negative if belonging to \(C\) decreases per-SNP heritability accounting for all other annotations included, and will be positive if belonging to \(C\) increases per-SNP heritability, accounting for all other factors (ref). Then we computed \(p\) values that tested whether \(\tau_C\) was positive. When reporting quantitative results, we normalized the coefficient \(\tau_C\) by our estimate of the mean per-SNP heritability \(\sum \text{Var}(\hat{\beta}) / M\) to make it comparable across phenotypes, where \(\text{Var}(\hat{\beta})\) is per-SNP
heritability of SNP \( j \), \( M \) is the total number of SNPs. The normalized coefficient can be interpreted as the proportion by which the per-SNP heritability of an average SNP would increase if \( \tau_c \) were added to it.

We collected GWAS summary statistics for 17 human complex traits from public database (Details in Table S8). We ranked (from the largest to smallest) genes according to their degree of differential expression (Figure 4B, measured by \( p \)-value), the top 1000 genes were identified as transcriptionally conserved genes between human and a specific species. ANOVA was applied for each gene with species variable as factor in the following model: \( \text{lm} = \text{Species} \). Then we sorted genes according to \( p \)-values, the top 1000 genes with the largest \( p \)-value were identified as transcriptionally conserved genes across all species. Genes with absolute log2-transformed fold-change values >1 and FDR <0.05 and shared between humans and the other five species were considered as human-specific genes.

The code for LD score regression was from https://github.com/bulik/ldsc. LD score was calculated using data for the European population in the 1000 Genomes project (Consortium, 2015) (https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_plinkfiles.tgz) and the minor allele frequency of SNPs in this population was downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_frq.tgz. Only HapMap3 (Gibbs et al., 2003) SNPs were included. Next, we performed LD score regression for each trait. The regression model used in the test included human-species’ transcriptionally conserved genes (Human-Mouse conserved group, Human-Pig conserved group, Human-Cattle conserved group, Human-Sheep conserved group, and Human-Chicken conserved group), human-specific genes and the annotations in the pretrained baseline model (https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_baselineLD_v2.2_lscores.tgz). We extended 50 kb of gene regions. In total, we performed 1,071 tests (17 traits \( \times \) 63 groups). A threshold of 5% FDR was used to call significant enrichment.

A total of 16,539 QTLs of 10 health-relevant traits in animals were downloaded from AnimalQTLdb database (https://www.animalgenome.org/cgi-bin/QTLdb/index), including 7,922, 6,761, 896, 960 health-relevant QTLs for cattle, pig, sheep and chicken, respectively.

Gene Ontology enrichment analyses

The GO term enrichment analysis was performed separately for each set of genes. The enrichment was tested with the hypergeometric test implemented in the R package clusterProfiler v3.14.3 (Yu et al., 2012). Ensembl gene IDs were converted to entrez gene IDs via the R package org.Hs.eg.db v3.10.0, and enrichGO function was applied to detect the enriched GO terms. The GO terms associated with the biological process hierarchy were sorted by their \( p \) values corrected for multiple testing (Benjamini–Hochberg correction), and the top three or five significantly enriched terms were shown for each group of genes.

QUANTIFICATION AND STATISTICAL ANALYSIS

In this study, heritability partition of human complex traits was conducted in LDSC software using \( \chi^2 \) statistic. One-way analysis of variance (ANOVA) was applied to identify transcriptionally conserved genes. Unless otherwise noted, differential gene expression and the differences between two groups were tested using Student’s t-test. The corrected \( p \)-values for multiple testing were calculated using Benjamini–Hochberg correction. Statistical details can be found in the figure legends, results and method details. All statistical analyses were performed using R.