Mao et al. *Zool. Res.* 2024, 45(2): 299−310
https://doi.org/10.24272/j.issn.2095-8137.2023.322

ZR | Zoological Research

Article Open Access

Comparative transcriptome analysis between rhesus macaques (*Macaca mulatta*) and crab-eating macaques (*M. fascicularis*)

Yu-Xiang Mao[1,2,3,4], Yamei Li[1,2,4], Zikun Yang[5,6], Ning Xu[1,2,3,4], Shilong Zhang[5], Xuankai Wang[5], Xiangyu Yang[5], Qiang Sun[1,2,4,*], Yafei Mao[5,7,*]

[1] *Institute of Neuroscience*, *State Key Laboratory of Neuroscience*, *Center for Excellence in Brain Science & Intelligence Technology*, *Chinese Academy of Sciences*, *Shanghai* 200031, *China*
[2] *Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology*, *Shanghai* 201210, *China*
[3] *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*
[4] *Key Laboratory of Genetic Evolution & Animal Models*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650201, *China*
[5] *Bio-X Institutes*, *Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders*, *Ministry of Education*, *Shanghai Jiao Tong University*, *Shanghai* 200030, *China*
[6] *Zhiyuan College*, *Shanghai Jiao Tong University*, *Shanghai* 200240, *China*
[7] *Center for Genomic Research*, *International Institutes of Medicine*, *Fourth Affiliated Hospital*, *Zhejiang University*, *Yiwu*, *Zhejiang* 322000, *China*

ABSTRACT

Understanding gene expression variations between species is pivotal for deciphering the evolutionary diversity in phenotypes. Rhesus macaques (*Macaca mulatta*, MMU) and crab-eating macaques (*M. fascicularis*, MFA) serve as crucial nonhuman primate biomedical models with different phenotypes. To date, however, large-scale comparative transcriptome research between these two species has not yet been fully explored. Here, we conducted systematic comparisons utilizing newly sequenced RNA-seq data from 84 samples (41 MFA samples and 43 MMU samples) encompassing 14 common tissues. Our findings revealed a small fraction of genes (3.7%) with differential expression between the two species, as well as 36.5% of genes with tissue-specific expression in both macaques. Comparison of gene expression between macaques and humans indicated that 22.6% of orthologous genes displayed differential expression in at least two tissues. Moreover, 19.41% of genes that overlapped with macaque-specific structural variants showed differential expression between humans and macaques. Of these, the *FAM220A* gene exhibited elevated expression in humans compared to macaques due to lineage-specific duplication. In summary, this study presents a large-scale transcriptomic comparison between MMU and MFA and between macaques and humans. The discovery of gene expression variations not only enhances the biomedical utility of macaque models but also contributes to the wider field of primate genomics.

Keywords: Crab-eating macaques; Rhesus macaques; Comparative transcriptomics; Biomedical models; Non-human primates; RNA-seq; Duplicated genes

INTRODUCTION

Cross-species transcriptomic comparisons play a crucial role in advancing our understanding of gene expression patterns across species (Breschi et al., 2017; Cardoso-Moreira et al., 2019), as well as the relationships between genetic variation and gene expression disparities (Wang et al., 2020). When applied to primates, such comparisons offer valuable insights into primate evolution, human diseases, and primate genome heterogeneity (Guo et al., 2023; He & Mao, 2023; He et al., 2017; Krienen et al., 2020; Yan et al., 2020).

The primate genus *Macaca*, belonging to the Old World monkeys, is of notable importance due to its wide geographic distribution and strong adaptability (Fan et al., 2017; Shao et al., 2023). Rhesus macaques (*M. mulatta*, MMU) and crab-

Received: 09 October 2023; Accepted: 28 December 2023; Online: 29 December 2023
Foundation items: This work was supported by the National Natural Science Foundation of China (82021001 and 31825018 to Q.S., 32370658 to Y.M., 82001372 to X.Y.), National Key Research and Development Program of China (2022YFF0710901) and National Science and Technology Innovation 2030 Major Program (2021ZD0200900) to Q.S., and Shanghai Pujiang Program (22PJ1407300) and Shanghai Jiao Tong University 2030 Initiative (WH510363001-7) to Y.M.
*Corresponding authors, E-mail: qsun@ion.ac.cn; yafmao@sjtu.edu.cn

eating macaques (*M. fascicularis*, MFA) are two extensively utilized non-human primate (NHP) models in biomedical research (Juan et al., 2023; Lan et al., 2020; Ma et al., 2022; Zhang et al., 2020). Nonetheless, prior comparative transcriptomic studies on these macaques have been confined to a limited number of tissues (Blake et al., 2020; Yan et al., 2011), thus restricting our understanding of the gene expression patterns within and between these two species.

Macaques share a close genetic lineage with humans, diverging approximately 25 million years ago, whereas rodents diverged from humans around 70 million years ago (Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007). Consequently, macaques exhibit similarities to humans in physiology, neurobiology, and disease susceptibility. Despite this, the specific gene expression differences between macaques and humans have yet to be fully explored. To address this gap, we sequenced the transcriptomes of 84 samples (41 MFA and 43 MMU samples) across 14 tissues from both macaque species. This comprehensive sequencing effort aimed to (1) establish the global gene expression patterns and identify tissue-specific genes in both species, (2) identify differentially expressed genes (DEGs) between the two species, and (3) identify DEGs between humans and macaques and explore the correlation between DEGs and lineage-specific structural variations.

MATERIALS AND METHODS

Ethics statement

The use of experimental animals in this study was subject to thorough review and approval by the Primate Life Sciences Ethics Committee of the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences (CEBSIT-2021074). Appropriate anesthesia and euthanasia protocols were employed to reduce discomfort during surgical interventions and to ensure pain-free euthanasia methods for animals when necessary. Prior to euthanasia by exsanguination, monkeys were anesthetized using a Zoletil 50 injection at a dose of 25 mg/kg.

Collection of animal tissues

A total of eight female MFA and eight female MMU were involved in this study. The use and care of animals complied with the guidelines of the Center for Excellence in Brain Science and Intelligence Technology. The animal tissues used in this study were collected from naturally deceased and euthanized monkeys for teaching purposes, in accordance with ethical guidelines. All tissues were sourced from adult monkeys, and samples were checked at the time of collection to verify the absence of any pathological indications, thus minimizing the potential influence of age, sex, and physiological conditions on gene expression. Most tissues used in the study originated from the same individuals, except for blood samples. Blood samples were collected from six healthy monkeys of similar age. To ensure sample integrity, tissue collection was carried out within eight hours of death (Supplementary Table S1). Following collection, the samples were immediately snap frozen in liquid nitrogen and stored in a liquid nitrogen environment until RNA extraction.

RNA-seq library preparation and sequencing

In total, 84 samples from 14 different tissues were collected for Illumina next-generation sequencing. Total RNA was extracted with TRIzol reagent and assessed using a NanoDrop spectrophotometer. Subsequently, mRNA was isolated from 3 μg of total RNA. After fragmentation, synthesis of both first- and second-strand cDNA was carried out. Adapters were ligated to cDNA fragments, 400–500 bp in length, which were then purified using the AMPure XP system. The final library was amplified using polymerase chain reaction (PCR), quantified using the Bioanalyzer 2100 system (Agilent, USA), and sequenced on the NovaSeq 6000 platform (Illumina, USA).

RNA-seq samples in humans

Normalized gene expression data (transcripts per million, TPM) of all human RNA-seq samples, previously analyzed by the GTEx (v.8) consortium, were obtained from the GTEx portal (https://gtexportal.org/home/datasets) (The GTEx Consortium, 2015). For comparison with the macaque samples, a total of 84 human samples corresponding to the same 14 tissues were selected. Analysis of tissue-specific gene expression patterns from the 84 human samples showed consistency with the findings from the entire dataset of 4 650 samples (Supplementary Figure S1A, B). In addition, to ensure robustness, 84 human samples were randomly selected multiple times from the downloaded dataset of 4 650 samples across the 14 different tissues. The DEGs in brain tissues between humans and macaques from three independent sample selections demonstrated consistency (Supplementary Figure S1C). Using the BioMart tool in Ensembl (v.110), 15 000 one-to-one orthologous genes were identified between humans and macaques. Given the similarity in gene expression profiles across the five brain tissues, including neocortical areas (prefrontal and visual cortices) and subcortical regions (striatum, hippocampus, and thalamus), they were consolidated into one aggregated brain tissue group for subsequent analyses. However, we could not completely exclude sex and other biases when integrating the human data, a common issue in current comparative transcriptomic studies (Yao et al., 2022; Zheng-Bradley et al., 2010).

RNA-seq data processing

A total of 380 billion raw reads obtained from the 84 libraries were used for analysis. First, fastp (v.0.22.0) was used to control the quality of the raw reads and generate clean reads (Chen et al., 2018). All clean reads were then mapped to the rheMac10(*M. mulatta*) reference genome using HISAT2 (v.2.2.0) with the parameters "--dta --new-summary -p 10 -x" (Kim et al., 2015; Warren et al., 2020). Aligned reads were converted to BAM files and sorted based on genomic position using SAMtools (v.1.6.0) (Li et al., 2009). SAMtools was also used to remove mitochondrial RNA in the BAM files. The BAM files were then used to generate a count matrix using the R package Rsubread (v.2.12.0) with the "featurecount" function (Liao et al., 2019). The "DGEList" and "rpkm" functions in the R package edgeR were used to normalize the count matrix (Robinson et al., 2010). Subsequently, TPM values were calculated from Fragments Per Kilobase of transcript per Million mapped reads (FPKM) using the formula $\log(\mathrm{FPKM}) - \log(\mathrm{sum}(\mathrm{FPKM})) + \log(1e^6)$.

Sample clustering and quality control

To validate the reproducibility of individuals within each tissue type and to examine outliers with substantial within-group variations, gene expression correlations between samples were calculated. During quality control, 10 unqualified tissue

samples were removed, with 84 samples ultimately retained for subsequent analyses due to their similar gene expression patterns within each group and minimal individual variations, including 41 MFA and 43 MMU samples (Supplementary Figure S2 and Table S2). Median gene expression in each MFA and MMU tissue was then calculated to represent the gene expression value in specific tissues. Hierarchical clustering of the median expression correlation matrix was performed using pheatmap (v.1.0.12)(Kolde, 2019) to illustrate the correlations among MFA and MMU tissues. In addition, dimensionality reduction was applied to the gene expression data from different samples using PCAtools (v.2.8.0) with the parameters "mat=expression, metadata=sample_df, removeVar=0.1". The pheatmap package was also used to determine the relationships between human and macaques tissues based on median expression values, as described above (Blighe et al., 2019).

Differential gene expression and tissue-specific gene analysis

To identify tissue-specific genes, raw count data of samples from each tissue were compared to those in other tissues using DESeq2 (v.1.36.0) (Love et al., 2014). The DESeq2 package was also used to detect species-specific genes between MFA and MMU. Genes with expression levels of zero were removed using the code "dds<-dds[rowSums (counts(dds))>1,]". DESeq2 was used to calculate adjusted *P*-values using the Benjamini-Hochberg method, with $\log_2$fold-change ($\log_2$FC)>1.5 and false discovery rate (*P*-adjusted)<0.05 applied as the threshold criteria to identify tissue- and species-specific genes. Hypergeometric tests were performed to evaluate the significance of overlapping tissue-specific genes in the two species using the "phyper" function, with *P*-values adjusted using the Benjamini-Hochberg method. Gene Ontology (GO) analysis of DEGs was performed using the "enrichGO" function in clusterProfiler with the parameters "pAdjustMethod='BH', ont='BP', pvalueCutoff=0.05, qvalue Cutoff=0.05)" (Wu et al., 2021).

Differential gene expression between humans and macaques across 10 tissues was analyzed using the limma (v.3.52.4) package (Ritchie et al., 2015), as the database-obtained human TPM matrix was pre-normalized, not permitting further normalization through DESeq2. The TPM matrix of the 15 000 human-macaque orthologous genes was analyzed using the "model.matrix", "lmFit", "contrasts.fit", "eBayes", and "topTable" functions in the limma package. The TPM matrix was used to identify tissue- and species-specific genes in both humans and macaques. To identify tissue-specific genes, $\log_2$FC>1.5 and *P*-adjusted<0.05 threshold criteria were utilized. In contrast, more stringent threshold criteria ($\log_2$FC>1.5 and *P*-adjusted<0.001) were used to identify species-specific genes to minimize intra-group variances and select for genes exhibiting higher significance. All other analyses followed similar aforementioned procedures.

Intersection of DEGs and macaque-specific structural variants (SVs)

Macaque-specific SVs in comparison to the human genome (hg38)weredownloadedfromapublicdatabase(https://eichlerlab.gs.washington.edu/help/yafmao/public_html/ NHP_Genomes_ scripts/lineage_specific_SV_with_annotation/) (Mao et al., 2024). Summary data of DEGs identified in at least two tissues were also generated. The overlap between this gene set and SVs was examined using Bedtools (v.2.30.0) (Quinlan & Hall, 2010), with the significance of this overlap determined using a permutation test. Initially, 5 000 sets of randomly permuted SVs were generated, while preserving original SV lengths, using the "shuffle" function in Bedtools. Bedtools was then used to intersect the genes overlapping the macaque-specific SVs. Finally, the statistical significance (*P*-values) of the findings was determined by ranking the observed and permuted values.

Human-specific duplicated gene analysis

Human cortex RNA-seq data were downloaded from the NCBI database (GSE174409 and GSE157428) (Supplementary Table S3) (Otero et al., 2021; Seney et al., 2021). The RNA-seq data were mapped to the complete human genome (T2T-CHM13) (Nurk et al., 2022). The data analysis workflow was consistent with previous analysis. The duplicated gene set was obtained from previous study (Supplementary Table S4) (Dennis et al., 2017). Gene expression between ancestral duplicated genes in macaques and orthologous duplicated genes in humans (e.g., *CD8B*_macaques vs. *CD8B*_humans) was first compared, with read counts of the human-specific duplicated genes (e.g., *CD8B_CD8B2*_humans) then aggregated and compared to gene expression of ancestral orthologous duplicated genes in macaques (e.g., *CD8B*_macaques) and aggregated duplicated genes in humans (e.g., *CD8B_CD8B2*_humans).

Gene expression validation with qPCR

Briefly, total RNA was extracted using TRIzol reagent following the manufacturer's instructions and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher, USA). Gene expression levels were measured using the Roche 480II Real-Time PCR System (Roche, Germany). Reverse transcription quantitative PCR (RT-qPCR) was then carried out with specific primers for target genes, using *GAPDH* as an internal control. Finally, relative expression levels were determined using the $2^{(-\Delta\Delta CT)}$ method (Livak & Schmittgen, 2001). The RT-qPCR primers are listed in Supplementary Table S5.

Evolutionary history of *FAM220A* analysis

Copy number variation summary data were downloaded from a public database. Bedtools and custom scripts were then used to select genes with copy number variations between humans and macaques and gene expression differences in at least two tissues. Minimiro was also used to determine the syntenic relationship between apes and macaques in the *FAM220A* locus. The *FAM220A* region was extracted and mafft was used to align the sequences. Finally, BEAST2 (v.2.7.5) was used to examine the time of duplicated events (Bouckaert et al., 2014).

Statistical analysis

Statistical analysis was performed using R software (v.4.2.3) (https://www.R-project.org/). Most figures in the paper were generated using ggplot2 (Villanueva & Chen, 2019). *P*-adjusted<0.05 was considered significant for all analyses.

RESULTS

Global gene expression patterns in two macaque species

In total, approximately 570 Gb of transcriptomic data were generated from the 84 samples of 14 different tissue types collected from eight female MFA individuals and eight female

MMU individuals (Supplementary Tables S1, S2). The 14 tissues included six viscera (heart, liver, lung, kidney, ovary, and spleen), six brain tissues (prefrontal cortex, visual cortex, striatum, hippocampus, thalamus, and cerebellum), muscle, and blood. To reduce bias related to age, physical conditions, and other factors, all included macaques were female adults older than 4 years of age. Additionally, all tissue samples were carefully selected to ensure there were no signs of disease present. The RNA integrity number (RIN) and GC content of the samples were comparable to standard values, indicating the reliability and quality of our data (Supplementary Table S6). After data filtering (see Methods), 99.55% of clean reads were retained and each sample showed a proper duplication rate (mean: 32.58%), except for heart tissues in rhesus macaques ($mean_{MMU}$=55.05%), potentially owing to the higher content of mitochondrial RNA in heart tissues (mean=3.98%) (Supplementary Table S6). All sequencing reads, excluding mitochondrial RNA, were aligned to the high-quality long-read assembly (rhesus macaque: Mmul_10), with an average alignment rate of 88.36% (Table 1) (Warren et al., 2020). Finally, a TPM matrix was generated containing 84 samples and 33 960 genes (Supplementary Table S7).

Hierarchical clustering analysis of median gene expression across the 14 tissues showed that most samples were grouped by different tissues, except for the prefrontal and visual cortices, which were grouped by species. Interestingly, brain tissues, including neocortical regions (prefrontal and visual cortices), subcortical regions (striatum, hippocampus, and thalamus), and the cerebellum, showed the most similar expression patterns between the two species, as well as distinct expression patterns compared to other tissues (Figure 1A). Low variability was observed within each cluster (Supplementary Figure S2). In addition, the number of expressed genes in each tissue was significantly correlated between the two macaque species (Spearman, R^2=0.76, P<2.2e^{-16}) (Figure 1B). Principal component analysis (PCA) further showed that PC1 accounted for the differences between brain tissues and other tissues (MFA: 34.80% variation, MMU: 36.64% variation), while PC2 explained the variation among different visceral organs (MFA: 14.75% variation, MMU: 13.05% variation) in both species (Figure 1C).

Based on gene models from the long-read macaque genome assembly Mmul_10 (Warren et al., 2020), 33 634 genes (99.12%) were expressed in our dataset. Nonetheless, in the 14 tissues analyzed, 349 genes in MFA and 356 genes in MMU were found to have no detectable expression (TPM<0.1). Of these, 296 non-expressed genes shared by both species (84.8% in MFA vs. 83.1% in MMU, hypergeometric test, P<2.2×10^{-16}) were significantly enriched in male reproductive processes (e.g., spermatogenesis and male gamete generation) (P-adjusted=0.01), defense responses to bacteria (P-adjusted=0.003), and sensory perception (P-adjusted=0.01) (Supplementary Figure S3 and Table S8). The non-detectable expression of these genes may be due to the lack of skin tissues or male reproductive tissues collected in this study. Consistent with previous research (Huh et al., 2012; Yao et al., 2022), ovarian (#gene mean=21 586) and brain tissues (#gene mean=21 714) expressed a greater number of genes, while muscle (#gene mean=18 086) and liver tissues (#gene mean=17 845) expressed fewer genes (Figure 1D).

Tissue specificity of gene expression in macaques

Despite the remarkably similar gene expression patterns observed between MFA and MMU, these distinct species diverged approximately 2–3 million years ago (Zhang et al., 2023). This evolutionary separation raises questions about the differences in their gene expression regulation. Our examination of these differences initially focused on the degree of divergence in tissue-specific gene expression between these two species, revealing significant overlap in tissue-specific genes within the same tissues. Notably, among the 14 tissues examined, blood demonstrated the highest diversity in tissue-specific genes (n_{MFA}=2 381, n_{MMU}=1 635, $n_{overlap}$=1 309), while the prefrontal cortex demonstrated the lowest tissue-specific expression (n_{MFA}=15, n_{MMU}=20, $n_{overlap}$=5) (Figure 2A). Tissue-specific genes in each tissue were almost identical between the two macaque species (Figure 2B). The number of tissue-specific genes was also significantly correlated between MFA and MMU (Spearman, R^2=0.92, P<2.2e^{-16}) (Figure 2C).

Relatively similar gene expression patterns were observed in the five different brain tissues, explaining the reduction in the identification of tissue-specific genes within each brain tissue. Consequently, the five brain tissues were integrated for

Table 1 Macaque transcriptome data

Average	Sample size (n)		Total data (Gbp)		Overall aligned rate (%)		Read duplication rate (%)	
Tissue/Species	MFA	MMU	MFA	MMU	MFA	MMU	MFA	MMU
Blood	3	3	7.01	6.57	90.59	94.63	25.65	26.62
Lung	3	3	6.47	6.39	90.14	94.44	21.25	34.18
Liver	3	3	6.86	7.19	88.85	92.02	29.25	52.66
Hippocampus	3	3	6.41	7.45	83.38	92.69	22.25	36.92
Muscle	3	3	6.09	7.15	84.47	95.22	30.57	42.27
Ovary	2	3	6.70	6.71	91.09	93.88	22.19	40.02
Spleen	3	3	6.16	7.14	88.64	91.98	25.36	48.65
Prefrontal cortex	3	3	6.82	7.07	81.70	92.26	26.17	37.38
Thalamus	3	3	6.71	7.42	82.83	93.34	24.74	37.38
Kidney	3	3	7.07	6.63	83.50	93.06	26.57	34.81
Visual cortex	3	3	6.48	6.81	79.84	90.52	26.33	41.26
Striatum	3	3	6.09	7.80	79.18	92.37	23.60	38.31
Cerebellum	3	3	6.55	7.26	85.73	94.43	23.22	32.85
Heart	3	4	6.47	6.58	69.80	83.40	31.09	51.68

MMU: Rhesus macaques (*Macaca mulatta*); MFA: Crab-eating macaques (*M. fascicularis*).

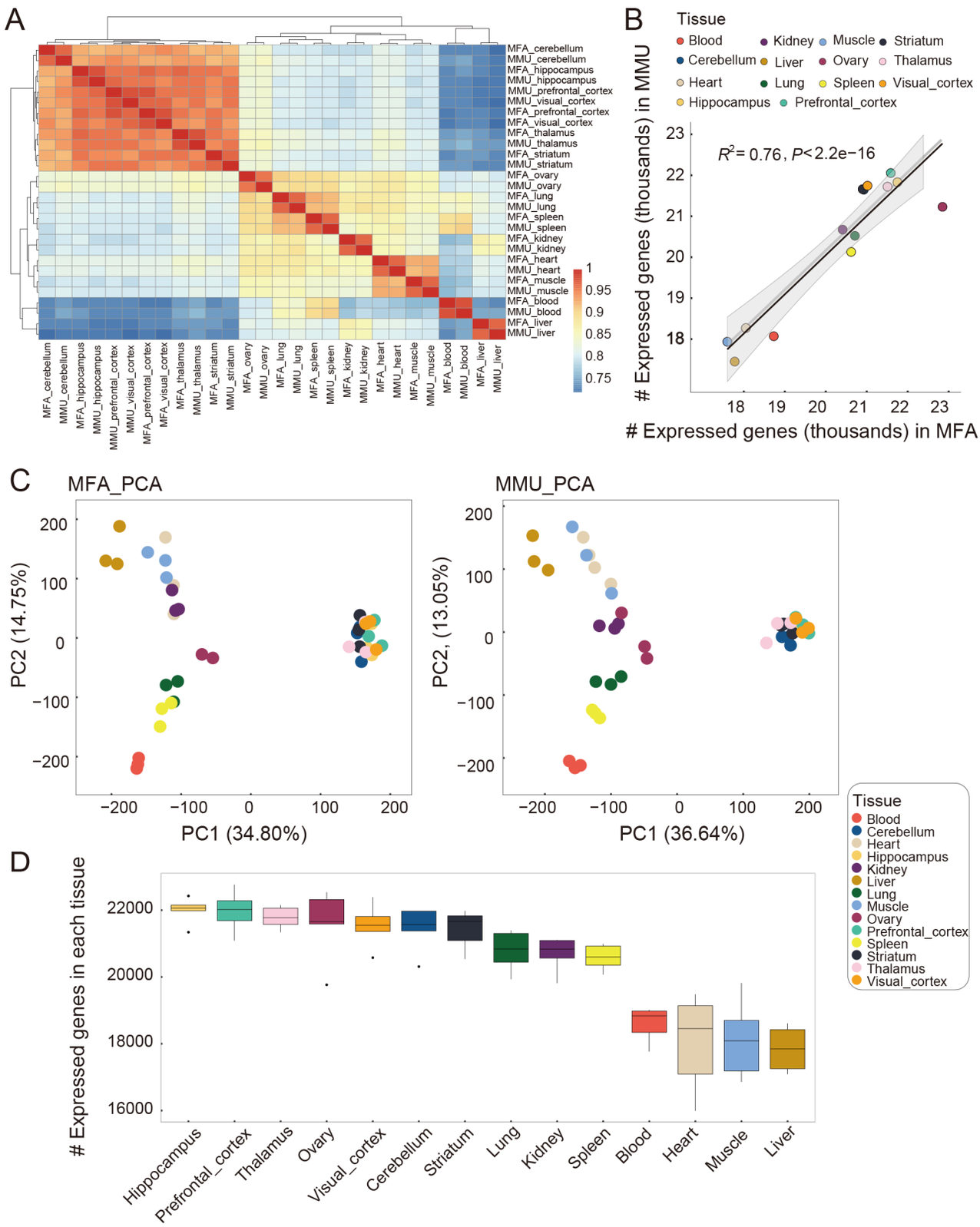

Figure 1 Quality control and global gene expression patterns in macaque tissues

A: Hierarchical clustering of 14 tissues in crab-eating macaques (MFA) and rhesus macaques (MMU) based on Pearson's correlation of median gene expression value. B: Spearman correlation of number of expressed genes (median TPM>0.1) between MFA and MMU. Gray area represents 95% confidence interval. Each dot represents a tissue. C: Principal component analysis of gene expression. Each dot represents a sample, colored by tissue types. Left panel refers to 41 MFA samples, right panel refers to 43 MMU samples. D: Number of expressed genes (TPM>0.1) identified in different tissues. In each boxplot, black line represents median; gray dots represent outliers.

further analysis, resulting in the identification of a significant number of brain-specific genes (n_{MFA}=3 467, n_{MMU}=3 477), including 2 778 genes shared between MFA and MMU (79.89% and 80.12%). Visceral tissues also showed

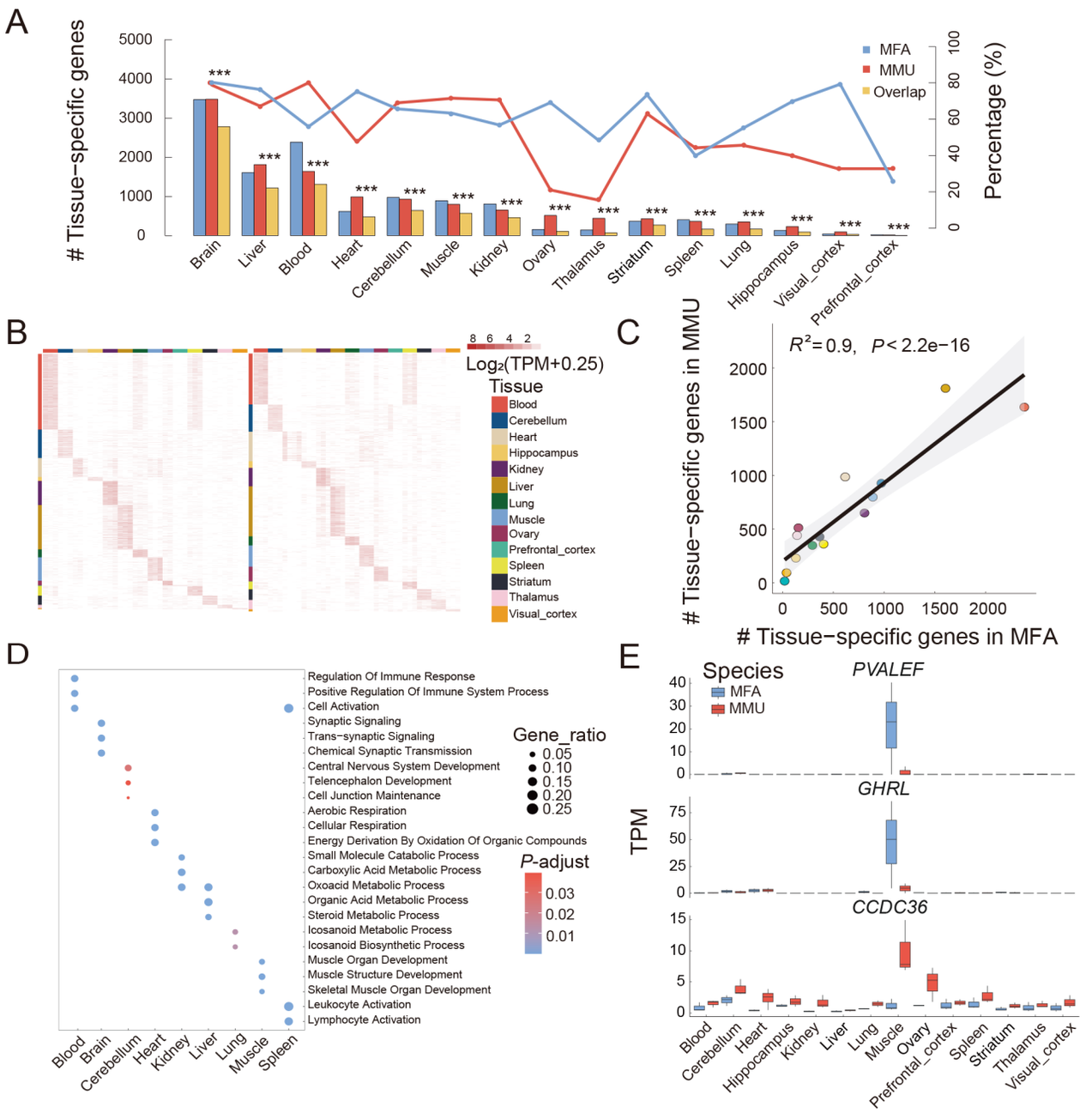

Figure 2 Comparison of tissue specificity of gene expression in macaques

A: Number of tissue-specific genes ($log_2FC>1.5$ and *P*-adjusted<0.05) in MFA (blue) and MMU (red). Overlapping tissue-specific genes are in yellow. ***: *P*<0.001. Lines show percentage of overlapping tissue-specific genes in each species (blue: MFA; red: MMU). B: Expression profiles of tissue-specific genes in MFA (left) and MMU (right). Each row represents a gene, and each column represents a sample from the corresponding tissue. C: Spearman correlation of number of tissue-specific genes across 14 tissues between MFA and MMU. Gray area represents 95% confidence interval. Each dot represents a tissue. D: Gene Ontology (GO) enrichment analysis of overlapping tissue-specific genes. Top three most significant functional enrichment terms are shown in each tissue. E: Expression profiles of *NTRK2*, *SYN1*, and *SNAP25*.

substantial overlap, with rates ranging from 40% to 80%. In contrast, ovarian tissues showed lower overlap (69.23% in MFA, 21.14% in MMU) (Figure 2A), which may stem from differences in the physiological status of the ovaries, such as the estrus phases and ovulation timing. GO enrichment analysis of ovary-specific genes revealed divergent patterns between MMU and MFA, with enrichment in extracellular matrix organization and collagen fibril organization in MFA and enrichment in reproductive pathways such as cell division and cell cycle process in MMU (Supplementary Figure S4) (Niwa et al., 2015).

The tissue-specific genes shared between the two macaques closely mirrored those in humans, indicating evolutionary conservation (Supplementary Figure S5). Notably, two different troponin genes, *TNNT2* and *TNNC2*, were specifically expressed in the heart and muscle tissues, respectively, of the two macaque species, as also observed in humans (Supplementary Figure S6). The tissue-specific genes shared by these species reflect the conserved biological function of these tissues (Figure 2D; Supplementary Table S9). For example, brain-specific genes in MFA and MMU were enriched in chemical synaptic transmission and neuron

differentiation (Supplementary Figure S7).

Notably, 3 152 and 2 916 tissue-specific genes were uniquely expressed in MMU and MFA, respectively. For example, *PVALEF* and *GHRL* were uniquely expressed in the muscle tissues of MFA, while *CCDC36* was uniquely expressed in the muscle tissues of MMU (Figure 2E). Interestingly, MFA-unique brain-specific genes were enriched in odontogenesis and dentin formation, whereas MMU-unique brain-specific genes were enriched in potassium ion import across the plasma membrane (Supplementary Table S10).

DEGs between macaque species

Analysis of DEGs between MFA and MMU revealed that the brain, blood, and lung tissues contained the highest number of DEGs (Figure 3A; Supplementary Table S11). In contrast, the heart tissue exhibited the lowest number of DEGs (Supplementary Figure S8), consistent with previous research suggesting slower evolution of heart tissues in mammals (Yao et al., 2022). Blood and spleen tissues demonstrated higher intra-group variability compared to other tissues (Supplementary Figure S9). These findings suggest that differences in gene expression in these tissues may be influenced by individual immune status variations rather than by species divergence alone. Although lung tissues exhibited significant differences in immune-related gene expression, the macaques in this study were free from viral infections and their lung tissues showed no anatomical lesions (Supplementary Figure S10). Thus, we propose that the DEGs in lung tissues may reflect potential species-level differences between MMU and MFA or their different responses to viral infection (Flynn et al., 2015; Hewitt et al., 2020).

Combining all brain tissue data for DEG analysis revealed 441 genes up-regulated in MFA (e.g., *ATP1B2*) and 356 genes up-regulated in MMU (e.g., *PDE11A*, and *NCR3LG1*) (Figure 3B). Notably, several of these genes are related to neurogenic diseases in humans. For instance, *ATP1B2* is involved in maintaining membrane potential and is down-regulated in schizophrenic patients (Pan et al., 2022), while *PDE11A* expression is significantly decreased in the brains of Alzheimer's patients (Qin et al., 2021). These findings underscore the importance of considering expression divergence in disease-associated genes in brain tissues when choosing between MFA and MMU as an animal model for specific neurological disorders.

MFA-up-regulated genes in the liver and kidney, two metabolism-related tissues, were significantly enriched in

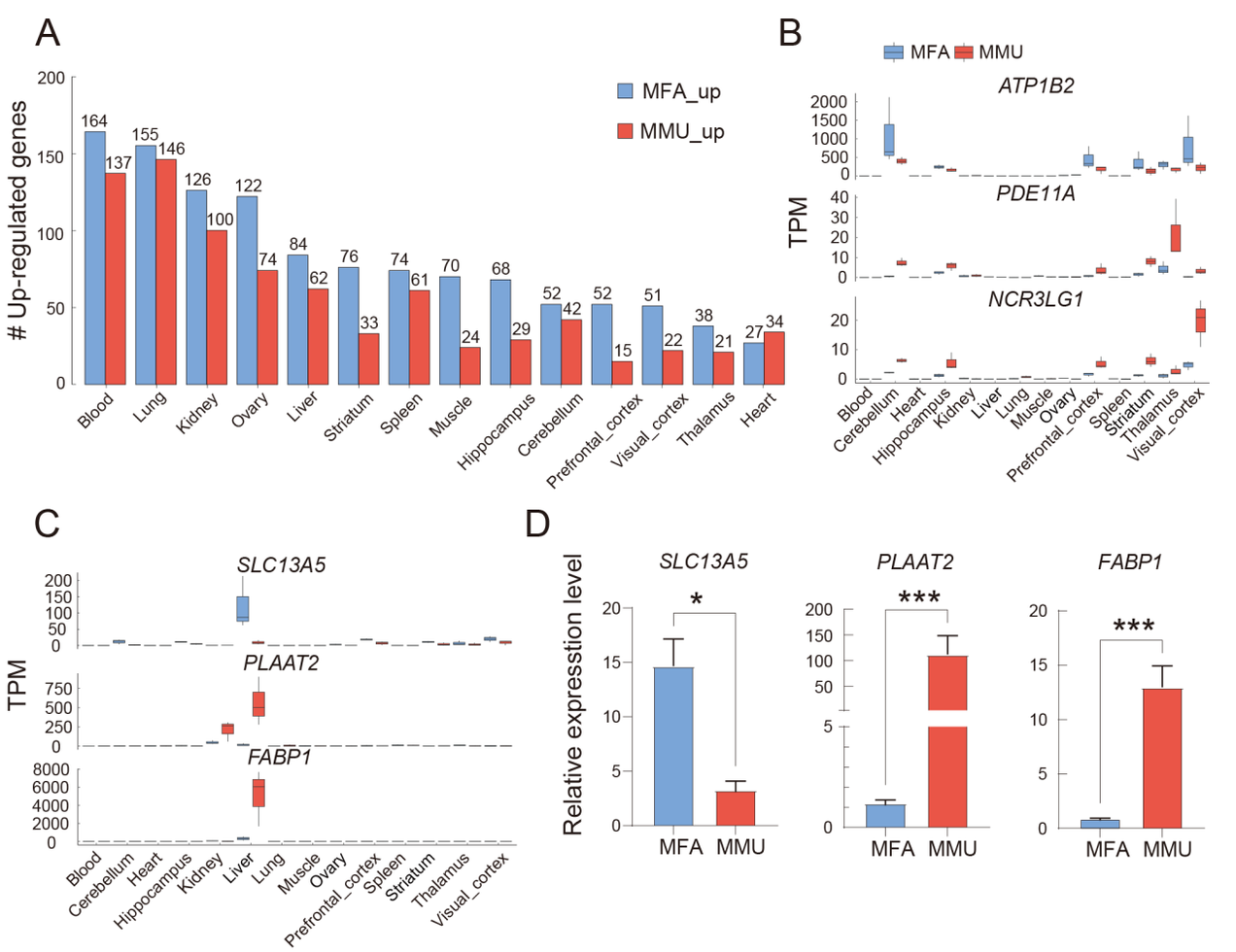

Figure 3 Comparison of gene expression across 14 tissues between two macaques

A: Number of significantly up-regulated genes in MFA (blue) and MMU (red) (log_2FC>1.5 and *P*-adjusted<0.05). B: Expression profiles of brain DEGs (*ATP1B2*, *PDE11A*, and *NCR3LG1*) in MFA (blue) and MMU (red). C: Expression profiles of liver DEGs (*SLC13A5*, *PLAAT2*, and *FABP1*) in MFA (blue) and MMU (red). D: Quantitative real-time PCR validation of three liver DEGs (MFA: *n*=3; MMU: *n*=2). Ctrl: MFA (*PLAAT2* and *FABP1*), MMU (*SLC13A5*). Data represent three biological repeats, shown as mean±standard error of the mean (SEM). Statistical significance was calculated using Wilcoxon test. *: *P*<0.05; ***: *P*<0.001.

amino acid and organic acid transport pathways, while MMU-up-regulated genes were enriched in pathways related to organic hydroxy compound transport (Supplementary Table S12). These findings suggest that MFA and MMU may exhibit differences in metabolism. For example, *SLC13A5*, which is involved in citrate metabolic processes, was significantly up-regulated in the MFA liver compared to the MMU liver (Milosavljevic et al., 2022). In addition, *PLAAT2* and *FABP1*, related to phospholipids and fatty acid metabolic processes (Zare-Feyzabadi et al., 2022; Zhou et al., 2020), respectively, were highly expressed in the MMU liver but not in the MFA liver (Figure 3C). The RNA-seq results of these three genes were validated by quantitative reverse transcription PCR (RT-qPCR) (Figure 3D). These findings highlight distinct metabolic differences between MFA and MMU, further emphasizing the importance of considering species-specific metabolic variations when using these macaque models in metabolism-related studies.

Comparative transcriptomics between macaques and humans

To better understand the utility of macaques as biomedical models (Zhang et al., 2014), the transcriptomes of 84 human samples from the same 14 tissues (Supplementary Figure S1) were downloaded (GTEx Analysis Release v.8) and compared to our macaque transcriptome data (The GTEx Consortium, 2015). Brain, lung, spleen, heart, and muscle tissues were clustered by species rather than by tissue type (Figure 4A). In contrast, ovary, kidney, and liver tissues were clustered by tissue type rather than by species. Gene expression between macaques and humans across the 14 tissues was highly correlated (correlation coefficient: 0.86–0.92) (Figure 4B). These findings highlight the potential of macaque models to closely mimic human gene expression patterns across various tissues (Bakken et al., 2016).

Brain tissues showed the largest number of tissue-specific genes in macaques, while heart tissues had the fewest tissue-

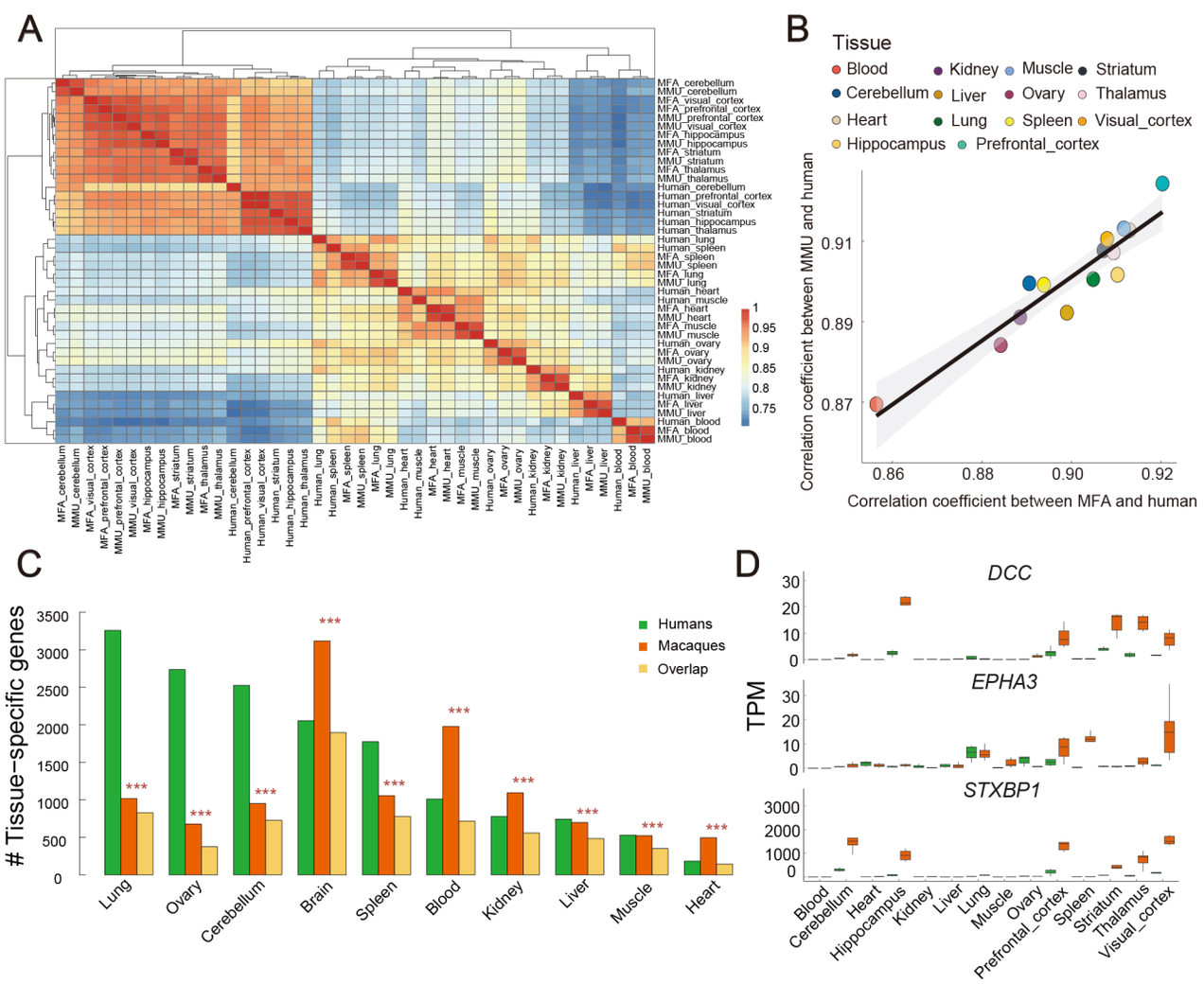

Figure 4 Correlation and tissue specificity humans and macaques

A: Hierarchical clustering of 14 tissues in humans and macaques based on Pearson correlation of median gene expression. B: Scatter plot of correlation coefficient between MFA and humans and between MMU and humans. Gray area represents 95% confidence interval. Each dot represents a tissue. C: Number of tissue-specific genes (log_2FC>1.5 and *P*-adjusted<0.05) and overlapping genes between humans (green) and macaques (orange). Overlapping tissue-specific genes are in yellow. Overlap significance was tested using hypergeometric test. ***: *P*<0.0001. D: Gene expression profiles of three human neurological disease-associated genes (*DCC*, *EPHA3*, and *STXBP1*) in humans (green) and macaques (orange).

specific genes both in humans and macaques (Figure 4C). The overlapping brain-specific genes included many critical human neurological disease-associated genes (e.g., *DCC*, *EPHA3*, and *STXBP1*) (Figure 4D). Notably, *DCC* guides axonal growth during neural development and is associated with various neurological diseases (Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders, 2019), while *STXBP1* is crucial for neurotransmitter release, with MFA carrying *STXBP1* mutations showing focal epilepsy as a primate model of human genetic disorders (Lu et al., 2022; Yang et al., 2023). These findings further underscore the relevance of macaques as models for studying neurological diseases and advancing our understanding of human health.

Gene expression differences between macaques and humans were further explored. Comparative analysis between macaques and humans indicated that brain tissues exhibited the highest number of DEGs, while ovarian tissues exhibited the fewest (Figure 5A). From a panel of 15 000 one-to-one orthologous genes between humans and macaques, 7 652 DEGs were identified, including 3 394 genes showing differential expression in at least two tissues (Supplementary Table S13). Certain genes were differentially expressed across multiple tissues, including *PSMB2*, which is involved in protein catabolism (Liu et al., 2022); *POLR2F* and *KAT14*, which are associated with transcriptional regulation (Antonacopoulou et al., 2008; Viita et al., 2019); and *FAM220A*, which is related to positive regulation of protein binding activity (Figure 5B) (Yong et al., 2023).

Analysis of the intersection between macaque-specific SVs and DEGs showed that approximately 19.41% of genes overlapping with macaque-specific SVs were more likely to be differentially expressed (n_{DEG}=2 601, permutation test, P=0, see methods) (Figure 5C). These DEGs were primarily

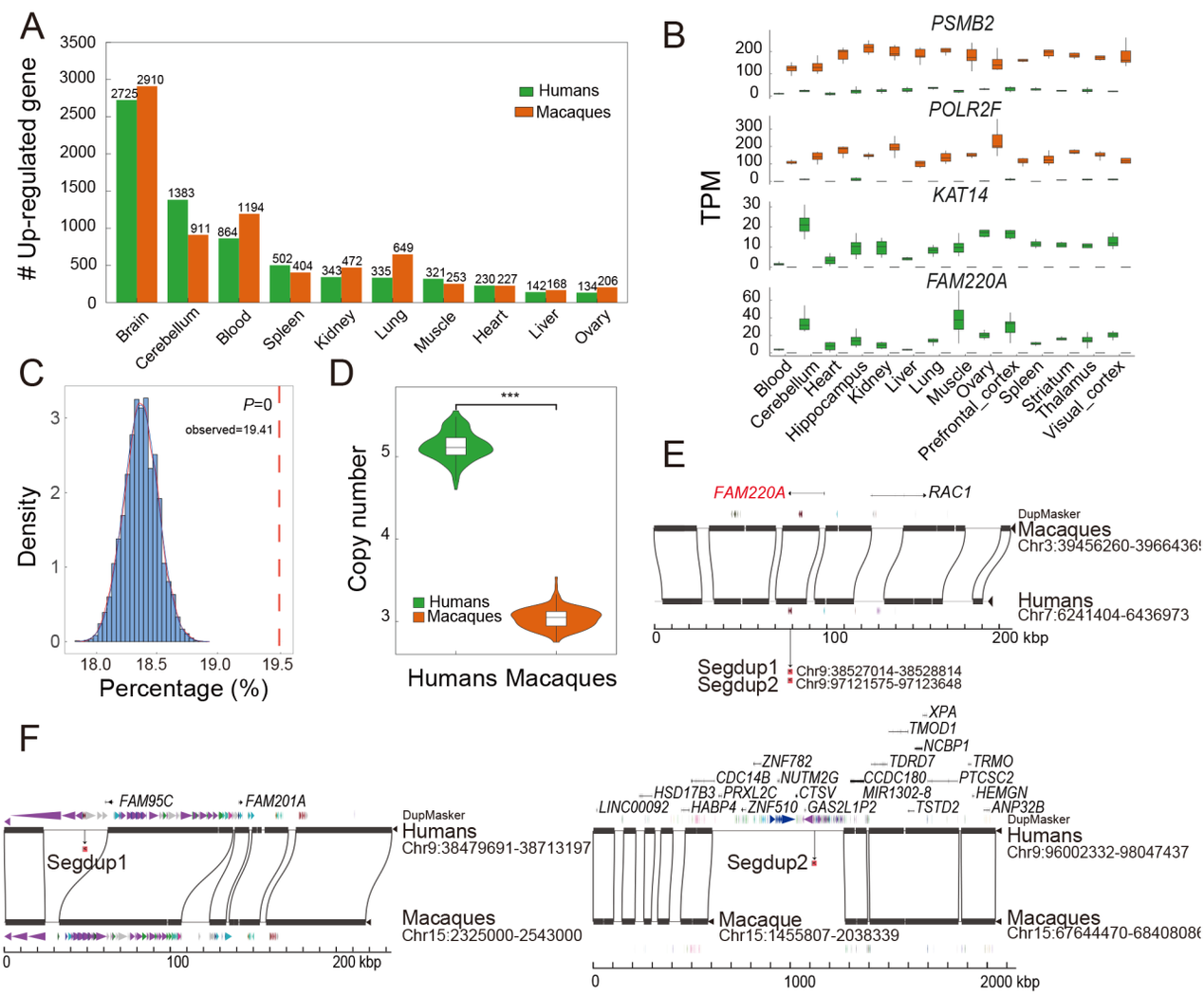

Figure 5 Differential gene expression between humans and macaques

A: Number of significantly up-regulated genes in humans (green) and macaques (orange) ($\log_2$FC>1.5 and P-adjusted<0.001). B: Expression profiles of DEGs (*PSMB2*, *POLR2F*, *KAT14*, and *FAM220A*) between humans (green) and macaques (orange). C: Percentage distribution of intersected DEGs with lineage-specific SVs over total number of genes intersected with lineage-specific SVs (n=5 000). Red dashed line represents observed value: percentage of DEGs intersected with lineage-specific SVs over number of genes intersected with lineage-specific SVs (empirical P=0). D: Copy number distribution of *FAM220A* in humans (green, n=208) and macaques (orange, n=132). Statistical significance was calculated using Wilcoxon test. ***: P<0.001. E: Syntenic relationship of *FAM220A* locus between humans (GRCh38) and macaques (Mmul10), displaying gene models and segmental duplications. F: Syntenic analysis of segmental duplication at *FAM220A* locus, highlighting duplication of *FAM220A* in humans compared to macaques (GRCh38 vs. Mmul10).

enriched in RNA regulation, protein synthesis, and metabolism (Supplementary Figure S11). A significant difference was observed in the copy number of *FAM220A* between humans (mean copy number=5.13) and macaques (mean copy number=3.04) (Figure 5D). Further analysis revealed two additional ape-specific duplications of *FAM220A* in humans (Figure 5E, F; Supplementary Figure S12; chr9: 38527014-38528814, chr9: 97121575-97123648), highlighting the impact of SVs on gene expression differences between humans and macaques.

To investigate the expression patterns of human-specific duplicated genes, 173.95 Gb of human cortex bulk RNA-seq data were downloaded (Otero et al., 2021; Seney et al., 2021), along with 20 human-specific duplicated genes identified in previous study (Dennis et al., 2017). Initially, gene expression between ancestral genes in macaques and orthologous duplicated genes in humans was compared. Read counts of human-specific duplicated genes were then aggregated for comparison against the expression of ancestral genes in macaques and aggregated duplicated genes in humans. Among the 20 human-specific duplicated genes, eight duplicated genes showed significant differential expression compared to the ancestral genes in macaques and their orthologous duplicated genes in humans (two up-regulated and six down-regulated in humans) (Supplementary Figure S13A). Comparing ancestral genes in macaques with aggregated duplicated genes in humans, *CD8B* was identified as an up-regulated DEG and *SERF1A* was identified as a down-regulated DEG in humans (Supplementary Figure S13B). These results suggest that the expression patterns of human-specific genes are complex, and these genes potentially evolved rapidly with different fates (Lynch & Conery, 2000).

DISCUSSION

Investigating the whole-body gene expression atlas in MFA and MMU can markedly improve our understanding of both the conservation and heterogeneity of these species (Yao et al., 2022). In the present study, we conducted a comprehensive transcriptomic comparison across 14 tissues in two macaque species, as well as between humans and macaques. This study represents an important initial step in profiling the transcriptomes of macaques.

While MFA and MMU are both prominent NHP biomedical models, they exhibit several phenotypic differences, such as tail length (Rowe, 1996). Our data indicated that 96.3% of genes were consistently expressed between MMU and MFA, with most tissue-specific genes being shared across the 14 tissues. Notably, brain tissues contained the highest number of DEGs between the two macaque species, including several associated with human neurogenic diseases (e.g., *ATP1B2*, *PDE11A*, and *NCR3LG1*). In addition, differences in gene expression related to metabolism (e.g., *SLC13A5*, *PLAAT2*, and *FABP1* in livers) were also observed between the two species. These findings highlight the importance of understanding the divergence in gene expression, particularly in disease-associated genes. However, determining which macaque species is more appropriate as a biomedical model for studying certain human diseases cannot solely be based on transcriptomic differences. Integrated comparison including gene expression levels, gene networks, physiological characteristics, and other factors is essential for reliably determining the most suitable model.

In addition to comparing the two macaque species, comparative transcriptome analysis was also conducted between macaques and humans. Results showed that approximately 49.0% of the one-to-one orthologous genes were coincidently expressed between humans and macaques. Furthermore, tissue-specific genes were highly conserved between humans and macaques. These results underscore the substantial overlap in biological processes between these species, supporting the utility of macaques as biomedical models (Qiu et al., 2019). Of note, brain tissues exhibited the highest number of DEGs between humans and macaques, highlighting potential differences in the evolutionary development of brains in primates (Bakken et al., 2016). These findings contribute to a more comprehensive understanding of the complex interactions among gene expression, evolution, and neurological trait development in humans and macaques.

Based upon prior studies showing the proliferation of segmental duplications in great apes (Mao et al., 2024; Sudmant et al., 2013), we examined the relationship between structural variation and DEGs. Interestingly, 2 601 genes (19.41%) that overlapped with lineage-specific SVs exhibited a propensity for differential gene expression. This observation suggests that SVs are a major contributor to gene regulation, potentially driving the observed expression disparities (He et al., 2019). Of note, analysis of the *FAM220A* gene identified a correlation between copy number variation and gene expression between humans and macaques.

Our study provides a comparative perspective on the transcriptomes of two macaque species, as well as of macaques and humans. However, for a deeper understanding of macaque transcriptomes, it is imperative to expand the number of samples and include a broader range of tissues (Han et al., 2022). Moreover, while bulk RNA-seq has been widely used, it struggles to capture the complex cellular heterogeneity within tissues and identify differences in alternative splicing between species (Stark et al., 2019). Additionally, the inherent limitations of short-read RNA-seq hinder the comprehensive detection of isoform information within transcripts (Leung et al., 2021; Li et al., 2021). The incorporation of long-read sequencing and single-cell RNA-seq technologies holds immense promise for unraveling the complexities of macaque transcriptomes.

DATA AVAILABILITY

The raw RNA-seq data reported in this study are available from the NCBI database (BioProject PRJNA1004471), China National Center for Bioinformation database (PRJCA023919), and Science Data Bank database (DOI: 10.57760/sciencedb.16465).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

AUTHORS' CONTRIBUTIONS

Y.M. and Q.S. conceived the project. Y.X.M. performed the RNA-seq analyses. Y.L. and N.X. helped with sample collection and RNA-sequencing. Z.Y., S.Z., X.Y., and X.W. performed the analysis of *FAM220A*. Y.X.M., X.Y., and N.X. performed the RT-qPCR experiments. Y.X.M. and X.Y. contributed to the editing of the figures. Y.X.M., Q.S., and Y.M. drafted the manuscript. Y.X.M., Q.S., and Y.M. finalized the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Dr. Kaiyue Ma for proofreading help. We thank the Non-Human

Primate Research Facility at the Center for Excellence in Brain Science and Intelligence Technology for providing all macaque samples, and the veterinarians for sample collection. The computations in this paper were run on the cluster supported by the Center for Data and Computing in Brain Science (CDCBS) at the Institute of Neuroscience, Chinese Academy of Sciences; and on the Siyuan-1 and π 2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University.

REFERENCES

Antonacopoulou AG, Grivas PD, Skarlas L, et al. 2008. POLR2F, ATP6V0A1 and PRNP expression in colorectal cancer: new molecules with prognostic significance. *Anticancer Res*, **28**(2B): 1221−1227.

Bakken TE, Miller JA, Ding SL, et al. 2016. A comprehensive transcriptional map of primate brain development. *Nature*, **535**(7612): 367−375.

Blake LE, Roux J, Hernando-Herraez I, et al. 2020. A comparison of gene expression and DNA methylation patterns across tissues and species. *Genome Res*, **30**(2): 250−262.

Blighe K, Lewis M, Lun A, et al. 2019. Package 'PCAtools. '. REAGENT or RESOURCE SOURCE IDENTIFIER Deposited data GitHub repository with code and data for analysis GitHub https: //github. com/CharlotteEPage/NFI_Disease_Ecology Raw data tables Mendeleyhttps://data.mendeley.com/datasets/y869bhhmzr/1 Software and algorithms R version, **4**(1): 9–11.

Bouckaert R, Heled J, Kühnert D, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*, **10**(4): e1003537.

Breschi A, Gingeras TR, Guigó R. 2017. Comparative transcriptomics in human and mouse. *Nat Rev Genet*, **18**(7): 425−440.

Cardoso-Moreira M, Halbert J, Valloton D, et al. 2019. Gene expression across mammalian organ development. *Nature*, **571**(7766): 505−509.

Chen S, Zhou Y, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884−i890.

Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. 2019. *Cell*, **179**(7): 1469–1482. e1411.

Dennis MY, Harshman L, Nelson BJ, et al. 2017. The evolution and population diversity of human-specific segmental duplications. *Nat Ecol Evol*, **1**(3): 69.

Fan P, Liu Y, Zhang Z, et al. 2017. Phylogenetic position of the white-cheeked macaque (*Macaca leucogenys*), a newly described primate from southeastern Tibet. *Mol Phylogenet Evol*, **107**: 80−89.

Flynn JL, Gideon HP, Mattila JT, et al. 2015. Immunology studies in non-human primate models of tuberculosis. *Immunol Rev*, **264**(1): 60−73.

Guo YT, Shao Y, Bi XP, et al. 2023. Harvesting the fruits of the first stage of the Primate Genome Project. *Zoological Research*, **44**(4): 725−728.

Han L, Wei XY, Liu CY, et al. 2022. Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. *Nature*, **604**(7907): 723–731.

He YX, Luo X, Zhou B, et al. 2019. Long-read assembly of the Chinese rhesus macaque genome and identification of ape-specific structural variants. *Nature Communications*, **10**(1): 4233.

He YX, Mao YF. 2023. Exploring the primate genome: Unraveling the mysteries of evolution and human disease. *The Innovation*, **4**(4): 100467.

He ZS, Han DD, Efimova O, et al. 2017. Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nature Neuroscience*, **20**(6): 886−895.

Hewitt JA, Lutz C, Florence WC, et al. 2020. ACTIVating resources for the COVID-19 pandemic: *in vivo* models for vaccines and therapeutics. *Cell Host & Microbe*, **28**(5): 646−659.

Huh JW, Kim YH, Park SJ, et al. 2012. Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research. *BMC Genomics*, **13**(1): 163.

Juan D, Santpere G, Kelley JL, et al. 2023. Current advances in primate genomics: novel approaches for understanding evolution and disease. *Nature Reviews Genetics*, **24**(5): 314−331.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357−360.

Kolde R. 2019. Pheatmap: pretty heatmaps. R Package Version 1.0.12. https://CRAN.R-project.org/package=pheatmap.

Krienen FM, Goldman M, Zhang QG, et al. 2020. Innovations present in the primate interneuron repertoire. *Nature*, **586**(7828): 262−269.

Lan Y, Wang J, Yang Q, et al. 2020. Blood transcriptome analysis reveals gene expression features of breast-feeding rhesus macaque (*Macaca mulatta*) infants. *Zoological Research*, **41**(4): 431−436.

Leung SK, Jeffries AR, Castanho I, et al. 2021. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Reports*, **37**(7): 110022.

Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Li YM, Shen QS, Peng Q, et al. 2021. Polyadenylation-related isoform switching in human evolution revealed by full-length transcript structure. *Briefings in Bioinformatics*, **22**(6): bbab157.

Liao Y, Smyth GK, Shi W. 2019. The R package rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, **47**(8): e47.

Liu ZM, Yu CD, Chen ZB, et al. 2022. PSMB2 knockdown suppressed proteasome activity and cell proliferation, promoted apoptosis, and blocked NRF1 activation in gastric cancer cells. *Cytotechnology*, **74**(4): 491−502.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**(4): 402−408.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.

Lu ZY, He ST, Jiang J, et al. 2022. Base-edited cynomolgus monkeys mimic core symptoms of STXBP1 encephalopathy. *Molecular Therapy*, **30**(6): 2163−2175.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*, **290**(5494): 1151−1155.

Ma Q, Ma WJ, Song TZ, et al. 2022. Single-nucleus transcriptomic profiling of multiple organs in a rhesus macaque model of SARS-CoV-2 infection. *Zoological Research*, **43**(6): 1041−1062.

Mao YF, Harvey WT, Porubsky D, et al. 2024. Structurally divergent and recurrently mutated regions of primate genomes. *Cell*, Feb 23: S0092-8674(24)00121-1. doi: 10.1016/j.cell.2024.01.052.

Milosavljevic S, Glinton KE, Li XQ, et al. 2022. Untargeted metabolomics of Slc13a5 deficiency reveal critical liver-brain axis for lipid homeostasis. *Metabolites*, **12**(4): 351.

Niwa T, Murayama N, Imagawa Y, et al. 2015. Regioselective hydroxylation of steroid hormones by human cytochromes P450. *Drug Metabolism Reviews*, **47**(2): 89−110.

Nurk S, Koren S, Rhie A, et al. 2022. The complete sequence of a human genome. *Science*, **376**(6588): 44−53.

Otero BA, Poukalov K, Hildebrandt RP, et al. 2021. Transcriptome alterations in myotonic dystrophy frontal cortex. *Cell Reports*, **34**(3): 108634.

Pan B, Wang YT, Shi YW, et al. 2022. Altered expression levels of miR-144-3p and ATP1B2 are associated with schizophrenia. *The World Journal of Biological Psychiatry*, **23**(9): 666−676.

Qin W, Zhou AH, Zuo XM, et al. 2021. Exome sequencing revealed *PDE11A* as a novel candidate gene for early-onset Alzheimer's disease. *Human Molecular Genetics*, **30**(9): 811−822.

Qiu PY, Jiang J, Liu Z, et al. 2019. BMAL1 knockout macaque monkeys display reduced sleep and psychiatric disorders. *National Science Review*, **6**(1): 87−100.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6): 841−842.

Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**(5822): 222−234.

Ritchie ME, Phipson B, Wu D, et al. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7): e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1): 139−140.

Rowe N. 1996. The Pictorial Guide to the Living Primates. Pogonias Pr.

Seney ML, Kim SM, Glausier JR, et al. 2021. Transcriptional alterations in dorsolateral prefrontal cortex and nucleus accumbens implicate neuroinflammation and synaptic remodeling in opioid use disorder. *Biological Psychiatry*, **90**(8): 550−562.

Shao Y, Zhou L, Li F, et al. 2023. Phylogenomic analyses provide insights into primate evolution. *Science*, **380**(6648): 913−924.

Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nature Reviews Genetics*, **20**(11): 631−656.

Sudmant PH, Huddleston J, Catacchio CR, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. *Genome Research*, **23**(9): 1373−1382.

The GTEx Consortium. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**(6235): 648−660.

Viita T, Kyheröinen S, Prajapati B, et al. 2019. Nuclear actin interactome analysis links actin to KAT14 histone acetyl transferase and mRNA splicing. *Journal of Cell Science*, **132**(8): jcs226852.

Villanueva RAM, Chen ZJ. 2019. ggplot2: elegant graphics for data analysis (2nd ed.). *Measurement:Interdisciplinary Research and Perspectives*, **17**(3): 160−167.

Wang ZY, Leushkin E, Liechti A, et al. 2020. Transcriptome and translatome co-evolution in mammals. *Nature*, **588**(7839): 642−647.

Warren WC, Harris RA, Haukness M, et al. 2020. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science*, **370**(6523): 1434.

Wu TZ, Hu EQ, Xu SB, et al. 2021. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **2**(3): 100141.

Yan CC, Zhang XS, Zhou L, et al. 2020. Effects of aging on gene expression in blood of captive Tibetan macaques (*Macaca thibetana*) and comparisons with expression in humans. *Zoological Research*, **41**(5): 557−563.

Yan GM, Zhang GJ, Fang XD, et al. 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature Biotechnology*, **29**(11): 1019−1023.

Yang XY, Mao YF, Wang XK, et al. 2023. Population genetics of marmosets in Asian primate research centers and loci associated with epileptic risk revealed by whole-genome sequencing. *Zoological Research*, **44**(5): 837−847.

Yao YL, Liu SL, Xia C, et al. 2022. Comparative transcriptome in large-scale human and cattle populations. *Genome Biology*, **23**(1): 176.

Yong YX, Wu QF, Meng XL, et al. 2023. Dyrk1a phosphorylation of *α*-synuclein mediating apoptosis of dopaminergic neurons in parkinson's disease. *Parkinson's Disease* **2023**: 8848642.

Zare-Feyzabadi R, Mozaffari M, Ghayour-Mobarhan M, et al. 2022. FABP1 gene variant is associated with risk of metabolic syndrome. *Combinatorial Chemistry & High Throughput Screening*, **25**(8): 1355−1360.

Zhang BL, Chen W, Wang ZF, et al. 2023. Comparative genomics reveals the hybrid origin of a macaque group. *Science Advances*, **9**(22): eadd3580.

Zhang P, Chen JS, Li QY, et al. 2020. Neuroprotectants attenuate hypobaric hypoxia-induced brain injuries in cynomolgus monkeys. *Zoological Research*, **41**(1): 3−19.

Zhang XL, Pang W, Hu XT, et al. 2014. Experimental primates and non-human primate (NHP) models of human diseases in China: current status and progress. *Dongwuxue Yanjiu*, **35**(6): 447−464.

Zheng-Bradley X, Rung J, Parkinson H, et al. 2010. Large scale comparison of global gene expression patterns in human and mouse. *Genome Biology*, **11**(12): R124.

Zhou J, Mock ED, Al Ayed K, et al. 2020. Structure-activity relationship studies of α-ketoamides as inhibitors of the phospholipase A and acyltransferase enzyme family. *Journal of Medicinal Chemistry*, **63**(17): 9340−9359.