PNAS
www.pnas.org

Supplementary Information for

Disruption of MeCP2-TCF20 complex underlies distinct neurodevelopmental disorders.

Jian Zhou, Hamdan Hamdan, Hari Krishna Yalamanchili, Kaifang Pang, Amy E. Pohodich, Joanna Lopez, Yingyao Shao, Juan A. Oses-Prieto, Lifang Li, Wonho Kim, Mark A. Durham, Sameer S. Bajikar, Donna J. Palmer, Philip Ng, Michelle L. Thompson, E. Martina Bebin, Amelie J. Müller, Alma Kuechler, Antje Kampmeier, Tobias B. Haack, Alma L Burlingame, Zhandong Liu, Matthew N. Rasband, Huda Y. Zoghbi*

* Correspondence: Huda Y. Zoghbi: address: 1250 Moursund St., Suite 1350, Houston, TX 77030, USA; phone: 713-798-6558, **Email:** hzoghbi@bcm.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Supplementary Materials and Methods

In vitro BioID

Cultured rat primary neurons (≅ $20x10^6$ primary hippocampal neurons for each virus used) were transduced with each Helper Dependent Adenoviruses; HDAd-hEno2-MycBioID2, HDAd-hEno2-MeCP2-MycBioID2, HDAd-hEno2-MeCP2 (R111G)-MycBioID2 and HDAd-hEno2-MeCP2ΔNLS-MycBioID2 at DIV14 with titer concentration 7 MOI for each virus. Biotin was added at DIV14 at a final concentration of 50 μM (Sigma, B4639-1G). Cells were collected at DIV19 and were lysed in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, and 1% NP-40). Biotinylated proteins were pulled down by Streptavidin magnetic sepharose beads (GE Healthcare, 28985799) overnight at 4 °C and then beads washed three times in 50mM Tris buffered saline pH7.5 as described previously (1).

BioID Plasmids construction and adenovirus production

PLPBL-1 shuttle vector was modified by adding human neuron-specific enolase promoter (hENO2) purchased from Addgene (Plasmid #11606 deposited by the Rosalyn Adam Lab), hEno2 promoter was amplified by PCR and inserted into PLPBL-1 and named PLPBL-1-hENO2-MCS, BioID2 was amplified by PCR from pcDNA3.1 Myc-BioID2-MCS purchased from Addgene (Plasmid #74223 deposited by the Kyle Roux Lab) to generate shuttle template named PLPBL-1-hENO2-MCS-BioID2. PLPBL-1-hEno2-MeCP2-MycBioID2, PLPBL-1-hEno2-MeCP2 (R111G)-MycBioID2 and PLPBL-1-hEno2-MeCP2ΔNLS-MycBioID2 were constructed by amplifying PCR fragments from MeCP2, MeCP2(R111G) and MeCP2ΔNLS plasmids. PLPBL-1-hEno2-MycBioID2, PLPBL-1-hEno2-MeCP2-MycBioID2, PLPBL-1-hEno2-MeCP2 (R111G)-MycBioID2 and PLPBL-1-hEno2-MeCP2ΔNLS-MycBioID2 were inserted in the viral vector pdelta28E4LacZ-2 to generate helper dependent viruses HDAd-hEno2-MycBioID2, HDAd-hEno2-MeCP2-MycBioID2, HDAd-hEno2-MeCP2 (R111G)-MycBioID2 and HDAd-hEno2-MeCP2ΔNLS-MycBioID2 respectively. Viruses were produced and characterized as described (2). Production of helper-dependent adenoviral vectors were previously described (3).

Mass spectrometry

Sample-incubated streptavidin magnetic sepharose beads were resuspended in 5 mM DTT in 100mM NH_4HCO_3 and incubated for 30 min at room temperature. After this, iodoacetamide was added to a final concentration of 7.5 mM and samples incubated for 30 additional minutes. 0.5ug of sequencing grade trypsin (Promega) was added to each sample and incubated at 37 °C overnight. Supernatants of the beads were recovered, and beads digested again using 0.5ug trypsin in 100mM NH_4HCO_3 for 2 hrs. Peptides from both consecutive digestions were recovered by solid phase extraction using C18 ZipTips (Millipore), and resuspended in 0.1% formic acid for analysis by LC-MS/MS. Peptides resulting from trypsinization were analyzed either on a QExactive Plus (Thermo Scientific), connected to a NanoAcquity™ Ultra Performance UPLC system (Waters). A 15-cm EasySpray C18 column (Thermo Scientific) was used to resolve peptides (90-min gradient with 0.1% formic acid in water as mobile phase A and 0.1% formic acid in acetonitrile as mobile phase B). MS was operated in data-dependent mode to automatically switch between MS and MS/MS. The top 10 precursor ions with a charge state of 2+ or higher were fragmented by HCD. Peak lists were generated using PAVA in-house software (4). All generated peak lists were searched against the rat subset of the UniProt database (UniprotKB 2017.11.01) using Protein Prospector (5). The database search was performed with the following parameters: a mass tolerance of 20 ppm for precursor masses; 30 ppm for MS/MS, cysteine carbamidomethylation as a fixed modification and acetylation of the N terminus of the protein, pyroglutamate formation from N terminal glutamine, and oxidation of methionine as variable modifications. All spectra identified as matches to peptides of a given protein were reported, and the number of spectra (Peptide Spectral Matches, PSMs) used for label free quantitation of protein abundance in the samples.

DNA constructs

Full-length cDNA clones for human *MECP2*, *TCF20*, *PHF14* were cloned into pcDNA3.1(+) by Gibson reaction using NEBuilder® HiFi DNA Assembly kit (New England Biolabs, E2621). The *MECP2, TCF20, PHF14* deletion mutants were constructed from the full-length cDNA and subcloned into pcDNA3.1(+) by Gibson reaction using NEBuilder® HiFi DNA Assembly kit. The point mutations of *MECP2* and *PHF14* were introduced using QuikChange XL Site-Directed Mutagenesis Kit (Aglient, 200517).

pAAV-YFP/miRE construct generation for neuron infection

pAAV-YFP/mire-shRNA were constructed as previously described (6). Briefly, An AAV8 vector containing both YFP and a miRE cassette-containing shRNA under the control of the chicken β actin promoter was generated using Gibson cloning (New England Biolabs, E5520S). Oligonucleotides containing the shRNA target sequences are as follows: (*Tcf20*-sh1: 5'-TGCTGTTGACAGTGAGCGCTCCAATGACAATGAAATTTAATAGTGAAGCCACAGATGTATTAAATTTCATTGTCATTGGAATGCCTACTGCCTCGGA-3', *Tcf20*-sh2: 5'-TGCTGTTGACAGTGAGCGAAAAGACTTTGTTTAAATGAAATAGTGAAGCCACAGATGTATTTCATTTAAACAAAGTCTTTCTGCCTACTGCCTCGGA-3')

Cell and primary neuron cultures

HEK293T cells and NIH-3T3 cells (obtained and certified from ATCC) were cultured in DMEM (Thermo Fisher Scientific, MT10013CV) containing 10% FBS and Antibiotic-Antimycotic (Gibco, 15240096). Mouse hippocampal neurons were prepared from postnatal day 0 male mice with indicated genotypes and plated on poly-D-lysine/mouse laminin-coated coverslips (Corning, 354087) in 24-well plates (2×10^5 cells per well) in Neurobasal medium supplemented with GlutaMAX (Thermo Fisher Scientific, 35050-061), B-27 (Thermo Fisher Scientific, A3582801), and antibiotics (Penicillin/Streptomycin).

Transfection and virus infection for cultured cells and primary neurons

Plasmids were transfected using TransIT-293 (Mirus, MIR 2704) (for HEK293T cells) or lipofectamine 3000 (Invitrogen, L3000001) (for NIH-3T3 cells) and left to express for 48-72 hrs.

HEK293T cells are infected with GIPZ Lentiviral shRNA (Horizon Discovery) to knock down TCF20 and PHF14, media was changed to puromycin-containing media (1 μg/ml) to select for infected cells. After 9-10 days of selection, cells were used for immunoprecipitation or western experiments.

At days in vitro (DIV) 1, cultured mouse hippocampal neurons were infected with Adeno-associated viruses (AAV) containing shRNA of the target genes and control. At DIV 11-13, neurons were used for immunostaining and RT-qPCR analyses.

Co-Immunoprecipitation

Protein lysates were extract from cortical tissue or HEK293T cells in NP-40 buffer (10 mM HEPES, pH 7.9, 3 mM MgCl2, 5 mM KCl, 140 mM NaCl, 0.1mM EDTA, 0.5% NP-40) supplemented with phosphatase and protease inhibitors (GenDEPOT, P3200 and P3100) and nuclease (Thermo Fisher, 88700). Protein lysates were incubated on a rotator at 4 °C for 1 hr and centrifuged (30 min at 15,000 rpm, 4 °C) to remove cell debris. Lysates were mixed with GFP-Trap Dynabeads (ChromoTek), anti-Flag Dynabeads (Sigma, M8823), or protein G Dynabeads (Invitrogen, 10003D) conjugated with anti-MeCP2 (Sigma, M6818), anti-TCF20 (Invitrogen, PA5-57816), anti-PHF14 (Sigma, HPA000538), anti-RAI1 (Bethyl, A302-317A), and Rabbit IgG (Millipore, 12-370) at 4 °C for 1 hr. Beads were then washed 4 times with NP-40 buffer lysis buffer (without nuclease) before being eluted in NuPAGE LDS Sample buffer (Invitrogen, NP0007) with NuPAGE Sample Reducing Agent (Invitrogen, NP0009) at 95 °C for 10 minutes. Input and IP elutes were analyzed by SDS-PAGE and western blot.

Western blot

Protein lysates were extract from tissue or cells in NP-40 buffer. NuPAGE LDS Sample buffer (Invitrogen, NP0007) with NuPAGE Sample Reducing Agent (Invitrogen, NP0009) was added to protein lysate followed by a 10 min incubation at 95 °C. Samples were spun down and run on Bis-Tris gels (Invitrogen), transferred to a nitrocellulose membrane (BIO-RAD, 1620145) and blocked for 1 hr with 5% non-fat milk in TBST (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Tween-20). Primary antibodies anti-MeCP2 (1:2000, CST, 3456), anti-TCF20 (1:100, Invitrogen, PA5-57816), anti-PHF14 (1:500, Sigma, HPA000538), anti-RAI1 (1:500, Bethyl, A302-317A), anti-HMG20A (1:2000, Proteintech, 12085-2-AP), anti-PCNA (1:10000, Proteintech, 10205-2-AP), anti-Flag (1:5000, Millipore, F7425), anti-GFP (1:5000, Abcom, ab290), anti-HDAC2 (1:1000, Abcom, ab7029), anti-TBL1 (1:1000, Abcom, ab24548), anti-GAPDH (1:10000, Advanced Immunochemical, 2-RGM2) were diluted in 3% BSA in TBST and incubated at 4°C overnight. The membranes were washed for three times in TBST, and incubated with secondary antibody in 5% milk in TBST before being developed and imaged using a GE ImageQuant LAS 4000.

Immunostaining of mouse brain slices, mouse hippocampal neurons, 3T3 cells

The brains were fixed by transcardial perfusion of 4% paraformaldehyde (PFA) dissolved in

PBS. Brains were dissected out, kept in 4% PFA for overnight, cryoprotected in 25% sucrose solution, and frozen in optimal cutting temperature medium (O.C.T.). Sagittal brain sections were obtained using a Leica CM3050S cryostat at 50 μm thickness. The slices were incubated in a PBS-buffered blocking solution containing 3% normal goat serum and 0.1% Triton X-100 for 1 hr, then incubated in a primary antibody solution containing anti-TCF20 (1:50, Invitrogen, PA5-57816), anti-NeuN (1:250, Millipore, MAB377) and anti-MeCP2 (1:250, Sigma, M6818) overnight at 4°C. Slices were then washed three times with PBS and were incubated with secondary antibodies conjugated with either Alexa Fluor 488, Alexa Fluor 555, or Alexa Fluor 633 (1:250, Invitrogen) and DAPI (2.5 μg/ml) for 2 hrs at room temperature. After three washes with PBS, the slices were mounted onto Superfrost Plus microscope slides (Fisher Scientific, 12-550-15) with ProLong Gold Antifade mounting medium (Thermo Fisher, P10144).

At DIV 11–13, mouse primary hippocampal neurons were fixed with 4% formaldehyde and 4% sucrose in PBS on ice for 30 min and wash with PBS for 3 times, then permeabilized/blocked with 3% goat serum and 0.1% Triton X-100 in PBS (blocking buffer) for 1 hr at room temperature. Samples were then incubated in a primary antibody solution containing anti-PSD95 (1:100, CST, 36233), anti-Synapsin I (1:200, CST, 5297) and an anti-MAP2 (1:500, Abcom, ab5392) at 4 °C overnight, then secondary Alexa-conjugated antibodies (1:250, Invitrogen) at room temperature for 2 hrs. Z-stack images were acquired by LSM880 (Zeiss) confocal microscope. Synapse density was measured from secondary dendrites.

3T3 cells were fixed with 4% formaldehyde in PBS for 20min at room temperature and wash with PBS for 3 times, then permeabilized/blocked with 3% goat serum and 0.1% Triton X-100 in PBS (blocking buffer) for 1 hr at room temperature. Samples were then incubated in a primary antibody solution containing anti-Flag (1:200, Millipore, F7425) anti-HA (1:750, BioLegend, 901514) at 4 °C overnight, then secondary Alexa-conjugated antibodies (1:250, Invitrogen) and DAPI (2.5 μg/ml) for 2 hrs at room temperature. Z-stack images were acquired by LSM880 (Zeiss) confocal microscope with 64x magnificence. In order to compare PHF14 and TCF20 enrichment to heterochromatic foci, 3T3 cells with overexpression of GFP, WT and mutant MeCP2-GFP were imaged and analyzed using the same parameters. Quantitative analysis of PHF14 and TCF20 enrichment to heterochromatin foci was performed using ImageJ as following: For each nucleus, DAPI image was used to determine the presence of heterochromatin foci. Mean gray values within foci area and outside of foci of HA-tagged PHF14, Flag-tagged

TCF20, and GFP-tagged MeCP2 area were measured by ImageJ. The heterochromatin foci enrichment fold was calculated as the ratio of mean gray value within foci area to mean gray value outside of foci area.

RNA isolation and RT-qPCR

The cortex, hippocampus, or prefrontal cortex of *Tcf20*$^{+/-}$, *Phf14*$^{+/-}$ and control mice were dissected on ice, flash frozen in liquid nitrogen, and stored at -80 °C until further use. RNA was harvested using the RNeasy Mini Kit (Qiagen, 74104). 500 ng of purified RNA was reverse transcribed using M-MLV Reverse Transcriptase (ThermoFisher Scientific, 2802501) and random hexameters primers (ThermoFisher Scientific, 48190011). Newly synthesized cDNA was analyzed by qPCR using PowerUp SYBR Green Master Mix (ThermoFisher Scientific, A25742) on a CFX96 Real-Time PCR Detection System (Bio-Rad, 1855195).

The following primers were used: mTcf20: Forward 5′-CTTCAGAAGGTGGGCCTGAGCTGG-3′; Reverse 5′-GCCTCCTGGCAGTGGGAGCA-3′; mPhf14: Forward 5′-GCTATCAGAGGCAGCAGCAGAAGAGG-3′; Reverse 5′-TGATCCTTGCCTGTGCTTCAGGGG-3′; mBdnf: Forward 5′-GGCTGGTGCAGAAAAGCAACAA-3′; Reverse 5′-TCGCCAGGTAAGAAACCCTTCG-3′; mGapdh: Forward 5′-GGCATTGCTCTCAATGACAA -3′; Reverse 5′-CCCTGTTGCTGTAGCCGTAT -3′; hHMG20A: Forward 5′-GCCTCCTGAGGAAAAACAGCGCTAC-3′; Reverse 5′-CCTTTCTGACGGTCCTGGGTTTTCC-3′; hPHF14: Forward 5′-GTCTGGGAGATAATAGTGAGGACGC-3′; Reverse 5′-GAGAAACACCACATTTACAGGCATCAC-3′; hGAPDH: Forward 5′-GTCGGAGTCAACGGATTTGGTCG-3′; Reverse 5′-CTTCCCGTTCTCAGCCTTGACG-3′;

Chromatin Immunoprecipitation

Cortical tissues were dissected on ice, homogenized and crosslinked in lysis buffer (320 mM sucrose, 5 mM $CaCl_2$, 3 mM Mg acetate, 0.1 mM EDTA, 10mM Tris-HCl pH8, 0.1% Triton X-100, 1 mM DTT) with 1% formaldehyde for 10 min. After quenching and washing in 125 mM glycine, the lysates were centrifuged at 4000 rpm for 5 min, then resuspend the pellet in lysis buffer with protease inhibitors (GenDEPOT, P3100). The resuspended lysates were then mixed

with an equal volume of 50% Optiprep solution (82.6% OptiPrep™ Density Gradient Medium [Sigma, D1556], 5mM $CaCl_2$, 3mM Mg acetate, 10mM Tris-HCl pH8, 1 mM DTT, protease inhibitors), and then gently layered (now in 25% Optiprep solution) onto the top of the 29% Optiprep solution layer. Nuclei were collected by centrifuging at 4 °C for 30 min at 10,000 x g, followed by 3 washes with sonication buffer (5 mM EDTA, 20mM Tris-HCl pH8, 50 mM NaCl, 0.1% SDS, protease inhibitors). Samples were sonicated (Covaris M220) with the following settings: duty cycle 20%, intensity peak incident power 75 Watts, 200 cycles per burst, 7°C bath temperature. Soluble chromatin was diluted 1:5 in IP buffer (5 mM EDTA, 20mM Tris-HCl pH8, 50 mM NaCl, 0.1mM PMSF, 0.1mM benzamidine, protease inhibitors). An aliquot of chromatin was stored as input and the remaining supernatant was incubated overnight at 4°C with 10 µl of anti-TCF20 (Invitrogen PA5-57816). 50 µl of Protein A Dynabeads was then added to each reaction and allowed to bind for 2 hrs at 4°C. Bound complexes were then washed in IP buffer, Low-Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris-HCl, pH 8, 0.1 mM PMSF, protease inhibitors), High-Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, pH 8, 500 mM NaCl, 20 mM Tris-HCl, pH 8, 0.1 mM PMSF, protease inhibitors), LiCl Wash Buffer (10 mM Tris-HCl pH 8, 250 mM LiCl, 1mM EDTA, 1% NP-40, 1% sodium deoxycholate, 0.1 mM PMSF, protease inhibitors), and 1xTE. Complexes were then eluted twice in 1% SDS/100 mM NaHCO3. Precipitated chromatin and input samples were then reverse crosslinked and treated with proteinase K (GoldBio, P-480-2). DNA was recovered using a QIAGEN PCR Purification Kit (QIAGEN, 28104).

For qPCR analysis, the input and immunoprecipitated chromatin were diluted and stored at −20°C until needed. PowerUp SYBR Green Master Mix (ThermoFisher Scientific, A25742) on a CFX96 Real-Time PCR Detection System (Bio-Rad, 1855195). The following primers were used: BDNF I: Forward 5′-CGCCCTTTGTTCAGGCCAGG-3′; Reverse 5′-ATAACGCCCCTCACCCGCATC-3′; BDNF IV: Forward 5′-GGTCTTGGCGCCTTCTGTCCTC-3′; Reverse 5′-GTTCCGCAGACCCTTTCAGGC-3′; Non-specific Intergenic control (NC): Forward 5′-GTGTACCACAAGTTCTGAAGCCTTGGC-3′; Reverse 5′-GCCCATTGGTAACAGTCACACCCAC-3′.

Behavioral assays

All behavioral studies were performed on mice at adult stage (8 weeks of age or older). All the tests were performed during the light period. Mice were habituated to the test room for 30

min before each test. At least one day was given between assays for the mice to recover. All the tests were performed as previously described (7–9) with few modifications. Behavioral analyses were performed blind to genotypes. The Baylor College of Medicine Institutional Animal Care and Use Committee approved all research and animal-care procedures.

Open field test

Animals were habituated in the test room (150 lux, 60 dB white noise) for 30 min. After habituation, animals were placed in the center of a 40 × 40 × 30-cm chamber equipped with photobeams (Accuscan, Columbus, OH, USA) to record activity during a 30 min test period. General locomotor activity was automatically analyzed using AccuScan Fusion software (Omnitech) by counting the number of times mice break the laser beams (activity counts). In addition, rearing activity, the time spent in the center of the arena, entries to the center and distance travelled were analyzed.

Elevated plus maze

Animals were habituated to the test room (700 lux, 60 dB white noise) for 30 min. After habituation, animals were placed in the center of a maze consisting of two arms (each arm 25 × 7.5 cm) enclosed by approximately 15 cm high walls, and two open arms (each arm 25 × 7.5 cm, with a raised 0.5 cm lip around the edges) elevated 50 cm above ground level. The amount of time animals spent in the open arms, the number of arm entries and the total distance traveled were recorded for 10 min using a camera and detection software (ANY-maze video tracking system, Stoelting Co., Wood Dale, IL, USA).

Light–dark box test

Animals were habituated to the test room (700 lux, 60 dB white noise) for 30 min. After habituation, animals were placed in the light side of a chamber separated by dark side, which was a covered black plastic chamber containing a 10.5 × 5 cm opening. The amount of time animals spent in the light side and the number of transitions were recorded for 10 min using the Fusion software (AccuScan Instrument).

Accelerating rotarod test

Animals were habituated to the test room (700 lux, 60 dB white noise) for 30 min. After habituation, motor coordination was measured using an accelerating rotarod apparatus (Ugo Basile). Mice were tested for four consecutive days, four trials each, with an interval of 30 min between trials to rest. Each trial lasted for a maximum of 5 min (10 min for *MECP2*^{Tg1}; *Tcf20*$^{+/-}$ mice and control littermates); and the rod accelerated from 4 to 40 rpm in the first 5 min. The time that it took for each mouse to fall from the rod (latency to fall) was recorded.

Three-chamber test

Age- and gender-matched novel partner mice were placed within a small wire cage to habituate them to their test environment. These animals were placed randomly in either the left or right chambers of the three-chamber apparatus, which consists of a clear Plexiglas box (24.75 × 16.75 × 8.75) with removable partitions that separate the box into three chambers, for 1 hr per day for at least 2 consecutive days prior to the actual test day. On the day of testing, test animals were first habituated to the test room for 30 min (700 lux, 60 dB). After habituation, mice were placed in the central chamber and allowed to explore the three chambers for 10 min. Next, a novel partner mouse was placed into a wire cup in either the left or the right chamber. An inanimate object was placed as control in the wire cup of the opposite chamber. The location of the novel mouse was randomized between left and right chambers across subjects to control for side preference. The mouse tested was allowed to explore again for an additional 5 min.

Fear conditioning

Animals were habituated to the test room (150 lux, 60 dB white noise) for 30 min. After habituation, animals were trained in a fear-conditioning chamber with a grid floor that can deliver an electric shock (Med Associates Inc.). Each mouse was initially placed in the chamber and left undisturbed for 2 min, after then a tone (30 s, 5 kHz, 80 dB) coincided with a scrambled foot shock (2 s, 0.7 mA). The tone/foot shock stimuli were repeated after 2 min. The mouse was then returned to its home cage. The context test was assessed in 24 hrs. The mice were placed in exactly the same environment and observed for 5 min. The cued fear test was assessed 2 hrs after the context test. The mice were placed in a novel environment for 3 min, followed by a 3-min tone. Mouse behavior was recorded and scored automatically by Freeze Frame (Actimetrics). The percentage of time spent freezing during the tests serves as an index of fear memory.

Barnes Maze

The Barnes Maze protocol comprised two phases: training from Days 1 to 4, and the probe test on Day 5. For each day, animals were habituated to the test room (700 lux, 60 dB white noise) for 30 min. In the training phase, mice were allowed to freely explore the circular platform with 18 circular holes evenly spaced around the periphery for 3 min until they entered the escape box. Once the experimenter confirmed that the mouse had successfully entered the escape box, the recording was turned off. Each mouse was subjected to two training trials per day. Twenty-four hours after the last day of training (Day 5), mice were subjected to the probe test. In the probe test, mice explored the arena without the escape box for 3 min. This trial was done once per mouse. The maze was thoroughly cleaned with 70% ethanol and dried after every trial.

Tube test

The tube test was performed as previously described (10) with modifications. We used transparent plexiglass tubes with 30.5 cm length/3 cm inner diameter for male mice, and 30.5 cm length/2.5 cm inner diameter for female mice. Mice were habituated to walk through the tube two sessions per day for two consecutive days before testing. On the day of testing, mice with different genotypes (same age and gender, body weight ±15%) were pushed to the middle of the tube and released simultaneously. The mouse that completely retreated first from the tube was defined as the loser, and the other as the winner. Each mouse was tested against up to four different mice of a different genotype. The tubes were cleaned with 75% ethanol between trials. A two-tailed binomial test was used to determine the significance of test score between mice.

RNA Sequencing

Total RNA was isolated from dissected tissue using the RNeasy Mini Kit (Qiagen, 74104). A double-stranded DNA library was created using 250ng of total RNA. First, cDNA was created using the fragmented 3' poly (A) selected portion of total RNA and random primers. Libraries were created from the cDNA by first blunt ending the fragments, attaching an adenosine to the 3' end and finally ligating unique adapters to the ends. The ligated products were then amplified using 15 cycles of PCR. The resulting libraries were quantitated using the NanoDrop spectrophotometer and fragment size assessed with the Agilent Bioanalyzer. A qPCR

quantitation was performed on the libraries to determine the concentration of adapter ligated fragments using Applied Biosystems ViiA7 Real-Time PCR System and a KAPA Library Quantification Kit (Roche, KK4824). Equimolarly pooled library is loaded onto two lanes of the NovaSeq S1 v1.0 flowcell (Illumina, 20012864) and amplified paired-end 100 bp cycle using the Illumina NovaSeq 6000 sequencing instrument. An average of 76 million read pairs per sample was sequenced.

RNA-seq data analysis

For each sample lane-wise raw reads are pooled by appending respective fastq files. Hypothalamus MeCP2 knockout and NS-DADm RNA-Seq data are downloaded from GEO database (GSE66870 and GSE92450 respectively). Sequencing quality and adapter contamination are assessed using FastQC v0.10.1. First 10 base pairs of each read except for NS-DADm Hypothalamus data are trimmed using fastp v0.20.0 (11).

Reads are aligned to the mouse reference genome GRCm38v23 using STAR v2.7.2c (12). Raw FASTA sequence and annotations are downloaded from GENCODE portal (https://www.gencodegenes.org/mouse/). Genome is indexed using STAR by setting --runMode to genomeGenerat.

Gene expression values from each sample were quantified as the number of reads mapped (to a specific gene) by setting --quantMode to GeneCounts in STAR. Sample clustering is assessed by principle component analysis (PCA) on normalized read counts (13). Raw read counts are normalized and genes with an average read count >50 across all the samples are considered for differential gene expression analysis using DESeq2 (14). An adjusted *p* value threshold of 0.05 was used to call DEGs. Both Frontal cortex and Hypothalamus datasets are processed similarly. Litter is included a covariate for *Tcf20*$^{+/-}$ *and Phf14*$^{+/-}$ datasets. Genes with no published symbols (LOC, Gm and RIKEN) are excluded in the overlap analysis. Overlap significance is computed using http://nemates.org/MA/progs/overlap_stats.html.

Gene Ontology analysis

Functional enrichment analysis of the dysregulated genes in Figure 5A was performed on gene lists of interest using WEB-based GEne SeT AnaLysis Toolkit (15) with an FDR cutoff of 0.05. Minimum number of genes per category was set to 5.

Co-expression and network analysis

The scRNA-seq data of the 565 cell types across nine regions in the adult mouse brain (16) were downloaded from the DropViz website (http://dropviz.org/). To normalize the scRNA-seq data, the aggregate UMI count of each gene in each cell type was divided by the total aggregate UMI count of the cell type and multiplied by 100K. To explore brain cell type-specific expression pattern of *Mecp2*, *Tcf20*, and *Phf14*, the scRNA-seq data were grouped into 6 major cell types, with the cell types from the same major cell type being grouped together. Using the normalized gene expression of the 565 cell types, the Spearman's correlation coefficients between *Mecp2* and all the genes expressed in at least one cell type were calculated and then ranked.

Gene co-expression networks of the overlapping up-regulated genes and the overlapping down-regulated genes between $Tcf20^{+/-}$ and $Mecp2^{-/y}$ mouse models were constructed as previously described (17). To construct the co-expression networks, we used the union of the overlapping up-regulated genes and the overlapping down-regulated genes which are also expressed in at least one cell type as background genes. We first computed the pairwise Spearman's rank correlation coefficients between background genes and sorted all the pairwise Spearman's correlation coefficients in descending order. Then, we determined the correlation threshold for the top 10% highest pairwise Spearman's correlation coefficients and used the correlation threshold to construct gene co-expression networks of the overlapping up-regulated/down-regulated genes. The genes which are in the gene co-expression network of the overlapping up-regulated/down-regulated genes were used as input for Gene Ontology and pathway enrichment analyses. Gene Ontology and pathway enrichment analyses were performed using Enrichr (18) (http://amp.pharm.mssm.edu/Enrichr). The co-expression networks were visualized using Cytoscape (19).

Genome/Exome sequencing and Sanger confirmation

Patient and parent peripheral blood samples were collected in EDTA tubes and were sent to the HudsonAlpha Clinical Services Laboratory (CSL; CAP/CLIA certified lab) for DNA extraction (QIAsymphony) and storage. Sequencing libraries were constructed from patient genomic DNA using the CSL's custom genome library preparation protocol. DNA library

fragments were sequenced from both ends (paired) with a read length of 150 base pairs using NovaSeq 6000, with targeted mean coverage depth of 30X with >80% of bases covered at 20X. Sequence reads were aligned to GRCh38 using DRAGEN (20). SNVs/indels were called using DRAGEN and GATK (21). Identified SNVs and indels were annotated, filtered, and visualized using an in-house custom-built software package called Codicem. Variants that survived filtration were manually curated and classified in accordance with ACMG guidelines (22). GS testing was conducted in a CAP/CLIA-certified laboratory, while variant analysis/interpretation were conducted as part of a research protocol. All variants deemed to be returnable to patients/families were validated by Sanger sequencing in a CAP/CLIA clinical laboratory (HudsonAlpha CSL) to confirm variant presence and generate a clinical report that contained clinical interpretation.

In patient GER01, exome sequencing and variant filtering was performed as described previously (23). In brief, coding genomic regions were enriched using a SureSelect XT Human All Exon Kit V7 (Agilent Technologies) for subsequent sequencing as 2x100 bp paired end reads on a NovaSeq6000 system (Illumina). Generated sequences were analyzed using the megSAP pipeline (https://github.com/imgag/megSAP). Clinical variant prioritization included different filtering steps (e.g., MAF<0.1 % in 1000g, ExAC or gnomAD, in-house database) and was conducted independently by two analysts according to an in-house standard operating procedure. Prioritized variants were validated by Sanger sequencing.

Quantification and statistical analysis

Statistical parameters including the exact value of n and measures (mean ± SEM) and statistical significance are reported in the Figure Legends. Data were judged to be statistically significant using Prism 8 when $p < 0.05$ by two-tailed Student's T-Test, two-tailed binomial test, one-way or two-way ANOVA with post-hoc Tukey's tests, where appropriate.

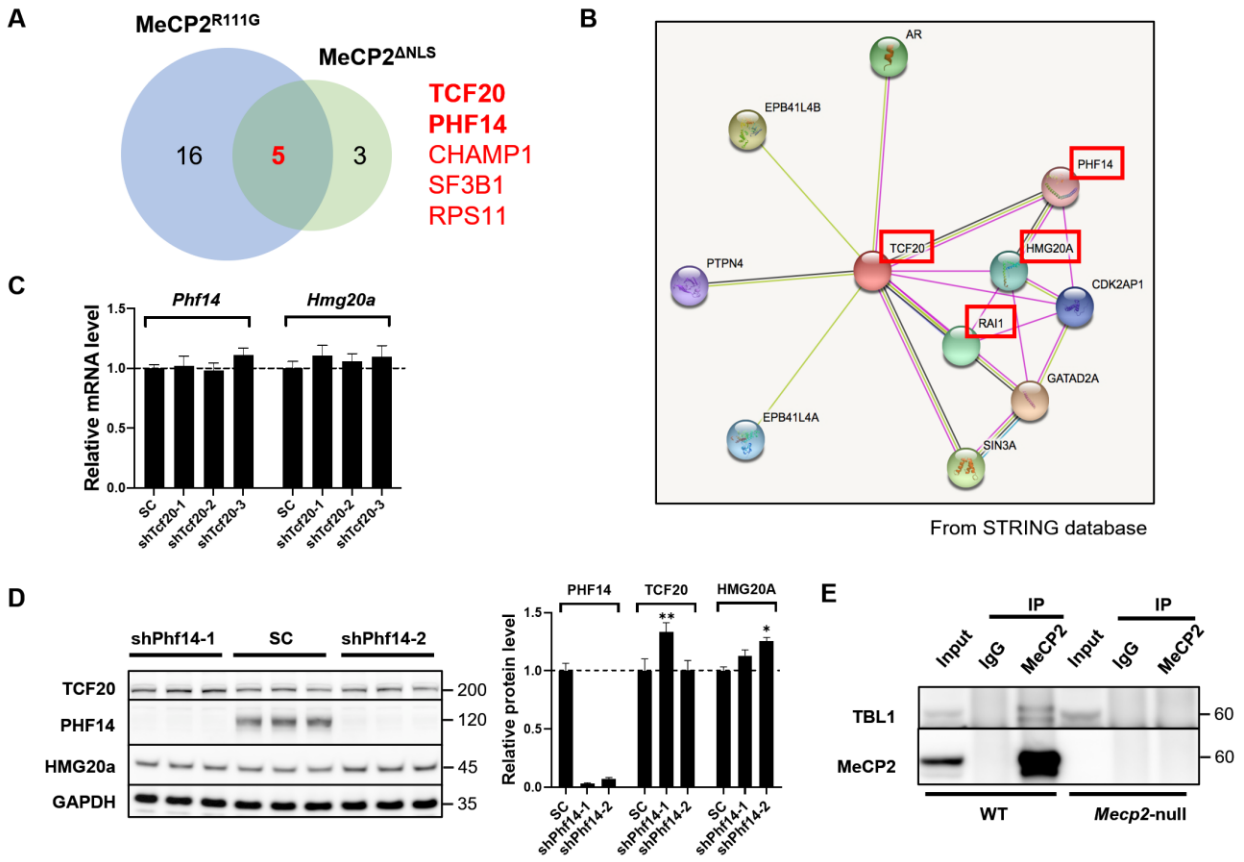

Figure S1. BioID-MS in rat primary hippocampal neurons identified the TCF20 complex as a novel MeCP2 interactor. (A) Diagram showing shared MeCP2 interactor candidates identified by BioID-MS in rat primary neurons using MeCP2^{R111G} and MeCP2$^{\Delta NLS}$ as negative controls. (**B**) Protein-protein interaction network enrichment analysis showing TCF20, PHF14, HMG20A and RAI1 potentially interact with each other in a protein complex. (**C**) Quantification of *Phf14* and *Hmg20a* mRNA levels by RT-qPCR upon knockdown of *Tcf20* in HEK293T cells. (**D**) Representative immunoblot (left) and quantification (right) of TCF20 and HMG20A protein levels following *PHF14* knockdown in HEK293T cells (n=3/group, two-way *ANOVA* with post-hoc Tukey's tests). **(E)** Representative immunoblots of TBL1 protein levels following IP of MeCP2 from WT and *Mecp2* null cortical lysates. *p < 0.05, **p < 0.01, Data are mean ± SEM.

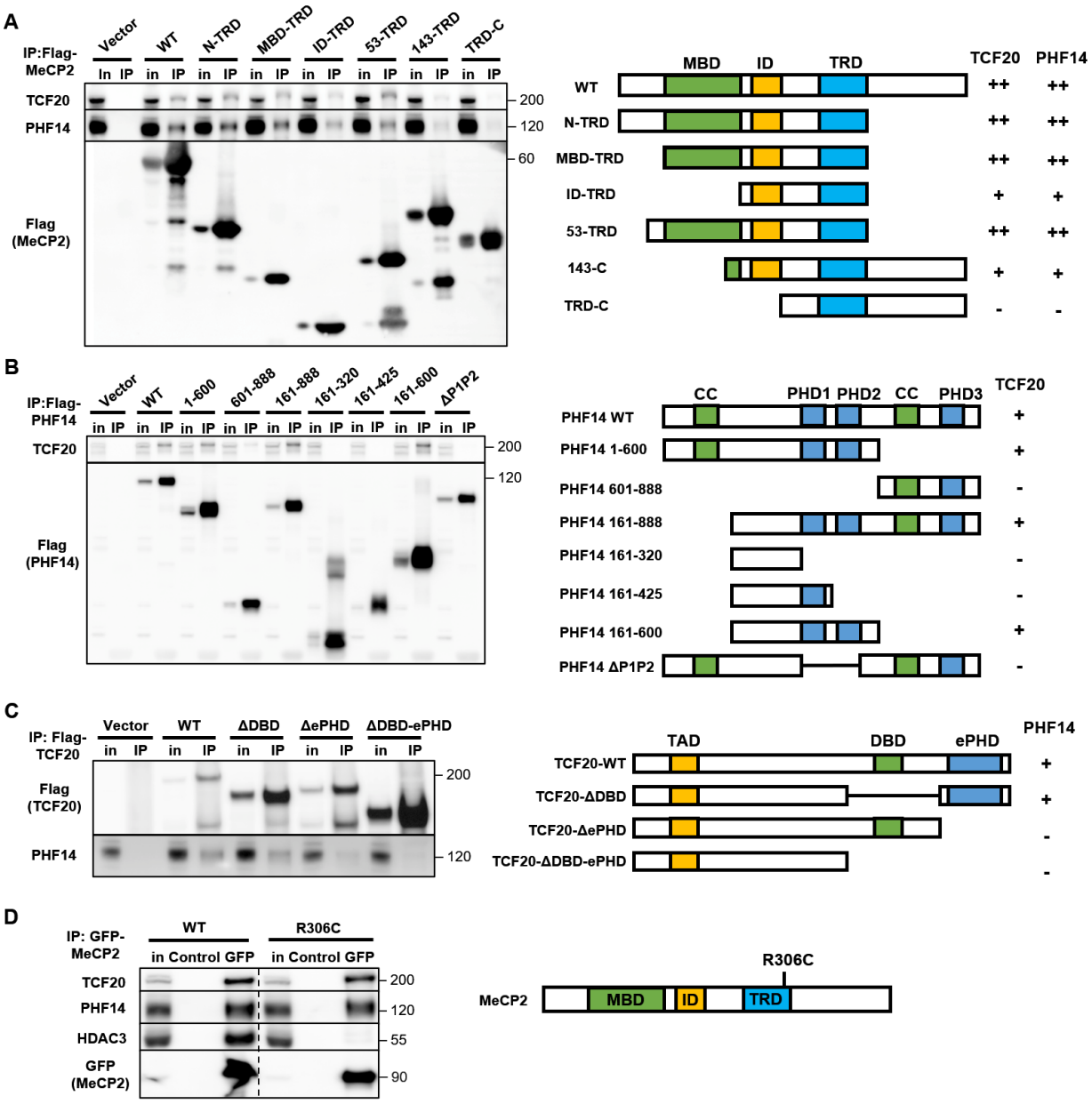

Figure S2. MeCP2, PHF14, and TCF20 interact via multiple binding domains. (A) Representative immunoblot (left) and summary of co-IP results (right) of TCF20 and PHF14 protein levels following IP of Flag-tagged WT and truncated MeCP2 in HEK293T cells. TCF20 and PHF14 interactions are denoted by strong (++), weak (+), or absent (-). **(B)** Representative immunoblot (left) and summary of co-IP results (right) of TCF20 protein levels following IP of Flag-tagged WT and truncated PHF14 in HEK293T cells. The +/- denotes the presence (+) or absence (-) of TCF20 interaction. **(C)** Representative immunoblot of PHF14 (left) and summary of co-IP results (right) following IP of Flag-tagged WT and truncated TCF20 in HEK293T cells.

The +/- denotes the presence (+) or absence (-) of PHF14 interaction. **(D)** Representative immunoblot (Left) and schematic (Right) of TCF20, PHF14, and HDAC3 protein levels following IP of Flag tagged WT MeCP2 or MeCP2 with R306C variant in HEK293T cells.

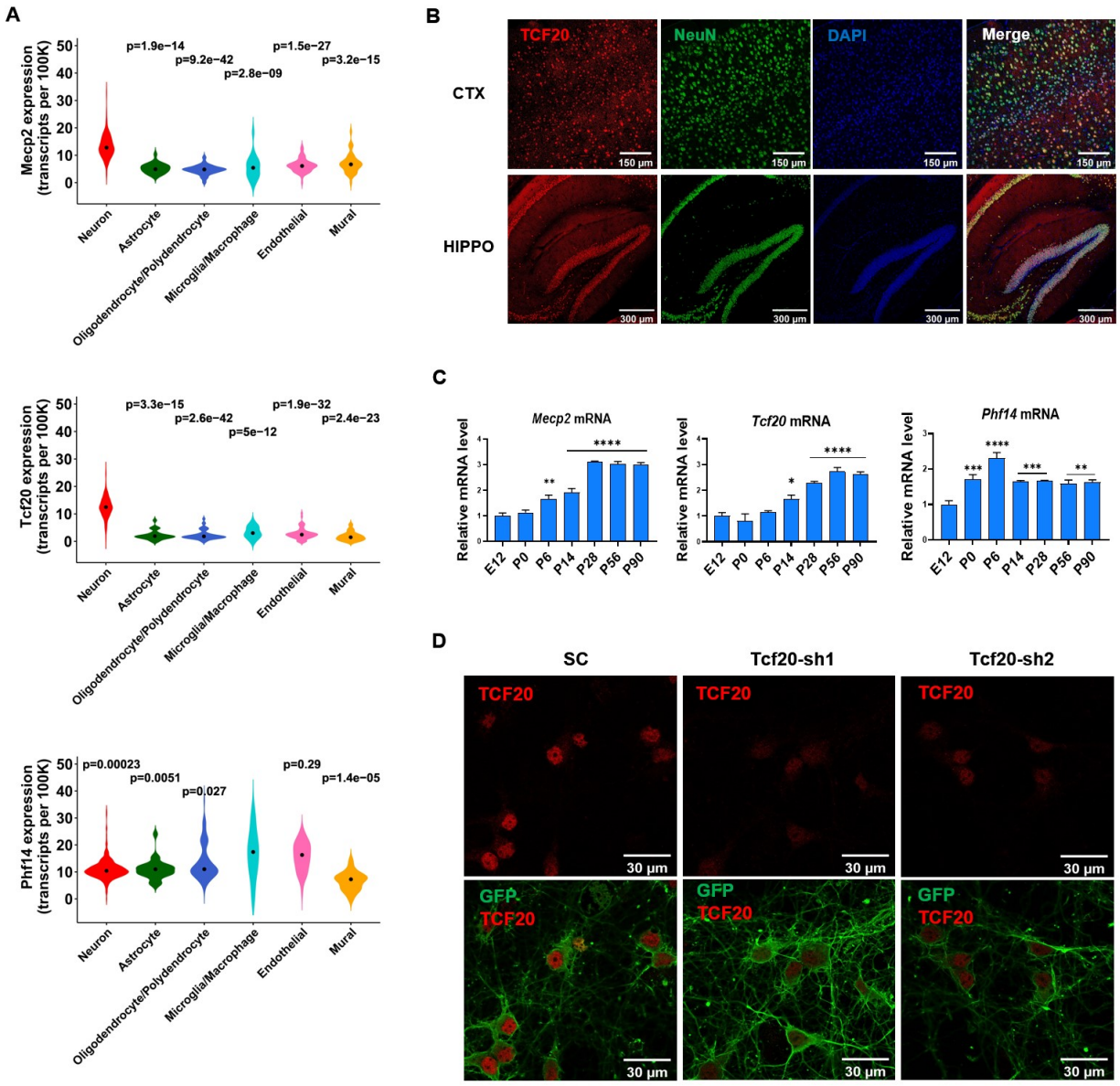

Figure S3. *Tcf20* is co-expressed with *Mecp2* in mouse neurons. (A) Expression of *Mecp2*, *Tcf20*, and *Phf14* across adult mouse brain cell types. Violin plot shows the median value (point). P-value indicates whether the expression of *Mecp2* and *Tcf20* in the corresponding cell type is lower than that in neuron, and the expression of *Phf14* in the corresponding cell type is lower than that in microglia/macrophage (one-sided Wilcoxon rank-sum test). **(B)** Representative immunocytochemical images of mouse cortex (CTX) and hippocampus (Hippo) showing TCF20 (red), NeuN (green) and DAPI (blue) proteins. **(C)** Quantification of *Mecp2*, *Tcf20* and *Phf14* mRNA levels by RT-qPCR during mouse cortical development as neurons mature when compared to E12. (n=3/group, one-way ANOVA with post-hoc Tukey's tests). **(D)**

Representative images showing TCF20 (red) protein levels following *Tcf20* knockdown. GFP (green) represents the transduction efficiency of non-targeting scramble (SC) AAV virus or AAV-shRNAs targeting *TCF20*. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, Data are mean ± SEM.

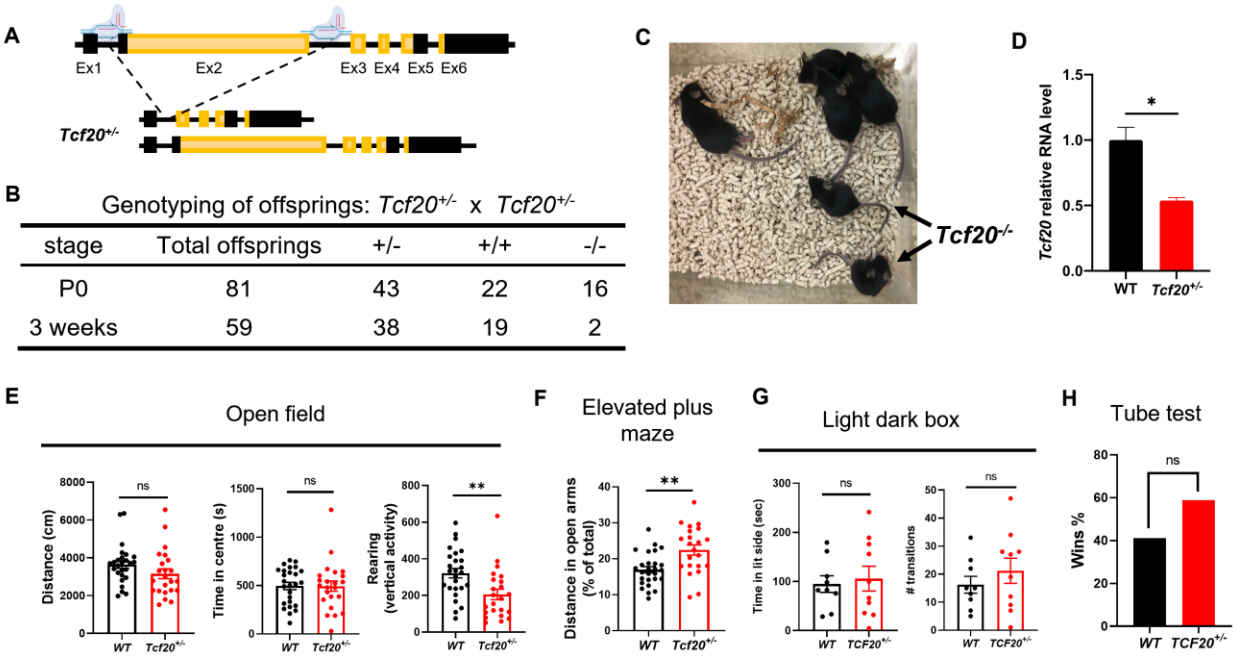

Figure S4. *Tcf20*$^{+/-}$ female mice show milder behavioral deficits. (A) Strategy to generate the *Tcf20* knockout allele. In the WT *Tcf20* genomic locus (first row), yellow boxes indicate the *Tcf20* exons, black boxes indicate untranslated region (UTR). Cas9-gRNA complex binds to and cut out exon 2, resulting in a knockout allele (second row). **(B)** Table showing the numbers of survived offspring of indicated genotypes at P0 or 3 weeks after birth. Chi square analysis shows no significant difference between actual and expected at P0 and significant difference between actual and expected at 3 weeks (p=0.0027). **(C)** A photograph showing the reduced body size of *Tcf20*$^{-/-}$ mice. **(D)** Quantification of *Tcf20* mRNA levels by RT-qPCR in WT and *Tcf20*$^{+/-}$ mice (n=3/group, unpaired two-tailed Student's t-test). **(E)** Statistical analysis of open field test for WT and *Tcf20*$^{+/-}$ female mice. Left, total mouse movement in 30 min; middle, time spent in the center area; right, time spent for vertical exploring (rearing) (WT, n=26, *Tcf20*$^{+/-}$, n=23, unpaired two-tailed Student's t-test). **(F)** Time spent in the open arm for WT and *Tcf20*$^{+/-}$ female mice in the elevated plus maze (WT, n=25, *Tcf20*$^{+/-}$, n=22, unpaired two-tailed Student's t-test). **(G)** Statistical analysis of light-dark box test for WT and *Tcf20*$^{+/-}$ female mice. Left, time spent in the light side; right, the number of transitions between the two compartments (WT, n=9, *Tcf20*$^{+/-}$, n=10, unpaired two-tailed Student's t-test). **(H)** Percentage of wins in test pairs between WT and *Tcf20*$^{+/-}$ female mice in the tube test (n = 80 matches, two-tailed binomial test). *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Data are mean ± SEM.

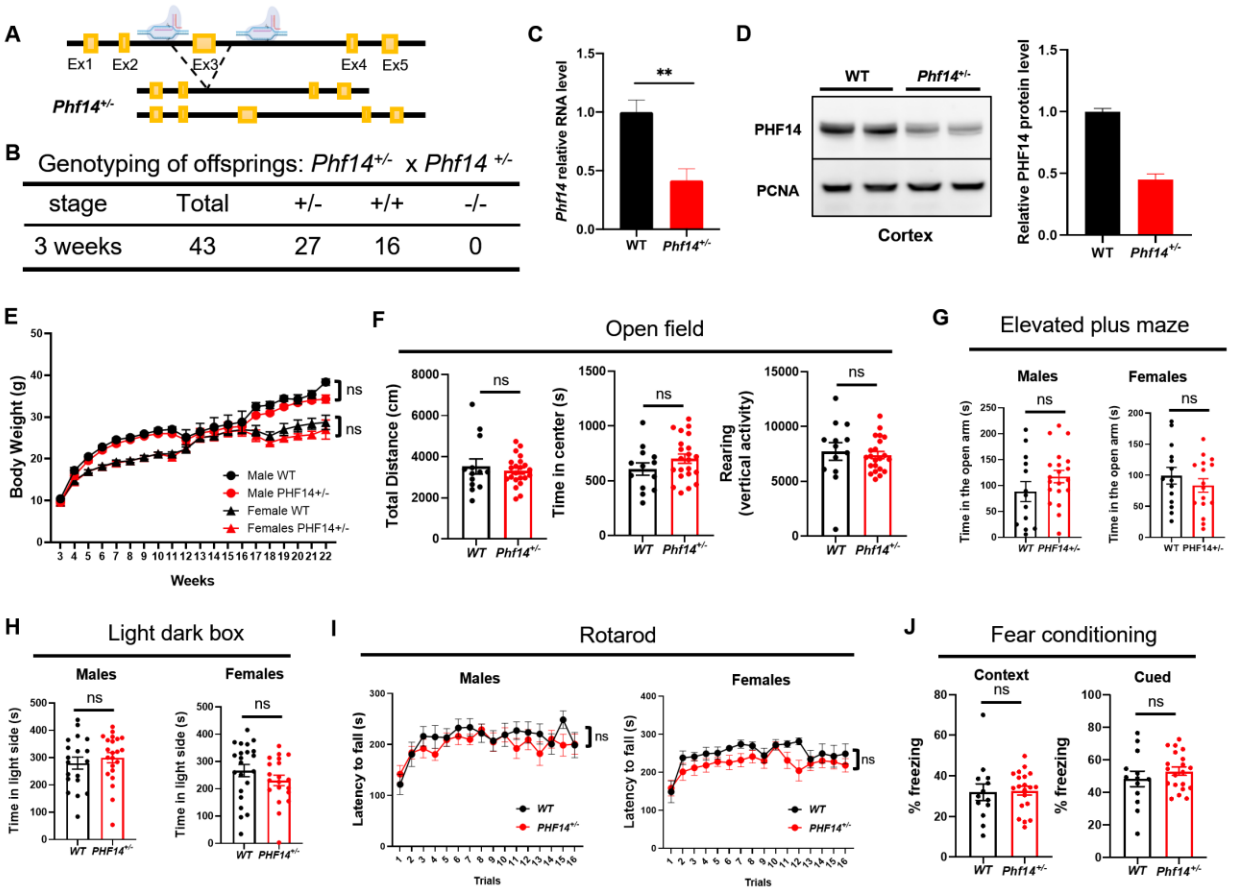

Figure S5. *Phf14*$^{+/-}$ mice show no behavioral deficits. (A) Strategy to generate the *Phf14* knockout allele. In the WT *Phf14* genomic locus (first row), yellow boxes indicate the *Phf14* exons. Cas9-gRNA complex binds to and cut out exon 3, resulting in a knockout allele (second row). **(B)** Table showing the numbers of survived offspring of indicated genotypes at 3 weeks after birth. Chi square analysis shows significant difference between actual and expected at 3 weeks (p= 0.0006). **(C)** Quantification of *Phf14* mRNA levels by RT-qPCR in WT and *Phf14*$^{+/-}$ mice (n=3/group, unpaired two-tailed Student's t-test). **(D)** Representative immunoblot (left) and quantification (right) of PHF14 protein levels in the cortex from WT and *Phf14*$^{+/-}$ mice. (n=2/group). **(E)** Body weights of WT and *Phf14*$^{+/-}$ mice over the course of 22 weeks (For males, WT, n=5, *Phf14*$^{+/-}$, n=9; for females, WT, n=6, *Phf14*$^{+/-}$, n=7; two-way ANOVA with post-hoc Tukey's tests). **(F)** Statistical analysis of open field test for WT and *Phf14*$^{+/-}$ mice. Left, total mouse movement in 30 min; middle, time spent in the center area; right, time spent for vertical exploring (rearing) (WT, n=13, *Phf14*$^{+/-}$, n=22; unpaired two-tailed Student's t-test). **(G)** Statistical analysis of time spent in the open arm for WT and *Phf14*$^{+/-}$ mice in the elevated plus maze (For males, WT, n=13, *Phf14*$^{+/-}$, n=20; for females, WT, n=14, *Phf14*$^{+/-}$, n=16; unpaired

two-tailed Student's t-test). **(H)** Statistical analysis of time spent in the light side in light-dark box test for WT and *Phf14*$^{+/-}$ mice (For males, WT, n=19, *Phf14*$^{+/-}$, n=22; for females, WT, n=22, *Phf14*$^{+/-}$, n=20; unpaired two-tailed Student's t-test). **(I)** Statistical analysis of rotarod test for WT and *Phf14*$^{+/-}$ mice (For males, WT, n=12, *Phf14*$^{+/-}$, n=15; for females, WT, n=8, *Phf14*$^{+/-}$, n=10). **(J)** Statistical analysis of context (left) and cued (right) learning and memory in conditioned fear test for WT and *Phf14*$^{+/-}$ mice (WT, n=13, *Phf14*$^{+/-}$, n=20, unpaired two-tailed Student's t-test). *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Data are mean ± SEM.

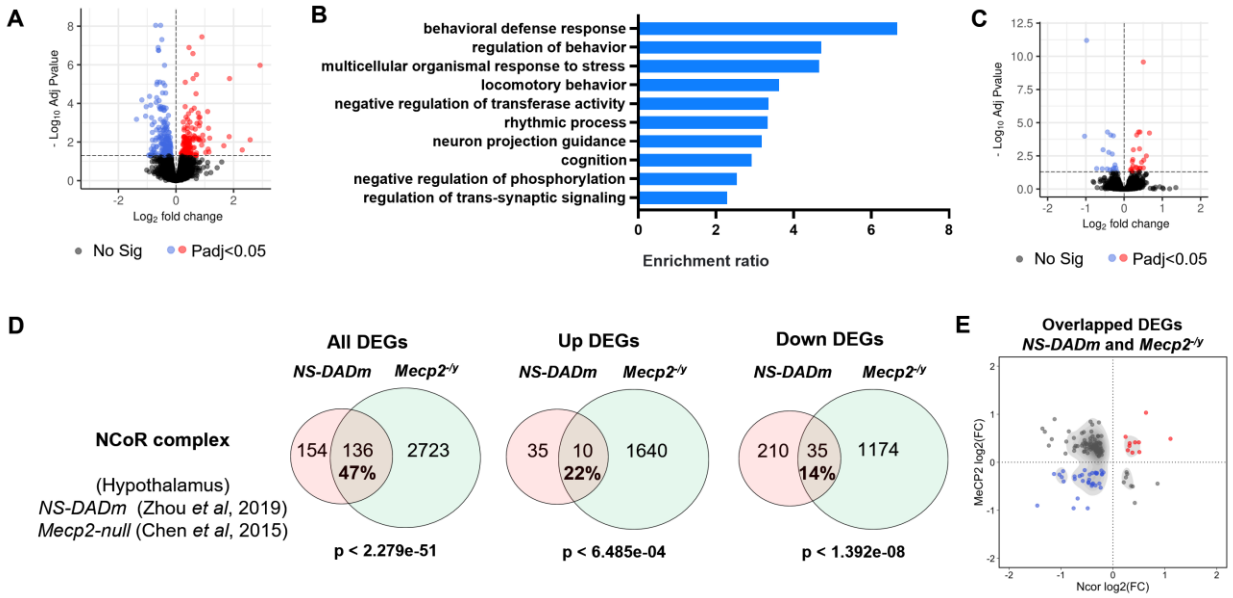

Figure S6. MeCP2 and TCF20 share common downstream genes. (A) Volcano plot showing genome-wide gene expression change within a fixed scale in 7-week-old *Tcf20*$^{+/-}$ prefrontal cortex (PFC) compared to WT control. **(B)** Gene Ontology analysis of differentially expressed genes (DEGs) in *Tcf20*$^{+/-}$ PFC. Gene Ontology terms are ranked by enrichment ratio. **(C)** Volcano plot showing genome-wide gene expression change in 7-week-old *Phf14*$^{+/-}$ prefrontal cortex compared to WT control (Full list is given in supplementary dataset S6). **(D)** Venn diagrams showing hypothalamus DEGs that overlap between *NS-DADm* and *Mecp2*$^{-/y}$ mouse models; left, all DEGs, middle, up-regulated DEGs, right, downregulated DEGs (Full lists of *NS-DADm* and *Mecp2*$^{-/y}$ DEGs are given in supplementary dataset S4 and S5, respectively). The percentage rate indicates the ratio of the number of overlapped DEGs to the number of *NS-DADm* DEGs. **(E)** Scatter plot showing $\log_2$ fold-change for overlapped DEGs in *NS-DADm* vs *Mecp2*$^{-/y}$ mouse models.

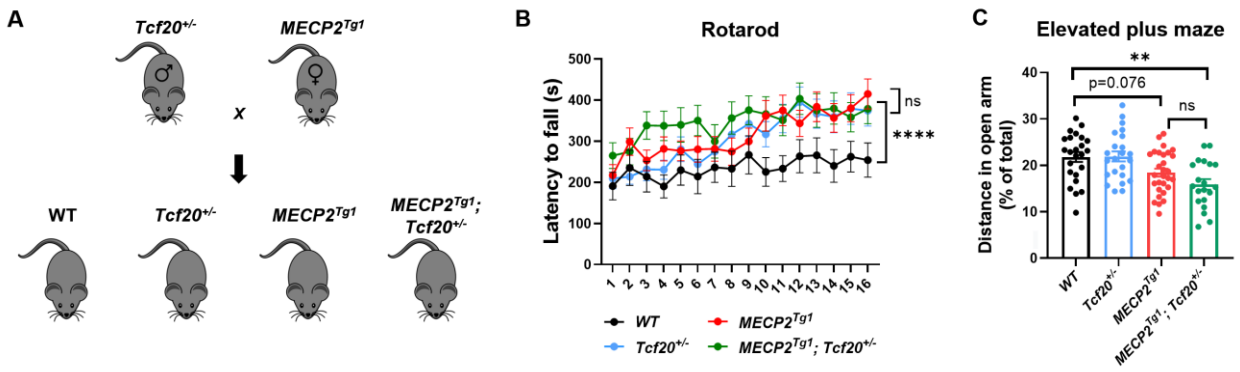

Figure S7. Genetic reduction of *Tcf20* shows no improvement of rotarod and elevated plus maze tests in adult *MECP2^{Tg1}* mice. (A) Breeding scheme with male *Tcf20*$^{+/-}$ and female *MECP2*^{Tg1} mice to generate WT, *Tcf20*$^{+/-}$, *MECP2*^{Tg1}, and *MECP2*^{Tg1}*; Tcf20*$^{+/-}$ mice. **(B)** Statistical analysis of rotarod test for mice with the indicated genotypes. (n=16-27/genotype, two-way ANOVA with post-hoc Tukey's tests). **(C)** Statistical analysis of time spent in the open arm for mice with the indicated genotypes in the elevated plus maze (n = 24-29/genotype, one-way ANOVA with post-hoc Tukey's tests). **p < 0.01, ****p < 0.0001, Data are mean ± SEM.

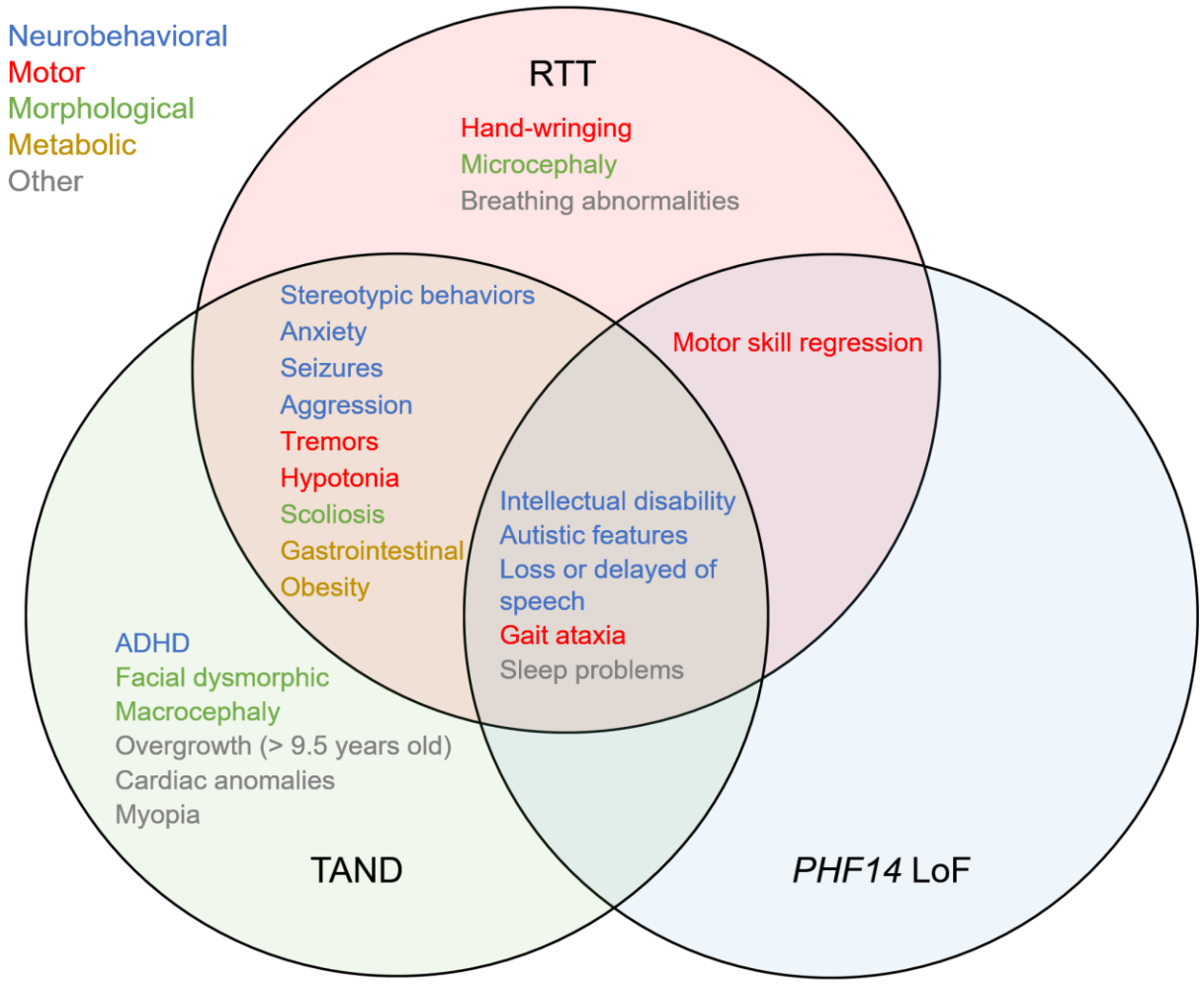

Figure S8. Patients with variants in *PHF14* display neurodevelopmental deficits that overlap with those of RTT or TAND. Venn diagrams showing the shared and distinct clinical features between RTT patients, TAND patients, and patients with LoF variants in *PHF14*.

Supplementary datasets

Dataset S1. Full list of peptides collected in the BioID mass spectrometry experiments.

Dataset S2. Full list of genome-wide gene expression changes in the prefrontal cortex of *Tcf20*$^{+/-}$ mice compared to WT control.

Dataset S3. Full list of genome-wide gene expression changes in the prefrontal cortex of *Mecp2*$^{-/y}$ mice compared to WT control.

Dataset S4. Full list of genome-wide gene expression changes in the hypothalamus of *NS-DADm* mice compared to WT control.

Dataset S5. Full list of genome-wide gene expression changes in the hypothalamus of *Mecp2*$^{-/y}$ mice compared to WT control.

Dataset S6. Full list of genome-wide gene expression changes in the prefrontal cortex of *Phf14*$^{+/-}$ mice compared to WT control.

References

1. H. Hamdan, *et al.*, Mapping axon initial segment structure and function by multiplexed proximity biotinylation. *Nat. Commun.* **11**, 1–17 (2020).

2. D. Palmer, P. Ng, Improved system for helper-dependent adenoviral vector production. *Mol. Ther.* **8**, 846–852 (2003).

3. D. J. Palmer, P. Ng, Methods for the production of helper-dependent adenoviral vectors. *Methods Mol. Biol.* **433**, 33–53 (2008).

4. S. Guan, J. C. Price, S. B. Prusiner, S. Ghaemmaghami, A. L. Burlingame, A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. *Mol. Cell. Proteomics* **10** (2011).

5. K. R. Clauser, P. Baker, A. L. Burlingame, Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871–2882 (1999).

6. M. W. C. Rousseaux, *et al.*, A druggable genome screen identifies modifiers of α-synuclein levels via a tiered cross-species validation approach. *J. Neurosci.* **38**, 9286–9301 (2018).

7. R. C. Samaco, *et al.*, Female Mecp2+/- mice display robust behavioral deficits on two different genetic backgrounds providing a framework for pre-clinical studies. *Hum. Mol. Genet.* **22**, 96–109 (2013).

8. Y. Sztainberg, *et al.*, Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature* **528**, 123–126 (2015).

9. Y. Shao, *et al.*, Antisense oligonucleotide therapy in a humanized mouse model of MECP2 duplication syndrome. *Sci. Transl. Med.* **13** (2021).

10. L. Wang, *et al.*, An autism-linked missense mutation in SHANK3 reveals the modularity of Shank3 function. *Mol. Psychiatry* **25**, 2534–2555 (2020).

11. S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor in *Bioinformatics*, (2018), pp. i884–i890.

12. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

13. H. K. Yalamanchili, Y. W. Wan, Z. Liu, Data analysis pipeline for RNA-seq experiments: From differential expression to cryptic splicing. *Curr. Protoc. Bioinforma.* **2017**, 11.15.1-11.15.21 (2017).

14. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** (2014).

15. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).

16. A. Saunders, *et al.*, Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030.e16 (2018).

17. K. Pang, *et al.*, Coexpression enrichment analysis at the single-cell level reveals convergent defects in neural progenitor cells and their cell-type transitions in neurodevelopmental disorders. *Genome Res.* **30**, 835–848 (2020).

18. M. V. Kuleshov, *et al.*, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

19. P. Shannon, *et al.*, Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

20. N. A. Miller, *et al.*, A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* **7**, 100 (2015).

21. G. A. Van der Auwera, *et al.*, From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).

22. S. Richards, *et al.*, Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical

Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).

23. T. Froukh, *et al.*, Genetic basis of neurodevelopmental disorders in 103 Jordanian families. *Clin. Genet.* **97**, 621–627 (2020).