Diversified expression of gamma-protocadherins regulates synaptic specificity in the mouse neocortex

Yijun Zhu^{1,2,4}, Caiyun Deng^{1,4}, Yaqian Wang¹, Hua-Tai Xu^{*1,3}

1 Institute of Neuroscience and State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031, China.

2 University of Chinese Academy of Sciences, Beijing 100049, China.

3 Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, Shanghai 201210, China.

4 These authors contributed equally to this work.

* Correspondence should be addressed to H. -T. X. (xuhuatai@ion.ac.cn)

23 **Abstract**

24 Synaptic specificity is the basis of forming neural microcircuits. However, how a
 25 neuron chooses which neurons out of many potentials to form synapses remains largely
 26 unknown. Here we identified that the diversified expression of clustered protocadherin
 27 γ s (cPCDH γ s) plays an essential role in regulating such specificity. Our 5-prime end
 28 single-cell sequencing data revealed the diversified expression pattern of cPCDH γ s in
 29 neocortical neurons. Whole-cell recording of neuron pairs in developing mouse brain
 30 slices showed that knocking out PCDH γ s significantly increased the local connection
 31 rate of nearby pyramidal neurons. By contrast, neurons overexpressing the same group
 32 of clustered PCDH γ isoforms through *in utero* electroporation dramatically decreased
 33 their synaptic connectivity. Finally and more importantly, decreasing the similarity
 34 level of PCDH γ isoforms over-expressed in neuron pairs through sequential *in utero*
 35 electroporation led to a progressive elevation of synaptic connectivity. Our
 36 observations provide strong evidence to support that the existence of diversely
 37 expressed cPCDH γ s allows a neuron to choose which neurons not to form a synapse,
 38 rather than choosing which neurons to make synapses.

39

40 **Introduction**

41 The specificity of synaptic connections is critical for the function of neural
 42 circuits(Yogev and Shen 2014). There are two levels of specificity to be
 43 considered(Sanes and Zipursky 2020): first, a self-nonsel specificity, in which the axon
 44 of a neuron distinguishes the surface of its own (or the same neuron type) from that of

45 others. Second, the synaptic target specificity, in which the axon distinguishes different
 46 types of neurons for selective formation of synapses. Cell adhesion molecules play
 47 critical roles for the specificity of synapse formation(Duan et al. 2014, Serizawa et al.
 48 2006, Tan et al. 2015, Rawson, Martin, and Williams 2017, Sytnyk, Leshchyns'ka, and
 49 Schachner 2017, Berns et al. 2018, Jang, Lee, and Kim 2017, Courgeon and Desplan
 50 2019), but how the specificity at the microcircuit level was achieved remains to be
 51 clarified. Due to their combinatorial and stochastic expression pattern, the homophilic
 52 adhesion molecules clustered protocadherins (cPCDHs) made them a strong candidate
 53 for promoting synaptic specificity(Yagi 2012, Kohmura et al. 1998, Toyoda et al. 2014).
 54 This unique pattern could lead to millions of possible alternative patterns of cPCDH
 55 isoforms on the neuron surface(Kaneko et al. 2006, Esumi et al. 2005), suggesting
 56 cPCDHs could function as a unique barcode for each neuron(Esumi et al. 2005,
 57 Mountoufaris et al. 2017). However, homophilic interactions mediated by large
 58 overlapping patterns of cPCDH isoforms on opposing cell surfaces may lead to
 59 intercellular repulsion(Rubinstein et al. 2015, Brasch et al. 2019, Honig and Shapiro
 60 2020). The intracellular mechanism underlying the repulsive effect on synaptic
 61 formation triggered by homophilic interaction of cPCDH proteins may involve the
 62 protein kinase C signal pathway(Garrett et al. 2012) and the WAVE regulatory
 63 complex(Chen et al. 2014). Whether homophilic cPCDHs could act as an attractive
 64 binding force in the neocortex for specifying synaptic targets at the microcircuit level
 65 remains unclear. Based on 5-prime end (5'-end) single-cell sequencing and multiple
 66 whole-cell patch-clamp recordings, our data indicated that diversely expressed

cPCDH γ s played a vital role for synaptic specificity at the level of self-nonself discrimination.

Results

5'-end single-cell sequencing demonstrated the diversified expression pattern of cPCDH γ s in neocortical neurons

The gamma isoform of cPCDHs (cPCDH γ) is critical for synaptic connectivity (Kostadinov and Sanes 2015, Tarusawa et al. 2016). To determine whether and how cPCDH γ might influence synaptic specificity in the neocortex, we first examined their expression in the cortical tissue of postnatal mice. Previous studies have suggested that cPCDHs were stochastically expressed in cortical neurons (Hirayama et al. 2012, Toyoda et al. 2014). Here we developed a 5'-end single-cell RNA sequencing approach for identifying the exon 1 of cPCDH γ isoforms since they differ only here (Kohmura et al. 1998, Wu and Maniatis 1999). After reverse transcription and cDNA amplification from the neocortex of a postnatal day 11 (P11) mouse (Fig. 1-S1A and B), cDNA was split into two parts: one for specific amplification of cPCDH γ mRNAs and the other for the 5' gene expression library construction (Fig. 1-S1C). In total, we collected 6505 neurons out of 17438 cells from a P11 neocortex (Fig. 1A and Fig1-S1D). We analyzed neurons with more than 10 UMI (cutoff>1) for all cPCDH γ (Fig. 1B, C) and found that “constant” isoforms were indeed expressed in significantly more cells compared to “variable” isoforms (Fig. 1D and Fig. 1-S1E), consistent with a previous study (Toyoda et al. 2014). Pair-wise analysis on similarity in the single-cell expression pattern of cPCDH γ isoforms revealed that most neocortical neurons expressed a distinct

Figure 1: Diversified expression of *pcdhg* isoforms in neocortical neurons.

(A) UMAP of all cells (n=17438) collected from 5'-end single-cell sequencing after the cleanup and doublets removing. Neurons were labeled as green dots and non-neurons as red. (B) UMAP clusters of all neurons further labeled by the UMI cutoff. Red dots represented cells with no more than 10 total UMIs of *pcdhg* (n=3671), and green dots represented cells with more than 10 UMIs (n=2834). (C) The fraction of neurons with more than 10 UMIs in different clusters. (D) Fraction of cells expressing different *pcdhg* isoforms in neocortical neurons. (E) Fraction distribution for different similarity levels in the combinatory expression of *pcdhg* isoforms among neurons. (F) Fraction distribution of the number of the expressed isoforms per cell for all neurons. (G) The difference between observed and expected variance of the fraction distribution from (F).

89 combination of cPCDHγ isoforms (Fig. 1E). Based on the variance analysis(Wada,
90 Wallerich, and Becskei 2018) of fraction distribution of the number of the expressed
91 isoforms per cell for all neurons (Fig. 1F and Fig. 1-S2), we found that there was a weak
92 but significant co-occurrence for the expression of cPCDHγ isoforms in most of cells
93 (Fig. 1G). We did not detect obvious differences among clusters except that cluster 1
94 with much less expression and no co-occurrence of cPCDHγ isoforms. Overall, our 5'-
95 end single-cell revealed that cPCDHγs were diversely expressed in neocortical neurons
96 with a weak but significant co-occurrence.

97

Absence of cPCDH γ increased synaptic connectivity among nearby pyramidal neurons

To examine the function of cPCDH γ in the synaptic specificity, we first used cPCDH γ conditional knockout mice, obtained by crossing *Pcdhg flox/flox* mice (Lefebvre et al. 2008) with *Nes-cre* mice (Goebbels et al. 2006), to determine whether synapse formation was affected by deleting all cPCDH γ isoforms. Using multiple whole-cell patch-clamp recordings from acute brain slices of P9-P32 mice, we measured the connectivity among nearby pyramidal cells (<50 μ m between cell somas) in the neocortex by the presence of evoked monosynaptic responses. As shown in the sample traces by the connectivity matrix among six recorded neurons, 2 cell pairs exhibited monosynaptic connections (indicated by yellow stars) out of 15 possibilities (Fig. 2A). Overall, we found that the percentage of connected pairs was significantly higher in *Pcdhg* KO mice (20.2%, 103/511) than in wild-type (WT) mice (15%, 122/813), and no such difference existed between *Pcdha* KO mice (11.3%, 38/337) and WT mice (14.3%, 26/182) (Fig. 2B). We performed an extra set of recordings for *Pcdhg* KO mice to separate neuron pairs along the axis vertical vs. horizontal to the pial surface and found a more significant difference in the connectivity for vertically aligned cells (18.3%, 94/515 in *Pcdhg* KO mice vs. 11.2%, 73/651 in WT mice for vertical aligned neuron pairs; 12.2%, 27/221 in *Pcdhg* KO mice vs. 9.52%, 28/294 in WT mice for horizontal aligned neuron pairs) (Fig. 2C). In the more detailed analysis of these vertical pairs, we found that the absence of cPCDH γ expression appeared to significantly increase synapse formation between cells separated vertically by 50 to 150 μ m (24%,

Figure 2: The absence of cPCDH γ increases the synaptic connectivity probability between nearby pyramidal cells.

(A) Sample traces (red/green, average traces; gray, 10 original traces) of a multiple-electrode whole-cell patch-clamp recording on six neurons in layer 2/3 of the barrel cortex. Positive evoked postsynaptic responses were indicated by yellow stars. Scale bars, 100mV (green), 40pA (red), and 10ms (black). (B) Connectivity probability among nearby pyramidal cells in layer 2/3 of the barrel cortex in *pcdhg* KO mice or *pcdha* KO mice and their littermate WT controls. (C) Connectivity probability among vertically or horizontally aligned neurons in layer 2/3 of the barrel cortex in *pcdhg* KO mice and their littermate WT controls. (D) Connectivity probability among vertically aligned pyramidal cells along with the distance between recorded pairs in *pcdhg* KO mice or WT mice. (E) Developmental profiling of the connectivity probability among vertically aligned neurons in *pcdhg* KO mice or WT mice. Chi-square tests were used in B-E to calculate the statistical difference.

50/208 in *Pcdhg* KO mice vs. 9.62%, 20/208 in WT mice for the distance range 50-100μm; 11.4%, 12/105 in *Pcdhg* KO mice vs. 5.59%, 10/179 in WT mice for the distance range 100-150μm (Fig. 2D). Furthermore, we found that the absence of cPCDH γ increased synapse formation started from P10 (16.1%, 32/199 in *Pcdhg* KO mice vs. 8.70%, 20/230 in WT mice for P10-12; 21.6%, 22/102 in *Pcdhg* KO mice vs. 12.8%, 29/227 in WT mice for P13-15), when chemical synapses between cortical neurons become detectable, and such tendency appeared to be maintained for as long as measurements were performed (up to P20) (Fig. 2E). Thus, cPCDH γ may play a role in preventing synapse formation during early neural development.

Diversified expression of cPCDH γ s regulates synaptic specificity in the mouse neocortex

To further detect the effect of diversified cPCDH γ s on synapse specificity, we overexpressed one or multiple cPCDH γ isoforms tagged with fluorescent proteins through *in utero* electroporation and then performed multiple whole-cell recording to address the effect on pair-wise synaptic connectivity. Since the localization of neurons plays a critical role in their synaptic connection(He et al. 2015), we first examined the effect of overexpressing cPCDH γ on this matter and found that overexpressing cPCDH γ isoforms did not affect neuronal localization in the postnatal neocortex (Fig. 3-S1). To estimate how many isoforms were expressed in a given individual neuron when genes for multiple isoforms were electroporated simultaneously, we tagged five isoforms with mNeongreen and one with mCherry (Fig. 3-S2A). Based on the rates of yellow and red-only cells in the total electroporated neuron population (Fig.3-S2B), the probability analysis revealed that each positive neuron expressed an average of 5.6 types of isoforms (Fig. 3-S2C). This estimate was further supported by data from single-cell RT-PCR analysis in which an average of 5.3 types of isoforms was detected out of 6 electroporated plasmids from 19 neurons (Fig. 3-S2D). The result from single-cell RT-PCR also suggested that electroporation-introduced isoforms were dominant ones in these neurons (Fig. 3-S2D). Multiple whole-cell recordings revealed that overexpression of 1 or 6 cPCDH γ isoforms in neurons significantly reduced the connectivity probability among them from the control level 10.3% (15/146) to 1.85%

151 (4/216) and 4.42% (8/181), respectively (Fig. 3A). The impact of the diversified
 152 expression pattern for cPCDH γ isoforms on the synapse formation was further
 153 examined by sequential *in utero* electroporation at E14.5 and E15.5. We electroporated
 154 the same or different combinations of cPCDH γ isoforms, with plasmids tagged with
 155 different fluorescent proteins mNeongreen and mRuby3(Bajar et al. 2016) at different
 156 days (Fig. 3B). The sequential-electroporation allowed us to manipulate the similarity
 157 level in cPCDH γ isoform combinations among neurons. We set the similarity level of
 158 cPCDH γ isoforms electroporated at E14.5 and E15.5 to ranges from 0% (“no overlap”)
 159 to 100% (“complete overlap”) (Fig. 3C). Whole-cell recordings were made on acute
 160 brain slices containing both green and red cells from mice at P10-P14. Each set of
 161 recordings included at least one mNeongreen⁺, one mRuby3⁺, and one nearby non-
 162 fluorescent control neuron (Fig. 3D). We found that neurons electroporated with the
 163 same batch of isoforms exhibited significantly lower connectivity (Fig. 3E, the left
 164 panel), consistent with the result obtained from recordings after the single
 165 electroporation (Fig. 3A). The connectivity rate between neurons expressing the same
 166 combination of isoforms was also significantly lower for neuronal pairs electroporated
 167 at different days (100% group in the middle panel of Fig. 3E). However, the
 168 connectivity level progressively increased to the control level between neurons
 169 electroporated with partially overlapped or completely different cPCDH γ isoforms (no
 170 overlap groups). The connectivity probability dropped from 12.5% (19/151) at the
 171 control level to 3.7% (5/134) for neuronal pairs with the similarity level of 100%. The
 172 probability recovered to 8.4% (14/165) for the pairs with the similarity level of 33%,

Figure 3: Diversified expression of cPCDH γ isoforms is critical for synapse formation in cortical neurons.

(A) The effect of overexpressing cPCDH γ on the synaptic connectivity probability after a single *in utero* electroporation. (B) Diagram of a sequential *in utero* electroporation at E14.5 and E15.5. (C) The overexpressed isoforms in different sets of experiments. S.I., Similarity level; EP, Electroporation; P., plasmids mixture; s./d., same/different isoforms in two electroporations. (D) A sample image of recorded neurons after two times of electroporation at E14.5 with plasmids carrying mNeogreen and E15.5 with plasmids carrying mRuby3. In this sample, cells 3 and 4 are mNeogreen positive, cell 6 is mRuby3 positive, and cells 1, 2, and 5 are without fluorescence. Neurobiotin was added into the internal solution to indicate recorded neurons. (E) The effect of overexpressing cPCDH γ s on the synaptic connectivity probability after a sequential *in utero* electroporation. Left panel: the connectivity probability for neuron pairs overexpressed same cPCDH γ s combinations (with same fluorescence); middle panel: the connectivity probability for neuron pairs overexpressed different batch of cPCDH γ s isoforms (with different fluorescence); right panel: the connectivity probability for neuron pairs contained one overexpressed cell and one adjacent cell without fluorescence. Chi-square test and false discovery rate (FDR) correction were used to determine the statistical difference in A and E. $0.01 < p < 0.05$ labeled as “*”; $0.001 < p < 0.01$ as “**”; $p < 0.001$ as “***”. (F) The correlation between the similarity level of overexpressed cPCDH γ s combination and the probability of synaptic connection rate (The source data were the same as the middle panel of E).

173 and to 10.3% (12/117) and 10.5% (12/114) for the pairs with the similarity level of 11%

or 0%, respectively (Fig. 3E, the middle panel). Similar to that found in the single electroporation experiment (Fig. 3A), there was no significant difference in the synaptic connectivity between electroporated and nearby control neurons (Fig. 3E, the right panel). Overall, the probability of forming synaptic connection negatively correlated with the similarity level of cPCDH γ isoforms expressed in the neuron pair (Fig. 3E, F). Our findings strongly demonstrated that diversified expression of cPCDH γ isoforms was critical for synapse formation between nearby pyramidal cells, and more similar cPCDH γ isoforms patterns expressed in neurons resulted in a lower probability of forming synapses between them.

Discussion

Homophilic proteins cPCDHs are strong candidates for promoting synaptic specificity due to their combinatorial and stochastic expression pattern (Yagi 2012, Kohmura et al. 1998, Toyoda et al. 2014). Our 5'-end single-cell sequencing data revealed the diversified expression pattern of cPCDH γ isoforms in neocortical neurons. We further demonstrated that this diversity was critical for the specificity of synapse formation at the level of self-nonself discrimination. This conclusion was based on three lines of evidence. First, the absence of cPCDH γ significantly elevated functional connectivity between adjacent neocortical neurons. Second, electroporation-induced overexpression of the same cPCDH γ isoforms in developing neurons markedly decreased the connectivity among them. Finally and more importantly, using sequential expression of a different group of isoforms, we found that decreasing the similarity level of cPCDH γ

isoforms expressed on the neuron elevated the connectivity among them. The findings strongly suggest that cPCDH γ s regulate the specificity of synapse formation at the first level for recognizing cells expressing similar combinations, but not at the second level for choosing specific targets to make synapses. It remains to be studied whether the diversified patterns of cPCDH γ isoforms expressed in different neurons allow for additional coding functions for neurons beyond their homophilic interaction. Meanwhile, since there is a weak but significant co-occurrence for the expression of cPCDH γ isoforms, the existence of cPCDH γ might prevent a certain degree of synapse formation in the neocortex.

In summary, our data demonstrated that the similarity level of cPCDH γ isoforms between neurons was critical for their synapse formation, more similar cPCDH γ isoforms patterns expressed in neurons resulting in a lower probability of forming synapses between them. These findings strongly suggest that the existence of cPCDH γ s allows a neuron to choose which neurons not to form synapse, rather than choosing which neurons to make synapses. Whether there are specific attractive forces between cells to promote synaptic specificity is still an open question.

Materials and methods

Mice

All mice were handled according to the protocol approved by the Animal Care and Use Committee of the Center for Excellence in Brain Science and Intelligence Technology/ Institute of Neuroscience, Chinese Academy of Sciences. All the mice are group-housed. *Pcdhg flox/flox*, *Nex-cre* mice are maintained as heterozygotes at the background of C57BL6. F2 mice after the cross of these two lines contained *Pcdhg flox/flox/Nex-cre*, and their littermates as controls were used for experiments. ICR mice ordered from Vital River Laboratories were used in the overexpression experiments.

Single cell dissociation

An ICR mouse at p11 were perfused with dissociation-artificial cerebrospinal fluid (aCSF) (choline-aCSF with 50 μ M DL-APV (Abcam/ab120271), 10 μ M DNQX (Sigma/D0540), bubbled thoroughly with 95% O₂/5% CO₂ before use.). The brain was stripped out immediately on ice and then immersed in 0°C dissociation-aCSF. Under the protection of buffer, a half neocortex was stripped off and cut into ~100 μ m diameter small chunks. The chunks were transferred into the digestion buffer (dissociation-aCSF contains 100 units papain and 500 units DNase I) and incubated at 37°C for 60 min. During the incubation, 95% O₂/5% CO₂ was gently puffed above the buffer. Afterward, the sample was centrifuged for 5 min at 300 g at room temperature. The precipitate was collected and resuspended with 2 ml Buffer II (dissociation-aCSF contains 0.05% BSA). We gently triturated the sample sequentially with 600, 400, and 200 μ m fire-polished Pasteur pipettes until no large visible chunks. Then we used a debris removal solution (130-109-398, Miltenyi Biotec) to remove the debris according to the manufactory's protocol. The sample was resuspended with 500 μ l Buffer II containing 2.5 μ g DAPI and filtered with a 40 μ m cell strainer before FACS sorting. FACS sorting was used to sort the alive cells according to the 350nm fluorescence. The living rate was higher than 85%. The cells were collected into Buffer II by FACS and centrifuged at 0°C 500g for single-cell suspension (Fig. 1-S1A).

Library construction for 5' end single-cell sequencing

We used Chromium Single Cell V(D)J Enrichment Kit, Mouse T Cell, PN-1000071 Kit (10X

Genomics) to construct the *pcdhg* library. The Kit was originally designed for sequencing TCR in immune cells. We used specific primers for *pcdhg* to sequence 5' end of different isoforms of this cluster gene.

The nanoliter-scale GEMs (Gel Beads-in-emulsion) were generated with Chromium Next GEM Chip G. The procedures of reverse transcription, template switching, barcoding, and transcript extension occurred inside GEMs (Fig. 1-S1B). GEMs were broken for cDNA amplification (Fig. 1-S1B). After cDNA amplification, the sample was split into two parts, one for the expression library construction of other genes at the 5' end, the other for *pcdhg* enriched library construction (Fig. 1-S1C). For the *pcdhg* enriched library construction, we replaced the original Reverse Outer & Inner Primers in Mouse T Cell Mix 1 and 2 with the *pcdhg* nested-PCR primers mix (target to the constant region). After enriching *pcdhg*, the rest of the procedures followed the manufactory's protocol (Fig. 1-S1C1). For the 5' gene expression library construction, the amplified cDNA was immediately used in the rest procedure according to the kit manual (Fig. 1-S1C2).

The two libraries were sequenced respectively and intersected based on the cell barcode (Fig. 1-S1D).

Sequencing data profiling

We acquired 212.66 GB clean reads for the 5' gene expression library and 183.30 GB for the 5' *pcdhg* expression library from the sequencing platform NovaSeq 6000 (Illumina, Novogene). 5' gene expression sequencing data were aligned by Cellranger v3.0.2 (10X Genomics). There were 23753 cells from the output of Cellranger. The mean reads per cell were 29843, and the sequencing saturation was 51.1%.

Cells were sorted by the range of 7000 > feature RNA > 1800 & mitochondrial RNA < 7%. After the cleanup, the cell number became 20419. Next, we used doubletfinder (McGinnis, Murrow, and Gartner 2019) to remove 2981 potential doublets (15%) and further clustered the rest cells by Seurat2 (resolution=2.4, optimized by Clustree algorithm, Fig. 1-S1D). After clustering, the identity of each cluster was determined by marker genes and selected genes (*Slc17a7*, *Neurod2*, *Neurod6*, *Slc32a1*, *Gad1*, *Satb2* for neurons and *Sox2*, *Car2*, *Ascl1*, *Aldh1l1*, *Pdgfra*, *Pdgfrb*, *Olig1*, *C1qa*, *Dlx2* for non-neurons). Neurons and other types of cells could be separated. Neurons were further clustered into 15 clusters (cluster 0-14) (resolution=0.4, optimized by Clustree algorithm). Cluster12

was classified as atypical neurons due to their high expression of *Clqa*. Therefore, it was not included in the following analysis.

We aligned 5' *pcdhg* expression sequencing data according to Bowtie(Langmead et al. 2009) by using the reference sequence of *pcdhg* variable axons of all isoforms. We further detected cell barcodes and UMI sequences from the aligned reads with the custom script written by Matlab (R2018b, Mathworks). The *pcdhg* expression matrix was abstracted from the intersection of data from the 5' gene expression library and *pcdhg* enrichment library. In the output matrix, the *pcdhg* expression level was indicated by the number of UMI. The *pcdhg* expression matrix was filtered by UMI>1 and binarized used in the following analysis (Fig. 1-S1D). Different thresholds of UMI number for total *pcdhg* were applied to see the fraction of cells for all isoforms in individual cells. UMI>10 was chosen for the similarity level analysis in 2834 neurons to ensure the quality of analysis.

Similarity level

Similarity level was calculated as $\frac{A \cap B}{A \cup B}$. In which, A and B indicate the numbers of PCDHγ isoforms (except constant isoforms C3, C4, and C5) expressed in two compared cells; $A \cap B$ represents the number of isoforms expressed in both cells and $A \cup B$ equals to the total number of isoforms expressed in these two cells.

Co-occurrence

We generated an expected matrix by shuffling 100 times from different clusters (or total population) based on the hypothesis that all the isoforms expressed randomly following their expression frequencies. Then the fraction of cells with various numbers of isoform types from both the observed expression matrix (observed data) and the expected matrix (shuffled data) in the whole population were plotted. The variance of this distribution was further computed for cells in different clusters or the total population(Wada, Wallerich, and Becskei 2018). The expected variance was determined by the following function with the expression possibility of each isoform (p_i):

$$\sigma_{expected}^2 = \sum_{i=1}^{N_{isoforms}} p_i(1 - p_i)$$

The statistical difference between the observed variance and the expected one was then measured

by the Z-test. If the observed variance is higher than expected, the genes are determined to be co-occurrent.

Generation of *PCDHa* cKO mice

Pcdha conditional knockout (cKO) mice were generated by Biocytogen Pharmaceuticals (Beijing) based on the CRISPR/Cas9 approach. The candidate sgRNAs corresponding to the parts of intron 1 and intron 2 of *Pcdha1* were designed by the CRISPR design tool (<http://crispr.mit.edu>) and screened for their on-target activity with the Universal CRISPR Activity Assay kit (UCATM, Biocytogen Pharmaceuticals (Beijing) Co., Ltd). A circular donor vector was employed to minimize random integrations. The gene-targeting template containing 5' homologous arm, target fragment (exon2), and 3' homologous arm was utilized to repair the double-strand breaks generated by Cas9/sgRNA. Cas9 mRNA, the template, and sgRNAs were co-injected into the cytoplasm of the one-cell stage fertilized C57BL/6 egg. Afterward, injected eggs were implanted into oviducts of Kunming pseudopregnant females to produce F0 mice. The F0 mice with the correct genotype were crossed with C57BL/6 mice to establish germline-transmitted F1 heterozygous mice. The PCR-positive F1 mice were further confirmed with Southern blot and DNA sequencing. Exon 2 (ENSMUSE00001345888) of *Pcdha* genes will be removed with the help of Cre, which results in an 813 aa (796aa native EGE-LZX-015, 17aa in-frame shifted) truncated protein that may be subject to non-sense mediated decay.

Acute slice preparation

The brain of P9-P14 mice was gently stripped out immediately on ice after the sacrifice. The brain was coronally sectioned with a vibratome (VT1200S, Leica) into 300µm-thickness slices in the choline modified artificial cerebrospinal fluid (choline-aCSF containing 120mM choline chloride, 2.6mM KCl, 26mM NaHCO₃, 1.25mM NaH₂PO₄, 15mM D-glucose, 1.3mM ascorbate acid, 0.5mM CaCl₂ and 7mM MgCl₂, 300-310 mOsm). The slices were quickly transferred into a chamber with normal aCSF (containing 126mM NaCl, 3mM KCl, 26mM NaHCO₃, 1.2 Mm NaH₂PO₄, 10mM D-Glucose, 2.4mM CaCl₂ and 1.3mM MgCl₂, 300-310 mOsm) and incubated at 32°C for 30min and then maintained at room temperature before recording.

Mice aged over P14 was anesthetized first with pentobarbital sodium (intraperitoneal injection, 50-

90mg/kg) and then perfused with 0°C NMDG-aCSF (containing 93mM N-Methyl-D-glucamine, 93mM HCl, 2.5mM KCl, 1.25mM NaH₂PO₄, 30mM NaHCO₃, 25mM D-glucose, 20mM HEPES, 5mM Na-ascorbate, 3mM Na-pyruvate, 2mM Thiourea, 10mM MgSO₄, 0.5mM CaCl₂ and 12mM N-Acetyl-L-cysteine, 300-310 mOsm). The brain was coronally sectioned in the NMDG-aCSF with the method described above. The slices were transferred into a chamber containing 37°C NMDG-aCSF for 12-15min and then kept in normal aCSF at room temperature for 1 hour before recording. All the solutions described above were constantly oxygenated with 95% O₂/5% CO₂.

Electrophysiology

The slice was constantly perfused with 30°C normal aCSF oxygenated by 95% O₂/5% CO₂ during the recording. A differential interference contrast microscope (BX51W1, Olympus) with a 5X/NA 0.1 objective and a 40X/NA 0.8 W objective was used to visualize cells in slice. The glass pipettes were pulled by a Flaming/Brown micropipette puller (P-97, Sutter instrument), and the pipettes with resistance at 8-12MΩ were used. The internal pipettes solution contains 126mM potassium gluconate, 2mM KCl, 2mM MgCl₂, 10mM HEPES, 0.2mM EGTA, 4mM Na₂-ATP, 0.4mM Na₃-GTP, 10mM creatine phosphate and 10ng/ml Alexa Fluor 405 Cadaverine). Up to 6-channel, whole-cell patch-clamp recording was performed with the amplifier (Multiclamp 700B, Molecular Devices) and Digidata (Digidata 1550A, Molecular Devices) and analyzed with pClamp9 (Molecular Devices). The electrophysiological data were low-pass Bessel filtered at 2kHz and digitized at 20kHz. The feedback resistor was set at 5 GΩ. All the recordings had a series resistance of less than 50MΩ. We assessed neuronal connectivity by injecting current to one of all the recorded cells under the current-clamp mode to evoke action potentials and testing for evoked EPSC in other cells under the voltage-clamp mode at -70mV. All the neuron pairs in each recording were tested at least 10 times in both directions.

***In utero* electroporation**

The date with plug detection was considered embryonic day 0.5 (E0.5) for pregnant ICR mice. The pregnant mice at the E14.5 or E15.5 gestation stage were initially anesthetized with 5% isoflurane and maintained with 1.5% isoflurane (0.8-1L/min). Uterine horns were exposed after the cut. The plasmid mixture with the total concentration of 5mg/ml plasmid (equal molar mixture if there are

multiple plasmids) containing 1mg/ml Fast green (Sigma) was injected into the embryonic lateral ventricle through a glass pipette. After injection, the utero was clasped with electrodes (7mm diameter, 45-0118, Tweezertrodes) and given five 30V pulses of 50 ms with 1s intervals.

In the sequential *in utero* electroporation experiment, the surgeries were performed at both E14.5 and E15.5. We injected the plasmids into the same lateral ventricle and delivered electroporation at a similar angle. Lincomycin hydrochloride and lidocaine hydrochloride gel were applied to the wound after stitching.

Estimating the numbers of overexpressed *pcdh* isoform

Five plasmids with mNeongreen each containing one of the following isoforms *pcdhga2*, *pcdhga8*, *pcdhga10*, *pcdhgb1*, and *pcdhgb2*, as well as the sixth plasmid pCAGGS-*pcdhga6*-p2A-mCherry were equimolarly mixed and used for the electroporation at E15. Mice were sacrificed at P7. The brains were coronally sectioned on a vibratome at 60μm thickness. The mCherry signal was enhanced by using an anti-RFP antibody (Rockland 600-401-379). The images were taken by a confocal microscope (Nikon C2, 20X, NA 0.75, Fig. 3-S2A).

We counted green/red/yellow cells and calculated their percentages, respectively (Fig. 3-S2B). We first simulated under the condition in which the number of plasmids electroporated into neurons was identical. However, the result derived from this assumption couldn't fit with the observed data. Thus, we hypothesized that the number of plasmids electroporated into neurons would follow a normal distribution. To fit the percentages of "red in total" and "red only" in the total electroporated population, we simulated hundreds of normal distributions of plasmids within 10000 cells. We made simulations by enumerating mean and standard error with integers and found that the mean of 18 and the standard error of 6 fitted the data well (Fig. 3-S2C). Under this distribution, each neuron took up about 18 plasmids which contained ~5.6 isoforms.

Single cell RT-PCR

We used the SuperScript II Kit (18064014, Invitrogen) for single-cell RT-PCR. The neuron was aspirated from an acute brain slice by glass pipette with resistance at 3-4 MΩ prefilled with ~1μL aCSF. Cell contents were then gently blown into a PCR tube containing 0.5μL Rnaseout (10777019, Invitrogen). Afterward, the sample was chilled immediately with liquid nitrogen. RT

mix1 (~1μL sample, 0.5μL Rnaseout, 10mM dNTP, 1μL primer mix (GR1 and β-actin R, 2.5μM), 1.5μL ddH₂O) was added into the sample tube and kept at 65°C for 5min, and then quickly chilled on ice. Then the RT mix2 (1μL 10X RT buffer, 2μL 25mM MgCl₂, 1μL 0.1M DTT, 0.5μL Rnaseout, and 0.5μL SuperScript II) was added to sample/mix1 for the reverse transcription which occurred in 50°C for 60min followed by 85°C for 5min.

LATaq Kit (RR002B, TaKaRa) was used in nested PCR. The primer mix containing forward primers Gb F1a, Gb F1b, Ga F1a, Ga F1b, Ga F1c, Ga F1d, and reverse primer GR1 was used for the first round of nested PCR. Another primer mix containing forward primer corresponding to each isoform and reverse primer GR2 was used for the second round of nested PCR. The sequences of each primer were listed in the following table. Due to the high sensitivity after two rounds of PCR, false-positive signals showed up in some negative controls with very faint bands (indicated by red stars in Fig. 3-S2D). A band with at least 5 times more intensity than the false-positive band will be considered a real positive signal.

Primer list

Primer name	Sequence (5'-3')
Gb F1a	GACACAATGCCTGGCTGTC
Gb F1b	GACACAATGCCTGGCTATC
GR1	tcgtttgccagcgccattg
G b1 f2	CACATTCTACGACTATGGCAGG
G b2 f2	ACCAGGTACTCTTGGAGACAC
G b4 f2	CTACAATTTGCAGATTCAGCG
G b5 f2	ATGGTGCACATGCTGGAAAGAC
G b6 f2	CGTTGGGCAAACGGGAAAGG
G b7 f2	GCCGGGAGATCAACTTAAAC
G b8 f2	GAGTGTCGCTGAGGAGAATA
GR2	attggtcagcgtggcattgct
Ga F1a	CCTCTCTCCGCCACTGTCA
Ga F1b	TCTACCTGGTGGTGGCAGT

Ga F1c	TGGTGGCAGTGGACAGAGA
Ga F1d	TGGTGGCAGTGGACAAAGA
G a1 f2	CAGGAATTTTTGTCAGCACC
G a2 f2	CTGATTTCTCTCAGCACCTC
G a3 f2	ACGAAAGAAGACCCACGCTG
G a4 f2	ATAAAAGGAGACTCCAGTCTGCAG
G a5 f2	CACAAAGAAGAGCCCGGAGATG
G a6 f2	GAAAGGTGTGTCAAATATGTCC
G a7 f2	TTCAGGTGGTGGCCTGGAAGAT
G a8 f2	GATGAAGATGCTTGCGCTCCG
G a9 f2	GCAGTTCAGTTGAGCCCAAGT
G a10 f2	CAAGTGTCTGTAGAAGACG
G a11 f2	GACTCAACTACAGTTTTAGGC
G a12 f2	GGCAGATCTAGACAATCTAG
β-actin F	GTCGTACCACAGGCATTGTGATGG
β-actin R1	ctagaagcacttgcggtgc
β-actin R2	gcaatgcctgggtacatggtgg
Outer R1	gtaaactggggtccgtatc
Inner R1	cattttgggatccgctcgt

415

416

417

418

419

420

421

Acknowledgements

We thank Dr. Yifeng Zhang for providing the *Pcdhg flox/flox* mouse line and Dr. Zilong Qiu for *Nex-cre* mouse line, Dr. Jun Chu for sharing plasmids with mNeongreen and mRuby3, and other members in Xu lab for their discussions and technique supports. We are grateful to Prof. Mu-ming Poo and Song-hai Shi for critical reading of the manuscript. This work was supported by grants from the Training Program of the Major Research Plan of the National Natural Science Foundation of China, grant No. 91632101; Strategic Priority Research Program of the Chinese Academy of Sciences, grant No. XDB32010100; National Natural Science Foundation of China project 31671113; Shanghai Municipal Science and Technology Major Project, grant No. 2018SHZDZX05 and the State Key Laboratory of Neuroscience.

Author contributions

H. -T. X. conceived the project. Y. Z. performed most recordings and single-cell analysis. C. D. performed part of recordings. Y. W. prepared PCDH plasmids. Y. Z. and H. -T. X. wrote the manuscript.

References

- Bajar, B. T., E. S. Wang, A. J. Lam, B. B. Kim, C. L. Jacobs, E. S. Howe, M. W. Davidson, M. Z. Lin, and J. Chu. 2016. "Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting." *Sci Rep* 6:20889. doi: 10.1038/srep20889.
- Berns, D. S., L. A. DeNardo, D. T. Pederick, and L. Luo. 2018. "Teneurin-3 controls topographic circuit assembly in the hippocampus." *Nature* 554 (7692):328-333. doi: 10.1038/nature25463.
- Brasch, J., K. M. Goodman, A. J. Noble, M. Rapp, S. Manneppalli, F. Bahna, V. P. Dandey, T. Bepler, B. Berger, T. Maniatis, C. S. Potter, B. Carragher, B. Honig, and L. Shapiro. 2019. "Visualization of clustered protocadherin neuronal self-recognition complexes." *Nature* 569 (7755):280-283. doi: 10.1038/s41586-019-1089-3.
- Chen, B., K. Brinkmann, Z. Chen, C. W. Pak, Y. Liao, S. Shi, L. Henry, N. V. Grishin, S. Bogdan, and M. K.

449 Rosen. 2014. "The WAVE regulatory complex links diverse receptors to the actin cytoskeleton."
450 *Cell* 156 (1-2):195-207. doi: 10.1016/j.cell.2013.11.048.

451 Courgeon, M., and C. Desplan. 2019. "Coordination between stochastic and deterministic specification
452 in the Drosophila visual system." *Science* 366 (6463). doi: 10.1126/science.aay6727.

453 Duan, X., A. Krishnaswamy, I. De la Huerta, and J. R. Sanes. 2014. "Type II cadherins guide assembly of a
454 direction-selective retinal circuit." *Cell* 158 (4):793-807. doi: 10.1016/j.cell.2014.06.047.

455 Esumi, S., N. Kakazu, Y. Taguchi, T. Hirayama, A. Sasaki, T. Hirabayashi, T. Koide, T. Kitsukawa, S. Hamada,
456 and T. Yagi. 2005. "Monoallelic yet combinatorial expression of variable exons of the
457 protocadherin-alpha gene cluster in single neurons." *Nat Genet* 37 (2):171-6. doi:
458 10.1038/ng1500.

459 Garrett, A. M., D. Schreiner, M. A. Lobas, and J. A. Weiner. 2012. "gamma-protocadherins control cortical
460 dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway."
461 *Neuron* 74 (2):269-76. doi: 10.1016/j.neuron.2012.01.028.

462 Goebbels, S., I. Bormuth, U. Bode, O. Hermanson, M. H. Schwab, and K. A. Nave. 2006. "Genetic
463 targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice." *Genesis* 44
464 (12):611-21. doi: 10.1002/dvg.20256.

465 He, S., Z. Li, S. Ge, Y. C. Yu, and S. H. Shi. 2015. "Inside-Out Radial Migration Facilitates Lineage-
466 Dependent Neocortical Microcircuit Assembly." *Neuron* 86 (5):1159-66. doi:
467 10.1016/j.neuron.2015.05.002.

468 Hirayama, T., E. Tarusawa, Y. Yoshimura, N. Galjart, and T. Yagi. 2012. "CTCF is required for neural
469 development and stochastic expression of clustered Pcdh genes in neurons." *Cell Rep* 2
470 (2):345-57. doi: 10.1016/j.celrep.2012.06.014.

471 Honig, B., and L. Shapiro. 2020. "Adhesion Protein Structure, Molecular Affinities, and Principles of Cell-
472 Cell Recognition." *Cell* 181 (3):520-535. doi: 10.1016/j.cell.2020.04.010.

473 Jang, S., H. Lee, and E. Kim. 2017. "Synaptic adhesion molecules and excitatory synaptic transmission."
474 *Curr Opin Neurobiol* 45:45-50. doi: 10.1016/j.conb.2017.03.005.

475 Kaneko, R., H. Kato, Y. Kawamura, S. Esumi, T. Hirayama, T. Hirabayashi, and T. Yagi. 2006. "Allelic gene
476 regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic
477 expression in single Purkinje cells." *J Biol Chem* 281 (41):30551-60. doi:
478 10.1074/jbc.M605677200.

479 Kohmura, N., K. Senzaki, S. Hamada, N. Kai, R. Yasuda, M. Watanabe, H. Ishii, M. Yasuda, M. Mishina,
480 and T. Yagi. 1998. "Diversity revealed by a novel family of cadherins expressed in neurons at a
481 synaptic complex." *Neuron* 20 (6):1137-51. doi: 10.1016/s0896-6273(00)80495-x.

482 Kostadinov, D., and J. R. Sanes. 2015. "Protocadherin-dependent dendritic self-avoidance regulates
483 neural connectivity and circuit function." *Elife* 4. doi: 10.7554/eLife.08964.

484 Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. "Ultrafast and memory-efficient alignment
485 of short DNA sequences to the human genome." *Genome Biol* 10 (3):R25. doi: 10.1186/gb-
486 2009-10-3-r25.

487 Lefebvre, J. L., Y. Zhang, M. Meister, X. Wang, and J. R. Sanes. 2008. "gamma-Protocadherins regulate
488 neuronal survival but are dispensable for circuit formation in retina." *Development* 135
489 (24):4141-51. doi: 10.1242/dev.027912.

490 McGinnis, C. S., L. M. Murrow, and Z. J. Gartner. 2019. "DoubletFinder: Doublet Detection in Single-Cell
491 RNA Sequencing Data Using Artificial Nearest Neighbors." *Cell Syst* 8 (4):329-337 e4. doi:
492 10.1016/j.cels.2019.03.003.

- Mountoufaris, G., W. V. Chen, Y. Hirabayashi, S. O'Keeffe, M. Chevee, C. L. Nwakeze, F. Polleux, and T. Maniatis. 2017. "Multiclust er Pcdh diversity is required for mouse olfactory neural circuit assembly." *Science* 356 (6336):411-414. doi: 10.1126/science.aai8801.
- Rawson, R. L., E. A. Martin, and M. E. Williams. 2017. "Mechanisms of input and output synaptic specificity: finding partners, building synapses, and fine-tuning communication." *Curr Opin Neurobiol* 45:39-44. doi: 10.1016/j.conb.2017.03.006.
- Rubinstein, R., C. A. Thu, K. M. Goodman, H. N. Wolcott, F. Bahna, S. Mannepalli, G. Ahlsen, M. Chevee, A. Halim, H. Clausen, T. Maniatis, L. Shapiro, and B. Honig. 2015. "Molecular logic of neuronal self-recognition through protocadherin domain interactions." *Cell* 163 (3):629-42. doi: 10.1016/j.cell.2015.09.026.
- Sanes, J. R., and S. L. Zipursky. 2020. "Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits." *Cell* 181 (3):536-556. doi: 10.1016/j.cell.2020.04.008.
- Serizawa, S., K. Miyamichi, H. Takeuchi, Y. Yamagishi, M. Suzuki, and H. Sakano. 2006. "A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting." *Cell* 127 (5):1057-69. doi: 10.1016/j.cell.2006.10.031.
- Sytnyk, V., I. Leshchyn's'ka, and M. Schachner. 2017. "Neural Cell Adhesion Molecules of the Immunoglobulin Superfamily Regulate Synapse Formation, Maintenance, and Function." *Trends Neurosci* 40 (5):295-308. doi: 10.1016/j.tins.2017.03.003.
- Tan, L., K. X. Zhang, M. Y. Pecot, S. Nagarkar-Jaiswal, P. T. Lee, S. Y. Takemura, J. M. McEwen, A. Nern, S. Xu, W. Tadros, Z. Chen, K. Zinn, H. J. Bellen, M. Morey, and S. L. Zipursky. 2015. "Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila." *Cell* 163 (7):1756-69. doi: 10.1016/j.cell.2015.11.021.
- Tarusawa, E., M. Sanbo, A. Okayama, T. Miyashita, T. Kitsukawa, T. Hirayama, T. Hirabayashi, S. Hasegawa, R. Kaneko, S. Toyoda, T. Kobayashi, M. Kato-Itoh, H. Nakauchi, M. Hirabayashi, T. Yagi, and Y. Yoshimura. 2016. "Establishment of high reciprocal connectivity between clonal cortical neurons is regulated by the Dnmt3b DNA methyltransferase and clustered protocadherins." *BMC Biol* 14 (1):103. doi: 10.1186/s12915-016-0326-6.
- Toyoda, S., M. Kawaguchi, T. Kobayashi, E. Tarusawa, T. Toyama, M. Okano, M. Oda, H. Nakauchi, Y. Yoshimura, M. Sanbo, M. Hirabayashi, T. Hirayama, T. Hirabayashi, and T. Yagi. 2014. "Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity." *Neuron* 82 (1):94-108. doi: 10.1016/j.neuron.2014.02.005.
- Wada, T., S. Wallerich, and A. Becskei. 2018. "Stochastic Gene Choice during Cellular Differentiation." *Cell Rep* 24 (13):3503-3512. doi: 10.1016/j.celrep.2018.08.074.
- Wu, Q., and T. Maniatis. 1999. "A striking organization of a large family of human neural cadherin-like cell adhesion genes." *Cell* 97 (6):779-90. doi: 10.1016/s0092-8674(00)80789-8.
- Yagi, T. 2012. "Molecular codes for neuronal individuality and cell assembly in the brain." *Front Mol Neurosci* 5:45. doi: 10.3389/fnmol.2012.00045.
- Yogev, S., and K. Shen. 2014. "Cellular and molecular mechanisms of synaptic specificity." *Annu Rev Cell Dev Biol* 30:417-37. doi: 10.1146/annurev-cellbio-100913-012953.

Figure 1-Figure supplement 1: 5'-end single-cell sequencing for cPCDHys in neocortical neurons.

(A-D) Schematic of 5' end single-cell sequencing protocol. The details were described in the methods part.

(E) Fraction of cells expressing different *pcdhg* isoforms in neocortical neurons after the different UMI cutoff 1 (top) or 20 (bottom).

Figure 1-Figure supplement 2: A weak but significant co-occurrence of cPCDH γ isoforms in most of neocortical neurons (except neurons in cluster 1).

Fraction distribution of the number of the expressed isoforms per cell for neurons in different clusters.

Figure 3-Figure supplement 1: Overexpressing cPCDH γ did not affect neuronal localization in the neocortex.

(A) Sample confocal images for the distribution of neurons overexpressing PCDH γ A2 through electroporation at different embryonic stages; the bottom panels demonstrate the numbers of neurons in different evenly distributed zones of the neocortex along the pial surface to the start of white matter.

(B) The distribution of neurons after a sequential-electroporation with four plasmids PCDH γ A4, A9, A11, or B6 tagged by mNeogreen at E14.5 and other four PCDH γ A2, A6, A8, or B1 tagged by mRuby3 at E15.5; the bottom panel shows the distribution of cell numbers in different zones (Zone 1 starts on the top of the pial surface and zone 10 ends at the border with the white matter).

Figure 3-Figure supplement 2: Probability analysis for the number of overexpressed PCDHy isoforms after electroporating multiple plasmids.

(A) A sample confocal image of electroporated neurons with five different isoforms tagged with mNeongreen, and the sixth isoform with mCherry. Scale bar = 100um.

(B) The numbers of cells and corresponding probabilities in total population after one electroporation containing five different isoforms with green and the sixth isoform with mCherry.

(C) The normal (top) and logarithmic (bottom) probability distribution of plasmid numbers in electroporated neurons after the simulation of 10000 cells. The top panel is to show the correlation between the plasmid number and the probability of “red in total” (blue line), while the bottom panel is to show the correlation between the plasmid number and the likelihood of “red only” (red line).

(D) Single-cell RT-PCR for PCDHy isoforms after electroporation. Different PCDHy isoforms were amplified by their corresponding primers. Green numbers indicated neurons with fluorescence, and black ones represented neurons without fluorescence. N1-N6 were PCR negative controls. Isoforms used in electroporation were written in green color. Red stars indicate faint bands in negative controls.