

Title: CTCF barrier breaking by ZFP661 promotes protocadherin diversity in mammalian brains

Authors: Jinpu Jin^{1*}, Sherry Ralls¹, Elaine Wu^{1†}, Gernot Wolf^{1‡}, Ming-An Sun^{1,2}, Danielle A. Springer³, Rachel L. Cosby¹, Anna D. Senft¹, Todd S. Macfarlan^{1*}

Affiliations:

¹*The Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA.*

²Institute of Comparative Medicine, College of Veterinary Medicine, Yangzhou University, Yangzhou, Jiangsu 225009, China.

³The National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, 20892, USA.

*Corresponding author. Email: todd.macfarlan@nih.gov (T.S.M.); jinpujin@gmail.com (J.J.)

†Present address: Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

‡Present address: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna 1090, Austria.

Abstract: Mammalian brains are larger and more densely packed with neurons than reptiles, but the genetic mechanisms underlying the increased connection complexity amongst neurons are unclear. The expression diversity of clustered protocadherins (Pcdhs), which is controlled by CTCF and cohesin, is crucial for proper dendritic arborization and cortical connectivity in vertebrates. Here, we identify a highly-conserved and mammalian-restricted protein, ZFP661, that binds antagonistically at CTCF barriers at the *Pcdh* locus, preventing CTCF from trapping cohesin. ZFP661 balances the usage of Pcdh isoforms and increases Pcdh expression diversity. Loss of *Zfp661* causes cortical dendritic arborization defects and autism-like social deficits in mice. Our study reveals both a novel mechanism that regulates the trapping of cohesin by CTCF and a mammalian adaptation that promoted Pcdh expression diversity to accompany the expanded mammalian brain.

One sentence summary: ZFP661 blocks cohesin trapping by CTCF and increases protocadherin diversity for proper cortical dendritic arborization.

Main Text: Mammals have enlarged brains relative to reptiles, with greater numbers of neurons and increased neuronal density, especially in the cerebrum and cerebellum (1, 2). The increased number and density of neurons in mammals likely necessitated evolutionary adaptations to allow neurites to properly identify each other and intermingle in a more crowded space during development. However, the genetic mechanisms that underlie the adaptations are not well understood. In vertebrates, the clustered protocadherin genes (Pcdhs), which encode cell adhesion molecules, play a central role as neuronal identity tags (3-7). Pcdhs have dozens of alternative isoforms, and only a selected subset of isoforms is expressed in individual neurons, producing a large diversity of isoform combinations (5). Neurites with the same/similar Pcdh

combinations repel each other (6-8) whereas neurites with diverse Pcdhs can intermingle (5, 6). Repulsive signals amongst neurites are likely achieved from the formation of zipper-like chains mediated by isoform-specific homophilic interactions of Pcdh molecules, in such a way that isoform mismatch would terminate chain formation and reduce the chance of forming repulsive signals (9-11). At the transcriptional level, the combination of repression (12, 13) and stochastic demethylation (14) of alternative Pcdh promoters produces the diverse Pcdh isoforms that are accessible for activation in individual neurons. The enhancers downstream of the *Pcdh* locus further activate the expression of accessible Pcdh isoforms, dependent on CTCF and cohesin (14-16). CTCF and cohesin are two key players that control 3D genome structures and gene expression; ATP-dependent cohesin extrusion of chromatin promotes enhancer-promoter interactions, while CTCF acts as a barrier to stop (*i.e.*, trap) cohesin extrusion and confine most enhancer-promoter interactions within the CTCF boundary (17-21). Whether the trapping of cohesin by CTCF can be regulated is also unclear.

Krüppel-associated box zinc finger proteins (KRAB-ZFPs) represent the largest family of transcription factors in mammals. Although many of these factors are species-restricted and bind to and repress lineage specific retrotransposons (22-24), a large number also emerged in the last common ancestor of mammals (23, 25). Some of these factors, like ZFP57/ZFP445 and ZFP568 play critical roles in mammalian-specific phenomenon like genomic imprinting and placental *Igf2* suppression (26-28), although the vast majority have gone unexplored. Many KRAB-ZFPs are preferentially expressed in human fetal brains (29, 30) and a small number have been shown to overlap CTCF and RAD21 (a subunit of cohesin complex) binding (23), indicating their potential roles in brain development and evolution by rewiring chromatin loops. Here, we identified a conserved KRAB-ZFP in Theria, ZFP661 (ZNF2 in humans), that binds to a small subset of CTCF sites genome-wide, including a cluster of CTCF sites at the *Pcdh* locus. Using mouse genetics, we uncovered a critical function for ZFP661 in modulating the trapping of cohesin at CTCF barriers, diversifying neuronal Pcdh expression and shaping mammalian brain development.

ZFP661 is highly conserved in Theria and suppresses cohesin trapping at CTCF barriers without altering CTCF binding

To identify a candidate to explore the potential role of KRAB-ZFPs in rewiring 3D genome structures, we re-analyzed ChIP-seq data of 221 human KRAB-ZFPs (23) and found that 77.4% (171/221) of KRAB-ZFPs had binding peaks that significantly overlapped CTCF and RAD21 peaks (Fig. 1A). Amongst these KRAB-ZFPs, ZNF2 (known as ZFP661 in mice), was unique due to both the number and proportion of ChIP peaks that overlapped with CTCF and RAD21 (Fig. 1A). We determined that ZFP661 originated before the divergence of placental mammals and marsupials and that it is present in most (89/94) of the available therian genomes (Fig. 1B) according to a gain/loss tree at Ensembl. Notably both the DNA-binding domain (DBD) (90.1% identity across three representative species) and zinc fingerprint amino acids that are responsible for making specific DNA base contacts (97.2% identity) are highly conserved (fig. S1), suggesting it serves an important function in mammals.

In mice, *Zfp661* mRNA is broadly expressed at low levels with detectable expression in mouse embryonic stem cells (mESCs) and relatively higher expression in the developing brain compared to other organs (fig. S2). To explore the function of ZFP661 during embryonic development, we generated *Zfp661* knockout (KO) mESCs and mice, endogenously tagged

ZFP661-3HA mESCs and transgenic mice using CRISPR-Cas9 technology, and ZFP661-3HA stably overexpressing (OE) mESCs using lentiviral vectors (Fig. 1C). We determined ZFP661 binding locations in the mouse genome using ChIP-seq, revealing that over 80% of ZFP661 binding peaks overlapped with those of CTCF and cohesin in both endogenously tagged ZFP661-3HA mESCs (Fig. 1D) and ZFP661-3HA OE mESCs (fig. S3), consistent with its ortholog ZNF2 in human (Fig. 1A). We further identified a putative ZFP661 binding motif that closely matched a computationally predicted ZFP661 binding motif, which was present in all 200 peak regions that were used to detect motifs (fig. S4).

Unlike many KRAB-ZFPs that strongly recruit the co-factor KAP1 and subsequently SETDB1 to establish H3K9me3 heterochromatin at genomic targets (28, 31, 32), ZFP661 target regions only presented weak KAP1 and H3K9me3 signals after overexpression of ZFP661 (fig. S5), indicating a potentially atypical functional mechanism. To determine whether ZFP661 might influence CTCF or cohesin binding, we performed ChIP-seq for CTCF and RAD21 in *Zfp661*^{-/-} mESCs and ZFP661-3HA OE mESCs relative to wild-type controls. Surprisingly, we found that ZFP661 decouples CTCF and cohesin binding; *Zfp661* loss-of-function increased RAD21 binding, whereas overexpression of ZFP661 suppressed RAD21 binding at CTCF sites co-occupied with ZFP661, without altering CTCF binding (Figs. 1E and 1F).

ZFP661 binds exclusively inside CTCF loop anchors and allows cohesin to pass through CTCF barriers

To determine whether ZFP661 co-occupies targets simultaneously with CTCF, we performed ChIP-reChIP-seq assays with anti-HA (1st ChIP) and anti-CTCF (2nd ChIP) antibodies in ZFP661-3HA OE mESCs. These experiments confirmed that ZFP661 and CTCF are co-bound (Fig. 2A). Previous studies have demonstrated that chromatin loops occur primarily between two convergent CTCF binding sites (as shown in the upper portion of Fig. 2B) (33, 34) and that reversal of the orientation of CTCF binding sites prevents cohesin from being trapped at CTCF barriers (16). To gain insights into how ZFP661 might influence cohesin binding, we analyzed ZFP661 binding locations, indicated by ChIP summits, relative to CTCF barriers. We found that ZFP661 bound exclusively inside CTCF loop anchors, like cohesin (Fig. 2B). We confirmed this finding by analyzing the position of ZFP661 binding sites relative to CTCF binding sites, which were consistently located inside the loop anchors (fig. S6).

We hypothesized two possible mechanisms that could explain ZFP661 decoupling of CTCF and cohesin signals: 1) ZFP661 could allow cohesin to pass through CTCF barriers or 2) ZFP661 could promote cohesin release from chromatin. To distinguish amongst these hypotheses, we tested the insulation scores at CTCF barriers and the proportion of chromatin loops that passed through CTCF barriers, reasoning that cohesin release would not decrease the barrier function of CTCF. We generated high-resolution Hi-C maps (1.18 billion valid interaction pairs in total) in ZFP661-3HA OE relative to empty vector (EV) controls. This allowed us to calculate both the insulation scores at CTCF barriers and the distribution of loop anchors at and over CTCF barriers. We found that overexpression of ZFP661 reduced insulation function of CTCF barriers (*i.e.*, increased insulation scores) (Fig. 2C) and promoted loops to pass through CTCF barriers (rather than being trapped by CTCF) (Fig. 2D) at CTCF sites co-occupied with ZFP661. Moreover, this effect was not seen at CTCF barriers that were not co-occupied with ZFP661 (Fig. 2C), demonstrating that ZFP661 binding suppresses the trapping of cohesin at CTCF barriers (Figs. 1E, 1F, and 2C). In sum, these results demonstrate that ZFP661 suppresses

cohesin binding by allowing cohesin to pass through CTCF barriers. Considering that the interaction of the CTCF N-terminus (the portion facing cohesin extrusion and ZFP661 binding) with cohesin is critical for cohesin trapping (35), combined with our finding that ZFP661 binds at the same location where cohesin is typically trapped (Fig. 2B), we speculate that ZFP661 suppresses cohesin trapping by decreasing the probability of cohesin interaction with CTCF and/or competition with trapped cohesin (Fig. 2E).

ZFP661 promotes the interaction of downstream enhancers with distal *Pcdh* promoters

To determine the potential role of ZFP661 during development by modulating the strength of CTCF barriers, we generated *Zfp661* null (*Zfp661*^{-/-}) mice (Fig. 1C). *Zfp661*^{-/-} mice were viable and fertile, and displayed no overt phenotypes. Therefore, we suspected that ZFP661 might play a more subtle role in development. We performed Gene Ontology enrichment analysis of ZFP661 target genes (based on proximity to binding sites) and identified “homophilic cell adhesion via membrane adhesion molecules” as the most significantly enriched biological process (Fig. 3A). Interestingly, 21 of 25 target genes associated with this term belong to the protocadherin (*Pcdh*) gene clusters (the occupancy of ZFP661 and CTCF at the *Pcdh* locus is shown in Fig. 3B and fig. S7A). There are three *Pcdh* gene clusters (*Pcdha/a*, *Pcdhb/b* and *Pcdhy/g*) in mammalian genomes, representing 14 *Pcdha* and 21 *Pcdhy* isoforms (excluding *Pcdhgb8*, which is annotated as a non-coding gene), each with alternative promoters, and 22 *Pcdhb* genes in the mouse genome (fig. S7A). *Pcdhs* are activated by downstream enhancers dependent on the cooperation of CTCF and cohesin, where CTCF located at each promoter can stop cohesin extrusion to facilitate the contact of this promoter with downstream enhancers (14-16).

At the *Pcdh* locus, ZFP661 binding regions were co-bound by CTCF and were primarily located at the proximal promoters closest to the downstream enhancer cluster of *Pcdhy* (Fig. 3B and fig. S7A). We generated capture Hi-C maps in mouse E16.5 forebrains (when neurons occupy the largest proportion by cell population in the developing mouse brain (36)) and demonstrated that the downstream enhancers of *Pcdhy* interact with *Pcdhb* and *Pcdhy* clusters (Fig. 3B), indicating they control both *Pcdhb* and *Pcdhy* expression (37). We reasoned that ZFP661 binding at enhancer proximal promoters might promote cohesin to pass through CTCF barriers, allowing the enhancers to contact and activate more distal promoters of the *Pcdhb* and *Pcdhy* clusters. To test this hypothesis, we performed Capture Hi-C experiments of the *Pcdh* locus in E16.5 forebrains of *Zfp661*^{+/+} and *Zfp661*^{-/-} littermates. Importantly, we found that loss of *Zfp661* increased the contacts of enhancers with proximal promoter regions (4/5 increased) while decreasing contacts with more distal promoters (21/29 decreased; Figs. 3C and 3D). Consistent with this result, we found that ZFP661 overexpression (in ZFP661-3HA OE mESCs) decreased the contacts of these enhancers with proximal regions (fig. S7B) and promoted contacts with distal promoters (fig. S7C).

ZFP661 balances the usage of *Pcdh* isoforms and increases *Pcdh* expression diversity

Previous studies have demonstrated that the interaction between downstream enhancers and *Pcdh* alternative promoters is the basis for activating *Pcdh* expression (15, 16, 37). To investigate how the alteration of enhancer-promoter interactions affects the *Pcdh* expression repertoire in individual neurons, we performed single-cell 5' RNA-seq in mouse E16.5

forebrains to distinguish each Pcdh isoform. The number of unique molecular identifiers (UMIs) and genes detected per cell between *Zfp661*^{+/+} and *Zfp661*^{-/-} littermates were matched by adjusting for sequencing depth and by down-sampling UMIs to avoid possible biases in downstream analyses (figs. S8A and S8B). Cell types were assigned by projection onto a cell atlas of developing mouse brains at the corresponding ages (36)(fig. S8C and Fig. 4A). Compared with *Zfp661*^{+/+} brains, *Zfp661*^{-/-} brains had no obvious alterations in cell populations (Fig. 4A) or in the overall expression level of *Pcdhs* within cortical or hippocampal glutamatergic neurons (referred to as “glutamatergic neurons”) or forebrain GABAergic neurons (referred to as “GABAergic neurons”) (Fig. 4B). Compared with GABAergic neurons, glutamatergic neurons expressed higher levels of *Pcdhs* (Fig. 4B), perhaps due to their requirement for farther neurite projection (38). When comparing the usage of Pcdh isoforms, *Zfp661*^{-/-} brains had greater usage of enhancer proximal isoforms at the expense of distal isoforms in pooled neurons (Fig. 4C), glutamatergic neurons (fig. S9A) and GABAergic neurons (fig. S9B), consistent with the underlying chromatin-interaction results (Figs. 3C and 3D).

Diversity in the Pcdh expression repertoire allows neurites to co-share local space and form more complex connections (5). To better quantify Pcdh diversity and explore the impact of Pcdh usage alteration in *Zfp661*^{-/-} neurons, we calculated a similarity score between all pairs of cells within the population, which represents the probability that a Pcdh isoform taken from one cell will match an isoform from another cell (as shown in the upper portion of Fig. 4D). By comparing median similarity scores of 1000 subsamples from all pairs of cells within each population, we determined that *Zfp661*^{-/-} neurons possessed significantly greater similarity (*i.e.*, reduced diversity) than *Zfp661*^{+/+} neurons, both within glutamatergic and GABAergic neuron types and between the two types (Fig. 4D). To further investigate the impact of this reduced diversity on neuronal dendrite repulsion, we calculated repulsive signals, which represent the probability of sequential matching of Pcdh isoforms during continuous comparison by stimulating the chain formation of Pcdh molecules (9) (as shown in the upper portion of Fig. 4E). The results demonstrate that ZFP661 dramatically decreases repulsive signals from Pcdh chain formation, both within glutamatergic and GABAergic neuron types and between types (Fig. 4E). In sum, these data reveal that the loss of *Zfp661* leads to reduced complexity in Pcdh expression, which increases the likelihood that a cell will encounter another cell with the same or highly similar Pcdh expression repertoire. This subsequently results in increased repulsive signals during dendritic projection, potentially reducing the number of dendrites that can fit in a local space.

Loss of *Zfp661* causes deficits in dendritic arborization and social interaction

To determine whether the reduced Pcdh expression diversity in *Zfp661*^{-/-} mice might cause dendritic projection defects, we performed Golgi-Cox staining on early adult mouse brains (P60-66) and imaged pyramidal neurons at layer II/III of somatomotor areas (Bregma: 1.70 to 1.18 mm), where we could easily ensure consistency in neuron type and location during sampling. Sholl analysis, which measures the complexity of neuronal dendrites, confirmed that *Zfp661*^{-/-} neurons have deficits in arborization and the distribution of neuronal dendrites (fig. S10 and Fig. 5A). Compared with *Zfp661*^{+/+} neurons, *Zfp661*^{-/-} neurons displayed a significant decrease in both the branch number (Fig. 5B) and total branch length of neuronal dendrites (Fig. 5C). These data are consistent with the hypothesis that the loss of *Zfp661*, which causes greater repulsive signals during Pcdh tiling (Fig. 4E), might inhibit the elongation and arborization of neuronal dendrites. This is also consistent with previous studies that demonstrated that direct decreases in Pcdh

diversity (via genetic ablation) also cause deficits in dendritic arborization (39-41). Finally, we explored how *Zfp661* loss-of-function might impact behavior. Sociability tests revealed that *Zfp661*^{-/-} mice displayed deficits in social interactions (Fig. 5D), similar to mice that have reduced *Pcdh* diversity (42, 43).

Discussion

CTCF binding is well known to act as a barrier for cohesin extrusion, but how CTCF stops the extrusion of cohesin is still unclear. Genetic modifications, including reversal of the orientation of CTCF binding sites (16) and destruction of CTCF-cohesin interaction (35) can suppress cohesin trapping without altering CTCF binding. These highlight the roles of CTCF binding orientation and CTCF-cohesin interactions in cohesin trapping. Here, we uncovered a novel mechanism that decouples CTCF and cohesin binding; ZFP661 exclusively binds inside CTCF loop anchors (Fig. 2B) and allows cohesin to pass through CTCF barriers (Figs. 2D and 3C and fig. S7B), indicating that the accessibility of CTCF-cohesin interface is crucial for cohesin trapping. The chromatin loops established by CTCF and cohesin confine over 90% of enhancer-promoter interactions that occur within CTCF boundaries (21). Consequently, the need for enhancer-promoter interactions to occur across specific CTCF barriers may require the breaking of CTCF boundaries. Our findings demonstrate that ZFP661 directly regulates CTCF-mediated cohesin trapping at CTCF barriers, providing a potentially rapid and reversible way to open/close CTCF boundaries.

During neural development, 3D genome structures are dramatically changed to accompany altered gene expression patterns (44). Numerous KRAB-ZFPs that originated during mammal evolution (23, 25) are preferentially expressed in developing brains (29) and bind at CTCF barriers (Fig. 1A), similar to ZFP661. Our findings suggest that KRAB-ZFP-dependent CTCF/cohesin modulation may have played important roles in driving the evolution of 3D genome structures in mammalian brain development.

Nodes and edges (*i.e.*, connections amongst nodes) are two key components of a network. Increased connections per node are required after node expansion to maintain proper network architecture and avoid network fragmentation (as shown in the simplified cases in Fig. 5E), which also holds true in neuronal networks. During mammalian evolution, ZFP661 emerged as a factor that increases the diversity of *Pcdh* expression (Fig. 4D), which decreases the repulsive signals amongst neurons during dendritic projection (Fig. 4E). We hypothesize that this allows more neuronal dendrites to fit in a tighter space and promotes the arborization and elongation of neuronal dendrites during development (Fig. 5F), providing a possible solution that ensures proper neuronal connections and network preservation in the expanded mammalian brain. Multiple findings indicate this function is critical throughout Theria: 1) ZFP661 is retained in most therian genomes (Fig. 1B); 2) ZFP661 has a highly conserved DNA binding domain with 97.2% (35/36) of fingerprint amino acids that determine DNA bases to contact being conserved (fig. S1); 3) the ZFP661/ZNF2 binding motif (fig. S11A), its association with CTCF binding (Figs. 1A and 1D), and its binding at the *Pcdh* locus (fig. S11B) are conserved between human and mouse.

Both the loss of *Zfp661* (Figs. 5A to 5D) and reduced *Pcdh* diversity from direct genetic ablation (39-43) can induce cortical dendritic arborization defects and social deficits. Autism Spectrum Disorder (ASD) is one of most common brain developmental disorders in humans associated with social deficits, which impacts about 2.3% children in the United States (45). The

causes of ASD are largely unclear, but growing evidence suggests that defects in neuronal connections are a major underlying cause (46). Remarkably, ZNF2 binding at the *Pcdh* locus is conserved in humans (fig. S11B), and structural variations at *ZNF2* (fig. S12) and *Pcdh* loci (47) are associated with ASD in humans. These results highlight an important link between the regulation of *Pcdh* diversity by ZFP661(ZNF2), neuronal connections, and ASD.

References and Notes

1. T. B. Rowe, T. E. Macrini, Z. X. Luo, Fossil evidence on origin of the mammalian brain. *Science* **332**, 955-957 (2011).
2. K. Kverková *et al.*, The evolution of brain neuron numbers in amniotes. *Proceedings of the National Academy of Sciences* **119**, e2121624119 (2022).
3. N. Kohmura *et al.*, Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* **20**, 1137-1151 (1998).
4. Q. Wu, T. Maniatis, A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**, 779-790 (1999).
5. G. Mountoufaris *et al.*, Multiclusted *Pcdh* diversity is required for mouse olfactory neural circuit assembly. *Science* **356**, 411-414 (2017).
6. J. L. Lefebvre, D. Kostadinov, W. V. Chen, T. Maniatis, J. R. Sanes, Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517-521 (2012).
7. W. V. Chen *et al.*, *Pcdh*α2 is required for axonal tiling and assembly of serotonergic circuitries in mice. *Science* **356**, 406-411 (2017).
8. D. Kostadinov, J. R. Sanes, Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. *Elife* **4**, (2015).
9. R. Rubinstein *et al.*, Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* **163**, 629-642 (2015).
10. C. A. Thu *et al.*, Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell* **158**, 1045-1059 (2014).
11. D. Schreiner, J. A. Weiner, Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc Natl Acad Sci U S A* **107**, 14893-14898 (2010).
12. S. Toyoda *et al.*, Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* **82**, 94-108 (2014).
13. Y. Jiang *et al.*, The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet* **49**, 1239-1250 (2017).
14. D. Canzio *et al.*, Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin alpha Promoter Choice. *Cell* **177**, 639-653 e615 (2019).
15. Y. Guo *et al.*, CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc Natl Acad Sci U S A* **109**, 21081-21086 (2012).

16. Y. Guo *et al.*, CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910 (2015).
17. E. P. Nora *et al.*, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).
18. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017).
19. I. F. Davidson *et al.*, DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).
20. H. Huang *et al.*, CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat Genet* **53**, 1064-1074 (2021).
21. D. Hnisz, D. S. Day, R. A. Young, Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188-1200 (2016).
22. F. M. Jacobs *et al.*, An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242-245 (2014).
23. M. Imbeault, P. Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550-554 (2017).
24. G. Wolf *et al.*, KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *Elife* **9**, (2020).
25. A. D. Senft, T. S. Macfarlan, Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* **22**, 691-711 (2021).
26. X. Li *et al.*, A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell* **15**, 547-557 (2008).
27. N. Takahashi *et al.*, ZNF445 is a primary regulator of genomic imprinting. *Genes Dev* **33**, 49-54 (2019).
28. P. Yang *et al.*, A placental growth factor is silenced in mouse embryos by the zinc finger protein ZFP568. *Science* **356**, 757-759 (2017).
29. C. J. Playfoot *et al.*, Transposable elements and their KZFP controllers are drivers of transcriptional innovation in the developing human brain. *Genome Res* **31**, 1531-1545 (2021).
30. P. A. Johansson *et al.*, A cis-acting structural variation at the ZNF558 locus controls a gene regulatory network in human brain development. *Cell Stem Cell* **29**, 52-69 e58 (2022).
31. H. M. Rowe *et al.*, KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240 (2010).
32. M. Bruno, M. Mahgoub, T. S. Macfarlan, The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* **53**, 393-416 (2019).
33. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

34. Z. Tang *et al.*, CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627 (2015).
35. Y. Li *et al.*, The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476 (2020).
36. G. La Manno *et al.*, Molecular architecture of the developing mouse brain. *Nature* **596**, 92-96 (2021).
37. Z. Jia *et al.*, Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol* **21**, 75 (2020).
38. H. Peng *et al.*, Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174-181 (2021).
39. A. M. Garrett, D. Schreiner, M. A. Lobas, J. A. Weiner, gamma-protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* **74**, 269-276 (2012).
40. L. Suo, H. Lu, G. Ying, M. R. Capecchi, Q. Wu, Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. *J Mol Cell Biol* **4**, 362-376 (2012).
41. Z. Shao *et al.*, Dysregulated protocadherin-pathway activity as an intrinsic defect in induced pluripotent stem cell-derived cortical interneurons from subjects with schizophrenia. *Nat Neurosci* **22**, 229-242 (2019).
42. D. M. Steffen *et al.*, The gamma-Protocadherins Interact Physically and Functionally with Neuroligin-2 to Negatively Regulate Inhibitory Synapse Density and Are Required for Normal Social Interaction. *Mol Neurobiol* **58**, 2574-2589 (2021).
43. S. Balan *et al.*, A loss-of-function variant in SUV39H2 identified in autism-spectrum disorder causes altered H3K9 trimethylation and dysregulation of protocadherin beta-cluster genes in the developing brain. *Mol Psychiatry* **26**, 7550-7559 (2021).
44. B. Bonev *et al.*, Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e524 (2017).
45. D. L. Christensen *et al.*, Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveill Summ* **65**, 1-23 (2018).
46. T. Bourgeron, From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci* **16**, 551-563 (2015).
47. A. Anitha *et al.*, Protocadherin alpha (PCDHA) as a novel susceptibility gene for autism. *J Psychiatry Neurosci* **38**, 192-198 (2013).
48. J. Jin, Imaging data and 3D reconstruction of neuronal dendrites. *Mendeley Data*, (2023).
49. J. Jin, Calculation of similarity scores for clustered protocadherins amongst individual neurons. *Zenodo*, (2023).

Acknowledgments: We thank A. Grinberg, J. Yimdjio and V. Biggs for generating and maintaining transgenic mice; T. Li, J. Iben, R. Lazris, F. Faucz and S. Coon for next-generation

sequencing (NGS) supports; H. Potts and M. Peiravi for mouse behavioral tests; members of Macfarlan lab, K. Pfeifer, J. Kassis and P. Rocha for helpful discussion; J. Kassis, B. Shen, M. Bruno, M. Custance and G. Tricola for critically reading the manuscript and helpful suggestions. Most of the computational analyses were performed on the NIH Biowulf high-performance computing platform.

Funding: National Institutes of Health grant DIR 1ZIAHD008933 (TSM)

Author contributions:

Conceptualization: JJ, TSM

Methodology: JJ, TSM

Investigation: JJ, SR, EW, GW, DAS

Software: JJ

Formal analysis: JJ, MAS

Visualization: JJ

Funding acquisition: TSM

Project administration: TSM

Supervision: TSM

Writing – original draft: JJ

Writing – review & editing: JJ, TSM, RLC, ADS

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: NGS data are available at Gene Expression Omnibus (GSE231869). Raw imaging data are available upon request, and preprocessed imaging data and neuronal traces are available at Mendeley (48). Program codes for calculating similarity scores for Pcdhs amongst neurons are available at Zenodo (49). Cell and mouse models generated in this work are available under material transfer agreements with National Institutes of Health.

Supplementary Materials

Materials and Methods

Figs. S1 to S12

Tables S1 to S3

Fig. 1. ZFP661, a conserved KRAB-ZFP in Theria, suppresses cohesin binding at CTCF sites without altering CTCF binding. (A) Overlap of binding peaks for 221 human KRAB-ZFPs with those of CTCF and RAD21 for screening purposes. The proportion of random peaks overlapping with CTCF and RAD21 was used as background ($y=1.06$), and single-tailed exact binomial tests were performed (threshold: FDR adjusted $P \leq 0.01$). (B) Presence/absence of ZFP661 in different vertebrate clades. ZFP661 orthologs are detected in most therian genomes (89/94) but not monotremes or birds/reptiles. (C) Cell and mouse models generated to explore ZFP661 function. (D) Overlap of ZFP661 binding peaks with CTCF and RAD21 peaks in ZFP661-3HA mESCs. (E and F) Influence of ZFP661 on CTCF and RAD21 binding in *Zfp661* knockout mESCs (E) and ZFP661-3HA OE mESCs (F). CTCF and RAD21 peaks that do not overlap with ZFP661 peaks were used as controls. ZFP661 binding significantly suppresses the binding of RAD21 but not CTCF in both *Zfp661* knockout mESCs (E) and ZFP661-3HA OE mESCs (F). EV represents empty vector control. Single-tailed Wilcoxon rank sum tests were performed.

Fig. 2. ZFP661 binds exclusively inside CTCF loop anchors and allows cohesin to pass through CTCF barriers. (A) ChIP-reChIP assay in ZFP661 OE mESCs shows that ZFP661 and CTCF simultaneously co-bind to targets. The 1st ChIP was performed using anti-HA antibody, and the 2nd ChIP was performed using anti-CTCF and anti-IgG antibodies, respectively. ChIP-reChIP signals in the heatmap represent anti-CTCF signals subtracted by anti-IgG signals. ZFP661-only and CTCF-only peaks, which are not co-occupied with each other, were used as controls. (B) Binding locations of ZFP661 and RAD21, indicated by ChIP summits, relative to those of CTCF. CTCF binding orientation is shown as the motif in the upper portion. (C) Influence of ZFP661 binding on the insulation score at CTCF barriers that are co-occupied with ZFP661. CTCF barriers that are not co-occupied with ZFP661 were used as controls. Single-tailed paired t-tests were performed. (D) Comparison of the distribution of chromatin loops trapped around or passing over CTCF barriers that are co-occupied with ZFP661 between EV control and ZFP661 OE. (E) Schematic diagram illustrates the working model of ZFP661. ZFP661 binding at CTCF barriers blocks cohesin trapping and allows cohesin to pass through the barriers.

Fig. 3. ZFP661 promotes the interaction of *Pcdhy* downstream enhancers with distal protocadherin (*Pcdh*) promoters. (A) Gene Ontology (GO) enrichment analysis of ZFP661 target genes in “biological processes”. (B) Capture Hi-C map at the *Pcdh* locus in E16.5 mouse forebrains (upper) and ZFP661 and CTCF occupancy in mESCs (lower). The Capture Hi-C data demonstrate that the enhancer cluster downstream of *Pcdhy* locus interacts with both *Pcdhβ* and *Pcdhy* promoters. (C and D) Alterations in contacts at the *Pcdhβ* and *Pcdhy* regions (C) and contacts between the enhancer cluster and individual promoters (TSS \pm 2.5 kb) (D) after *Zfp661* knockout in the Capture Hi-C map of E16.5 forebrains. Loss of *Zfp661* leads to increased contacts of enhancers with proximal isoforms (C) and reduced contacts with distal promoters (D).

Fig. 4. ZFP661 balances the usage of *Pcdh* isoforms and increases the diversity of *Pcdh* expression. (A) UMAP visualization of the cell atlas in *Zfp661*^{+/+} and *Zfp661*^{-/-} E16.5 mouse forebrains. (B) Unique molecular identifier (UMI) counts of *Pcdhs* detected in glutamatergic and

GABAergic neurons of *Zfp661*^{+/+} and *Zfp661*^{-/-} brains. The y-axis was square-root transformed. (C) Usage alterations of Pcdh β/γ isoforms in pooled glutamatergic and GABAergic neurons of *Zfp661*^{+/+} and *Zfp661*^{-/-} brains. The lines represent isoform usage, and lollipops represent the difference in isoform usage between *Zfp661*^{+/+} and *Zfp661*^{-/-} neurons. ZFP661 allows for greater usage of distal Pcdh isoforms. (D) The distribution of median similarity scores of the Pcdh repertoire between all pairs of single neurons within glutamatergic (left), GABAergic (center), or between glutamatergic and GABAergic neurons (right) in *Zfp661*^{+/+} and *Zfp661*^{-/-} brains. Similarity scores within each type were subsampled 50% of the datapoints for 1000 times, and *P* values were calculated based on the 1000 subsamples. (E) Ratio of repulsive signals between *Zfp661*^{+/+} and *Zfp661*^{-/-} neurons by simulating Pcdh chain formation. Repulsive signal represents the probability that a Pcdh isoform from one neuron matches the isoform from another neuron during continuous comparisons. ZFP661 decreases Pcdh-based repulsive signals amongst neurons (data are shown as mean \pm SD).

Fig. 5. Loss of ZFP661 causes deficits in neuronal dendritic arborization and social interaction. (A) Sholl analysis of dendrites of pyramidal neurons in layer II/III of the cerebral cortex in *Zfp661*^{+/+} and *Zfp661*^{-/-} mouse brains (P60-66). Representative examples of 3D-reconstructions of *Zfp661*^{+/+} and *Zfp661*^{-/-} neurons are shown above (data are shown as mean \pm SEM; n=24, 8 neurons per mouse and 3 mice per genotype, one-way repeated measures ANOVA). (B and C) Comparison of the branch number (B) and total branch length of neuronal dendrites (C) between *Zfp661*^{+/+} and *Zfp661*^{-/-} neurons (n=24, single-tailed Wilcoxon rank sum test). (D) *Zfp661*^{-/-} mice show social interaction deficits in sociability tests (n=6, 3 male and 3 female mice per genotype, single-tailed paired Wilcoxon rank sum test). Littermate pairs between *Zfp661*^{+/+} and *Zfp661*^{-/-} mice are connected using dashed lines. (E) Two simplified diagrams demonstrating the necessity of additional connections (dashed edges) for new neurons (green nodes) to integrate well into original brain networks to avoid fragmentation. (F) Schematic diagram illustrates that ZFP661 increases Pcdh expression diversity and promotes intermingling and co-sharing of local space amongst neuronal dendrites by balancing the expression of Pcdh isoforms. This may have helped to accommodate the larger mammalian brain.

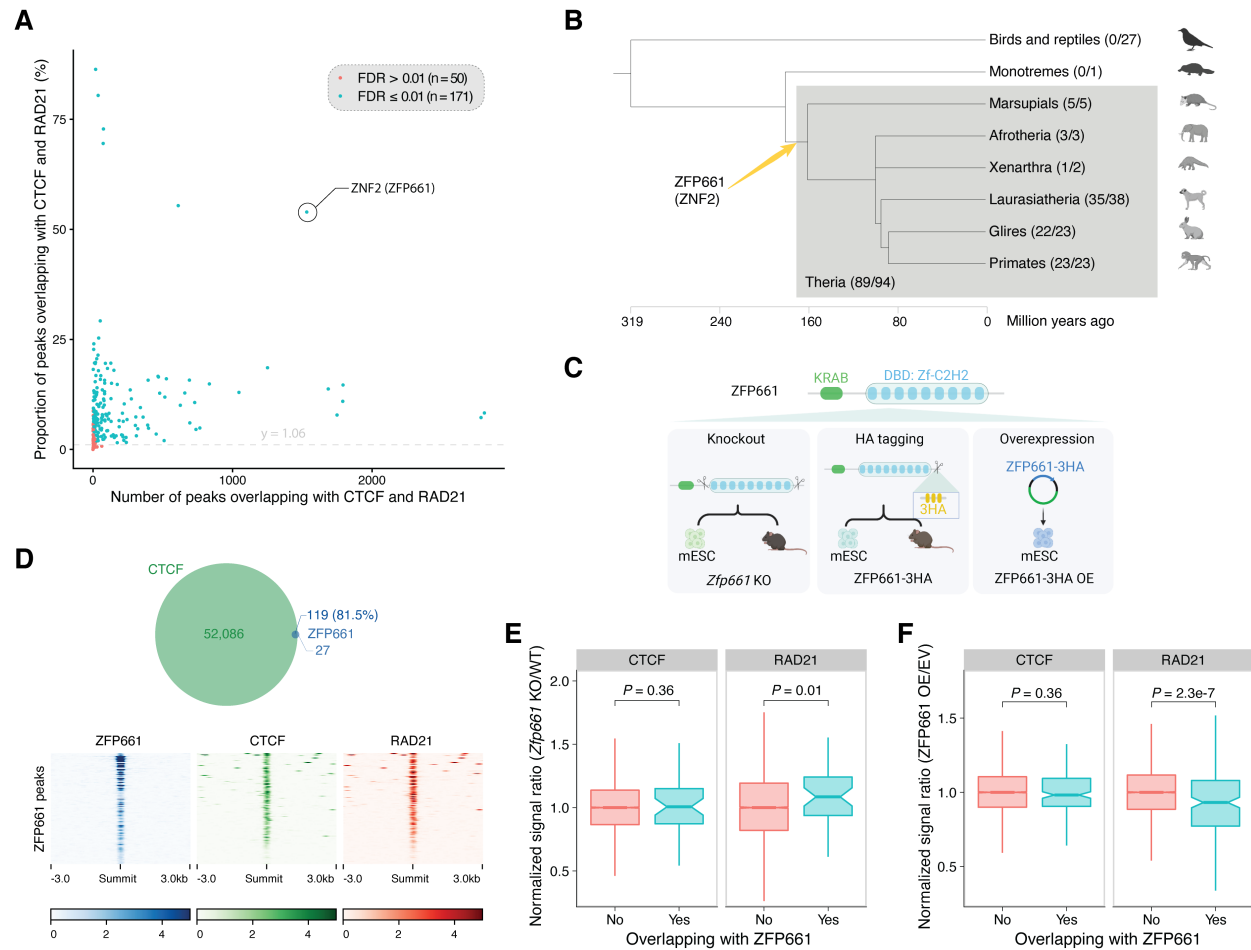


Fig. 1. ZFP661, a conserved KRAB-ZFP in Theria, suppresses cohesin binding at CTCF sites without altering CTCF binding.

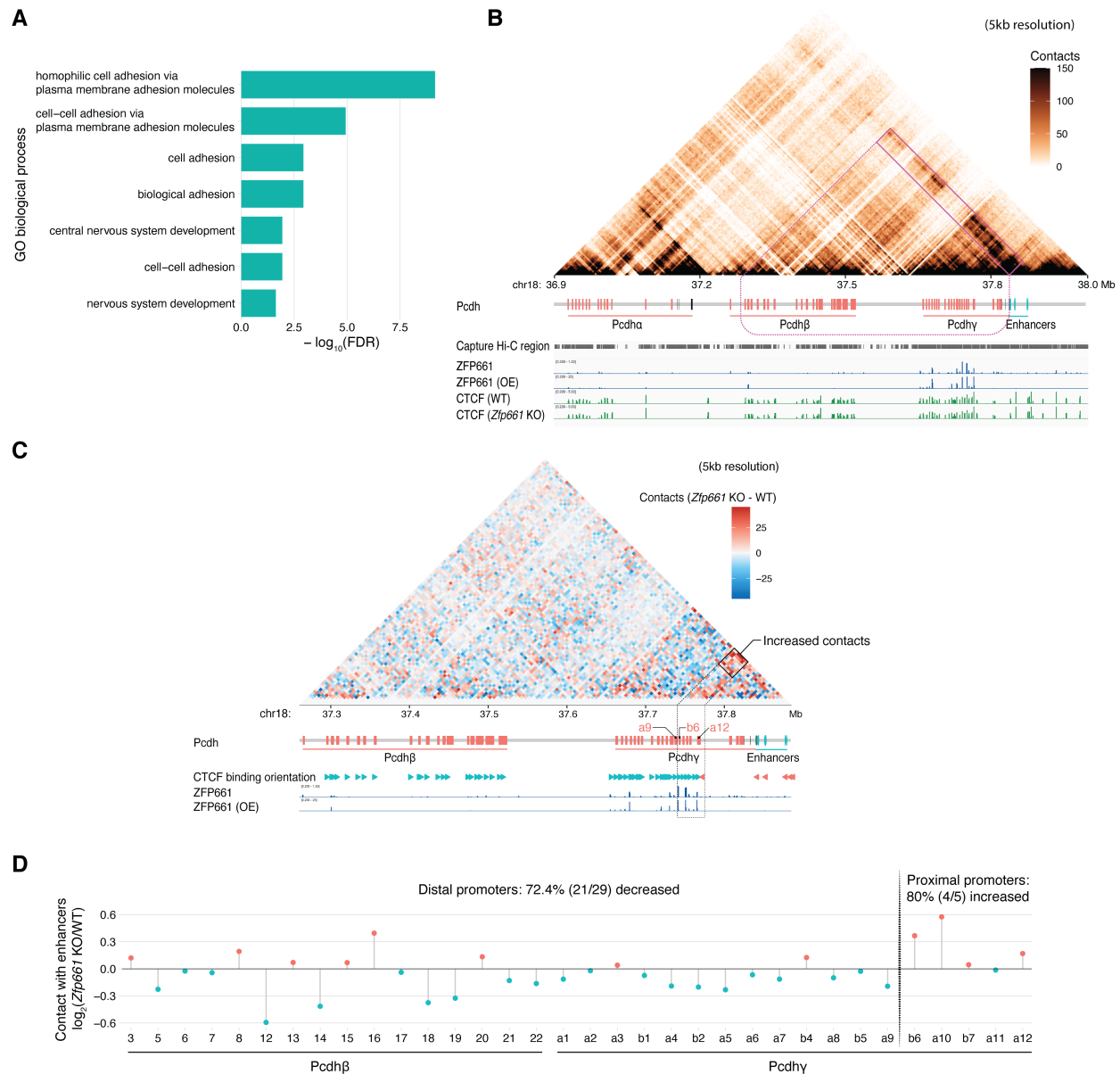


Fig. 3. ZFP661 promotes the interaction of *Pcdhy* downstream enhancers with distal protocadherin (*Pcdh*) promoters.

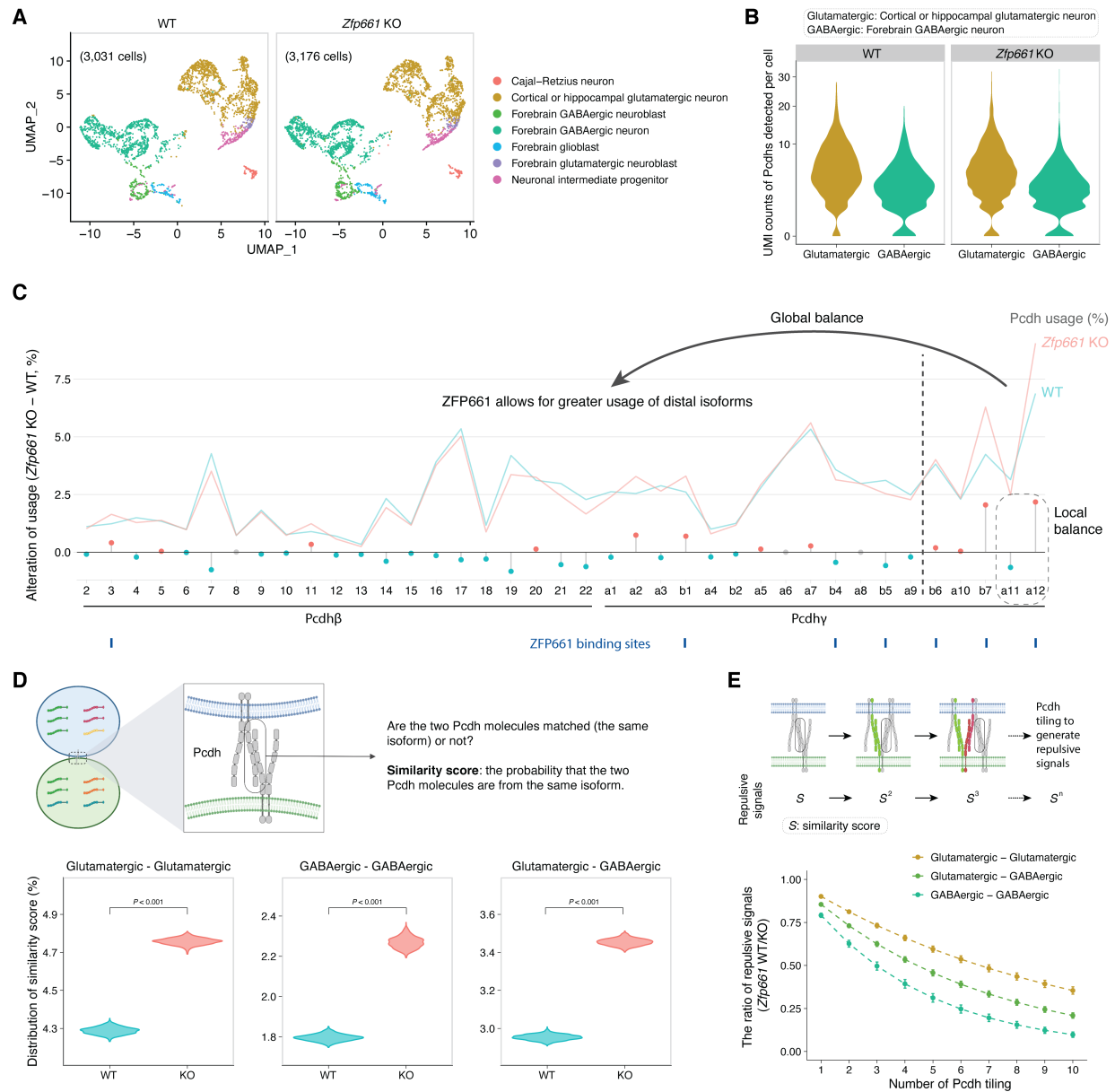


Fig. 4. ZFP661 balances the usage of Pcdh isoforms and increases the diversity of Pcdh expression.

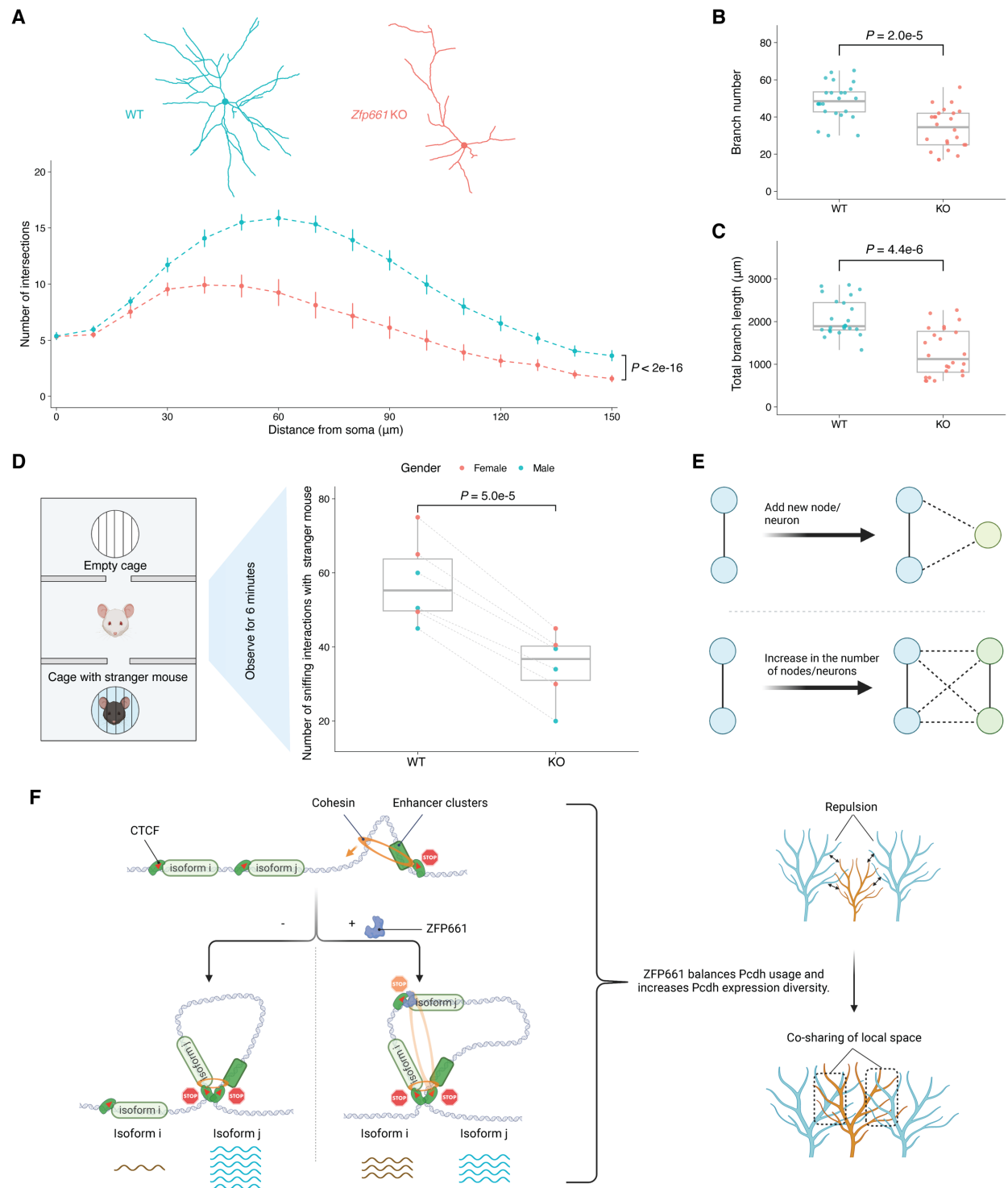


Fig. 5. Loss of ZFP661 causes deficits in neuronal dendritic arborization and social interaction.