**Supplemental methods and results**

**Methods**

**2.1Study population**

After informed consent was obtained from all participants, whole blood samples were collected before IUD insertion. In the screening stage, the mean age of the cases was 32.25 years at the date of IUD insertion, whereas the mean age of controls was 33.2 years. In validation stage, the mean age of the cases was 30.36 years at the date of IUD insertion, whereas the mean age of controls was 30.70 years at the date of IUD insertion.

**2.2Whole exome sequencing (WES)**

Genomic DNA was extracted from whole blood samples using a DNA extraction kit (Qiagen, Hilden, Germany), according to the manufacturer’s instructions. DNA was stored at −80 °C until further processing. Following the manufacturer's protocol, 200 ng of genomic DNA from each individual was sheared by Biorupter (Diagenode, Belgium) to acquire 150–200 bp fragments. The ends of the DNA fragments were repaired, and Illumina Adapter was added (Fast Library Prep Kit, iGeneTech, Beijing, China). After sequencing, libraries were constructed, and the whole exomes were captured using the AIExome Enrichment Kit V1 (iGeneTech,Beijing, China) and sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, CA) next generation sequencing platform, with 150 paired‐end reads.

WES data were analyzed using the standard pipeline. First, the raw reads were filtered to remove low-quality reads using FastQC. Clean reads were then mapped to the reference genome GRCh37 (hg19) using Burrows-Wheeler alignment (BWA). After removing duplications, single nucleotide variants (SNVs) and indels were identified using the Genome Analysis Toolkit (GATK), and the intersection of the mutation sets of all patients was analyzed.

**2.2.1 Identification of rare potentially-pathogenic variants**

The minor allele frequencies of the most promising candidate pathogenic variants were screened using public databases, including 1000 Genomes Project ([http://www.1000genomes.org](http://www.1000genomes.org/)), Exome Aggregation Consortium (ExAC) ([http://exac.broadinstitute.org](http://exac.broadinstitute.org/)), and the Genome Aggregation Database (gnomAD).

The significant SNVs (P<0.05) were classified according to damaging or probably damaging prediction. Predictive software tools were listed: PROVEAN (http://provean.jcvi.org/index.php), SIFT (http://sift.jcvi.org/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationAssessor, Combined Annotation Dependent Depletion (cadd.gs.washington.edu), LRT, FATHMM, GERP++, fathmm-MKL and MutationTaster (<http://www.mutationtaster.org/>). The mutations were categorized into four levels: high、likeyhigh、medium and low. High (within the exonic region; not within repeat region; All MAF was less than 0.01 by 1000g, ExAC, gnomAD; at least one software prediction result was damaging); Likely high (within the exonic or splicing region; not within repeat region; at least one MAF was less than 0.01 by 1000g, ExAC, gnomAD; at least one software prediction result was damaging); Medium (within the exonic or splicing region; not within repeat region; at least one MAF was less than 0.01 by 1000g, ExAC, gnomAD); Low represents all the other variants.

**2.3 Genotyping**

Genomic DNA was extracted from whole blood using a DNA extraction kit (Qiagen, Hilden, Germany), following the manufacturer’s instructions.

DNA was stored at −80 °C until further processing. Two microliters of DNA from each sample were used for genotyping. All SNVs were genotyped using the Sequenom MassArray platform, which involved polymerase chain reaction (PCR), shrimp alkaline phosphatase (SAP) reaction, single base extension, resin cleanup, and detection by mass spectrometry. The gene mass spectrogram was obtained using matrix-assisted laser desorption ionization time-of-flight mass spectrometry.

**2.4 Statistical analysis**

In the validation stage, we used the SHEsisPlus platform for the analyses of Hardy–Weinberg equilibrium (HWE), linkage disequilibrium (LD), and comparisons of genotype and allele frequencies [1, 2]. Results were expressed as odds ratios (ORs) with 95% confidence intervals (95% CI).

Functional enrichment analysis was performed using the DAVID web tool (<https://david.ncifcrf.gov/>). Protein–protein interaction analyses were conducted using STRING (https://string-db.org/).

The alteration mode of proteins of the significant genes was visualized using PyMOL (http://www.pymol.org)[3]. The prediction of structural change was performed using the online tool Hope (<https://www3.cmbi.umcn.nl/hope/>) [4].

**Results**

**3.4Enrichment and Protein-protein interaction network construction**

The online tool DAVID identified 10 enriched biological processes, 13 enriched molecular function clusters, 3 enriched cellular components, and two enriched KEGG pathways. The top-ranked biological processes with the smallest P-values showed that these 330 genes were strongly related to various aspects of G-protein coupled receptor signaling pathway, detection of chemical stimulus, cardiac muscle fiber development, microtubule-based movement, and substrate adhesion-dependent cell spreading, all of which are related to fibril functions. The top molecular function clusters showed that these differentially expressed genes between cases and controls were strongly related to functions involving microtubule motor activity and ATP pathways, including ATP binding and ATPase activity, and G-protein functions, such as G-protein coupled receptor activity and G-protein coupled serotonin receptor activity. The cell component GO category showed enrichment predominantly for the integral component of membrane, basement membrane, plasma membrane, and Golgi lumen. The two most highly enriched classes of KEGG pathways were the olfactory transduction, with 17 genes identified in our study, and the ABC transporters, with five genes defined (*ABCA10, ABCA6, ABCC6, ABCB5 and ABCA7*) and the Amoebiasis, with six genes defined (*COL27A1, CSF2, LAMC3, PIK3R3, LAMB1 and LAMC1*).

**3.5 Validation and Linkage disequilibrium analysis**

When analyzing the LD patterns of eight SNVs, we identified that *BRDT* rs10783071 with rs10747493, *OR2T35* rs78622116 with *NFASC* rs2802808 and *ZMYND12* rs1034268 showed strong relationships with D’ > 0.8 (Figure S3). Moreover, the gene interaction P values are shown in Table S8. TMEM201 rs4926472 with PBX1 rs2275558, BRDT rs10783071 with NFASC rs2802808 showed strong relationships with P < 0.05.

1. Shi YY, He L: **SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci**. *Cell research* 2005, **15**(2):97-98.

2. Li Z, Zhang Z, He Z, Tang W, Li T, Zeng Z, He L, Shi Y: **A partition-ligation-combination-subdivision EM algorithm for haplotype inference with multiallelic markers: update of the SHEsis (**[**http://analysis.bio-x.cn**](http://analysis.bio-x.cn)**)**. *Cell research* 2009, **19**(4):519-523.

3. Lilkova E, Petkov P, Ilieva N, Litov L: **The PyMOL Molecular Graphics System**. In*: 2015*; 2015.

4. Venselaar H, te Beek TAH, Kuipers RKP, Hekkelman ML, Vriend G: **Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces**. *BMC Bioinformatics* 2010, **11**(1):548.