Science Bulletin 66 (2021) 187–196

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

Article

# Whole genome analyses reveal significant convergence in obsessive-compulsive disorder between humans and dogs

Xue Cao [a,b,1], Wei-Peng Liu [c,d,1], Lu-Guang Cheng [e], Hui-Juan Li [c,d], Hong Wu [f], Yan-Hu Liu [a], Chao Chen [e], Xiao Xiao [c], Ming Li [c,d,g,h,*], Guo-Dong Wang [a,i,*], Ya-Ping Zhang [a,i,*]

[a] State Key Laboratory of Genetic Resources and Evolution and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[b] Department of Laboratory Animal Science, Kunming Medical University, Kunming 650500, China
[c] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[d] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650223, China
[e] Kunming Police Dog Base, Ministry of Public Security, Kunming 650204, China
[f] Laboratory for Conservation and Utilization of Bio-resource & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China
[g] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
[h] KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[i] Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

ARTICLE INFO

*Article history:*
Received 17 January 2020
Received in revised form 20 August 2020
Accepted 31 August 2020
Available online 15 September 2020

*Keywords:*
Obsessive-compulsive disorder
Circling behavior
Whole genome sequencing
Genetic convergence
Domestic dogs

ABSTRACT

Obsessive-compulsive disorder (OCD) represents a heterogeneous collection of diseases with diverse levels of phenotypic, genetic, and etiologic variability, making it difficult to identify the underlying genetic and biological mechanisms in humans. Domestic dogs exhibit several OCD-like behaviors. Using continuous circling as a representative phenotype for OCD, we screened two independent dog breeds, the Belgian Malinois and Kunming Dog and subsequently sequenced ten circling dogs and ten unaffected dogs for each breed. Using population differentiation analyses, we identified 11 candidate genes in the extreme tail of the differentiated regions between cases and controls. These genes overlap significantly with genes identified in a genome wide association study (GWAS) of human OCD, indicating strong convergence between humans and dogs. Through gene expressional analysis and functional exploration, we found that two candidate OCD risk genes, *PPP2R2B* and *ADAMTSL3*, affected the density and morphology of dendritic spines. Therefore, changes in dendritic spine may underlie some common biological and physiological pathways shared between humans and dogs. Our study revealed an unprecedented level of convergence in OCD shared between humans and dogs, and highlighted the importance of using domestic dogs as a model species for many human diseases including OCD.


## 1. Introduction

Obsessive-compulsive disorder (OCD) is a severe and disabling psychiatric disorder characterized by repetitive behaviors or mental acts to reduce distress from recurrent and persistent thoughts, urges, or images (obsessions) [1–3]. The lifetime prevalence of OCD is 2.5%–3% in world populations [2,4,5]. This disorder shows moderate heritability and constitutes a major health-economic burden [6–8]. The precise etiology of OCD remains unclear, but accumulating evidence suggests pivotal roles of structural abnormalities and/or dysregulations of synaptogenesis and neuronal circuits within particular regions of the brain in both OCD patients and affected animals [7,9–12]. Genetic analyses, including genome-wide association studies (GWAS), have been conducted to identify genes underlying risk of OCD in human populations [13–15]. However, results have been inconsistent, likely due to the phenotypic, genetic and etiologic heterogeneity of the disorder [16,17].

A small subset of domestic dogs (*Canis lupus familiaris*) exhibit some compulsive phenotypes that are similar to those observed in humans [18–20]. The canine compulsive disorder (CCD) manifests as repetition of normal canine behaviors such as:

* Corresponding authors.
*E-mail addresses:* limingkiz@mail.kiz.ac.cn (M. Li), wanggd@mail.kiz.ac.cn (G.-D. Wang), zhangyp@mail.kiz.ac.cn (Y.-P. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scib.2020.09.021
2095-9273/

(1) obsessions/checking including pacing/circling, tail chasing or fly snapping; (2) contamination/cleaning such as acral lick dermatitis; (3) eating disorders including pica, flank sucking (FS)/blanket sucking (BS) [21–23]. Given that most domestic dog breeds were generated from a limited number of founders who underwent rigorous selection for morphologic and behavior traits [24,25], this breeding program and its subsequent population bottlenecks have generated dogs of comparatively lower genetic diversity, increased frequencies of deleterious alleles, and an increased risk for multiple diseases [24,26–28]. Therefore, CCD may be less genetically complex than the corresponding human OCD, and utilization of the canine model is hypothesized to be an effective approach for identifying disease-related functional mutations [22,23,29,30].

We have analyzed dogs with and without stereotypical circling behaviors (one common phenotype of CCD) in two independent domestic breeds: the Belgian Malinois (BM), a western breed, and the Kunming dog (KMD), a Chinese working dog. While the two breeds are distinct, they are believed to share some common ancestry [31,32]. We sequenced 20 genomes from BM and KMD which demonstrated stereotypical circling behaviors and 20 from normal BM and KMD. We performed population differentiation analyses using re-sequencing data of dogs, GWAS statistics and mRNA profiling of humans, and identified several OCD candidate genes including *PPP2R2B* and *ADAMTSL3* which affected the density and morphology of dendritic spines.

## 2. Materials and methods

### 2.1. Ethical approval

All experimental protocols pertaining to animals have been reviewed and approved by the internal review board of the Kunming Institute of Zoology, Chinese Academy of Sciences.

### 2.2. Behavior test

Working dogs collected from the Kunming Police Dog Base underwent the behavior tests in their own kennel by the same researcher for more than two times. These tests were designed by our team based on practical experience and traits of the two breeds, and have been confirmed to effectively differentiate circling from normal behaviors. The test is split into three parts, and each part sustains at least 10 s. Part1: the tester stands away from the kennel about 3 m and keeps quiet; Part2: the tester stands beside the kennel and keeps quiet; Part3: the tester walks near the kennel, then intimidates and teases the dogs (Fig. S1, and Movie S1 online). The behavior presentation of dogs in all these parts of the test were recorded by camera. Behavior of the tested dogs in each part of the test was graded as 0, 1, and 2 according to their circling frequencies. If the dogs never showed circling behavior during the test, score 0 was recorded; if dogs showed intermittently circling behavior during the test, score 1 was assigned; if dogs showed invariably circling behavior during the test, it was given score 2. Subjects with the same behavioral presentation in at least two tests were used for further analyses. The age and sex were randomly selected in cases and controls. Finally, the gender ratio (female/male) was 5/5 for CCD-like circling dogs and 2/8 for control dogs in BM population, and 2/8 for CCD-like dogs and 5/5 for controls in KMD population. The average age for each breed and different behavior groups was about 5 (Table S1 online).

### 2.3. Genome sequencing

The genomes of 20 BMs (10 Circling, 10 Control) and 20 KMDs (10 Circling, 10 Control) were sequenced. Total genomic DNA was extracted from the venous blood according to a standard phenol–chloroform extraction procedure. Whole-genome sequencing of each subject was performed on the Illumina Hiseq 4000 platform, and more than 50 Gb of the sequence clean data was obtained. These sequencing data have been submitted to the Genome Sequence Archive (GSA, http://gsa.big.ac.cn/) under project number CRA001141.

### 2.4. Sequence data preprocessing and variant calling

Paired-end reads of our sequencing reads were all aligned to the dog reference genome assembly CanFam3.1 using the BWA-MEM version 0.7.10-r789 [33–35]. PICARD (version 1.87) was used to filter the reads with identical start/end points, and the Genome Analysis Tool Kit (GATK, version 3.7–0-gcfedb6) [36] was then used to local realignment and base-recalibration for sequences. Specifically, after genome alignment and removing PCR duplicates, the distribution of misincorporation near the ends of the reads were examined using mapDamage2.0 [37]. The calling of sequence variants were conducted using the UnifiedGenotypeCaller from GATK. During base and variant recalibration, a list of known single nucleotide polymorphisms (SNPs)/indels were downloaded from the Ensembl database to serve as the training set.

### 2.5. SNPs filtering and population structure

Indels were removed from further analyses. SNPs that had missing data, being triallelic or near the indels (no more than 5 bp) were also excluded. Principal component analysis (PCA) was then carried out on the remaining SNPs using the smartPCA program with the EIGENSOFT package v5.0.1 [38]. SNPs that were heterozygous in both populations were then collected for further analyses.

### 2.6. Population genetic analysis

The fixation index ($F_{ST}$) is a measure of population differentiation due to genetic structure [39]. The VCFtools (v0.1.13) [40] was used to estimate the $F_{ST}$ of each window to detect the differentiated genomic regions using the Weir and Cockerham's method comparing the circling and control dogs of the BM ($F_{ST}$ (BM)) and KMD ($F_{ST}$ (KMD)) breeds [41]. 50 kb was used as the window size and 10 kb was used as the stepwise size. The Root Mean Square of $F_{ST}$ (RMS_$F_{ST}$) from both groups were estimated using the following formula:

$$\mathrm{RMS}F_{ST} = \sqrt{\frac{1}{2}\left(F_{ST}(\mathrm{BM})^2 + F_{ST}(\mathrm{KMD})^2\right)}.$$

In addition, we also pooled the circling and control dogs from the BM and KMD breeds. Using the *Chi*-square test, SNPs with high differentiation between circling and control groups were detected using the VCFtools (v0.1.13) [40].

### 2.7. Annotation and detection for candidate genes

Extremely high $F_{ST}$ could drive behavior differences between cases and controls, so we firstly sorted the regions in descending order according to RMS_$F_{ST}$. Regions with extremely high RMS_$F_{ST}$ (top 1‰) were chosen as candidates, and genes (completely and partially) inside these regions were selected. Meanwhile, SNPs with highly significant $P$ values ($P \leq 0.001$) detected by *Chi*-square tests were used to annotate candidate genes with ANNOVAR [42]. Subsequently, genes harboring SNPs with significant $P$ values in the *Chi*-square tests ($P \leq 0.001$) were used to overlap with extreme RMS_$F_{ST}$ regions (top 1‰) to be defined as candidate genes for CCD.

The protein interaction analysis was conducted for our candidate genes, as well as previously reported CCD and human OCD relevant genes. The CCD relevant genes were defined through linkage disequilibrium-based clumping around SNPs with $P \leq 0.0001$ in a CCD GWAS dataset, which contained 92 rigorously phenotyped Doberman pinscher pica and flank sucking/blanket sucking cases and 68 controls [22,23]. The human OCD relevant genes were those whose related scores ranked top 20 of human OCD relevant gene list from GeneCards. The protein interaction analysis was performed by STRING (https://string-db.org/) [43], and only the experimental validated and co-expression active interaction sources were used.

### 2.8. Candidate genes analysis in GWAS data of human OCD

The GWAS result for human OCD was downloaded from the Psychiatric Genomics Consortium website (https://www.med.unc.edu/pgc/results-and-downloads) [16]. This dataset includes 2688 OCD patients and 7037 controls of European ancestry. All SNPs in this dataset were annotated using ANNOVAR [42], and the empirically significant SNPs ($P \leq 0.01$) residing in the candidate genes were retrieved.

### 2.9. Candidate genes analysis in expression data of human OCD

The expression profile data of human OCD (GSE60190) [44] was downloaded from the Gene Expression Omnibus (GEO) datdabase (https://www.ncbi.nlm.nih.gov/gds/) [45]. Gene expression data was obtained from the postmortem dorsolateral prefrontal cortex (DLPFC) tissues through the Illumina HumanHT-12 v3 microarray. The expression data for each candidate genes were retrieved, and the mean expression values of each gene, which were detected by different probes derived from the longest transcripts of the target gene, were calculated. After Shapiro-Wilk normality test, two-tailed *T* and MWU (Mann-Whitney *U*) tests were conducted for group comparisons when the data obeyed the normal distribution. In the comparative analysis, 102 non-psychiatric controls and 16 patients with obsessive–compulsive personality disorder or tics (named OC_mix, 6 individuals were pure OCD patients) were used.

### 2.10. Other analysis

The P-Match − 1.0 Public [46] was used to predict the transcriptional factor binding sites of the reference and derived sequences of dog genomes. The options of P-Match − 1.0 Public [46] for minimizing the sum of true positive and false positive error rates were selected for cut-off selection for matrix group, and nervous system expressed transcriptional factors were predicted.

### 2.11. Rat cortical neuronal cultures

Cortical neurons were prepared from E18 embryos of Sprague-Dawley rats, and cultured in serum-free medium following our previous studies [47,48]. Briefly, the frontal cortex tissues were dissected and minced in ice-cold sterile Hank's balanced salt solution. Dissociated cortical tissues were then digested, and gently pipetted to generate single cell suspension. Neurons were seeded at a density of $8 \times 10^5$ viable cells/well in 6 well culture-plates containing 25 mm coverslips, coated with poly-D-lysine (10 μg/mL, Sigma) and laminin (1.2 μg/mL, Invitrogen). Neurons were cultured in a 37 °C incubator supplemented with 5% $CO_2$. Half of the culture medium was changed with fresh medium every five days (Neurobasal media supplemented with 2% B27, 2 mol/L glutamax).

### 2.12. Vector construction

The gene-specific shRNA sequence designed for rat *Adamtsl3* was cloned into the lentiviral vector pSicoR-Ef1a-mCh-Puro (www.addgene.org/31845/) and verified through Sanger sequencing. The target sequence for silencing *Adamtsl3* was *Adamtsl3*-shRNA, 5′-GAATGGAAGTGCATGTATG-3′. The control shRNA sequence was: 5′-GATTTGCTGTTCGCCCAAG-3′.

For *Ppp2r2b* overexpression, the rat *Ppp2r2b* cDNA was respectively cloned into the pCAG vector (www.addgene.org/11150/) and the pLV lentiviral vector (www.addgene.org/22909/), with a C-terminus FLAG-tag. The recombinant constructs were verified by Sanger sequencing and restriction enzyme treatment.

Lentiviruses were generated by transfecting HEK293T cells with the above-described lentiviral recombinant vectors, the packaging plasmid psPAX2, and the envelope vector pMD2.G using Lipofectamine 2000 reagent (Invitrogen). Rat neurons were infected at 14–15 d *in vitro* (DIV) with lentiviruses and harvested for RT-qPCR and RNA-seq analysis 72 h post viral infection.

### 2.13. Real-time quantitative PCR (RT-qPCR)

RNA was isolated from cells using the TRIzol® reagent and reverse transcribed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RT-qPCR was performed using the FastStart Universal SYBR Green Master (Roche) to compare the relative expression levels of marker genes. ΔCt values were calculated by subtracting the *Sdha* Ct value from that of each target gene. Relative expression levels were calculated by using the $2^{-\Delta\Delta Ct}$ method. Primer sequences used for RT-qPCR were as follows:

*Sdha* rat F: 5′-CTCTTTCCTACCCGCTCACATAC-3′;
*Sdha* rat R: 5′-GCCAGTCAGAGCCTTTCACAGT-3′;
*Adamtsl3* rat F: 5′-GGGTAGTAGCGCCAAGAAGG-3′;
*Adamtsl3* rat R: 5′-AATGAAGAGCAGGGCAGGTC-3′;
*Ppp2r2b* rat F: 5′-GAAGAGTGTTTGCCAATGCG-3′;
*Ppp2r2b* rat R: 5′-GCTGGCTTGATGTCCACGAT-3′.

### 2.14. RNA-seq analysis

The total RNAs of cells infected with lentiviruses for *Adamtsl3*-shRNA (four biological replicates), control shRNA (four biological replicates), pLV-control (three biological replicates) or pLV-*Ppp2r2b* (three biological replicates) were extracted and sequenced with Illumina BGISEQ-500 using the PE151 sequencing strategy. More than 20 Gb clean data was obtained for each sample, and the sequencing quality was checked using FastQC (Babraham Institute, http://www.bioinformatics.babraham.ac.uk/projects /fastqc/). The first 15 bp of the reads were removed and only reads longer than 50 bp were retained for further analyses. Hisat2 was used to map the reads to the Rattus reference genome [49], and featureCounts and DEseq2 were then used to calculate the read counts and to identify differentially expressed genes [50,51]. After gene differential expression analysis, the FPKM (fragments per kilobase of exon per million fragments mapped) was used to quantify gene expression, and then the Pearson's correlation coefficients were calculated using cor function in *R* between *Ppp2r2b* or *Adamtsl3* and the respective remaining genes based on their FPKM values. Genes were considered differentially expressed when they had: Log2FoldChange $\geq$ 0.3 or Log2FoldChange $\leq$ −0.3, and P-adj $\leq$ 0.05, and correlation scores $\geq$ 0.5 or correlation scores $\leq$ −0.5 (correlation scores $\geq$ 0.85 or correlation scores $\leq$ −0.85 in *Adamtsl3* editing group). DAVID was used to perform the functional annotation and enrichment analysis for the differentially expressed genes [52].

### 2.15. Transfection in rat neurons and quantitative morphological analysis of dendritic spines

Analysis of dendritic spines in cultured neurons was performed following previous studies [47,53]. *Adamtsl3*-shRNA, control shRNA, pCAG-control or pCAG-*Ppp2r2b* was respectively transfected into cortical neurons together with Venus plasmid (which expressed EGFP) using Lipofectamine 2000 (Invitrogen) at 14–15 DIV. Immunofluorescence was performed 72 h after transfection. In brief, transfected neurons were fixed in PBS containing 4% paraformaldehyde and 4% sucrose for 15 min. Cells were then blocked in PBS containing 2% normal goat serum and 0.1% Triton-X-100 for 1 h at room temperature. Primary antibodies against mCherry (GeneTex), GFP (Abcam) or FLAG (Cell Signaling Technology) were incubated with cells overnight at 4 °C, after which the secondary antibodies were incubated with cells in the dark at room temperature for 1 h.

Images of the transfected neurons were captured using a LSM 880 Basic Operation (Carl Zeiss) confocal microscope with the 63 × oil-immersion objective as z series of 41 images taken (averaged 2 times, with 0.25 μm intervals, 1024 × 1024 pixel resolution). Only the secondary and tertiary dendrites (total length of 60–100 μm) were subjected to morphological analysis. An average of two dendrites per neuron on more than 20 neurons from each experimental group were analyzed. Dendritic branches were traced and measured using the Image-J (http://research.mssm.edu/cnic/tools-ns.html), and NeuronStudio [54] was used to quantify spine numbers and classify spines as thin, mushroom or stubby according to previous studies [55,56]. Statistical analyses between two groups were calculated using two-way ANOVA and two-tailed *t*-test.

## 3. Results

### 3.1. Behavior tests and whole genome sequencing

Both the BM and KMD dogs were obtained from the Kunming Police Dog Base. The present study focused exclusively on circling behavior which manifests in highly similar phenotypes in both the BM and KMD as described in the methods (Fig. S1 and Movie S1 online). We assigned scores 0, 1, or 2 for each of the three levels of circling behavior to quantify such phenotype in these dogs. For each breed, ten subjects whose performances were scored as 1 or 2 for all three levels were selected as cases, i.e., circling subjects, and ten with behavioral performances scored 0 in all three levels of the tests were defined as controls. The age and sex were randomly selected in cases and controls. All information of the samples was presented in Table S1 online.

All 40 subjects included in the current study were sequenced for an average of 75 Gb (60.36–87.60 Gb). All reads were mapped to the dog reference genome (CanFam3.1) [33,34] with an average mapping rate of 99%. Sequencing depth ranged from 17.65 to 27.38-fold coverage (Table S1 online). After genotyping using the Genome Analysis Toolkit (GATK) [57], strict filtering was carried out as described in the Methods. Finally, we identified ~ 9.6 million autosomal SNPs for further analysis.

### 3.2. Population stratification and differentiation analysis

PCA was performed based on whole-genome SNPs to examine population structure within each breed. The BM and KMD could be completely separated by the first and second PCs (Fig. S2a online), suggesting that they were independent populations.

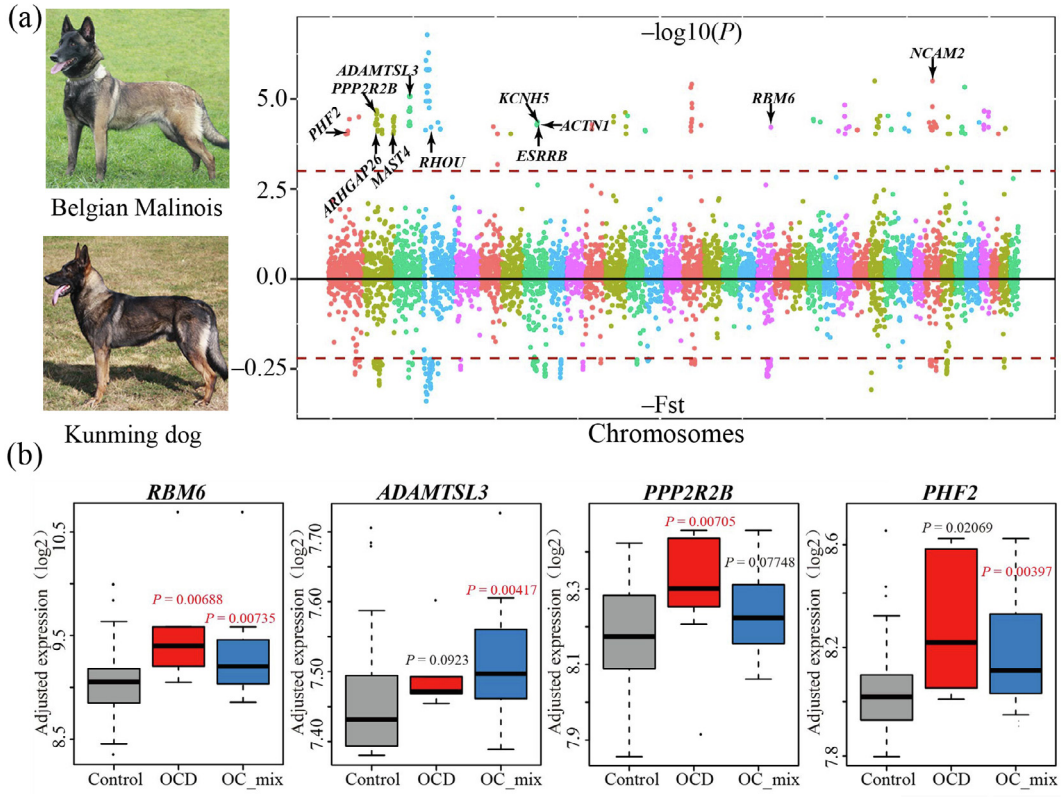



**Fig. 1.** The results display for whole genome sequencing data and candidate genes expression analysis. (a) The Manhattan figure for $-F_{ST}$ values and –log10(*P*) values of *Chi*-square test. The red lines indicated the cutoff of the $F_{ST}$ and *Chi*-square test. (b) Four genes differently expressed between non-psychiatric controls and OCD or OC_mix. The *P* value were calculated between OCD and non-psychiatric controls, or OC_mix and non-psychiatric controls.

Indeed, the divergence between the BM and KMD breeds (average $F_{ST}$ = 0.10, calculated using all autosomal SNPs) is greater than that between Han Chinese and Europeans estimated using HapMap and Perlegen samples (average $F_{ST}$ = 0.06) [58]. Although our results indicate that the breeds are independent from each other, and Guo et al. [59] have shown that the KMD breed is completely separated from its ancestry (German Shepherd dog), BM and KMD breeds are believed to share some common ancestry components [31,32]. Therefore, they are suitable for exploration of genetic basis underlying circling behavior both as separate populations and one single cohort (for the shared variants). Cases and controls in the BM and KMD did not show evident population stratification (Fig. S2b, c online).

To identify loci associated with circling behavior in both breeds, a set of 5.8 million informative SNPs were analyzed. We calculated the sliding window $F_{ST}$ values along the genomes between circling and control subjects in BM ($F_{ST}$ (BM)) and KMD ($F_{ST}$ (KMD)) populations, respectively. We also calculated the $F_{ST}$ (BM) and $F_{ST}$ (KMD) for randomly-selected subjects in each population. The results showed that $F_{ST}$ values obtained in the case-control setting (Median $F_{ST}$ (BM) = 0, $F_{ST}$ (KMD) = 0.001) was similar to that obtained in randomized settings (Median $F_{ST}$ (BM) = 0, $F_{ST}$ (KMD) = 0). This suggests that regions with elevated $F_{ST}$ values may not be due to random fluctuations, but rather be driven by behavior differences between the cases and controls. Moreover, the kinship analysis showed that even though the cases and controls had some relatedness (Table S2 online), they are well matched to each other. The root mean square of $F_{ST}$ (RMS_$F_{ST}$) was then calculated based on $F_{ST}$ (BM) and $F_{ST}$ (KMD) as described in the Methods. The values of RMS_$F_{ST}$ were distributed in all sliding windows from 0 to 0.340 (Fig. 1a). 67 genes in the sliding windows with the highest RMS_$F_{ST}$ (top 1‰) were identified (Table S3 online).

The genotype results from both BM and KMD subjects were then combined to perform a *Chi*-square test between cases and controls to validate the results of the $F_{ST}$ test. Significant SNPs ($P \leq 0.001$) between distinct behavior groups were used for gene annotation with ANNOVAR [42], and 190 genes were thereby chosen based on a stringent cutoff (Table S3 online). When combine the $F_{ST}$ and the *Chi*-Square test, 11 candidate genes were selected in both tests (Table 1). In order to assess the level of false positive (FP), the labels for cases and controls were randomly permuted and a similar analysis was carried out using both tests (Supplementary Methods online). Across all permutations, we only identified an average of 2.4 genes, which indicates a reasonably low FP rate in our analytical procedure. These results suggest that majority of these 11 genes are likely driving the circling behaviors in both breeds.

We also examined the previously reported CCD-associated genes, including neuronal cadherin (*CDH2*) [22,23,60], catenin alpha 2 (*CTNNA2*) [22], ataxin-1 (*ATXN1*)[22], and plasma glutamate carboxypeptidase (*PGCP*) [22] in our sample. We found that SNPs within these genes (22 SNPs in *CDH2*, 12 SNPs in *CTNNA2* and 81 SNPs in *PGCP*) exhibited significant differences between cases and controls at nominal *P* value of 0.05 (*Chi*-square tests), and the probability of observing normal significant SNPs in three of four genes was zero in randomly-selected genes (Supplementary Methods online). These results suggested the existence of both shared and distinct genetic mechanisms between subtypes (obsessions/checking, contamination/cleaning and eating disorders) of CCD showing different obsessive behaviors. Moreover, in protein interaction analysis, the proteins encoded by these 11 candidate genes showed direct physical protein–protein interactions (PPIs) with OCD related genes detected in previous studies [22,23,25] (Detailed information is shown in Methods and Fig. S3 online), supporting a putative role for these genes in the pathogenesis of OCD.

### 3.3. Significant convergence of the risk genes with human OCD

The similarity in repetitive phenotypes (obsessions/checking, contamination/cleaning and eating disorders) between CCD and human OCD suggests that they may share certain common risk genes. Using data from a recent GWAS for human OCD with 2688 cases and 7037 controls [16], a total of 1,986 human genes respectively contained more than 3 SNPs showing nominal associations ($P \leq 0.01$) were found. Among the orthologous gene pairs between humans and dogs (a total of 14,313 gene pairs), 9 out of the 11 genes overlapped with the GWAS hits existed in the humans (Table 1, Fig. S4 online). Using a random sampling based approach (Supplementary Methods online), we revealed a significant convergence (7.26-fold enrichment, $P < 0.0001$) on the genetic basis of the compulsive disorder between humans and dogs.

### 3.4. Expression analyses of the risk genes in human OCD

We then retrieved gene expression data of human OCD (GSE60190) [44] from the GEO [45] database. Notably, the expression of *RHOU* was not detected in this dataset, and we therefore examined expression of the other 10 candidate genes as described above. A total of 102 non-psychiatric controls and 16 patients with OC_mix were included, among whom 6 patients were pure OCD. We found that *RBM6* (RNA binding motif protein 6), *PPP2R2B* (protein phosphatase 2 regulatory subunit Bbeta), *PHF2* (PHD finger protein 2) and *ADAMTSL3* (ADAMTS like 3) were differentially expressed between non-psychiatric controls and OCD or OC_mix

**Table 1**
The summary list for the candidate genes.

| Symbol | Gene name | WGS[a] | GWAS[b] | OCD[c] | | OC_mix[c] | |
|---|---|---|---|---|---|---|---|
| | | | | *P* | FDR | *P* | FDR |
| *RBM6* | RNA binding motif protein 6 | 1 | 3 | 0.00688 | 0.0688 | 0.00735 | 0.0735 |
| *RHOU* | Ras homolog family member U | 2 | 0 | - | - | - | - |
| *PPP2R2B* | Protein phosphatase 2 regulatory subunit Bbeta | 27 | 9 | 0.00705 | 0.0353 | 0.07748 | 0.3870 |
| *PHF2* | PHD finger protein 2 | 4 | 6 | 0.02069 | 0.0690 | 0.00397 | 0.0132 |
| *NCAM2* | Neural cell adhesion molecule 2 | 124 | 27 | 0.10319 | 0.2060 | 0.11762 | 0.2350 |
| *MAST4* | Microtubule associated serine/threonine kinase family member 4 | 2 | 9 | 0.72227 | 0.7220 | 0.61215 | 0.6120 |
| *KCNH5* | Potassium voltage-gated channel subfamily H member 5 | 2 | 6 | 0.26850 | 0.3360 | 0.84728 | 1.0600 |
| *ESRRB* | Estrogen related receptor beta | 8 | 9 | 0.23523 | 0.3360 | 0.27284 | 0.3900 |
| *ARHGAP26* | Rho GTPase activating protein 26 | 4 | 4 | 0.44058 | 0.4900 | 0.47199 | 0.5240 |
| *ADAMTSL3* | ADAMTS like 3 | 37 | 4 | 0.09233 | 0.2310 | 0.00417 | 0.0104 |
| *ACTN1* | Actinin alpha 1 | 1 | 0 | 0.18359 | 0.3060 | 0.04916 | 0.0819 |

Note: [a] Number of significant SNPs in WGS analysis of circling dogs. [b] Number of significant SNPs in GWAS analysis of human OCD. [c] The differential expression analysis was performed in OCD and OC_mix samples respectively.

subjects ($P \leq 0.01$, Table 1, Fig. 1b). *PPP2R2B* and *ADAMTSL3* were still differentially expressed even after multiple testing correction (FDR $\leq 0.05$).

The mRNA expression profiles of those genes in dogs are not available. We noticed, however, that the significant SNPs detected in cases were in intronic or untranslated regions (UTRs) of the risk genes (Table S4 online). Functional predictions suggested that these variants might affect binding affinities of transcription factors (Table S5 online), especially for the variants within *PPP2R2B* and *ADAMTSL3*. Collectively, the above cumulative data identify several potential risk genes for OCD, among which *PPP2R2B* and *ADAMTSL3* are consistently highlighted across dogs and humans.

### 3.5. Dysregulation of Ppp2r2b or Adamtsl3 in rat neurons leads to altered dendritic spine density and morphology

Previous studies suggest that OCD pathogenesis is linked with abnormalities in dendritic spine development and synaptic plasticity [61–67] as well as dysregulation of excitatory synaptic gene expression in the brain [9]. Moreover, one human OCD relevant gene (*NCAM2*), which is not only related to OCD but also responsible for synapse maturation, stability and dendritic spine development [68], was detected by our population stratification and differentiation analysis. This gene also contained more SNPs significantly highlighted in our dog circling analysis compared with other candidate genes (Table 1).

We therefore sought to test whether *PPP2R2B* and *ADAMTSL3* were also associated with spine development using lentiviruses infection followed by RNA-seq in primary rat cortical neurons. Given that the *Adamtsl3* gene is too long to be successfully overexpressed, we used an shRNA to reduce *Adamtsl3* expression to examine its function in neurons. In brief, lentiviral constructs of *Adamtsl3*-shRNA, control shRNA, pLV-control and pLV-*Ppp2r2b* were established to infect the wild-type rat cortical neurons for 72 h. The expression levels of *Adamtsl3* and *Ppp2r2b* were confirmed via RT-qPCR and RNA-seq. The expression of *Adamtsl3* was reduced by 27% in neurons infected with *Adamtsl3*-shRNA relative to cells infected with the control shRNA, and the expression of *Ppp2r2b* in cells infected with pLV-*Ppp2r2b* increased 11.7 times compared with the controls (Fig. S5 online). The differentially expressed genes (DEGs), after manipulation of the risk genes (*Ppp2r2b* and *Adamtsl3*) in cultured neurons, were identified by RNA-seq. The 181 DEGs in *Ppp2r2b* overexpressing neurons (Table S6 online) and 297 DEGs in *Adamtsl3* knockdown neurons (Table S7 online) were enriched in neurogenesis, synapse and dendritic spine relevant function clusters by the DAVID analysis (Enrichment Score $\geq 1.33$, Table S8 online). We thus hypothesized that expressional changes in *Ppp2r2b* or *Adamtsl3* expression would lead to changes in the density and morphology of dendritic spines in neurons.

To test this hypothesis, we examined the dendritic spine density and morphology in rat cortical neurons transfected with *Adamtsl3*-shRNA, control shRNA, pCAG-control and pCAG-*Ppp2r2b* separately using confocal imaging analysis. We found that neurons overexpressing *Ppp2r2b* had a significantly lower density of total spines compared with control neurons (control, 3.16 spines per 10 μm; overexpression of *Ppp2r2b*, 2.50 spines per 10 μm; $P = 0.0015$, two-tailed *t*-test; Fig. 2a, b). In a further morphological analysis, a significant decrease in the density of thin spines was observed after overexpression of *Ppp2r2b* (control, 1.26 spines per 10 μm; overexpression of *Ppp2r2b*, 0.80 spines per 10 μm; $P = 6.76 \times 10^{-5}$, two-way ANOVA; Fig. 2b), while the density of mushroom or stubby spines were not altered. Through analyzing proportions of each subtype of dendritic spines (mushroom, stubby and thin), we saw that overexpression of *Ppp2r2b* induced a significant reduction of the proportion of thin spines ($P < 0.01$) and a concomitant increase in the proportion of stubby spines ($P < 0.05$) (Fig. 2b). We also studied the effect of *Adamtsl3* on dendritic spines using RNA interference. Plasmids expressing either control shRNA or shRNA targeting rat *Adamtsl3* were introduced into neurons at DIV-14. Neurons expressing *Adamtsl3* shRNA exhibited similar densities of total spines compared with the control neurons, whereas reduction of *Adamtsl3* resulted in a significant increase of density and proportion of mushroom spines compared with controls ($P < 0.05$, two-way ANOVA, Fig. 2c, d). Therefore, *Ppp2r2b* and *Adamtsl3* likely affect the development of dendritic spines, and *Ppp2r2b* may selectively affect spine formation and *Adamtsl3* might play roles in promoting dendritic spine maturation.

## 4. Discussion and conclusion

In this study we performed behavior tests for dogs with and without circling behavior from two independent breeds, and conducted whole genome sequencing in 10 cases and 10 controls from each breed to decipher the genetic mechanism of circling behavior as one analogous endophenotype of human OCD. Through population genetic analyses, we found 11 highly differentiated genes between circling and control populations, of which nine contained independent SNPs showing nominal associations for human OCD. Statistical analysis revealed a significant convergence (7.26-fold enrichment, $P < 0.0001$) on the genetic basis of the compulsive disorder between humans and dogs. Meanwhile, we found 4 of 11 genes showed significant differential gene expression between OCD and non-psychiatric controls in humans ($P \leq 0.01$). Interestingly, *PPP2R2B* and *ADAMTSL3* had statistical significance in expression data after multiple testing correction (FDR $\leq 0.05$) and 2 key variants in them were found to affect transcription factors binding affinities based on the computational prediction. Subsequently, we designed a series of functional experiments through lentiviruses infection, RNA-seq, and confocal imaging analysis. As we found out, neurons overexpressing *Ppp2r2b* had a significantly lower density of total spines, and neurons expressing *Adamtsl3* shRNA exhibited a significant increase of density and proportion of mushroom spines, consistent with the previous observations that OCD pathogenesis is linked with abnormalities in dendritic spine development and synaptic plasticity [61–67] as well as dysregulation of excitatory synaptic gene expression in the brains of OCD patients [9]. Taken together, we have identified a strong convergence in the genetic constitution of OCD in humans and dogs.

Although SNPs in the 11 candidate CCD risk genes did not reach genome-wide level of statistical significance in the human GWAS, it is not unexpected given that human OCD is presumed to have higher phenotypic and genetic heterogeneity than CCD. Indeed, genome-wide surveys for human OCD have identified only a limited number of risk genes demonstrating genome-wide level of statistical significance [16]. Meanwhile, aggregating and polygenic analyses suggest that there might be true associations among those markers exhibiting nominal significance in GWAS of complex diseases [69]. It has been shown that stringent multiple corrections may preclude the discovery of genuine biological risk-associated genes when GWAS sample sizes are not large enough, in which case such genes might only show nominal significance.

The current study revealed that functional manipulation of the risk genes (*Ppp2r2b* and *Adamtsl3*) resulted in differential expression of genes enriched in neurogenesis, synapse and dendritic spine relevant function clusters. The experiment further showed that *Ppp2r2b* and *Adamtsl3* could affect the density and morphology of dendritic spines in cultured neurons. These results provide compelling evidence for the potential link between OCD and dendritic spines, and are also in agreement with the recent studies showing spine and synaptic pathology in OCD [61–67,70,71]. For example, Vondervoort et al. [64] found that dendritic spine forma-

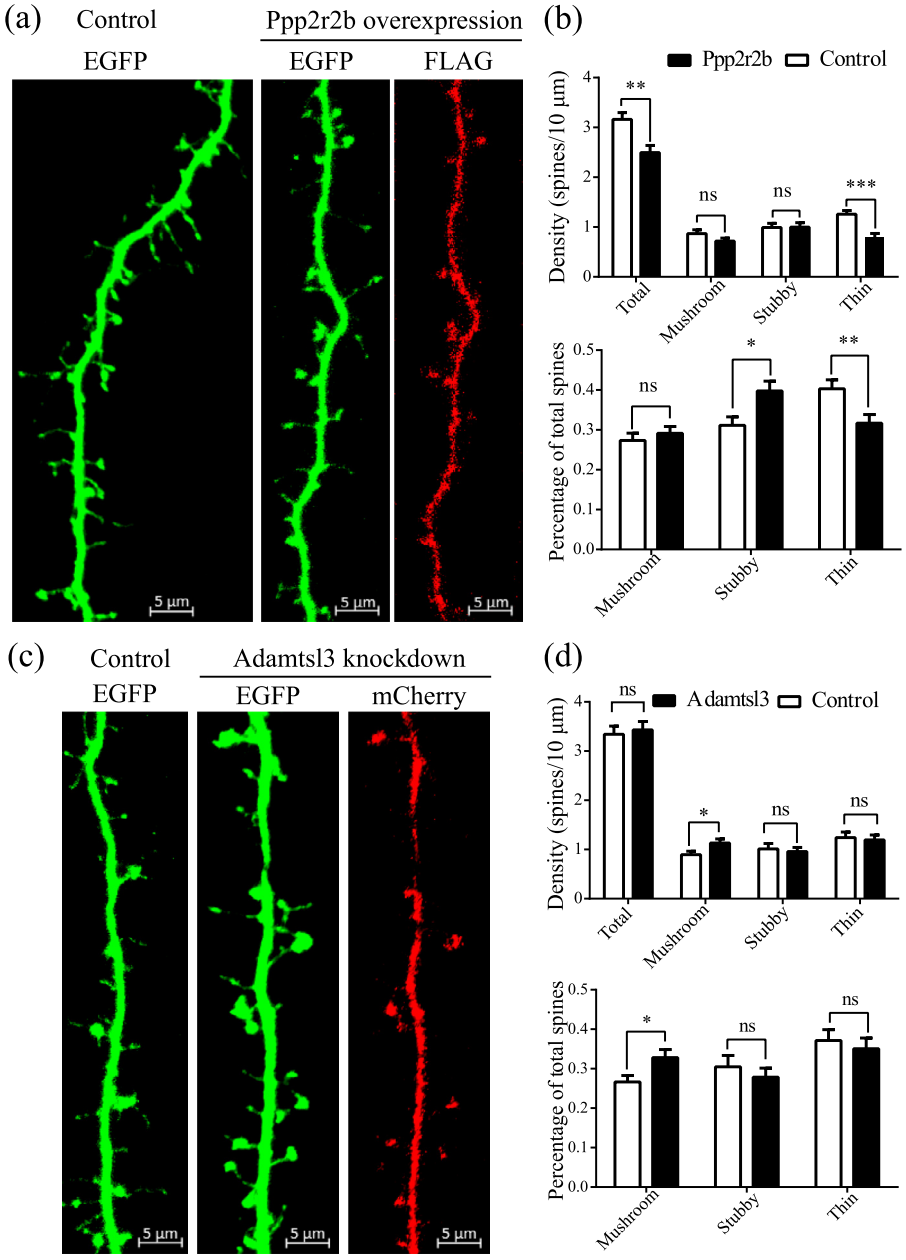



**Fig. 2.** Manipulation of *Ppp2r2b* and *Adamtsl3* expression in rat cortical neurons and dendritic spine analysis. (a) Representative figures of dendrites of rat cortical neurons transfected with pCAG-control and pCAG-*Ppp2r2b* respectively. (b) The density and percentage of spines for each condition. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and ns, not significant. $n = 25$ neurons per condition. (c) Representative figures of dendrites of rat cortical neurons transfected with *Adamtsl3*-shRNA and control shRNA separately. (d) The density and percentage of spines for each condition. $^{*}P < 0.05$ and ns, not significant. $n = 23$ neurons per condition.

tion could be regulated through insulin-related signaling and contributed to OCD. In addition, several synaptic proteins such as *SLC1A1*, *SLITRK5* and *SAPAP3* have also been reported to be associated with OCD or OCD-like behaviors [72–74]. We found that overexpression of *Ppp2r2b* in neurons reduced the density of dendritic spines, especially the thin spines which are immature and plastic [67,75]. Intriguingly, the thin spines are generally accepted as learning spines due to their greater potential for strengthening [76]. Hence, our results suggest that *Ppp2r2b* may selectively target immature spine structures and further affect synaptic plasticity in OCD pathogenesis, probably through a mechanism similar to that observed with the OCD risk gene *NCAM2* [77], which is also detected in our population stratification and differentiation analysis. We also found that reduced expression of *Adamtsl3* in neurons led to an increased density of mushroom spines, which are the most mature form of all spines and referred to as memory spines for their pivotal roles in long-term memory [76,78]. Therefore, *Adamtsl3* might play a role in the normal formation and function for dendritic spines. Taken these results together, we speculate that dendritic spine development may play a key role in the pathogenesis of dog and human OCD.

By synthesizing our results and previous studies [22,23,30], we found that our candidate genes and the CCD genes all contain significant SNPs identified in human OCD GWAS, suggesting that dogs may be an ideal model animal for studying OCD. The multiple shared risk genes between CCD and human OCD are also in agreement with the previously reported existence of many genes which have evolved in parallel between dogs and humans, especially for genes within the neurological system [79]. In addition, previous studies also found that dogs and humans shared many evolution-

ary mechanisms of environmental adaptation, such as high altitude [80] and malaria [81]. Therefore, the canine model of OCD may reflect similar convergence in humans. However, caution is needed when applying canine models in the research of human OCD as different genetic variants likely underlie different OCD-related symptoms. For example, previously identified CCD genes in Doberman Pinschers, German shepherd, Jack Russell terrier, and Shetland sheepdog, which showed flank sucking behavior or tail-chasers [22,23,30], were not highlighted in the current study after strict cutoff. Also, only 67 of 119 conserved sites reported in Tang et al. [22] were detected in our samples, and only 4 of the conserved sites showed significant association signals (Table S9 online). Moreover, the previously reported CCD genes (*CDH2, CTNNA2, ATXN1, PGCP*) [22,23,60] were not differentially expressed between non-psychiatric controls and OCD or OC_mix patients (GSE60190) [44]. However, genes (*RBM6, PPP2R2B, PHF2* and *ADAMTSL3*) defined in the current study exhibited significant expression differences between non-psychiatric controls and OCD or OC_mix patients (GSE60190) [44]. Therefore, potential inconsistencies should be considered when selecting models of different dog breeds and behavior types for investigating human OCD.

Our studies suggest that applying stringent behavior tests in dogs might be a suitable method for identifying phenotypic analogous to those of human diseases. Moreover, the reduced genetic background and phenotypic heterogeneity of dogs compared with humans may allow identification of genes underlying biological mechanisms for abnormal behaviors. These studies can ultimately promote the understanding of analogical diseases in humans. The current research focused on SNPs, nevertheless, indels, copy number variations and epigenetic modifications have also been reported related to OCD. Therefore, investigating the roles of these types of genetic variants in CCD is also necessary. Moreover, our behavior tests for circling behavior were designed according to practical experience and the trait of the two breeds, so the universality and validity still should be tested in more different breeds in the future.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Acknowledgments

We thank Lu Wang for helpful comments and suggestions. This work was supported by the National Key Research and Development Program of China (2019YFA0707101), the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (2019YFA0707101), the Innovative Research Team (in Science and Technology) of Yunnan Province (201905E160019), the National Natural Science Foundation of China (81860255, 31701133, 81722019, and 31201712) Yunnan Applied Basic Research Projects for Kunming Medicine University Special Fund (2017FE468(-134)), the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding), and the BIG Data Center at the Beijing Institute of Genomics, Chinese Academy of Sciences. Guo-Dong Wang is supported by the National Youth Talent Support Program. Xue Cao is supported by the Hundred-Talent Program of Kunming Medical University.

## Author contributions

Guo-Dong Wang, Ya-Ping Zhang, Ming Li, and Xue Cao designed the research. Xue Cao, Huijuan Li, and Yanhu Liu analyzed the genome data. Wei-Peng Liu performed neuronal experiment and analyzed the experimental data. Xue Cao, Luguang Cheng, Hong Wu, and Chao Chen performed behavior tests, scored and collected the samples. Ming Li, Guo-Dong Wang, Ya-Ping Zhang, and Xiao Xiao provided advice and discussion during the design and practice of the current study. Xue Cao and Wei-Peng Liu prepared figures and tables of this paper and drafted the manuscript, and all authors contributed to the final version of the paper.

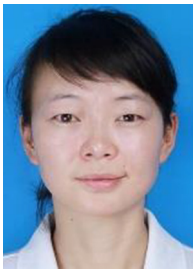
## Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2020.09.021.

## References

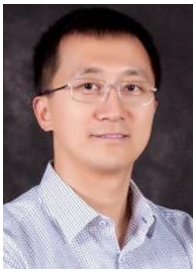
[1] Stein DJ, Costa DLC, Lochner C, et al. Obsessive–compulsive disorder. Nat Rev Dis Primers 2019;5:52.
[2] Robbins TW, Vaghi MM, Banca P. Obsessive-compulsive disorder: puzzles and prospects. Neuron 2019;102:27–47.
[3] Abramowitz JS, Taylor S, McKay D. Obsessive-compulsive disorder. Lancet 2009;374:491–9.
[4] Ruscio AM, Stein DJ, Chiu WT, et al. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010;15:53–63.
[5] Karno M, Golding JM, Sorenson SB, et al. The epidemiology of obsessive-compulsive disorder in five US communities. Arch Gen Psychiatry 1988;45:1094–9.
[6] Brander G, Kuja-Halkola R, Rosenqvist MA, et al. A population-based family clustering study of tic-related obsessive-compulsive disorder. Mol Psychiatry 2019. https://doi.org/10.1038/s41380-41019-40532-z.
[7] Pauls DL, Abramovitch A, Rauch SL, et al. Obsessive–compulsive disorder: an integrative genetic and neurobiological perspective. Nat Rev Neurosci 2014;15:410–24.
[8] van Grootheest DS, Cath DC, Beekman AT, et al. Twin studies on obsessive–compulsive disorder: a review. Twin Res Hum Genet 2005;8:450–8.
[9] Piantadosi SC, Chamberlain BL, Glausier JR, et al. Lower excitatory synaptic gene expression in orbitofrontal cortex and striatum in an initial study of subjects with obsessive compulsive disorder. Mol Psychiatry 2019. https://doi.org/10.1038/s41380-41019-40431-41383.
[10] Ullrich M, Weber M, Post AM, et al. OCD-like behavior is caused by dysfunction of thalamo-amygdala circuits and upregulated TrkB/ERK-MAPK signaling as a result of SPRED2 deficiency. Mol Psychiatry 2018;23:444–58.
[11] Radua J, Mataix-Cols D. Voxel-wise meta-analysis of grey matter changes in obsessive–compulsive disorder. Br J Psychiatry 2009;195:393–402.
[12] Menzies L, Chamberlain SR, Laird AR, et al. Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofronto-striatal model revisited. Neurosci Biobehav Rev 2008;32:525–49.
[13] Alemany-Navarro M, Cruz R, Real E, et al. Looking into the genetic bases of OCD dimensions: a pilot genome-wide association study. Transl Psychiatry 2020;10:151.
[14] Yilmaz Z, Halvorsen M, Bryois J, et al. Examination of the shared genetic basis of anorexia nervosa and obsessive–compulsive disorder. Mol Psychiatry 2020;25:2036–46.
[15] Sidorchuk A, Kuja-Halkola R, Runeson Bo, et al. Genetic and environmental sources of familial coaggregation of obsessive–compulsive disorder and suicidal behavior: a population-based birth cohort and family study. Mol Psychiatry 2019. https://doi.org/10.1038/s41380-019-0417-1.
[16] International Obsessive Compulsive Disorder Foundation Genetics Collaborative, OCD Collaborative Genetics Association Studies. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol Psychiatry, 2018, 23:1181–1188.
[17] Costas J, Carrera N, Alonso P, et al. Exon-focused genome-wide association study of obsessive-compulsive disorder and shared polygenic risk with schizophrenia. Transl Psychiatry 2016;6:e768.
[18] Ogata N, Gillis TE, Liu X, et al. Brain structural abnormalities in Doberman pinschers with canine compulsive disorder. Prog Neuro-Psychopharmacol Biol Psychiatry 2013;45:1–6.
[19] Vermeire S, Audenaert K, De Meester R, et al. Serotonin 2A receptor, serotonin transporter and dopamine transporter alterations in dogs with compulsive behaviour as a promising model for human obsessive-compulsive disorder. Psychiatry Res: Neuroimaging 2012;201:78–87.
[20] Overall KL. Natural animal models of human psychiatric conditions: assessment of mechanism and validity. Prog Neuro-Psychopharmacol Biol Psychiatry 2000;24:727–76.
[21] Gillan CM, Fineberg NA, Robbins TW. A trans-diagnostic perspective on obsessive-compulsive disorder. Psychol Med 2017;47:1528–48.
[22] Tang R, Noh H, Wang D, et al. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. Genome Biol 2014;15:R25.
[23] Dodman NH, Karlsson EK, Moon-Fanelli A, et al. A canine chromosome 7 locus confers compulsive disorder susceptibility. Mol Psychiatry 2010;15:8–10.

[24] Plassais J, Kim J, Davis BW, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun 2019;10:1489.
[25] Noh HJ, Tang R, Flannick J, et al. Integrating evolutionary and regulatory information with a multispecies approach implicates genes and pathways in obsessive-compulsive disorder. Nat Commun 2017;8:774.
[26] Ostrander EA, Wang G-D, Larson G, et al. Dog10k: an international sequencing effort to advance studies of canine domestication, phenotypes and health. Natl Sci Rev, 2019, 6: 810-824.
[27] Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, et al. Demographically-based evaluation of genomic regions under selection in domestic dogs. PLoS Genet, 2016, 12: e1005851.
[28] Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci USA 2016;113:152–7.
[29] Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protocols Bioinform 2016;54:1301–13033.
[30] Moya PR, Dodman NH, Timpano KR, et al. Rare missense neuronal cadherin gene (CDH2) variants in specific obsessive-compulsive disorder and Tourette disorder phenotypes. Eur J Hum Genet 2013;21:850–4.
[31] Parker HG, Kukekova AV, Akey DT, et al. Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. Genome Res 2007;17:1562–71.
[32] Parker HG, Kim LV, Sutter NB, et al. Genetic structure of the purebred domestic dog. Science 2004;304:1160–4.
[33] Penso-Dolfin L, Swofford R, Johnson J, et al. An improved microRNA annotation of the canine genome. PLoS One, 2016, 11: e0153453.
[34] Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 2005;438:803–19.
[35] Li H. Aligning sequence reads, clone sequences and assembly contigs with. BWA-MEM. 2013;arXiv:1303.3997.
[36] Van der Auwera GA, Carneiro MO, Hartl C, et al. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protocols Bioinform 2013;43:111011–33.
[37] Jónsson H, Ginolhac A, Schubert M, et al. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013;29:1682–4.
[38] Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. Genetics 2012;192:1065–93.
[39] Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 1965;19:395–420.
[40] Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–8.
[41] Cockerham BSWC. Estimating F-statistics for analysis of population structure. Evolution 1984;38:1359–70.
[42] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38: e164.
[43] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607–13.
[44] Jaffe AE, Deep-Soboslay A, Tao R, et al. Genetic neuropathology of obsessive psychiatric syndromes. Transl Psychiatry 2014;4:e432.
[45] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013;41:D991–5.
[46] Chekmenev DS, Haid C, Kel AE. P-Match: transcription factor binding site search by combining patterns and weight matrices. Nucleic Acids Res 2005;33: W432–7.
[47] Yang Z, Zhou D, Li H, et al. The genome-wide risk alleles for psychiatric disorders at 3p21.1 show convergent effects on mRNA expression, cognitive function, and mushroom dendritic spine. Mol Psychiatry 2020;25:48–66.
[48] Li H, Zhou D-S, Hu Z, et al. Interactome Analyses implicated CAMK2A in the genetic predisposition and pharmacological mechanism of Bipolar Disorder. J Psychiatr Res 2019;115:165–75.
[49] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60.
[50] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
[51] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
[52] Jiao X, Sherman BT, Huang DW, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 2012;28:1805–6.
[53] Srivastava DP, Woolfrey KM, Penzes P. Analysis of dendritic spine morphology in cultured CNS neurons. Neuroscience 2011;53:2794.
[54] Rodriguez A, Ehlenberger DB, Dickstein DL, et al. Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. PLoS One, 2008;3:e1997.
[55] Molumby MJ, Anderson RM, Newbold DJ, et al. γ-Protocadherins interact with neuroligin-1 and negatively regulate dendritic spine morphogenesis. Cell Reports 2017;18:2702–14.
[56] Smith K, Kopeikina K, Fawcett-Patel J, et al. Psychiatric risk factor ANK3/Ankyrin-G nanodomains regulate the structure and function of glutamatergic synapses. Neuron 2014;84:399–415.
[57] DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.
[58] Clark AG, Hubisz MJ, Bustamante CD, et al. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res 2005;15:1496–502.
[59] Gou X, Wang Z, Li N, et al. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res 2014;24:1308–15.
[60] Cao X, Irwin DM, Liu Y-H, et al. Balancing selection on CDH2 may be related to the behavioral features of the belgian malinois. PLoS One, 2014, 9: e110075.
[61] Delgado-Acevedo C, Estay SF, Radke AK, et al. Behavioral and synaptic alterations relevant to obsessive-compulsive disorder in mice with increased EAAT3 expression. Neuropsychopharmacology 2019;44:1163–73.
[62] Nagarajan N, Jones BW, West PJ. Corticostriatal circuit defects in Hoxb8 mutant mice. Mol Psychiatry 2018;23:1868–77.
[63] Forrest MP, Parnell E, Penzes P. Dendritic structural plasticity and neuropsychiatric disease. Nat Rev Neurosci 2018;19:215–34.
[64] van de Vondervoort I, Poelmans G, Aschrafi A, et al. An integrated molecular landscape implicates the regulation of dendritic spine formation through insulin-related signalling in obsessive–compulsive disorder. J Psychiatry Neurosci 2016;41:280–5.
[65] Wan Y, Ade KK, Caffall Z, et al. Circuit-selective striatal synaptic dysfunction in the Sapap3 knockout mouse model of obsessive-compulsive disorder. Biol Psychiatry 2014;75:623–30.
[66] Wan Y, Feng G, Calakos N. Sapap3 deletion causes mGluR5-dependent silencing of AMPAR synapses. J Neurosci 2011;31:16685–91.
[67] Penzes P, Cahill ME, Jones KA, et al. Dendritic spine pathology in neuropsychiatric disorders. Nat Neurosci 2011;14:285–93.
[68] Maia A, Oliveira J, Lajnef M, et al. Oxidative and nitrosative stress markers in obsessive–compulsive disorder: a systematic review and meta-analysis. Acta Psychiatr Scand 2019;139:420–33.
[69] International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–52
[70] Thompson SL, Welch AC, Ho EV, et al. Btbd3 expression regulates compulsive-like and exploratory behaviors in mice. Transl Psychiatry 2019;9:222.
[71] Ade KK, Wan Y, Hamann HC, et al. Increased metabotropic glutamate receptor 5 signaling underlies obsessive-compulsive disorder-like behavioral and striatal circuit abnormalities in mice. Biol Psychiatry 2016;80:522–33.
[72] Shmelkov SV, Hormigo A, Jing D, et al. Slitrk5 deficiency impairs corticostriatal circuitry and leads to obsessive-compulsive–like behaviors in mice. Nat Med 2010;16:598–602.
[73] Welch JM, Lu J, Rodriguiz RM, et al. Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. Nature 2007;448:894–900.
[74] Arnold PD, Sicard T, Burroughs E, et al. Glutamate transporter gene SLC1A1 associated with obsessive-compulsive disorder. Arch Gen Psychiatry 2006;63:769.
[75] Harris KM, Kater SB. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. Annu Rev Neurosci 1994;17:341–71.
[76] Bourne J, Harris KM. Do thin spines learn to be mushroom spines that remember? Curr Opin Neurobiol 2007;17:381–6.
[77] Sheng L, Leshchyns'ka I, Sytnyk V. Neural cell adhesion molecule 2 (NCAM2)-induced c-Src-dependent propagation of submembrane $Ca^{2+}$ spikes along dendrites inhibits synapse maturation. Cereb Cortex 2019;29:1439–59.
[78] Berry KP, Nedivi E. Spine dynamics: are they all the same? Neuron 2017;96:43–55.
[79] Wang G-D, Zhai W, Yang H-C, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. Nat Commun 2013;4:1860.
[80] Fan R, Liu F, Wu H, et al. A positive correlation between elevated altitude and frequency of mutant alleles at the EPAS1 and HBB Loci in Chinese Indigenous Dogs. J Genet Genomics 2015;42:173–7.
[81] Liu YH, Wang L, Xu T, et al. Whole-genome sequencing of African dogs provides insights into adaptations against tropical parasites. Mol Biol Evol 2018;35:287–98.



Xue Cao is an associate professor of Department of Laboratory Animal Science, Kunming Medical University. Her research interest is the diagnosis and TCM in preventing-treating of Autism, Alzheimer's disease and other mental disorders using bioinformatics, genetics, and network pharmacology methods.

Wei-Peng Liu is a research assistant of Institute of Medical Biology, Chinese Academy of Medical Sciences. He obtained his Ph.D. degree in 2020 from Kunming Institute of Zoology, Chinese Academy of Sciences. His research interest includes the genetic and molecular basis of complex diseases.



Guo-Dong Wang is a professor and a principal investigator of Genetics and Evolution of Animal Behavior Group, Kunming Institute of Zoology, Chinese Academy of Sciences. His research interest includes the population genetics, adaptive evolution, and the genetic basis of complex traits and behaviors in dogs.

Ming Li is a professor and a principal investigator of Translational Genomics Group, Kunming Institute of Zoology, Chinese Academy of Sciences. His research interest is the genetic and molecular basis of complex psychiatric disorders.

Ya-Ping Zhang is a professor and a principal investigator of Molecular Evolution and Genome Diversity, Kunming Institute of Zoology, Chinese Academy of Sciences. His research interest includes molecular phylogenetics, biodiversity, origin of domestic animals and artificial selection, and genome diversity and evolution.