

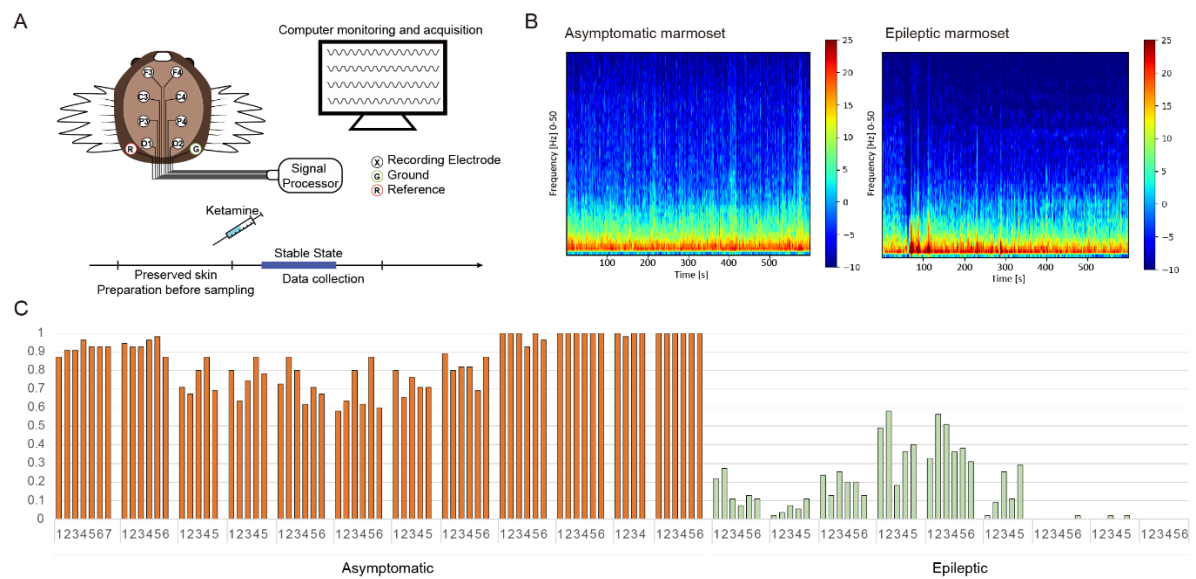
Supplementary Materials

Population genetics of marmosets in Asian primate research centers and loci associated with epileptic risk revealed by whole-genome sequencing

Xiang-Yu Yang^{1,2,3}, Ya-Fei Mao^{1,5}, Xuan-Kai Wang¹, Dong-Ni Ma^{1,3}, Zhen Xu⁶, Neng Gong⁶, Barbara Henning⁵, Xu Zhang^{1,3}, Guang He¹, Yong-Yong Shi^{1,7}, Evan E. Eichler^{5,8}, Zhi-Qiang Li^{1,7,*}, Eiki Takahashi^{1,9,*}, Wei-Dong Li^{1,2,3,4,*}

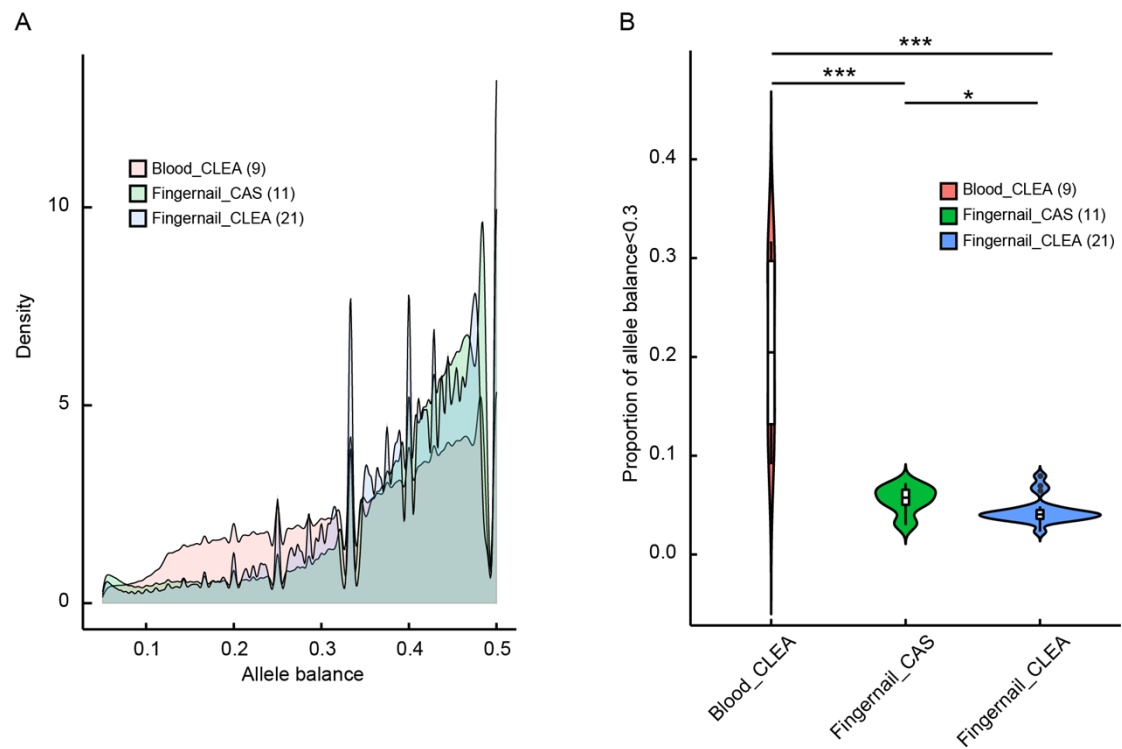
1. Bio-X Institutes, Key Laboratory for the Genetics of Development and Neuropsychiatric Disorders (Ministry of Education), Shanghai Key Laboratory of Psychotic Disorders, and Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai 200240, China
2. Institute of Psychology and Behavioral Science, Shanghai Jiao Tong University, Shanghai 200030, China
3. WLA Laboratories, World Laureates Association, Shanghai 201203, China
4. Global Institute of Future Technology, Shanghai Jiao Tong University, Shanghai 200240, China
5. Department of Genome Sciences, University of Washington School of Medicine, Seattle WA 98195, USA
6. Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
7. Affiliated Hospital of Qingdao University & Biomedical Sciences Institute of Qingdao University, Qingdao Branch of SJTU Bio-X Institutes, Qingdao University, Qingdao, Shandong 266003, China
8. Howard Hughes Medical Institute, University of Washington, Seattle WA 98195, USA
9. Department of Biomedicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan

*Corresponding authors, E-mail: lizqsjtu@163.com; eiki.takahashi.sjt@gmail.com; liwd@sjtu.edu.cn



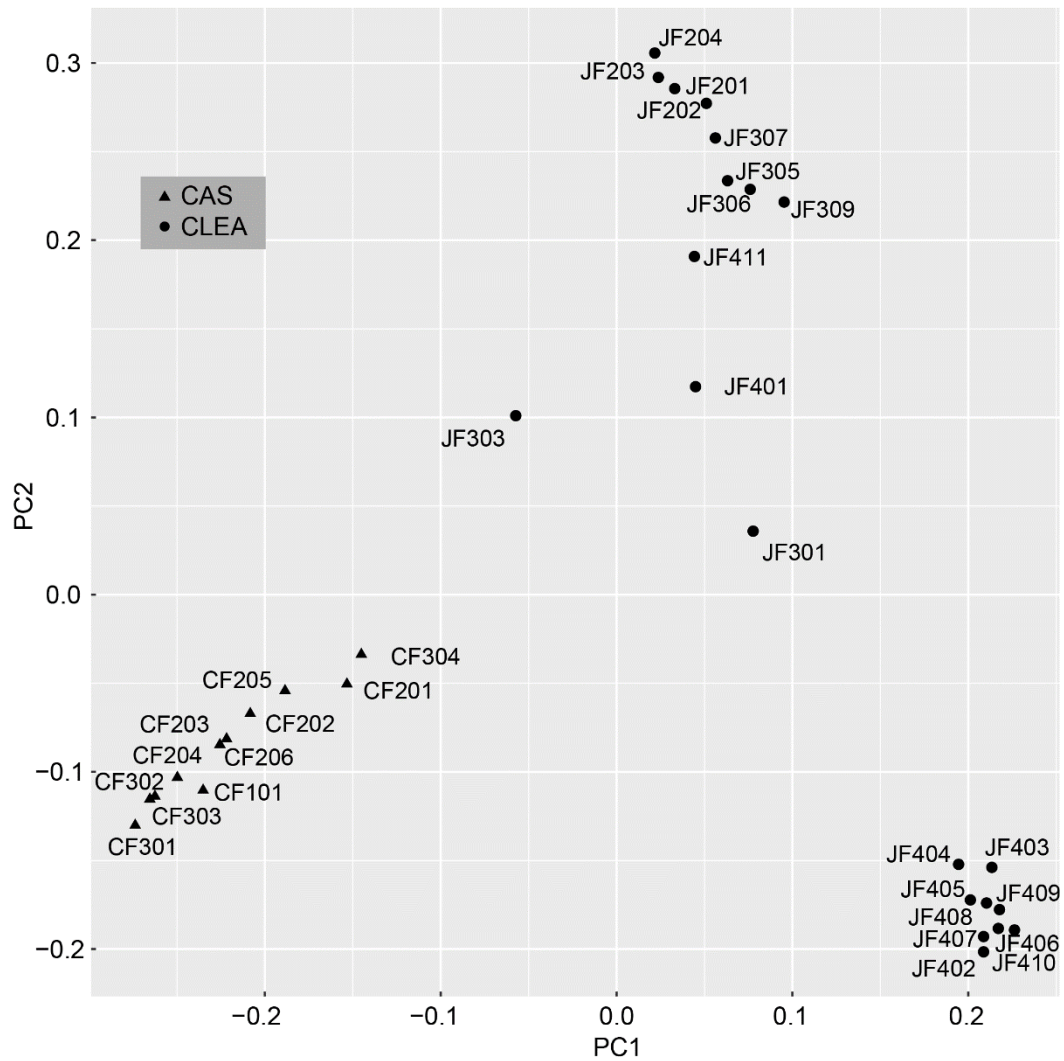
Supplementary Figure S1 Characteristics of neural activity in asymptomatic and epileptic marmosets

A: Schematic representation of EEG measurement in marmosets. B: Comparison of typical EEG spectra between asymptomatic and epileptic marmosets. C: Simple two-class classification results using support vector machines (SVM).



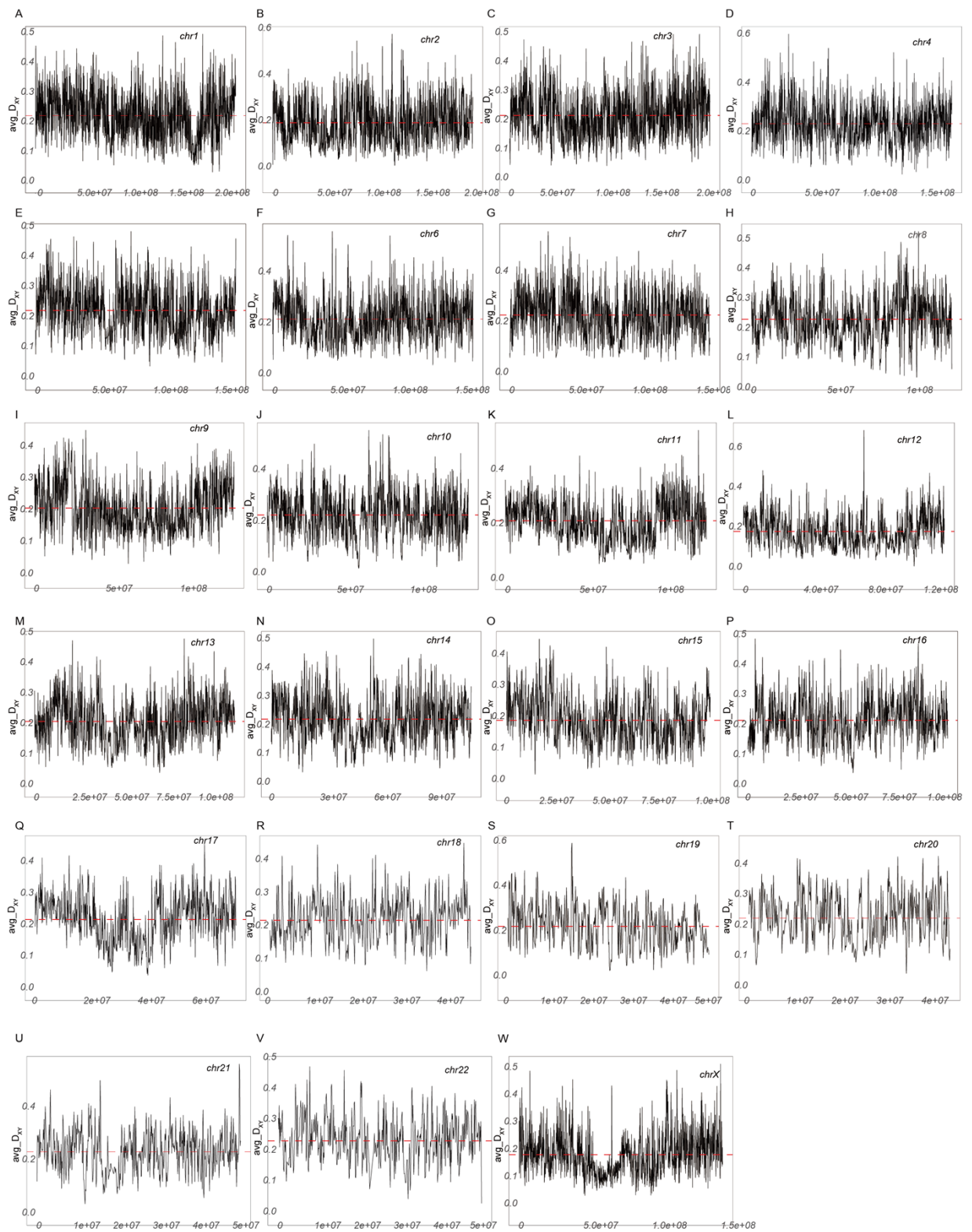
Supplementary Figure S2 Distribution of allele balance (alternative allele ratio)

A: Density plot of allele balance in blood and fingernail samples sequenced from CAS and CLEA colonies. B: Violin plot of allele balance < 0.3. Statistical analysis was conducted using Wilcoxon rank-sum test. Asterisk shows statistical significance (Blood_CLEA: Fingernail_CAS, $P=1.191 \times 10^{-5}$; Blood_CLEA: Fingernail_CLEA, $P=1.398 \times 10^{-7}$; Fingernail_CAS: Fingernail_CLEA, $P=0.0376$; *: $P < 0.05$; ***: $P < 0.001$).

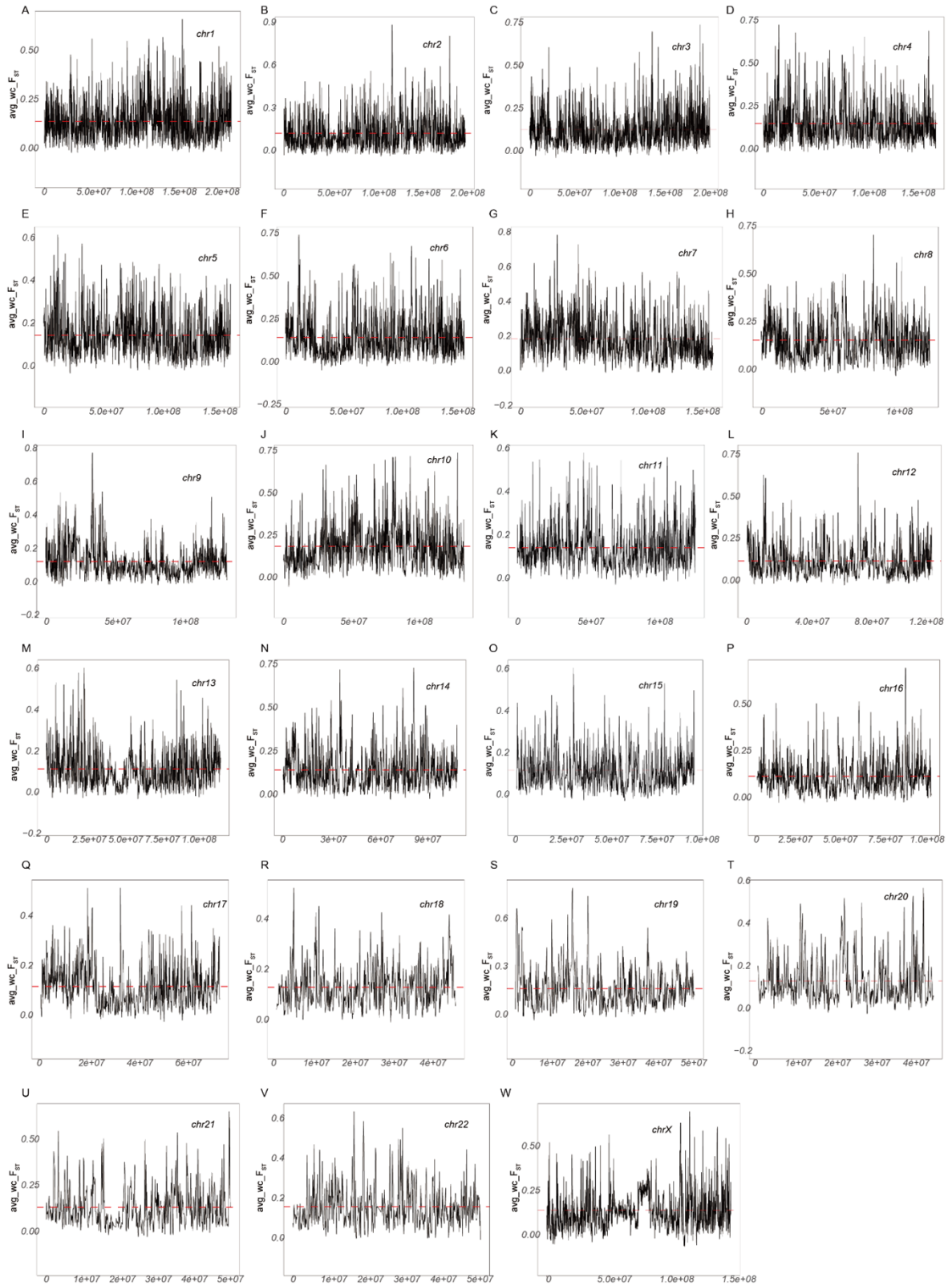


Supplementary Figure S3 Population structure of both marmoset colonies assessed through principal component analysis (PCA)

In total, 986 260 unlinked biallelic SNPs (10.86% of 9 084 947 biallelic SNPs) were used to perform PCA.

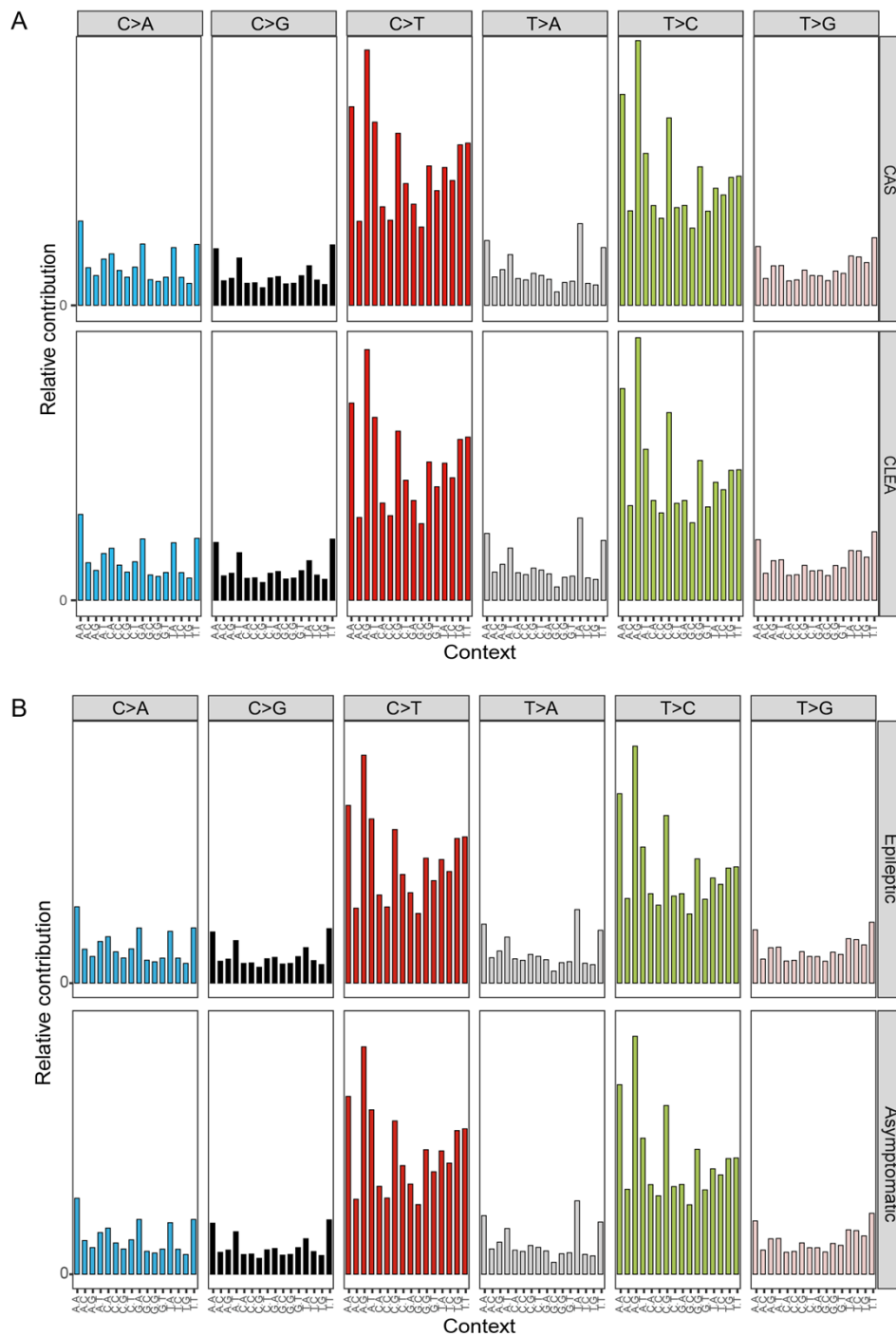


Supplementary Figure S4 Mean absolute nucleotide divergence (D_{XY}) between CAS and CLEA populations at chromosomal level
 Red dashed line indicates average D_{XY} .



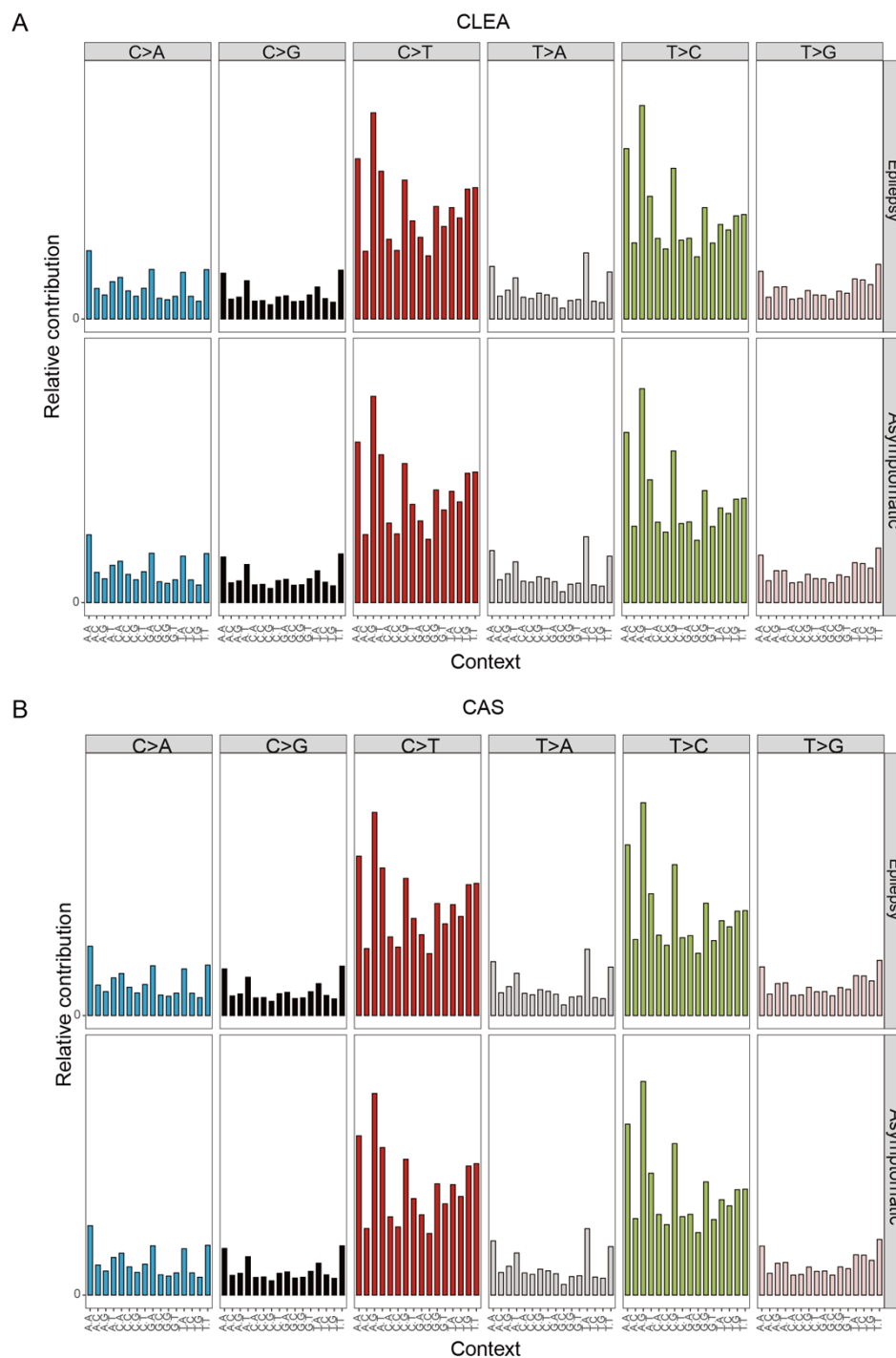
Supplementary Figure S5 Relative differentiation (F_{ST}) between CAS and CLEA populations

Red dashed line indicates average F_{ST} .



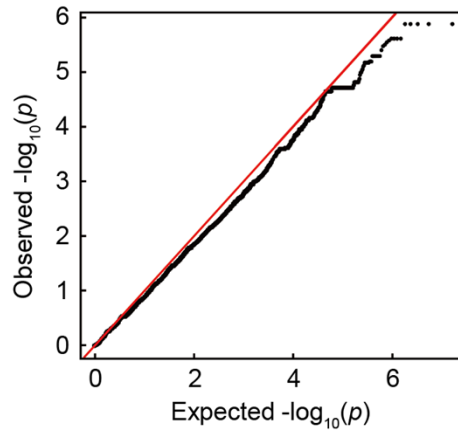
Supplementary Figure S6 Mutation spectrum comparison

A: Mutation spectrum between CAS and CLEA marmoset pedigrees. No significant difference was found in any mutation class between CAS and CLEA groups. B: Mutation spectrum between epileptic and asymptomatic marmoset groups. No significant difference was found in any mutation class between the epileptic and asymptomatic groups. Each mutation class was tested for each pair using the Wilcoxon rank-sum test (Supplementary Table S4).



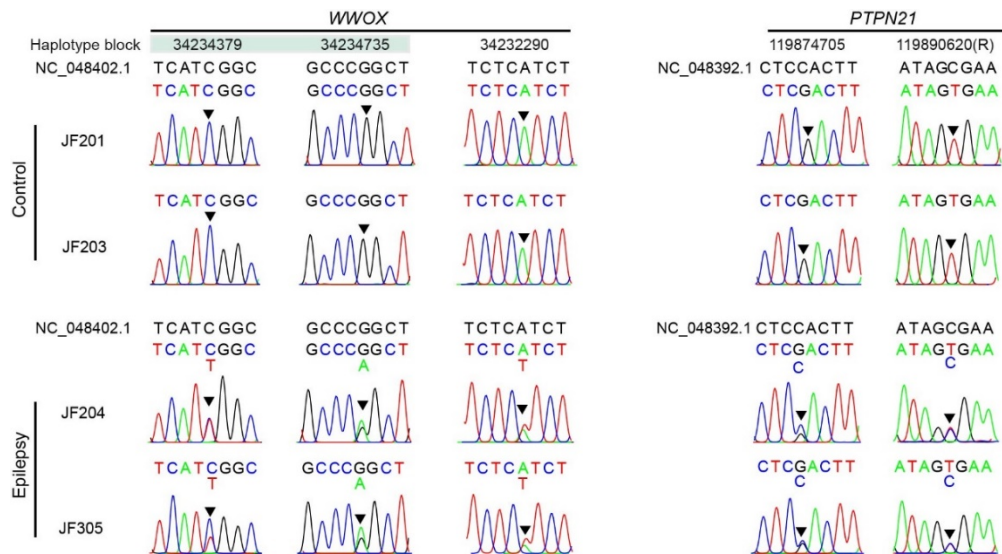
Supplementary Figure S7 Mutation spectrum comparison between CLEA and CAS colonies

Mutation spectrum comparison in CLEA and CAS. A: Mutation spectrum between epileptic and asymptomatic groups in CLEA pedigree (CTRL: CASE=9:11). B: Mutation spectrum between epileptic and asymptomatic marmoset groups in CAS pedigree (CTRL: CASE=6:5). Each mutation class was tested for each pair using the Wilcoxon rank-sum test (Supplementary Table S4).



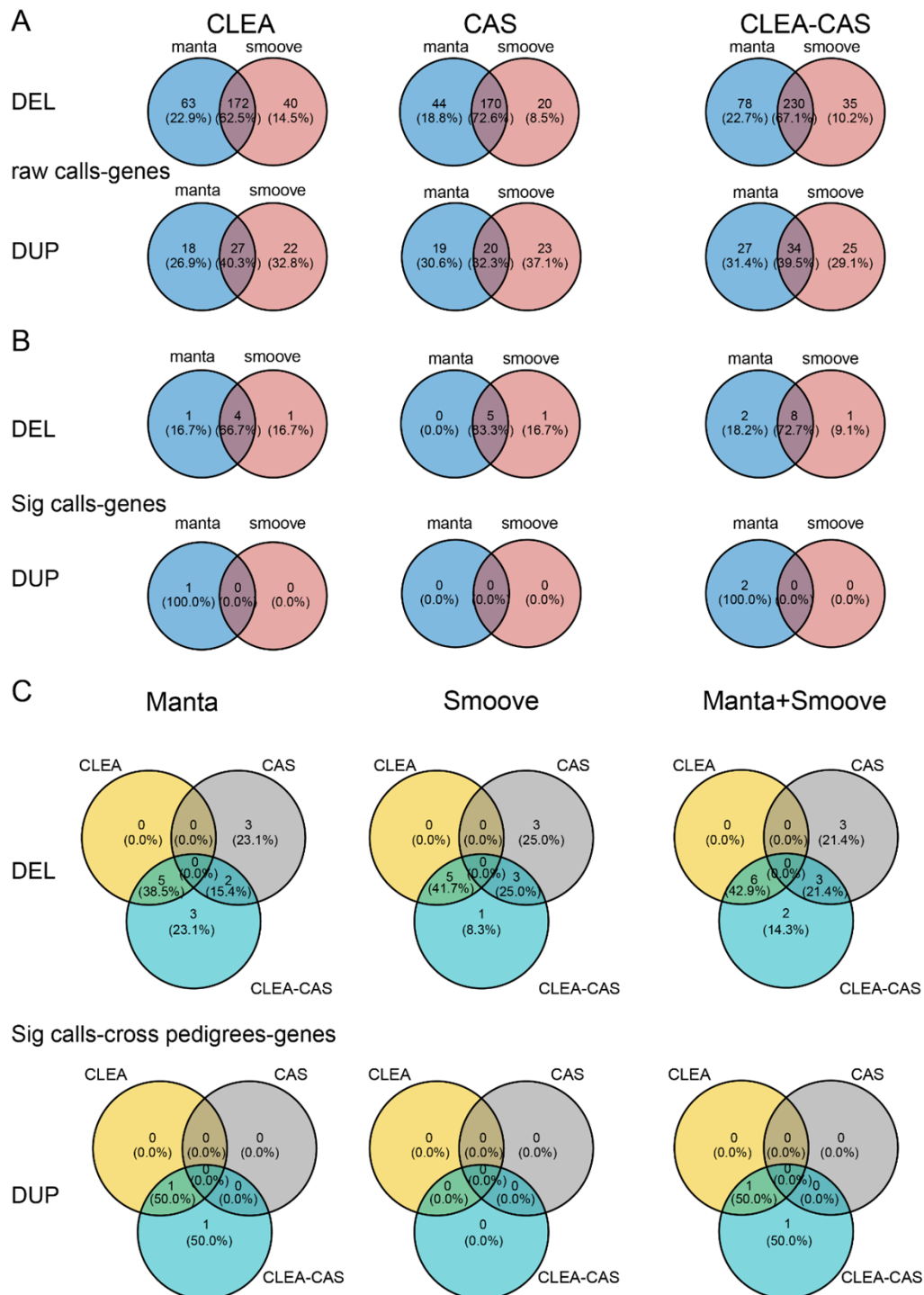
Supplementary Figure S8 Quantile-quantile (Q-Q) plot of GWAS analysis

Q-Q plot of observed (y-axis) vs. expected (x-axis) SNP P -value distributions. Expected P -values are those expected under the null hypothesis, with uniform null distribution marked with a red line. Genomic inflation factor ($\lambda=0.825$) showing evidence of substantial deflation, indicating that the test may be too conservative or, more likely, may be due to the small sample size.



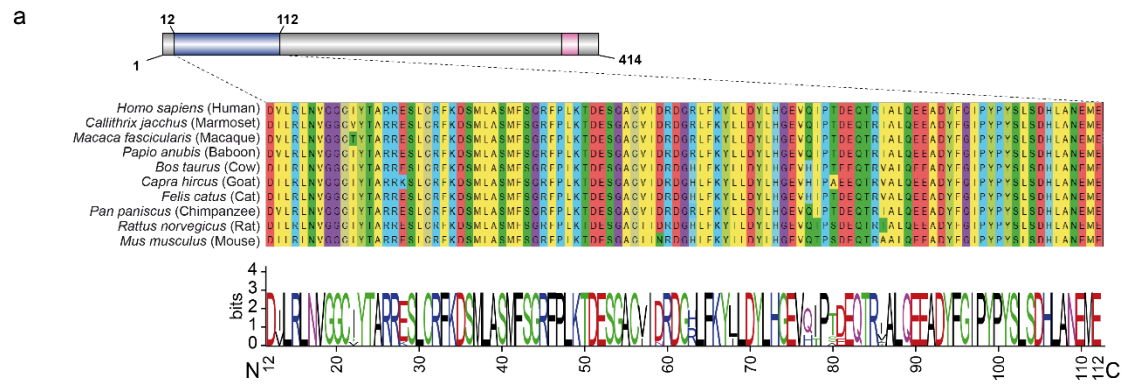
Supplementary Figure S9 Sanger sequencing validation results for control and epileptic samples in CLEA pedigree

Sanger sequencing results of two control and two epileptic marmoset samples. SNPs are marked with arrows. (R) indicates sequencing using reverse primer.



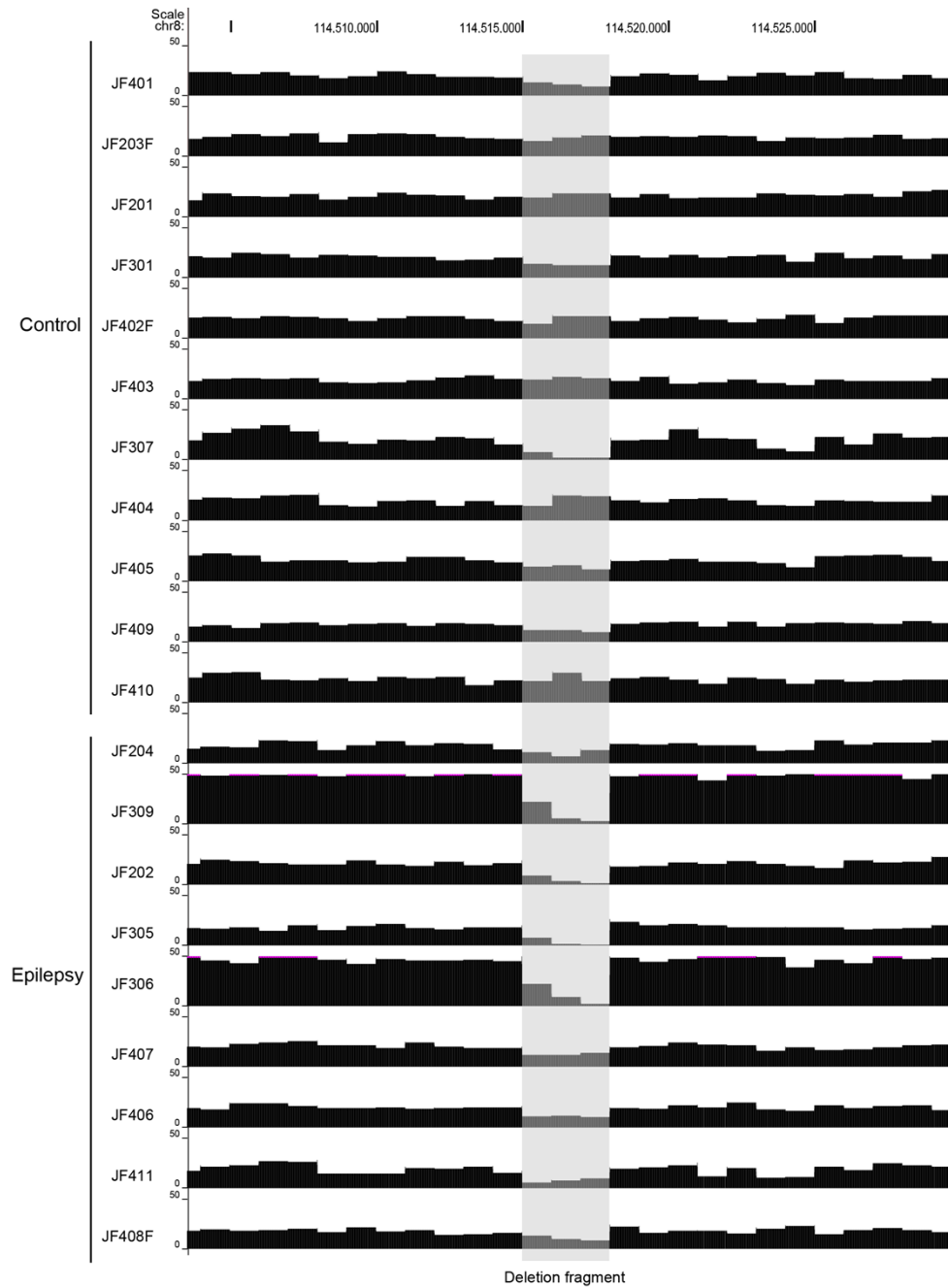
Supplementary Figure S10 CNV calling in different marmoset groups

A: Venn diagram showing raw CNV calls generated by two software packages. B: Venn diagram showing significant CNV calls generated by two software packages in different groups. C: Venn diagram showing significant CNV calls among and across different groups. CLEA indicates Japanese pedigree. CAS indicates Chinese pedigree. CLEA&CAS indicates integrated dataset of all marmosets from both pedigrees.



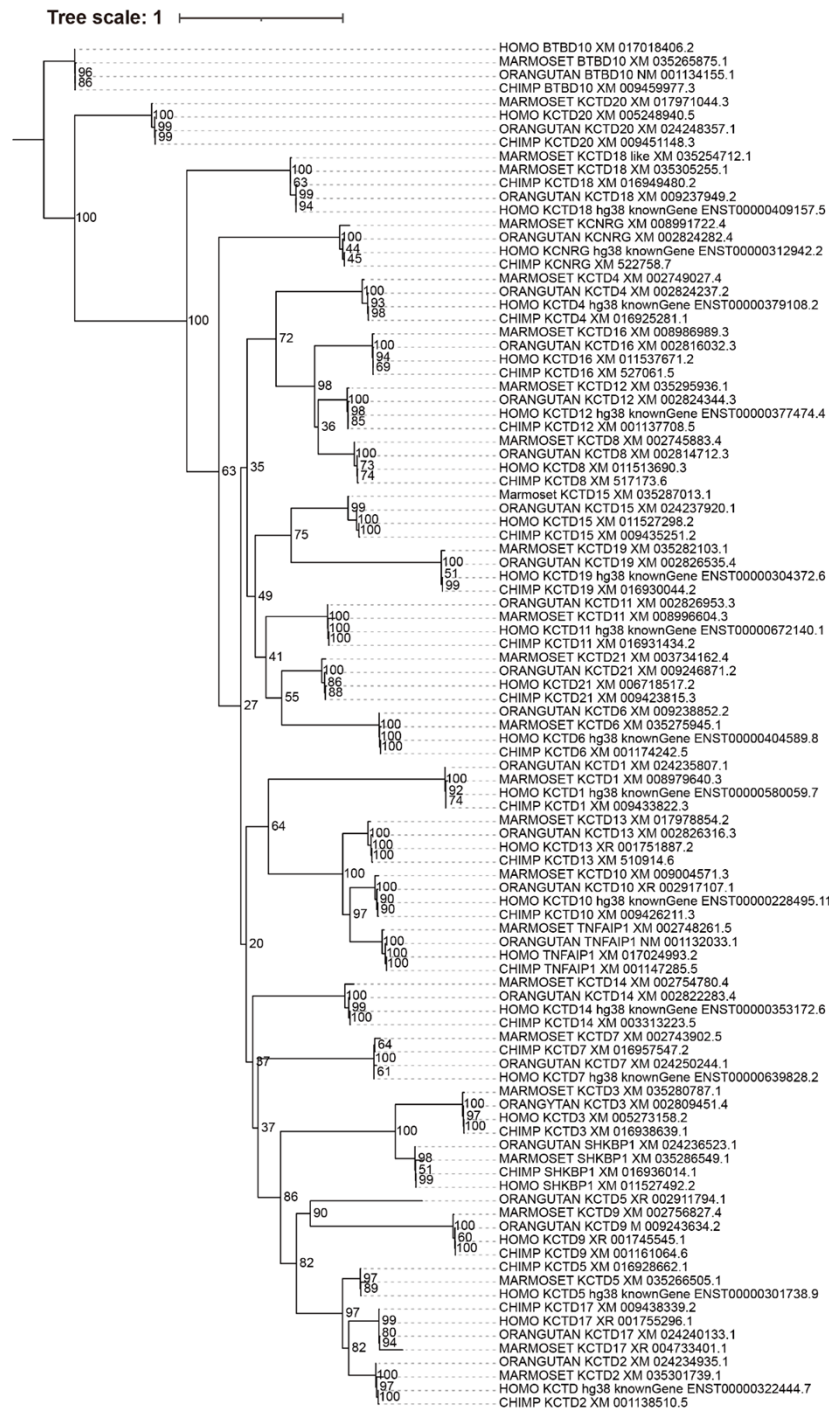
Supplementary Figure S11 KCTD18 protein structure

Predicted protein structure of human KCTD18 and multiple protein sequence alignments (MSAs) of 10 species for BTB domain of KCTD18, showing the same MSAs as a sequence logo.



Supplementary Figure S12 Read depth of fragment including *KCTD18-like* gene

Read depths of *KCTD18-like* gene region in 20 CLEA marmoset samples are listed (excluding mis-sampled JF303). Gray box indicates region of interest.



Supplementary Figure S13 Maximum-likelihood (ML) phylogenetic tree of *KCTD* family homologs in different species

Phylogenetic tree of *KCTD* family homologs in different species obtained by ML inference with IQTREE (v1.6.12). Numbers indicate bootstrap values.

Supplementary Table S1 Literature research results of worldwide geographic distribution of marmoset (*Callithrix jacchus*) research colonies

Supplementary Table S2 Sample information and sequencing depth of each sample from CLEA and CAS colonies

#	Sample ID	Sample tissue	Gender	Mean Depth (X)	Depth St.Dev.	Related/Unrelated	Phenotype	CAS/CLEA
1	JF303	Fingernail	Female	20.76	34.23	Unrelated	N/A	CLEA
2	JF204	Fingernail	Male	18.6	32.76	Related	Epilepsy	CLEA
3	JF401	Fingernail	Female	19.48	33.09	Related	N/A	CLEA
4	JF203	Fingernail	Male	19.31	33.53	Related	N/A	CLEA
5	JF309	Fingernail	Male	51.28	60.68	Related	Epilepsy	CLEA
6	JF201	Fingernail	Male	21.41	34.26	Related	N/A	CLEA
7	JF202	Fingernail	Male	20.89	29.21	Related	Epilepsy	CLEA
8	JF305	Fingernail	Female	17.85	28.07	Related	Epilepsy	CLEA
9	JF301	Fingernail	Male	20.87	31.21	Related	N/A	CLEA
10	JF306	Fingernail	Female	47.11	57.16	Related	Epilepsy	CLEA
11	JF402	Fingernail	Male	20.19	31.67	Related	N/A	CLEA
12	JF403	Fingernail	Female	19.86	29.47	Related	N/A	CLEA
13	JF307	Fingernail	Female	23.18	39.95	Related	N/A	CLEA
14	JF404	Fingernail	Male	19.83	32.5	Related	N/A	CLEA
15	JF405	Fingernail	Male	22.21	33.85	Related	N/A	CLEA
16	JF407	Fingernail	Female	21.45	29.17	Related	Epilepsy	CLEA
17	JF406	Fingernail	Male	20.05	29.74	Related	Epilepsy	CLEA
18	JF411	Fingernail	Male	19.95	34.71	Related	Epilepsy	CLEA
19	JF408	Fingernail	Male	17.2	27.76	Related	Epilepsy	CLEA
20	JF409	Fingernail	Female	18	26.96	Related	N/A	CLEA
21	JF410	Fingernail	Female	23.82	31.81	Related	N/A	CLEA
22	JF308	Blood	Male	29.57	41.28	Related	Epilepsy	CLEA
23	JF304	Blood	Female	28.72	35.49	Related	N/A	CLEA
24	JF205	Blood	Male	21.78	32.78	Related	Epilepsy	CLEA
25	JF302	Blood	Male	27.48	33.82	Related	Epilepsy	CLEA
26	JF203	Blood	Male	28.64	38.13	Related	N/A	CLEA
27	JF310	Blood	Male	29.81	38.08	Related	Epilepsy	CLEA
28	JF311	Blood	Male	31.4	34.95	Related	N/A	CLEA
29	JF402	Blood	Male	29.11	37.08	Related	N/A	CLEA
30	JF408	Blood	Male	31.17	39.74	Related	Epilepsy	CLEA
31	CF101	Fingernail	Male	35.72	39.24	Related	N/A	CAS
32	CF201	Fingernail	Male	36.48	39.05	Related	N/A	CAS
33	CF202	Fingernail	Male	30.73	37.11	Related	Epilepsy	CAS
34	CF205	Fingernail	Female	29.89	35.65	Unrelated	N/A	CAS
35	CF204	Fingernail	Male	29.59	33.5	Related	Epilepsy	CAS
36	CF206	Fingernail	Male	31.44	37	Related	Epilepsy	CAS

37	CF203	Fingernail	Female	29.5	33.74	Unrelated	N/A	CAS
38	CF304	Fingernail	Male	30.37	35.45	Related	N/A	CAS
39	CF301	Fingernail	Female	31.42	38.37	Related	N/A	CAS
40	CF302	Fingernail	Female	35.21	34.86	Related	Epilepsy	CAS
41	CF303	Fingernail	Female	43.26	41.95	Related	Epilepsy	CAS

Supplementary Table S3 Nucleotide diversity (π) of each chromosome between both colonies

CHR	CAS_100k_mean_pi	100k_SD_pi	CLEA_100k_mean_pi	100k_sd_pi
chr1	0.195135	0.0804887	0.191663	0.0767602
chr2	0.17142	0.0910328	0.168445	0.0809513
chr3	0.211431	0.0799409	0.172108	0.0808026
chr4	0.191375	0.0834907	0.208137	0.078158
chr5	0.19088	0.0816694	0.191837	0.0749524
chr6	0.185199	0.0784502	0.185322	0.0707416
chr7	0.164241	0.0890073	0.200995	0.0785714
chr8	0.21511	0.0856158	0.184062	0.0821621
chr9	0.190727	0.075559	0.176174	0.0758224
chr10	0.197067	0.0823796	0.17391	0.0773071
chr11	0.209269	0.073658	0.165919	0.0786482
chr12	0.160948	0.0774462	0.162406	0.0799391
chr13	0.185013	0.0766814	0.187249	0.0704824
chr14	0.180676	0.0821913	0.199173	0.0731149
chr15	0.167205	0.0745748	0.171954	0.0750715
chr16	0.185506	0.0809981	0.198813	0.076357
chr17	0.217313	0.0811892	0.174662	0.0768957
chr18	0.198531	0.0732442	0.187357	0.0737705
chr19	0.198484	0.0920783	0.178553	0.0932998
chr20	0.183498	0.0862446	0.205987	0.0744946
chr21	0.181226	0.0857026	0.21865	0.070194
chr22	0.22848	0.080824	0.171383	0.0791954
chrX	0.12814	0.0826756	0.175975	0.073522

Supplementary Table S4 Statistical results of mutation spectrum comparison

Supplementary Table S5 Raw data of genetic distance between CLEA and CAS colonies

Supplementary Table S6 Information on SNPs identified in our study and genotype with phasing results in CLEA and CAS colonies

Supplementary Table S7 Data of CNV calling in CLEA and CAS colonies

Supplementary Table S8 Read depth and electrophoretic quantitative results

ID	Phenotype	CNV region		
		depth	Average depth	Gray value
JF201	Asymptomatic	21.84	19.83	8.9
JF203	Asymptomatic	18.10	19.52	3.7
JF301	Asymptomatic	12.57	18.32	1.7
JF307	Asymptomatic	3.52	20.82	-0.9
JF401	Asymptomatic	10.99	19.28	4.2
JF402	Asymptomatic	19.52	20.87	12.9
JF403	Asymptomatic	20.37	20.93	8.4
JF404	Asymptomatic	20.81	17.57	20.4
JF405	Asymptomatic	13.50	22.67	3.7
JF409	Asymptomatic	10.60	17.19	1.4
JF410	Asymptomatic	23.95	22.24	10.2
JF202	Epileptic	4.40	21.02	-0.9
JF204	Epileptic	10.16	17.30	9.4
JF305	Epileptic	2.84	16.78	-0.6
JF306	Epileptic	10.75	45.64	-
JF309	Epileptic	9.65	50.91	-
JF406	Epileptic	10.87	19.65	3.2
JF407	Epileptic	12.06	18.87	8.7
JF408	Epileptic	10.28	14.44	0.3
JF411	Epileptic	7.41	18.39	-0.5

Notes:

1. CNV region depth indicates the depth of *Kctd-18 like* contained region. Average depth indicates the average of sequencing depth of upstream and downstream.
2. The gray value was analyzed and counted by *AlphaEaseFC* software.
3. Dash indicates no sample left for PCR validation.

Supplementary Table S9 Summary statistics of SNPs validated through Sanger sequencing

Supplementary Tables S1, S4-S7, S9 are listed as separate Excel files due to their large size.