

Cynomolgus-rhesus hybrid macaques serve as a platform for imprinting studies

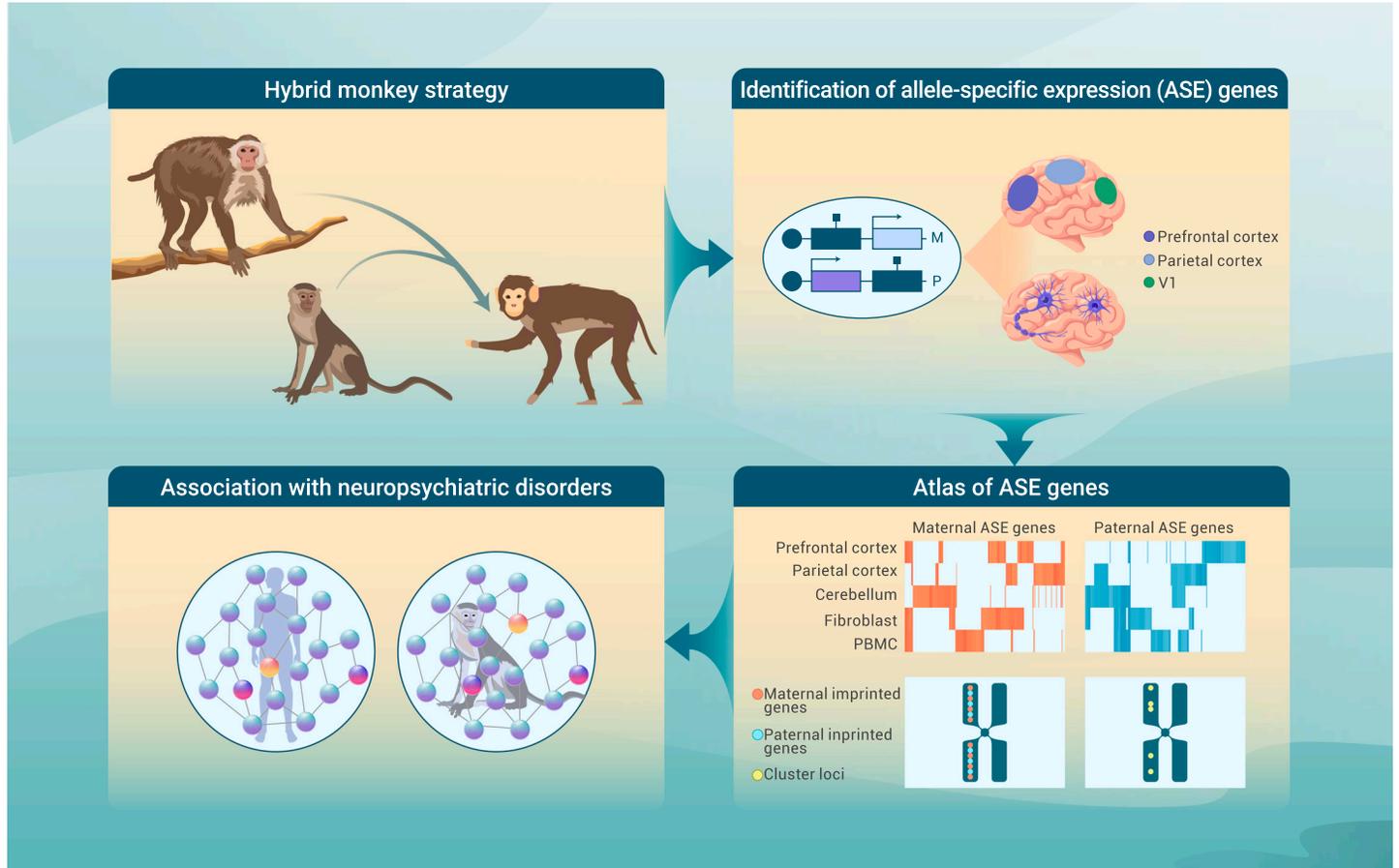
Zongyang Lu,^{1,3,4} Jie Li,^{2,4} Yong Lu,^{1,4} Ling Li,^{1,3} Wei Wang,² Chenchen Zhang,¹ Libing Xu,¹ Yanhong Nie,¹ Changshan Gao,¹ Xinyan Bian,¹ Zhen Liu,^{1,3,*} Guang-Zhong Wang,^{2,*} and Qiang Sun^{1,3,*}

*Correspondence: zliu2010@ion.ac.cn (Z.L.); guangzhong.wang@picb.ac.cn (G.-Z.W.); qsun@ion.ac.cn (Q.S.)

Received: February 10, 2023; Accepted: April 25, 2023; Published Online: May 2, 2023; <https://doi.org/10.1016/j.xinn.2023.100436>

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Cynomolgus-rhesus hybrid macaques is a powerful platform to investigate imprinting.
- Location information of ASE genes can be obtained using hybrid strategy.
- Identification of ASE genes benefits understanding the association between imprinting and neuropsychiatric disorders.



Cynomolgus-rhesus hybrid macaques serve as a platform for imprinting studies

Zongyang Lu,^{1,3,4} Jie Li,^{2,4} Yong Lu,^{1,4} Ling Li,^{1,3} Wei Wang,² Chenchen Zhang,¹ Libing Xu,¹ Yanhong Nie,¹ Changshan Gao,¹ Xinyan Bian,¹ Zhen Liu,^{1,3,*} Guang-Zhong Wang,^{2,*} and Qiang Sun^{1,3,*}

¹Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, CAS Key Laboratory of Primate Neurobiology, State Key Laboratory of Neuroscience, Chinese Academy of Sciences, Shanghai 200031, China

²CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

³Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, Shanghai 201210, China

⁴These authors contributed equally

*Correspondence: zliu2010@ion.ac.cn (Z.L.); guangzhong.wang@picb.ac.cn (G.-Z.W.); qsun@ion.ac.cn (Q.S.)

Received: February 10, 2023; Accepted: April 25, 2023; Published Online: May 2, 2023; <https://doi.org/10.1016/j.xinn.2023.100436>

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Lu Z., Li J., Lu Y., et al., (2023). Cynomolgus-rhesus hybrid macaques serve as a platform for imprinting studies. *The Innovation* 4(3), 100436.

Genomic imprinting can lead to allele-specific expression (ASE), where one allele is preferentially expressed more than the other. Perturbations in genomic imprinting or ASE genes have been widely observed across various neurological disorders, notably autism spectrum disorder (ASD). In this study, we crossed rhesus cynomolgus monkeys to produce hybrid monkeys and established a framework to evaluate their allele-specific gene expression patterns using the parental genomes as a reference. Our proof-of-concept analysis of the hybrid monkeys identified 353 genes with allele-biased expression in the brain, enabling us to determine the chromosomal locations of ASE clusters. Importantly, we confirmed a significant enrichment of ASE genes associated with neuropsychiatric disorders, including ASD, highlighting the potential of hybrid monkey models in advancing our understanding of genomic imprinting.

INTRODUCTION

Genomic imprinting is an epigenetic mechanism that results in the parent-of-origin monoallelic expression of specific genes in mammals. It plays a critical role in various aspects of brain function, including cell fate determination and behavior.^{1,2} The brain has thus become a key focus of imprinting-related research^{3–5} due to its potential significance. Imprinting defects have been identified in two prominent neurological disorders, Prader-Willi syndrome and Angelman syndrome, both of which are linked to the deletion of a paternally imprinted region.⁶ In addition, mutations in imprinted genes have been associated with cognitive disabilities, including myoclonus dystonia syndrome (related to paternal *SGCE* mutations),⁷ Birk-Barel syndrome (related to maternal *KCNK9* mutations),⁸ and intellectual disability (related to paternal *TRAPPC9* mutations).⁹

Allele-specific expression (ASE) is a type of imprinting characterized by subtle shifts in the expression ratio rather than complete silencing of one allele.^{5,10} In mouse studies, ASE can be detected using a method that distinguishes among alleles in crossed hybrids of inbred mouse strains based on technical and biological variation.^{3,5,10} However, this method is not applicable to human research. Current detection of human and monkey ASE relies on *in vitro* experimental systems^{11,12} or large-scale profiling of postmortem tissues.^{10,13} The former may not fully represent *in-vivo*-biased signatures, while the latter requires an extremely large number of samples (100–600)^{10,13,14} to overcome unrelated sources of noise, such as ancestry.

Given the similarities between macaque monkeys and humans in terms of their genome, development, physiology, and anatomy, macaques are a valuable animal model for identifying ASE genes in primates. In line with this, we used a hybrid monkey approach by breeding cynomolgus and rhesus monkeys, as hybrid monkeys derived from crab-eating and rhesus monkeys occur naturally.¹⁵ We analyzed a small group of brain samples from hybrid monkey individuals using a customized data analysis pipeline. As a proof-of-concept, we applied this strategy to identify ASE genes in three brain regions: prefrontal cortex, parietal cortex, and cerebellum. Our approach allowed us to identify 353 ASE genes in the hybrid monkey brains, some of which correspond to orthologs with aberrant expression patterns observed in patients with neuropsychiatric disorders. Importantly,

we identified several well-known genes associated with ASD susceptibility, including hub genes *NRXN1* and *NRXN3*, that exhibit ASE. Our findings highlight the significance of genetic variations and epigenetic dysregulation in the pathogenesis of ASD.

RESULTS

Workflow for identifying ASE genes in hybrid monkeys

Old-world monkeys, such as cynomolgus monkeys (*Macaca fascicularis*, Mfas) and rhesus monkeys (*Macaca mulatta*, Mmul), share similarities in their physiology and reproductive biology, and possess 21 pairs of chromosomes. We hypothesized that, by mating Mfas and Mmul, we could generate hybrid monkeys that would provide informative single-nucleotide polymorphisms (SNPs) distinguishable based on their parental origin, as previously demonstrated in hybrid mouse strains (C57BL/6J x CAST) (Figure 1A). To test this hypothesis, we performed whole-genome sequencing (WGS) on the parental animals to assess genetic variants and transcriptome sequencing on the hybrid offspring to estimate allelic bias. Because of the superior genome annotation of the Mmul genome assembly (Mmul_8.0.1)¹⁶ relative to that of Mfas (*Macaca fascicularis*_5.0, mfas5.0), we mapped all sequencing reads to the Mmul assembly.

To validate the hypothesis that hybrid monkeys harbor more informative SNPs for ASE detection, we sequenced the genomic DNA from 15 macaques (7 Mfas and 8 Mmul) and called only SNPs (no indels) using clean reads. Using all the WGS data, we simulated offspring of monkey mating and found that hybrid offspring resulting from mating between Mfas and Mmul had approximately 50% more informative SNPs than Mmul offspring, resulting from mating between Mmuls (Figure 1B). Furthermore, most informative SNPs detected in Mmul offspring (96.1%) were also detected in our hybrid offspring (Figure 1C). At the gene level, 93% of annotated genes harbored more informative SNPs in hybrid offspring compared with Mmul offspring (Figure 1C). These observations of an abundance of readily distinguishable SNPs provide strong support for the hypothesis that this hybrid strategy should enable straightforward ASE detection, which encouraged us to generate hybrid monkeys.

We performed intracytoplasmic sperm injection (ICSI) by microinjecting sperm from Mfas into oocytes from Mmul, and vice versa, and transplanted embryos at the 1- to 4-cell stage into 13 surrogates (10 Mfas, 3 Mmul) (Tables S1 and S2). The resulting F1 hybrids obtained from crosses between Mfas and Mmul were designated as F1i (F1 initial cross, with Mfas as the father and Mmul as the mother) or F1r (F1 reciprocal cross, with Mmul as the father and Mfas as the mother) (Figure 1A). We obtained 6 live pups and 11 aborted fetuses (Table S1). The birth rate for hybrids was 35.3% (6/17), similar to the birth rate for Mfas monkeys when ICSI was employed.¹⁷ Post-birth, the hybrids survived normally and displayed phenotypic characteristics that were a combination of cynomolgus and rhesus; for instance, exhibiting an intermediate body weight (5.92 ± 0.83 kg), head circumference (27.46 ± 1.15 cm), and tail length (36.98 ± 1.45 cm) (Figure 1D).

Given the well-characterized ASE gene *IGF2R* in fibroblasts from a mouse hybrid,^{4,5} we chose to use isolated fibroblasts as a model for our ASE detection framework (Figure 2A). Accordingly, we developed a customized pipeline, called

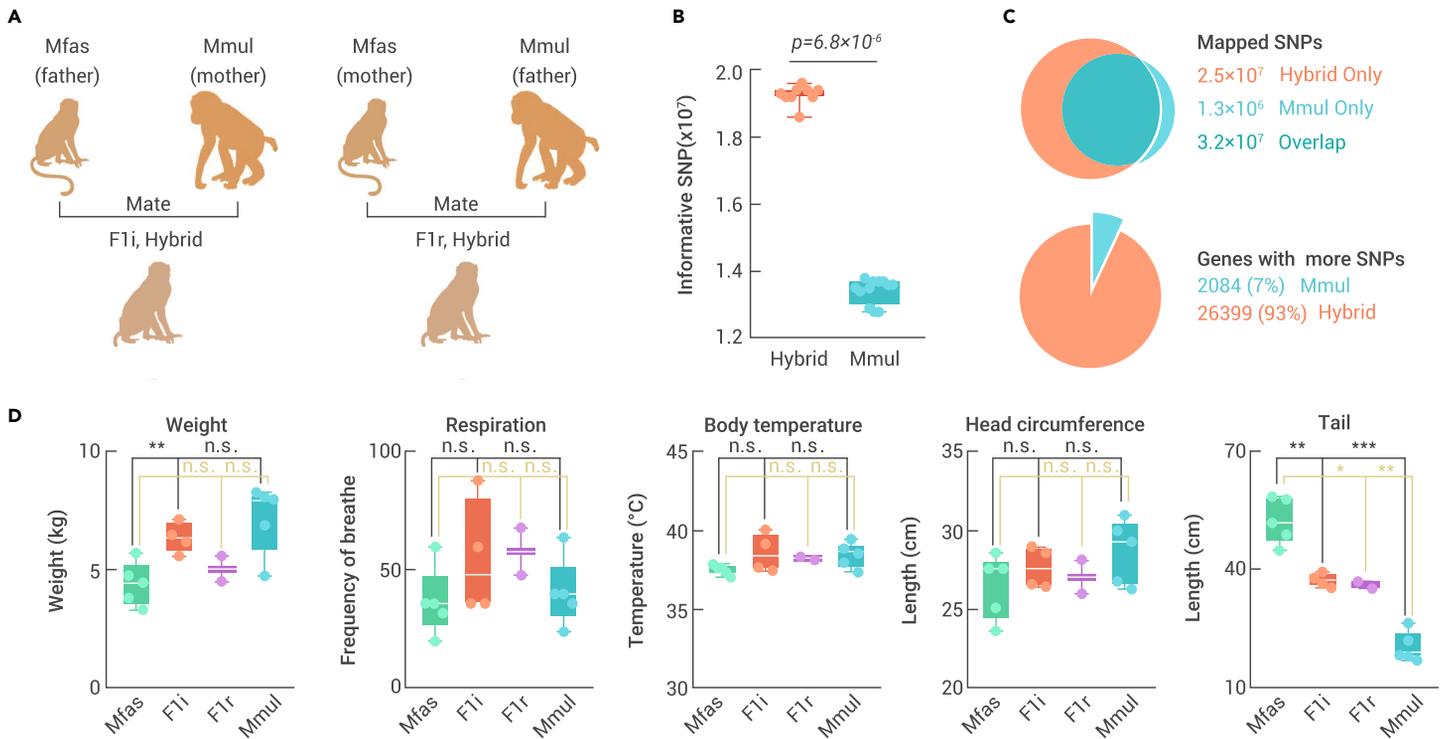


Figure 1. Generation of hybrid monkeys (A) Schematic for generation of hybrid monkeys. (B) The number of informative SNPs of hybrid offspring and Mmul offspring. Each plot represents one monkey. (C) Comparison of mapped SNPs (upper) and genes (bottom) between hybrid monkey and Mmul. Top: "hybrid only," hybrid-specific informative SNPs; "Mmul only," rhesus-specific informative SNPs; overlap, shared SNPs between hybrid and Mmul monkeys. Bottom: after SNP mapping to genes. (D) Feature measurements of Mfas, hybrid, and Mmul for weight, respiration frequency, body temperature, head circumference, and tail length. Significance is determined via unpaired Student's two-tailed test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the fast imprint detector (FID), based on a previous study of ASE in mice (Figure 2B).⁵ To identify informative SNPs, we utilized WGS reads from the parents of hybrid monkeys, consisting of four cynomolgus and five rhesus macaques (Figure S1A). Using these informative SNPs, we analyzed RNA sequencing (RNA-seq) reads from the fibroblasts of six hybrid monkeys (four F1i and two F1r) to estimate expression bias and distinguish parental origin. We generated an average of 34.3 million reads (34.3 ± 0.8) per sample, and the mean mapping rate ($85.91\% \pm 0.30\%$) was sufficient to support our subsequent analyses (Figures S1A and S1B).¹⁸

In light of obvious cross-individual variation of parental allelic bias reported in a previous study,¹⁴ we compared allelic bias between F1i and F1r hybrids at the individual level. Each F1i was compared with both F1r, resulting in eight comparisons. For each comparison, the bias was assessed using a chi-squared test, with a p value less than 0.05 in both F1i and F1r. We applied four filters to produce highly accurate and reproducible results: (1) a read count-based filter (coverage of RNA-seq ≥ 10), (2) a ratio-based filter (biased ratio ≥ 0.7), (3) a comparison-based filter (significant/total comparison $\geq 2/3$), and (4) an SNP-based filter (gene with ≥ 2 biased SNPs) (Figure 2B and methods). In total, 124 genes (ratio > 0.7) demonstrated significant ASE in fibroblasts, including the canonical paternally imprinted genes *IGF2R* and *H19* in monkeys,¹⁹ which had biased ratios of more than 0.95 (Figures 2C and S1C).

Afterward, our framework was applied to peripheral blood mononuclear cells (PBMCs) obtained from 6 hybrid individuals (4 F1i and 2 F1r) for ASE detection, resulting in the identification of 140 ASE genes (ratio > 0.7) (Figure 2B). Notably, previously reported ASE genes such as *GNAS*, *TRAPPC9*, *ZDF2*, *SNRPN*, and *KCNQ1* were among those identified. Encouraged by these results, we aimed to extend our framework to the nervous system.

Brain atlas of ASE genes

We collected samples from three different brain regions (prefrontal cortex, parietal cortex, and cerebellum) of 6 aborted fetuses (5 F1i, 1 F1r) (Table S2) and utilized FID for ASE gene detection. In total, 15 RNA-seq samples of the brain (4 F1i and 1 F1r per region) passed quality control, and we used 7 WGS datasets from their parents (4 Mfas and 3 Mmul) to identify inherited SNPs (Tables S1 and S2). After applying the aforementioned data-processing steps to the fibroblast and

PBMC samples, we detected a total of 353 ASE genes (ratio > 0.7) in the brain, including 144 in the prefrontal cortex, 162 in the parietal cortex, and 185 in the cerebellum (Figures 2B, S1D, and S1E). These genes included well-known brain-specific ASE genes, such as *GRB10*, *MEST*, *MAGEL2*, and *TRAPPC9*, as well as ASE genes associated with various diseases such as *KCNK9* and *SGCE* (Figure 2D; Table S3).

Remarkably, our brain atlas of ASE genes encompasses several genes that play vital roles in synaptic and epigenetic regulation, such as *NRXN1*, *SHANK2*, *DNMT3A*, and *SULT4A1*. Furthermore, our analysis identified ASE genes involved in diverse cellular processes, including cytoskeleton regulation (*ANK3*), potassium voltage-gated channel (*KCNQ3*), solute carriers (*SLC12A2*), and cyclin-dependent kinases (*CDK8*), which are essential for cell survival and signal transduction (Table S3). Therefore, our study successfully uncovered ASE genes originating from distinct types of cells and tissues, including fibroblasts, PBMCs, and three different brain regions. We observed a high degree of origin specificity among ASE genes, as demonstrated by the identification of only 16 maternal and 23 paternal ASE genes conserved across all three brain regions. Moreover, only three genes exhibited ASE in all five sample origins (Figure 2E; Table S3), highlighting the tissue-specific nature of ASE genes consistent with previous findings in humans and mice.^{10,20}

We searched for clustered ASE genes, defined as sets of >2 ASE genes located within a 1-megabase (MB) genomic region. This window size was based on previous reports of well-characterized ASE gene clusters such as *H19-IGF2*, *Dlk1-Mest*, and *Mest-Copg2*.⁵ We identified a total of 215 ASE cluster loci (Figure S1F), with the highest number of clusters (28) found on chromosome 19, while no clusters were found on chromosome 18, consistent with previous findings in mice that ASE clusters are non-randomly distributed throughout the genome (Figures 2F, S1F, and S1G).^{5,10} Our hybrid-based approach enabled us to generate an atlas of brain-specific ASE genes, which will prove invaluable in enhancing our understanding of the origins and underlying mechanisms of ASE.

ASE gene confirmation in cynomolgus monkeys

To assess the accuracy of ASE gene detection using our hybrid monkey strategy, we performed targeted deep sequencing analysis of SNPs. We utilized

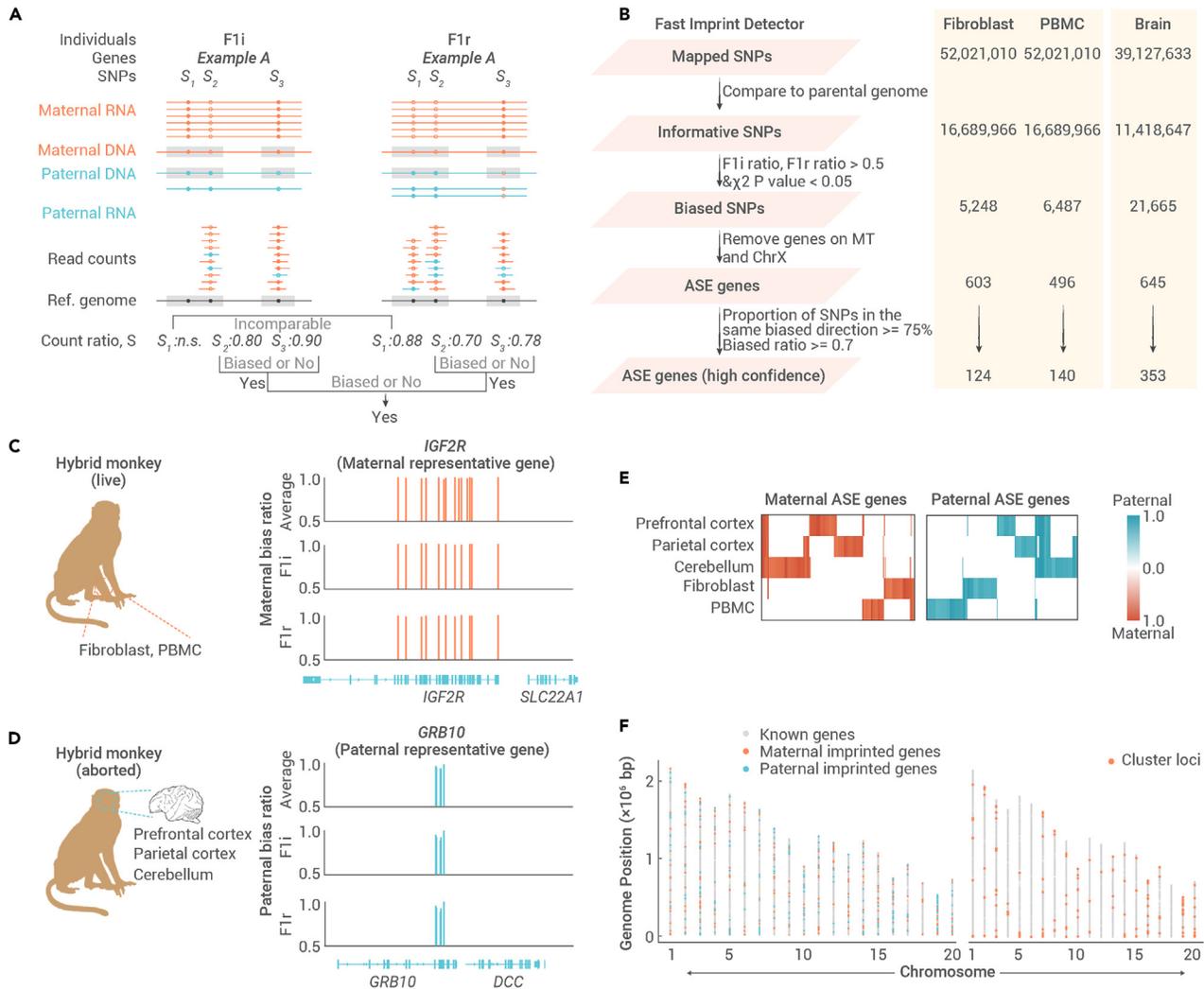


Figure 2. Establishment of a hybrid monkey strategy for ASE detection (A) Quantifying allelic bias of expression in hybrid individuals using an RNA-seq read count ratio statistic. The strength of bias toward the expression of the maternal (red) or paternal (blue) allele of a tested gene in individual (F1i or F1r) is gauged based on the count of RNA-seq reads carrying the reference allele (small closed circles) and the count of reads carrying an alternative allele (open squares), at all SNPs for which the individual is heterozygous. S_i , SNPs; gray rectangles, exons. (B) Framework for identifying ASE genes. The numbers of SNPs and genes for tissues are listed on the right of the flow chart. Fibroblast/PBMC and brain are separated by a solid line. For fibroblasts and PBMCs, 6 RNA-seq samples of live monkeys and 9 WGS of parents were used. For brain regions, 15 RNA-seq samples of aborted monkeys and 8 WGS datasets of parents were used. (C) Illustration showing the fibroblasts and PBMCs of live hybrid monkeys (left) and IGV viewer showing the maternal representative ASE genes, *IGF2R* (right). y axis, the maternal bias ratio (ranging from 0.5 to 1.0). Average indicated average bias ratio at each SNP. F1i, one representative F1i monkey; F1r, one representative F1r monkey. (D) Illustration showing the brain regions of aborted hybrid monkeys (left) and IGV viewer showing the paternal representative ASE genes, *GRB10* (right). y axis, the maternal, paternal, and average bias ratio at each SNPs (ranging from 0.5 to 1.0). F1i, one representative F1i monkey; F1r, one representative F1r monkey. (E) Heatmap showing detected ASE genes. Red rectangles, maternal genes; blue rectangles, paternal genes. (F) Genomic location of detected ASE genes across genome: maternal biased genes (red) and paternal biased genes (blue).

cDNA extracted from brain tissues of six Mfas fetuses, specifically the prefrontal cortex, parietal cortex, cerebellum, and V1 (Figure 3A; Table S4). Our analysis focused on 14 genes of interest (GOIs) that are associated with synaptic regulation (NRXN-NLGN-SHANK axis-related genes, e.g., *NRXN3* and *SHANK2*), epigenetic regulation (e.g., *KMT2C* and *MED13L*), and signal transduction (e.g., *GMP6A* and *VSNL1*) (Figure 3B). To determine allelic bias, we calculated the read count ratio of each GOI's individual SNPs (with at least two SNPs per GOI) based on deep sequencing reads using the read count ratio found in a previous human study.¹⁴ We then calculated the gene bias ratio by averaging the SNP bias ratios (Figure S2A; Table S5). Our analysis revealed that all 14 GOIs displayed biased expression of at least 0.7 in at least one brain region (Figure 3B; Table S5), confirming that these GOIs are ASE genes in Mfas monkey brains.

Given that DNA methylation is a well-established mechanism for imprinting and regulating parent-of-origin expression,²¹ we analyzed differentially methylated regions (DMRs) in the brains of Mfas monkeys using a published DNA methylation dataset. Our goal was to assess the validity of the ASE genes identified by our hybrid monkey strategy.¹² Using a 10-kilobase cutoff from a DMR to a gene body, we detected 2,359 DMR-related genes in the

cortex and cerebellum. Notably, 107 out of the 353 ASE genes in our atlas were found to overlap with DMR-related genes, including *SNRPN* and *PEG3* (Figures 3C and 3D), suggesting that their regulation may be controlled by their associated DMRs. These results, combined with our deep sequencing analysis, support the authenticity of the ASE genes identified by our hybrid strategy.

Enrichment of brain ASE genes in ASD signatures

As anticipated, the ASE gene set obtained from hybrids and their orthologs in humans exhibited similar expression patterns across the three brain regions (Figure 3E), indicating significant developmental similarities between the two species. Previous studies have suggested that dysregulation of ASE may contribute to neuropsychiatric disorders.² To investigate the potential roles of brain ASE genes in these disorders, we performed an enrichment analysis using risk genes curated from the DisGeNET database. Our finding showed that the ASE genes were significantly enriched for susceptibility genes associated with disorders such as ASD, BD, and epilepsy (Figures 3F, S2B, and S2C).

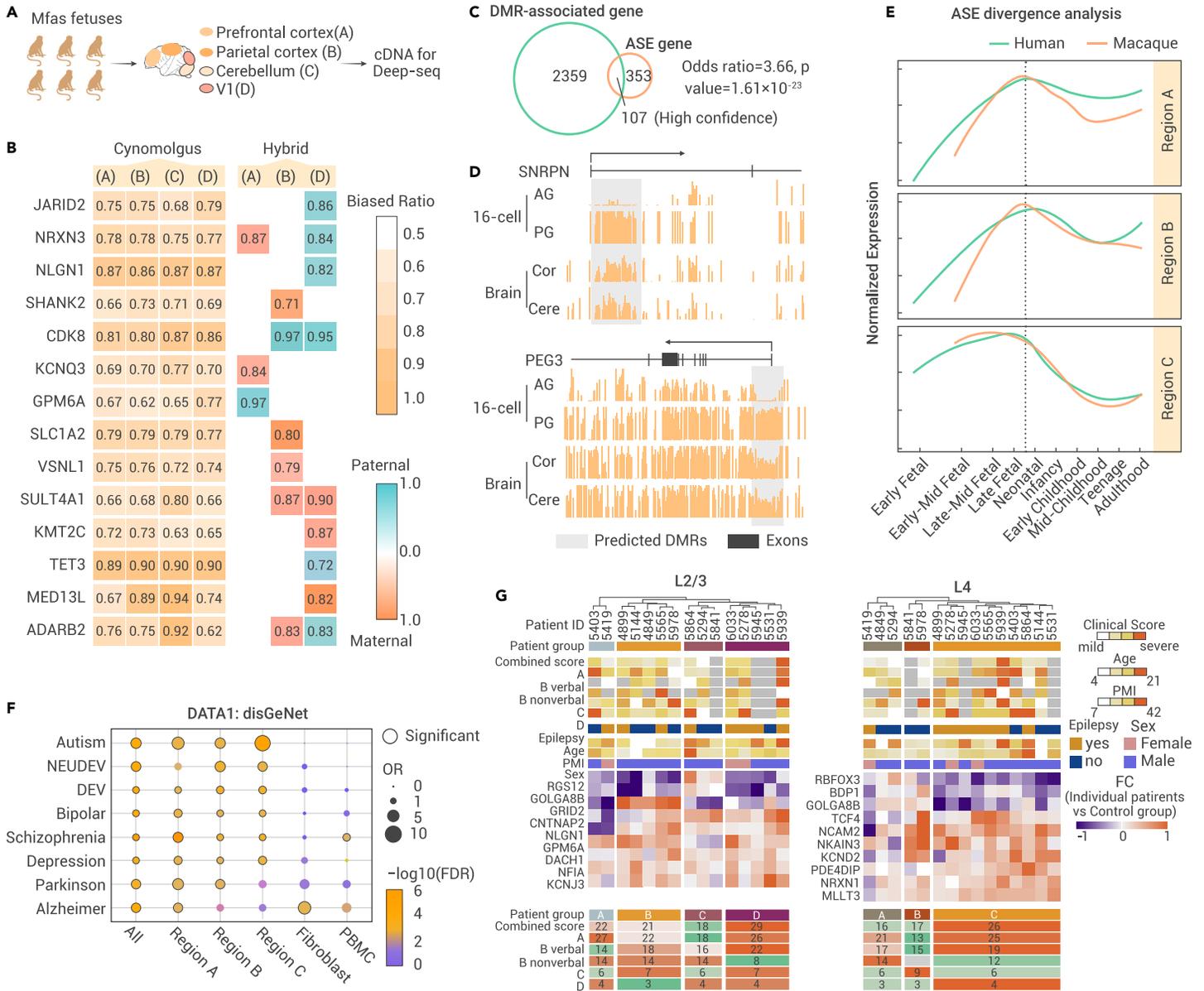


Figure 3. ASE genes are confirmed in cynomolgus and enriched in ASD signatures (A) Illustration showing four tested brain regions from six aborted Mfas. (B) Fourteen GOIs are confirmed as ASE genes in Mfas monkey brains. Numbers in boxes, the average biased ratios of each gene for Mfas or hybrids. (C) Venn diagram showing the enrichment between ASE genes and DMRs. (D) IGV viewer showing a representative ASE gene and its predicted DMR. AG/PG, an androgenetic/parthenogenetic haploid embryo at the 16-cell stage; Cor and Cere, the cortex and cerebellum, respectively. (E) Interspecies divergence, measured as difference in gene expression, between humans and macaques in each brain region across multiple stages. (F) Bubble matrix with ASE genes of different tissues showing differential enrichment on diseases. (G) Hierarchical clustering of ASD patients based on individual fold changes in gene expression level in L2/3 neurons and L4 neurons. Average clinical scores, prevalence of epilepsy, proportion of nonverbal subjects, and age and gender composition are shown below the heatmaps. A, social interactions; B, communication; C, repetitive behavior; D, abnormal development.

Since our ASE data were obtained from embryonic-stage hybrid monkeys, we focused on investigating the potential involvement of ASE genes in ASD, a disorder associated with brain development.²² We analyzed various ASD gene sets, including genes linked to autism susceptibility from the SFARI database²³ and differentially expressed genes between ASD patients and healthy controls at the bulk tissue level.²⁴ Our analysis revealed that ASE genes were significantly enriched for genes implicated in autism susceptibility (SFARI). Moreover, we identified 17 ASE genes with significantly decreased expression levels in the cortices of ASD patients, such as *SULT4A1*, and *KCNJ3*. These findings are consistent with previous reports that genes exhibiting ASE in ASD tend to be downregulated (Table S3).¹³

A recent study investigated the relationship between changes in gene expression related to ASD and clinical severity in specific types of brain cells. The study found that changes in upper-layer neurons are linked to the severity of ASD.²⁵ Building on this finding, we investigated whether brain ASE genes are dysregulated in specific cell types. Our study showed that brain ASE genes are more frequently over-represented among the sets of differentially expressed genes de-

tected in specific cell types, such as L4 excitatory neurons, L2/3 excitatory neurons, and VIP-expressing interneurons (IN-VIP). (Figures 3G and S2D). Moreover, we found a significant correlation between ASD-specific ASE genes and clinical phenotypes (Figure 3G). Interestingly, we also detected enrichment in non-neuronal brain cells, including microglia and protoplasmic astrocytes. This finding is consistent with a recent report suggesting that astrocyte development and survival are regulated by imprinting in a cell-type-specific manner (Figure S2D; Table S3).³

We utilized the STRING database²⁶ to conduct a protein-protein interaction (PPI) analysis of the ASE gene set and discovered a significant clustering of ASE genes. The PPI network consisted of two distinct modules: an imprinted module and a neuronal communication module (Figures S2F and S2G). Previous studies have implicated gene regulation and synaptic development in the pathogenesis of ASD^{27,28} and, as such, we further grouped the ASE genes into established functional modules based on gene ontology terms. These modules included "gene expression regulation" (54 genes), "neuronal communication" (41 genes), and "cytoskeleton" (19 genes) (Figure 3E; Table S3). These results

are in line with previous studies that have identified convergent molecular abnormalities in ASD.^{29,22}

In light of our detection of dysregulated ASE in ASD patients,^{13,30,31} we conducted a comparison between our ASE genes and those from a large-scale profiling study involving 263 postmortem brains of ASD patients.¹³ Our results revealed that our brain ASE genes were significantly enriched in the previous ASE dataset, with 108 out of 314 (34.39%) demonstrating ASE in ASD patients (Figures S2H and S2I). Notably, two of these genes, *NRXN1* and *NRXN3*, have roles in neuronal communication. In ASD patients, *NRXN1* was upregulated in L4 excitatory neurons and in the IN-VIP (Figure 3G),²⁵ while *NRXN3* was downregulated in the cortex (Figure S2J).¹³

Analysis of brain biopsies of macaque support our hybrid strategy

To confirm and clarify our findings, we performed RNA-seq analysis using published data from rhesus macaque brain.³² Initially, we collected datasets for the frontal lobe, parietal lobe, and cerebellum from three macaque brains (Table S6). In contrast to the FID pipeline applied for hybrid ASE detection, the rhesus macaque data did not have their parental WGS data available. To detect ASE, we employed an ASE pipeline based on SNP bias (Figure S3A). The number of ASE genes identified in the downloaded data overlapped among the three brain regions was significantly higher than the gene set produced by the hybrids (Figure S3B), suggesting that the hybrid approach generates a more precise ASE gene set. We also tested the enrichment level between gene sets generated from monkey brain biopsies and hybrids. The results indicated that the ASE genes detected in brain biopsies largely overlapped with our biased genes (Figures S3B and S3C), providing further support for the credibility of our results.

DISCUSSION

In summary, our study demonstrates the feasibility of *in vivo* primate ASE identification using a hybrid monkey strategy. Notably, our findings suggest that the limited overlap observed among existing human databases of imprinted genes underscores the significance of using an inbred strategy for ASE detection in non-human primates (Figure S1H). By leveraging this hybrid approach, we detected 353 genes exhibiting significant ASE signals in the brain, and our analysis supports previous profiling studies implicating ASE genes in ASD pathogenesis.

We summarize three advantages of our hybrid monkey strategy over previous approaches for primate ASE detection: (1) *in vivo* analysis allows for the detection of spatiotemporally resolved physiological ASE status in any tissue of interest, (2) our approach is less resource consuming, as we need fewer than 10 hybrid individuals compared with large-scale profiling (requiring 100–600 individuals), and (3) our strategy is reproducible, as the generated hybrid monkeys are models with a stable genetic background, while human sampling populations may exhibit significant genetic variations between different studies.

MATERIALS AND METHODS

Animal ethics statements

Healthy female cynomolgus and rhesus monkeys (*M. fascicularis* and *M. mulatta*) with regular menstrual cycles were selected for this study. All animals were housed in the sunny room. The use and care of animals complied with the guideline of Center for Excellent in Brain Science and Intelligence Technology, Chinese Academy of Science, which approved the ethics application (ER-SIBS-221106P and CEBSIT-2020014).

Monkey oocytes collection, ICSI, embryo injection, and embryo culturing

For monkey oocyte collection, procedures of ICSI, embryo injection, and culturing has been described in a previous publication.³³ In brief, from day 3 of menstrual cycle, healthy female monkeys began to receive 25 IU recombinant human follitropin twice daily for 7–8 days. On day 11 of the menstrual cycle, 1,000 IU human chorionic gonadotropin was injected, and oocytes were aspirated from ovarian follicles 36 h later.

Monkey embryos were cultured in HECM-9 medium at 37°C under 5% CO₂. The embryos were transferred to HECM-9+5% FBS medium after reaching the eight-cell stage and the medium was changed every other day until the embryo reached the blastocyst stage. For embryo transfer, females with synchronous menstrual cycle whose ovaries had a stigma or fresh corpus luteum were used as surrogates. Embryos (at two-cell to blastocyst stage) were transferred to the oviduct. After performing ICSI, we typically

perform embryo transplantation either 1 or 2 days after ICSI, depending on the surrogate availability. Thus, some embryos were transferred 1 day after ICSI at the one- to two-cell stage, and some embryos were transferred 2 days after ICSI at the two- to four-cell stage.

WGS process

Genomic DNA was extracted by the phenol-chloroform method. One hundred nanograms of DNA was fragmented using a Covaris LE220 (Covaris), size selected (300–550 bp), end repaired, A-tailed, and adapter ligated. Libraries were sequenced using the HiSeq X Ten platform (Illumina) as paired-end 100 base reads. For WGS data, adaptors of sequenced reads were firstly trimmed by Trimmomatic (v.0.39) with parameters PE ILLUMINACLIP: adaptors/TruSeq3-PE-2.fa:2:30:10:8:True SLIDINGWINDOW:5:15 LEADING:5 TRAILING:5 MINLEN:50.¹⁶ The variant calling pipeline of Genome Analysis Toolkit (GATK) was then applied following this protocol (<https://software.broadinstitute.org/gatk/best-practices>). The trimmed reads were mapped to the reference genome of *rheMac8* (*M. mulatta*) using BWA-mem with default parameters.³⁴ Aligned reads were converted to BAM files and sorted based on genome position after marking duplicates using the Picard toolkit. The raw BAM files were refined by base quality score recalibration using default parameters. The variant calling (SNPs and indels) was performed with the HaplotypeCaller module. Then, 15 gvcf files were combined into a single file and the SNPs and short insertions and deletions (indels) were selected and filtered. Only SNPs that provided information about parent and offspring (informative SNPs) were used for subsequent analysis.

Transcriptome profiling of monkey samples

First, 27 samples from fibroblasts, PBMCs, and brains were collected for next-generation sequencing. After total RNA extraction using TRIzol reagent (Vazyme Biotech, RC112), we produced cRNA, and then utilized TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme Biotech, TD501) and TruePrep Index Kit V2 for Illumina (Vazyme Biotech, TD202) according to standard protocol to generate bulk RNA-seq library. All samples were sequenced by Illumina XTen in the facility of Omics Core, Bio-Med Big Data Center, SINH, CAS.

RNA-seq processing

For RNA-seq data, adaptors of paired-end sequenced reads were trimmed by Trim Galore with parameters

```
-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
-a2 AGATCGGAAGAGCGTCGTGTAGGAAAGAGTGT
-paired -stringency 3 -phred33. Adaptor trimmed reads were then cut the first
10 bp reads using the FASTX-Toolkit. Then all the processed reads were mapped
to the reference genome of rheMac8 (M. mulatta) using STAR35 with parameters
-runMode alignReads -runThreadN 8 -outSAMattributes All -outSAMtype BAM
SortedByCoordinate -limitBAMsortRAM 62000000000 -outSAMunmapped Within
-outReadsUnmapped Fastx.35 Aligned reads were converted to BAM files and sorted
based on genomic position. The base frequency at each position was extracted using
the mpileup module of SAMtools.36 In addition, expression values represented by frag-
ments per kilobase per million and read counts were calculated by Cufflinks37 and fea-
tureCounts,38 respectively, based on the annotations from Ensemble database.
```

Targeted deep sequencing

Each allele-biased target site was amplified from monkey cDNA using Phanta Max SuperFidelity DNA Polymerase (Vazyme Biotech). Deep sequencing and analysis were performed using the Hi-TOM service (<http://www.hi-tom.net/hi-tom/>). Amplified sequences and primers were used for targeted deep sequencing and are listed in Table S7.

Identification of biased genes

We cross-referenced the F1 hybrid reads with the informative SNPs identified in WGS analysis of parents. For each cross of cynomolgus and rhesus monkeys, we defined that informative SNPs are loci where parental alleles can be distinguished based on Mendel's law of inheritance. Specially, we calculated the number of informative SNPs from each parent pair of hybrid monkeys and compared them to those in crosses between rhesus macaques. Then, SNP site base calls (from GATK) of informative SNPs were used to assign the reads as paternal and maternal specific. For each informative SNP with sufficient reads (number of reads ≥ 10), we calculated the number of reads harboring paternal alleles and maternal alleles in F1i and F1r. Then we estimated the bias of parental origin using chi-squared tests in both F1i and F1r monkeys. Only SNP sites with the same paternal/maternal bias (maternal/paternal reads/total reads, range from 0 to 1) and $p < 0.05$ both in F1i and F1r are considered as SNPs exhibiting a significant parental bias (biased SNPs). For each tissue, we performed multiple comparisons between F1i and F1r samples and only SNP

sites that are significant in two-thirds of the comparisons are considered as biased SNPs. Then, we mapped the biased SNPs to gene loci and genes with ≥ 2 biased SNPs are defined as biased gene.

PPI network analysis

PPIs of the ASE genes in brain were obtained from STRING database (v.11).²⁶ The cluster in the human ASE PPI network was identified via the Markov cluster algorithm³⁹ with default parameters, resulting in 298 nodes with 562 edges connected ($p < 1 \times 10^{-16}$).

DMRs analysis

The raw fastq files were downloaded from ENA database (<https://www.ebi.ac.uk/ena/browser/view/PRJNA668521?show=reads>) and the same pipeline was applied to identify the DMRs in two brain regions (cortex and cerebellum) of *M. fascicularis*.¹² The genes within 10 kilobase pairs of each DMR are considered as DMR-related genes.

Gene set enrichment analysis

Two disease-related gene sets were used to perform the over-represented disease analysis. One was collected from DisGeNET database (<https://www.disgenet.org/>)⁴⁰ and the other includes genes related to 64 diseases across the 6 categories.⁴¹ In addition, several autism gene sets were also accessed: (1) ASD risk genes in SFARI database (<https://gene.sfari.org/>), (2) differentially expressed genes between autism and healthy samples at both bulk tissue²⁴ and cell-type level,²⁵ and (3) allele-specific genes detected between ASD and control samples.¹³ Functional enrichments of gene sets were performed using a two-sided Fisher's exact test with $OR > 1$ and FDR-adjusted $p < 0.05$.

Statistical analysis

Statistical analysis was performed using R software (v.4.1.0) (<https://www.R-project.org/>). Adjusted $p < 0.05$ was considered significant for all the analysis.

REFERENCES

- Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. *Nat. Rev. Genet.* **15**, 517–530.
- Tucci, V., Isles, A.R., Kelsey, G., et al. (2019). Genomic imprinting and physiological processes in mammals. *Cell* **176**, 952–965.
- Laukoter, S., Pauler, F.M., Beattie, R., et al. (2020). Cell-Type specificity of genomic imprinting in cerebral cortex. *Neuron* **107**, 1160–1179.e9.
- Perez, J.D., Rubinstein, N.D., Fernandez, D.E., et al. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife* **4**, e07860.
- Gregg, C., Zhang, J., Weissbourd, B., et al. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643–648.
- Nicholls, R.D., Knoll, J.H., Butler, M.G., et al. (1989). Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature* **342**, 281–285.
- Zimprich, A., Grabowski, M., Asmus, F., et al. (2001). Mutations in the gene encoding epsilon-sarcoglycan cause myoclonus-dystonia syndrome. *Nat. Genet.* **29**, 66–69.
- Barel, O., Shalev, S.A., Ofir, R., et al. (2008). Maternally inherited Birk Barel mental retardation dysmorphism syndrome caused by a mutation in the genomically imprinted potassium channel KCN9. *Am. J. Hum. Genet.* **83**, 193–199.
- Mochida, G.H., Mahajnah, M., Hill, A.D., et al. (2009). A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. *Am. J. Hum. Genet.* **85**, 897–902.
- Babak, T., DeVeale, B., Tsang, E.K., et al. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat. Genet.* **47**, 544–549.
- Sagi, I., De Pinho, J.C., Zuccaro, M.V., et al. (2019). Distinct imprinting signatures and biased differentiation of human androgenetic and parthenogenetic embryonic stem cells. *Cell Stem Cell* **25**, 419–432.e9.
- Chu, C., Zhang, W., Kang, Y., et al. (2021). Analysis of developmental imprinting dynamics in primates using SNP-free methods to identify imprinting defects in cloned placenta. *Dev. Cell* **56**, 2826–2840.e7.
- Lee, C., Kang, E.Y., Gandal, M.J., et al. (2019). Profiling allele-specific gene expression in brains from individuals with autism spectrum disorder reveals preferential minor allele usage. *Nat. Neurosci.* **22**, 1521–1532.
- Gulyás-Kovács, A., Keydar, I., Xia, E., et al. (2018). Unperturbed expression bias of imprinted genes in schizophrenia. *Nat. Commun.* **9**, 2914.
- Hamada, Y., San, A.M., and Malaivijitnond, S. (2016). Assessment of the hybridization between rhesus (*Macaca mulatta*) and long-tailed macaques (*M. fascicularis*) based on morphological characters. *Am. J. Phys. Anthropol.* **159**, 189–198.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
- Liu, Z., Li, X., Zhang, J.T., et al. (2016). Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. *Nature* **530**, 98–102.
- Sims, D., Sudbery, I., Iltot, N.E., et al. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132.

- Cheong, C.Y., Chng, K., Ng, S., et al. (2015). Germline and somatic imprinting in the nonhuman primate highlights species differences in oocyte methylation. *Genome Res.* **25**, 611–623.
- Perez, J.D., Rubinstein, N.D., and Dulac, C. (2016). New perspectives on genomic imprinting, an essential and multifaceted mode of epigenetic control in the developing and adult brain. *Annu. Rev. Neurosci.* **39**, 347–384.
- Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, a002592.
- Willsey, H.R., Willsey, A.J., Wang, B., et al. (2022). Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341.
- Abrahams, B.S., Arking, D.E., Campbell, D.B., et al. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36.
- Rahman, M.R., Petralia, M.C., Ciurleo, R., et al. (2020). Comprehensive analysis of RNA-Seq gene expression profiling of brain transcriptomes reveals novel genes, regulators, and pathways in autism spectrum disorder. *Brain Sci.* **10**, 747.
- Velmeshev, D., Schirmer, L., Jung, D., et al. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689.
- Szklarczyk, D., Gable, A.L., Nastou, K.C., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/ measurement sets. *Nucleic Acids Res.* **49**, D605–D612.
- Parikshak, N.N., Luo, R., Zhang, A., et al. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021.
- Voineagu, I., Wang, X., Johnston, P., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384.
- Satterstrom, F.K., Kosmicki, J.A., Wang, J., et al. (2020). Large-Scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584.e23.
- Lin, C.Y., Chang, K.W., Lin, C.Y., et al. (2018). Allele-specific expression in a family quartet with autism reveals mono-to-biallelic switch and novel transcriptional processes of autism susceptibility genes. *Sci. Rep.* **8**, 4277.
- Ben-David, E., Shohat, S., and Shifman, S. (2014). Allelic expression analysis in the brain suggests a role for heterogeneous insults affecting epigenetic processes in autism spectrum disorders. *Hum. Mol. Genet.* **23**, 4111–4124.
- Yin, S., Lu, K., Tan, T., et al. (2020). Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat. Commun.* **11**, 474.
- Liu, Z., Cai, Y., Wang, Y., et al. (2018). Cloning of macaque monkeys by somatic cell nuclear transfer. *Cell* **174**, 245.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Dobin, A., Davis, C.A., Schlesinger, F., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
- Danecek, P., Bonfield, J.K., Liddle, J., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008.
- Trapnell, C., Williams, B.A., Pertea, G., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
- 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.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584.
- Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855.
- Zhou, Y., Hou, Y., Shen, J., et al. (2020). A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLoS Biol.* **18**, e3000970.

ACKNOWLEDGMENTS

We thank members of authors' laboratories for helpful discussions. This work was supported by the National Key Research and Development Program of China (2022YFF0710901 to Q.S.), the National Natural Science Foundation of China (grants 31825018 and 82021001 to Q.S.; 81827901 and 32170567 to G.W.), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX05 to Q.S. and Z. Liu), the Biological Resources Program of the Chinese Academy of Sciences (KFJ-BRP-005), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB32060000 to Q.S. and Z. Liu), the Basic Frontier Scientific Research Program of the Chinese Academy of Sciences From 0 to 1 original innovation project (ZDBS-LY-SM019 to Z. Liu), and the China National Postdoctoral Program for Innovative Talents (BX20200350 to Z. Liu).

AUTHOR CONTRIBUTIONS

Z. Liu, G.W., and Q.S. conceived, designed, and supervised the project. Z. Lu, J.L., and Y.L. performed most experiments with the help of C.Z., L.X., Y.N., and C.G. C.Z. performed the micro-injection. Z. Lu and Y.L. collected samples of monkeys. J.L. performed bioinformatics analysis with the help of W.W. and Z. Lu. C.Z., L.X., Y.N., and C.G. performed oocyte collection, embryo transfer, pregnancy examination, and infant care. Z. Lu and J.L. wrote the paper with input from all authors. Z. Liu, G.W., and Q.S. managed the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

The raw WGS and RNA-seq data reported in this paper are available in the SRA (Sequence Read Archive) under the accession number PRJNA855125 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA855125>). All original codes are available upon request.

SUPPLEMENTAL INFORMATION

It can be found online at <https://doi.org/10.1016/j.xinn.2023.100436>.

LEAD CONTACT WEBSITE

http://www.cebsit.cas.cn/ggpt/yjpt/lzlyjptsz/jyjs/http://www.sinh.cas.cn/rcdw/qtyjzz/201803/t20180328_4987188.htmlhttp://www.cebsit.cas.cn/yjz/lz/_yjfx/