CRISPR/Cas9 interrogation of the mouse *Pcdhg* gene cluster reveals a crucial isoform-specific role for *Pcdhgc4*

Andrew M. Garrett[1, 3,*], Peter J. Bosch[2], David M. Steffen[2], Leah C. Fuller[2], Charles G. Marcucci[2], Alexis A. Koch[1], Preeti Bais[3], Joshua A. Weiner[2,*], and Robert W. Burgess[3,*]

1 Wayne State University, Department of Pharmacology, Department of Ophthalmology, Visual, and Anatomical Sciences, Detroit, MI.

2 University of Iowa, Department of Biology and Iowa Neuroscience Institute, Iowa City, IA

3 The Jackson Laboratory, Bar Harbor, ME

* Co-corresponding authors

ABSTRACT

The mammalian *Pcdhg* gene cluster encodes a family of 22 cell adhesion molecules, the gamma-Protocadherins (γ-Pcdhs), critical for neuronal survival and neural circuit formation. The extent to which isoform diversity–a γ-Pcdh hallmark–is required for their functions remains unclear. We used a CRISPR/Cas9 approach to reduce isoform diversity, targeting each *Pcdhg* variable exon with pooled sgRNAs to generate an allelic series of 26 mouse lines with 1 to 21 isoforms disrupted *via* discrete indels at guide sites and/or larger deletions/rearrangements. Analysis of 5 mutant lines indicates that postnatal viability and neuronal survival do not require isoform diversity. Surprisingly, as it is the only γ-Pcdh that cannot independently engage in homophilic *trans*-interactions, we find that γC4, encoded by *Pcdhgc4*, is the only critical isoform. Because the human orthologue is the only *PCDHG* gene constrained in humans, our results indicate a conserved γC4 function that likely involves distinct molecular mechanisms.

INTRODUCTION

Cell-cell recognition *via* transmembrane cell adhesion molecules is essential for neural circuit formation. With trillions of exquisitely specific synapses in the human brain, it has been suggested that molecular diversity–achieved either *via* alternative gene splicing or combinatorial expression of large adhesion molecule families–plays an important role (Zipursky and Sanes, 2010). In *Drosophila*, this is exemplified by *Dscam1,* a single gene capable of generating 38,016 distinct protein isoforms through alternative splicing (Schmucker et al., 2000). Dscam1 isoform diversity is essential for neurodevelopmental processes including axon guidance, synapse specificity, and neurite self-avoidance (Hattori et al., 2009, 2007; Zhan et al., 2004, reviewed in Hattori et al., 2008). Mammalian Dscams, despite important roles in neurodevelopment, do not generate such isoform diversity (Fuerst et al., 2009, 2008; Garrett et al., 2018, 2016). In this sense, the clustered protocadherins (cPcdhs) are the

mammalian analogue to fly *Dscam1,* although their diversity is generated by differential isoform expression *via* promoter choice rather than alternative splicing (Zipursky and Sanes, 2010).

The cPcdhs, expressed broadly throughout the developing central nervous system (CNS), are cadherin superfamily molecules that engage in strictly homophilic *trans*-interactions. The cPcdh isoforms are encoded by three gene clusters arrayed in tandem at human chromosome 5q31 (encoding 53 cPcdh proteins) and on mouse chromosome 18 (encoding 58 cPcdh proteins) (Wu et al., 2001; Wu and Maniatis, 1999). In mouse, the *Pcdha* cluster encodes 14 α-Pcdhs, the *Pcdhb* cluster encodes 22 β-Pcdhs, and the *Pcdhg* cluster encodes 22 γ-Pcdhs (Figure 1A). While all three *Pcdh* gene clusters contribute to neural development to some extent (Chen et al., 2017; Emond and Jontes, 2008; Hasegawa et al., 2017, 2016; Ing-Esteves et al., 2018; Katori et al., 2009; Meguro et al., 2015; Mountoufaris et al., 2017), the *Pcdhg* locus is the only one required for postnatal viability, and disruption of this locus results in the strongest phenotypes (reviewed in Peek et al., 2017).

The *Pcdhg* cluster is comprised of 22 variable (V) exons, divided according to sequence homology into A, B, and C subgroups, and three constant exons. Each V exon is regulated by an individual promoter and, upon transcription, is spliced to the constant exons (Tasic et al., 2002; Wang et al., 2002)(Figure 1B). Each V exon encodes one γ-Pcdh isoform's extracellular domain, including 6 cadherin (EC) domains, a transmembrane domain, and a membrane-proximal variable cytoplasmic domain (VCD), while the constant exons encode a C-terminal cytoplasmic domain common to all 22 γ-Pcdh isoforms (Figure 1B). Single-cell RT-PCR from cerebellar Purkinje neurons indicated that each neuron expressed all three γC isoforms (*Pcdhgc3*, *Pcdhgc4*, and *Pcdhgc5*) from both alleles while stochastically expressing ~4 of the 19 γA and γB isoforms monoallelically (Kaneko et al., 2006). Interestingly, this does not hold in all neurons, as single-serotonergic neuron transcriptomics indicated that, of the γ-Pcdhs, many neurons expressed only γC4, and few if any expressed γC3 or γC5 (Chen et al., 2017); additionally, few olfactory sensory neurons expressed any of the C-type *Pcdha* or *Pcdhg* isoforms

(Mountoufaris et al., 2017). Clustered Pcdh isoforms interact strictly homophilically in *trans,* engaging in anti-parallel interactions involving EC1-EC4, while EC5 and EC6 mediate promiscuous *cis* dimer formation between γ isoforms, as well as with α- and β-Pcdhs (Goodman et al., 2016b, 2016a; Rubinstein et al., 2015; Schreiner and Weiner, 2010; Thu et al., 2014). These two types of interactions result in a multimeric lattice of dimers between cell membranes sharing the same isoform composition (Brasch et al., 2019; Rubinstein et al., 2017), and indeed, homophilic specificity is observed at the multimer level (Schreiner and Weiner, 2010; Thu et al., 2014). In this way, the 58 cPcdh isoforms generate thousands of distinct recognition signals.

While α, β and γ isoforms can all contribute to multimer formation (Rubinstein et al., 2017), the γ-Pcdhs are particularly critical for neural development (Peek et al., 2017). Mice lacking either the entire *Pcdhg* cluster or the γC3-C5 V exons exhibited neonatal lethality, with excessive apoptosis of neuronal subtypes in the spinal cord and hypothalamus (Prasad et al., 2008; Su et al., 2010; Wang et al., 2002b). *Pcdhg* mutants also exhibited reduced synapse number and disorganized synaptic terminals within the spinal cord, which were separable from the increased cell death (Garrett and Weiner, 2009; Prasad and Weiner, 2011; Weiner et al., 2005). When neonatal lethality was circumvented with a conditional *Pcdhg* allele, distinct phenotypes were observed in other parts of the CNS, including reduced dendrite arborization in forebrain neurons (Garrett et al., 2012; Keeler et al., 2015; Molumby et al., 2016; Suo et al., 2012) accompanied by an increase in morphologically immature dendritic spines (Molumby et al., 2017). In retina-restricted mutants, many neuronal subtypes exhibited excessive cell death without separable synaptic disorganization (Ing-Esteves et al., 2018; Lefebvre et al., 2008). Starburst amacrine cells (SACs), however, survived in normal numbers, but exhibited clumping of their dendritic fields, indicative of a failure of self-avoidance (Kostadinov and Sanes, 2015; Lefebvre et al., 2012).

The parallel molecular and phenotypic diversity of cPcdhs could, *a priori*, indicate three potential models, all of which may be correct for distinct subsets of neurons or distinct functions. First, isoform

diversity *per se* may be required. Second, there may be a high level of isoform redundancy such that any one (or few) isoform(s) may suffice. Third, there may be unique roles for *individual* isoforms, such that no other isoform can compensate for their loss. Some extant evidence exists for each of these possibilities. The γC3-5 isoforms, but not γA1-3, were required for postnatal viability in mice (Chen et al., 2012). As γC5 is expressed later in the postnatal period (Frank et al., 2005), and γC4 is unable to localize to the cell membrane alone (Rubinstein et al., 2015; Thu et al., 2014), it seemed most likely that this reflected a crucial role for γC3. While re-expressing a single γ-Pcdh isoform (γA1 or γC3) could rescue self-avoidance in SACs, it disrupted self/non-self recognition required for proper receptive field overlap and retinal circuit performance (Kostadinov and Sanes, 2015; Lefebvre et al., 2012). Similar re-expression of γA1 or γC3 in cortical neurons on a *Pcdhg* null background led to aberrantly increased or decreased dendrite arborization, depending on whether surrounding cells also expressed the same single isoform and could thus presumably engage homophilically (Molumby et al., 2016). A requirement for cPcdh diversity *per se* was discovered for the convergence of olfactory sensory neuron axons on particular glomeruli in the olfactory bulb: disruption of olfactory circuitry was mild in single gene cluster mutants, but devastating in mice lacking all three clusters, and it was not rescued by re-expressing a triad of single α-, β-, and γ-Pcdh isoforms (Hasegawa et al., 2016; Mountoufaris et al., 2017). Support for unique roles for individual cPcdh isoforms comes from the demonstration that the disrupted axonal branching observed in *Pcdha* mutant serotonergic neurons (Katori et al., 2009) is due entirely to the role of one isoform, αC2 (Chen et al., 2017), and from the demonstration of a unique role for γC3 in regulating Wnt signaling through Axin1 (Mah et al., 2016).

Thus, no single model is likely to encompass all of the γ-Pcdhs' diverse functions. Additionally, most prior studies have relied on mis- or over-expression of individual isoforms, which could in some cases result in new, distinct phenotypes. Paralleling studies establishing the necessity of *Dscam1* molecular diversity for specific neurodevelopmental roles in *Drosophila*, we used CRISPR/Cas9 genome

editing to simultaneously target the 22 *Pcdhg* variable exons in an unbiased manner, and created a new allelic series of mouse mutants with reduced isoform diversity from the endogenous gene cluster. We find, surprisingly, that only one isoform – γC4, which uniquely cannot mediate homophilic *trans*-interactions independently – is strictly required for postnatal viability and survival of the many neuronal subsets shown previously to depend on the γ-Pcdhs. Our results: 1) show that some γ-Pcdh functions do not require molecular diversity; 2) confirm that at least some γ-Pcdh isoforms have unique roles; and 3) suggest that the regulation of neuronal survival may require novel mechanisms of cPcdh interaction and/or signaling involving γC4.

RESULTS

CRISPR/Cas9 strategy for reducing Pcdhg isoform diversity

To assess the importance of γ-Pcdh isoform diversity to postnatal viability and neurodevelopmental functions, we used a shotgun CRISPR/Cas9 genome editing screen to disrupt varying numbers of *Pcdhg* variable exons (Figure 1B-D). We reasoned that by injecting pooled individual single guide RNAs (sgRNAs) targeting each V exon into many zygotes, we could generate a number of unique mutant mouse lines, each of which harbored distinct patterns of reduced isoform diversity due to variability in which sgRNAs efficiently bound and directed mutations. We designed sgRNAs to target within ~100-150 base pairs downstream of the start codon of each *Pcdhg* V exon, with the goal of creating frame-shifting mutations through non-homologous end joining (NHEJ) repair (Figure 1B-C)(Cong et al., 2013; Mali et al., 2013). A total of twenty guides were designed; the *PcdhgB4* and *PcdhgB5* V exons shared enough homology to allow a single guide to target both, as did *PcdhB6* and *PcdhgB7*. Pooled sgRNAs were concentrated, then combined with Cas9 mRNA and microinjected into C57BL/6J mouse zygotes at three different concentrations (each individual guide at 50 ng sgRNA/μl, 10 ng sgRNA/μl, or 5 ng sgRNA/μl: HI, MED, and LO, respectively), as described in Materials and Methods

(Figure 1D). Microinjected zygotes were transferred into a total of 20 pseudopregnant females, resulting in 100 live-born mice (16 from HI, 34 from MED, and 50 from LO). All 100 founders were screened at 7 *Pcdhg* exons by PCR with Sanger sequencing to detect frame-shifting mutations. From this initial screen, 15 founders exhibited some disruption (10 from HI and 5 from MED) and were designated for breeding. Most disruptions were found in pups from HI or MED injections, so the 50 founders resulting from the LO injections were not pursued further. The remaining 35 founders were further screened by PCR and Sanger sequencing at the remaining 15 *Pcdhg* variable exons. In this way, a total of 31 founders were identified that carried some constellation of mutations at the guide-targeted sites (14/16 HI, 17/34 MED).

Due to the likely mosaicism of the founders and the uncertainty of germline transmission of any given mutation, we did not characterize the 31 identified founders more extensively. Rather, each was crossed with wild-type C57BL/6J animals to generate G1 offspring for further analysis (Figure 1D). Sperm from male G1 mice was cryopreserved while somatic tissue was used for genotyping to identify lines of interest carrying reduced diversity of *Pcdhg* variable exons. Ninety-four G1 offspring were screened for heterozygous mutations using a custom amplicon assay from Illumina and Illumina MiSeq sequencing (see Materials and Methods for details, target coordinates are listed in Supplementary File 1). The amplicon assay was designed to sequence the 22 *Pcdhg* sgRNA target regions, all the analogous regions in the untargeted *Pcdha* and *Pcdhb* clusters, and the top 95 predicted potential off-target sites (Figure 1D). We screened for missense and nonsense mutations, small indels and, because our approach was intended to produce up to 22 double-stranded breaks within 160 kilobases, larger rearrangements between breakpoints. Smaller indels were identified using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010), while larger rearrangements were identified with BreaKmer (Abo et al., 2015). Twenty-six distinct lines carrying unique constellations of mutations are represented in Table 1, derived from 12 different founders. Most mouse lines (20 lines from 9 founders) resulted from microinjection of

158 the highest concentration of total sgRNAs (50 ng/μl/guide). Mutations ranged from only one disrupted V
159 exon (leaving 21 intact) to 21 disrupted V exons (leaving only 1 intact; Table 1). Each V exon was
160 disrupted in at least one mouse line, either by discrete indels, or through involvement in a
161 rearrangement between breakpoints.

162 *Whole genome sequencing reveals rearrangements undetected by amplicon sequencing*

163 Based on initial results from the custom amplicon assay, we chose three lines for cryo-recovery
164 and further analysis – *Pcdhg*^{em5}, *Pcdhg*^{em12}, and *Pcdhg*^{em35} (Table 1). Mice from each recovered line were
165 intercrossed to generate homozygous mutants. Once obtained, we used these homozygous mutants to
166 verify by PCR V-exon indels and rearrangements. We found that fewer exons were successfully amplified
167 by PCR than expected (Figure 1-figure supplement 1A), indicating that amplicon sequencing failed to
168 detect all the mutations harbored by the heterozygous G1 mutants. Therefore, we performed whole
169 genome linked-read sequencing on homozygous mutants, using the Chromium Genome Sequencing
170 Solution from 10X Genomics to detect large scale rearrangements (Figure 1-figure supplement 1, Figure
171 2-figure supplement 1). Paired-end reads from the whole genome sequencing were used to reconstruct
172 the rearrangements (Figure 2). We found that *Pcdhg*^{em5} contained frame-shifting indels in V exons A1,
173 B1, A5, A7, B7, and C4, all of which were accurately called by the prior amplicon sequencing. However,
174 there was an inversion and deletion that disrupted exons A9, B6, and A10 that was undetected by the
175 previous analysis (Figure 2-figure supplement 1A). Thus, *Pcdhg*^{em5} retained 13 intact V exons (Figure 2A),
176 and we renamed the allele *Pcdhg*^{13R1}; "13R" for the number of intact exons remaining and "1" as it was
177 the first allele identified with this number (referred to as 13R1 hereafter for simplicity).

178 Sequencing of *Pcdhg*^{em12} revealed extensive rearrangements. A fusion between exons A5 and B8
179 identified by the amplicon sequencing was confirmed, but additional junctions were found between A1
180 and A4, between A12 and A9, and between B6 and A6 (Figure 2A, Figure 1-figure supplement 1C).
181 Furthermore, linked reads revealed an insertion including a transposable element and coding sequence

from *Anp32a* which aligns to exons 4-7 of transcript Anp32a-201 without the intervening introns (Figure 2B, Figure 1-figure supplement 1D). This phenomenon of the insertion of a transposable element along with coding sequence from an early expressed gene has been previously described in CRISPR genome editing (Ono et al., 2015). As this sequence was inserted 3' to the inverted exon *Pcdhga5*, there is no associated transcription start site, and no protein product is expected. Altogether, at least 9 double-stranded breaks occurred, resulting in a frame-shifting 52 bp insertion into exon B2, and larger deletions and rearrangements disrupting all the other γA and γB isoforms. As only the 3 γC V exons remained intact in $Pcdhg^{em12}$, we renamed this allele $Pcdhg^{3R1}$ (3R1 hereafter; Figure 2B).

Whole genome sequencing from $Pcdhg^{em35}$ mutants confirmed small frame-shifting deletions at exons A12 and C5, as well as an in-frame deletion of 9 bp in exon C4. However, this more exhaustive sequencing also revealed a large rearrangement undetected by the prior amplicon sequencing analysis. Here, there was a ~94 kb deletion spanning the breakpoint from exon A1 to that of exon A11 (Figure 2-figure supplement 1B). Only exons B8 and C3 were unaffected by any mutation. As exon C4 still encoded a nearly full-length protein lacking only 3 amino acids (residues 27-29 in the signal peptide), we renamed the $Pcdhg^{em35}$ allele $Pcdhg^{3R2}$ (3R2 hereafter, Figure 2C), as it represented the second allele identified with 3 isoforms remaining.

In all three mutants analyzed thus far, rearrangements were identified by whole genome sequencing that were undetected by the amplicon analysis (Figure 2, Table 1 – Table supplement 1). With this information, we re-analyzed the amplicon sequencing data by visual inspection of the paired reads within the Integrative Genomics Viewer (IGV). We were able to find many, but not all, of the junctions identified by the whole genome sequencing, but missed by BreaKmer analysis of the amplicon sequencing (Figure 1-figure supplement 1B). Therefore, we manually inspected the alignments from each of the other frozen lines. Additional rearrangements were identified where each of the paired ends of multiple reads mapped to different exons (Table 1 – Table supplement 1).

206 *Isoform diversity* per se *is not required for postnatal viability*

207 The complete deletion of the Pcdhg cluster results in neonatal lethality (Wang et al., 2002b), as
208 does the deletion of the three γC V exons, whereas mice with deletion of the first three γA exons (A1,
209 A2, A3) had no reported phenotypes (Chen et al., 2012). Of the three lines initially characterized, 3R1
210 and 3R2 homozygous mutant mice were born at expected Mendelian ratios and survived into adulthood
211 without any overt differences from their wild-type or heterozygous littermates. In contrast, 13R1
212 homozygous mutants died within hours of birth with the hunched posture, tremor, and inability to right
213 themselves or nurse that is characteristic of *Pcdhg*$^{del/del}$ mutants lacking the entire *Pcdhg* cluster (Wang
214 et al., 2002b; Weiner et al., 2005). Whole genome sequencing from all three lines confirmed that there
215 were no off-target mutations in the *Pcdha* or *Pcdhb* locus, or any disruptions in the *Pcdhg* constant
216 exons. However, it remained formally possible that the particular rearrangements within 13R1 mutants
217 disrupted the expression of the other isoforms or V exon splicing to the constant exons, creating a
218 functional null mutation.

219 To exclude this possibility, we performed quantitative real-time PCR using cDNA from the
220 cerebral cortices of homozygous mutants and wild-type littermates (Figure 3A). To detect specific
221 isoform transcripts, forward primers targeting the 3' end of each V exon were used with a reverse
222 primer in constant exon 1 (product spans 1 intron) or exon 2 (product spans 2 introns). To monitor total
223 levels of *Pcdhg* locus transcription, a forward primer in constant exon1 was used with a reverse primer
224 in constant exon 2 (spanning 1 intron, primer sequences in Figure 3– figure supplement 2). Total locus
225 expression was significantly reduced in 3R2 homozygous mutants, but in both 3R1 and 13R1 mutant
226 cortex, expression levels were indistinguishable from controls. Furthermore, individual isoform
227 transcription was reduced only when mutations completely disrupted the exon by deletion or inversion.
228 Smaller indels generally had no effect on transcript expression level (e.g., A1, B1, A5, A7, B7, and C4
229 were not significantly reduced in 13R1 homozygous mutants, but A9, B6, and A10 were undetectable).

Additionally, transcription of several V exon fusions was detected, including A4 (fusion with A1) and A9 (fusion with A12) in 3R1 (predicted protein products encoded by these fused transcripts are listed in Supplementary File 2). We also asked the extent to which mutations within the *Pcdhg* locus altered isoform transcription from the *Pcdha* or *Pcdhb* clusters. While there were no significant changes detected in 13R1 mutants, isoforms from the 3' end of the *Pcdhb* locus were expressed at significantly higher levels in 3R1 and 3R2 homozygous mutants compared to controls (β11 in 3R1, β15 and β22 in both 3R1 and 3R2). Expression of *Pcdha* cluster genes appeared to be unchanged (Figure 3-figure supplement 1).

We also verified that these mRNA expression levels were reflected at the protein level, utilizing a series of antibodies specific for particular γ-Pcdh protein isoforms (Lobas et al., 2012). Western blot analysis of brain lysates from 13R1 neonates and 3R1 and 3R2 adults confirmed the presence of the expected isoforms at the appropriate molecular weights: 13R1 brains expressed γA isoforms, γB2, and γC3, but not γC4, while 3R1 and 3R2 brains expressed γC3 and γC4, but not γB2 or any γA isoforms (Figure 3B). Based on these analyses, we concluded that 13R1 homozygous mutants are not, in fact, complete *Pcdhg* functional nulls, and that the similarity of their neonatally lethal phenotype to that of nulls likely reflects the essential nature of a particular isoform lost in this line but present in both 3R1 and 3R2; that is, γC4, as described below. Before testing this conclusion in detail, we first asked whether the neonatal lethality of 13R1 was accompanied by cellular phenotypes previously described in *Pcdhg*$^{del/del}$ null mice, and whether the viable 3R1 and 3R2 lines lacked these phenotypes.

Excessive developmental neuronal apoptosis occurs in mutants exhibiting neonatal lethality

Pcdhg$^{del/del}$ mutants exhibit neonatal lethality with excessive cell death of interneurons in the ventral spinal cord and brainstem (Prasad et al., 2008; Wang et al., 2002b; Weiner et al., 2005). To ask if lethality in reduced diversity mutants was also accompanied by increased developmental apoptosis, we analyzed cryosections from spinal cords at P0. Indeed, we found significant cell death in 13R1

254 homozygous mutants, but not 3R1 or 3R2 mutants (Figure 4). Staining of transverse sections for the pan-
255 neuronal marker NeuN revealed that 13R1 mutant spinal cords were grossly smaller, with obvious
256 reductions in cell number primarily in the ventral spinal cord, as reported previously for *Pcdhg*$^{del/del}$ null
257 mutant neonates (Figure 4A-C)(Prasad et al., 2008; Wang et al., 2002b). This was accompanied by an
258 increase in reactive astrocytes, revealed by GFAP labeling within the gray matter (Figure 4D-F). Spinal
259 interneurons derive from 6 dorsal (dI1-6) and 4 ventral (V0-3) domains (reviewed by Lewis, 2006); we
260 used antibodies against two transcription factors, FoxP2 and Pax2, that label distinct subsets of
261 interneurons. We found significantly fewer FoxP2-positive ventral interneurons (derived from dI2, dI6,
262 and V1) in 13R1 mutants than in wild type littermates or 3R1 homozygous mutants (Figure 4G-I,P). Pax2
263 labels a broader subset of interneurons, derived from dI4, dI6, V0, and V1, that settle in both the dorsal
264 and ventral spinal cord. As we found previously for *Pcdhg* null animals (Prasad et al., 2008), 13R1
265 mutants had far fewer Pax2-positive ventral interneurons (Figure 4J-L,Q). To confirm that cell loss
266 resulted from excessive apoptosis as observed in *Pcdhg* nulls, we assayed for cleaved caspase 3 (CC3), a
267 marker of apoptotic cell death, and found significantly more CC3-labeled profiles in 13R1 mutants than
268 in wild type littermates or 3R1 (Figure 4M-O,R). An additional phenotype previously described in *Pcdhg*
269 null mutants is the clumping of parvalbumin-positive Ia afferent axon terminals around their motor
270 neuron targets in the ventral horn; this phenotype is worsened, though not entirely due to, increased
271 interneuron apoptosis (Prasad and Weiner, 2011). Again, we found that 13R1 mutants exhibited a null
272 mutant-like phenotype while 3R1 mutant Ia afferent projections appeared similar to those of controls.
273 (Figure 4-figure supplement 1).

274 Increased cell death of many neuronal subtypes is also a hallmark of *Pcdhg* loss of function in
275 the retina, with retinal cell death occurring largely postnatally (Ing-Esteves et al., 2018; Lefebvre et al.,
276 2008). To circumvent the neonatal lethality of 13R1 mutants, we made compound heterozygous
277 mutants with the conditional loss of function allele *Pcdhg*^{fcon3} (Prasad et al., 2008) crossed with Pax6α-

Cre to restrict recombination of this allele to the retina (Marquardt et al., 2001). These mutants are referred to as 13R1/cRKO (for conditional retinal knockout). Examination of immunostained retinal cross sections from 13R1/cRKO mutants at P14 revealed substantial thinning of the inner retina, including both cellular and synaptic layers, compared to wild type (Figure 5A-C). Neither 3R1 nor 3R2 mutant retinas were notably thinner (Figure 5C and data not shown). As reported for *Pcdhg* null mutants (Lefebvre et al., 2008), the retinal thinning in 13R1/cRKO mice was not accompanied by obvious disorganization of neurite stratification within the inner plexiform layer (Figure 5D-I). To verify that this resulted from cell loss, we measured the density of two amacrine cell types (tyrosine hydroxylase (TH)+ dopaminergic amacrine cells and VGLUT3+ amacrine cells) and two retinal ganglion cell types (Melanopsin+ RGCs and Brn3a+ RGCs) in whole-mount retinas from P14 animals. All four cell types were significantly less numerous in 13R1/cRKO mutants than in wild-type littermates or in 3R1 or 3R2 homozygous mutants, neither of which exhibited any abnormalities (Figure 6).

Pcdhgc4 is the necessary and sufficient Pcdhg *cluster gene for postnatal viability.*

Having confirmed that the cellular phenotypes of 13R1 resemble the complete deletion of the *Pcdhg* cluster, we next turned to utilizing the novel CRISPR mutant mouse lines to confirm which γ-Pcdh isoforms were critical. We noted that homozygous 3R1 mutants survived and did not exhibit exacerbated apoptosis despite lacking expression of any functional γA or γB isoforms, indicating that one or more of the C3-C5 isoforms must be critical. Furthermore, 3R2 mutants survived and were phenotypically normal without a functional *Pcdhgc5* gene, whilst 13R1 homozygous mutants, which exhibited neonatal lethality and exacerbated neuronal apoptosis uniquely harbored frame-shifting mutations in *PcdhgC4*. In a separate study, we have derived and are analyzing a CRISPR-targeted mouse line that specifically generated a *Pcdhgc3* loss-of-function allele, and have found that they are viable and fertile as adults (Figure 7-figure supplement 1, D. Steffen, K.M. Mah, P.J. Bosch, A.M. Garrett, R.W. Burgess, and J.A. Weiner, in preparation). Together with prior data indicating that mice lacking the

entire *Pcdha* and *Pcdhb* clusters are viable (Hasegawa et al., 2016; Mountoufaris et al., 2017), this suggests that *Pcdhgc4* encodes the sole cPcdh isoform essential for organismal survival. We generated two additional mouse lines to confirm this conclusion.

First, we chose *Pcdhg*^{em8} from our list of mutants for cryorecovery (Table 1). This line harbored a large deletion from within exon A1 to within exon C3 identified by visual inspection of the aligned reads from the original amplicon sequencing, as well as a 1 bp frame-shifting deletion in exon C5. Upon cryorecovery, we generated homozygous mutants and verified these mutations by linked-read whole genome sequencing (Figure 7A, Figure 7-figure supplement 2A). As only V exon C4 was left intact, this strain was renamed *Pcdhg*^{1R1} (1R1 hereafter). These homozygous mutants survived into adulthood and were fertile, with no overt differences from their wild type litter mates. Both quantitative RT-PCR and western blot analyses reflected the expected isoform expression (Figure 7C, Figure 7-figure supplement 3).

Second, we generated an entirely new mutant mouse line by specifically targeting *Pcdhgc4* for CRISPR/Cas9-directed gene disruption. This resulted in a 13 bp frame-shifting deletion 3' to the start codon (Figure 7B; Figure -figure supplement 2B). We named this new allele *Pcdhg*^{C4KO} (C4KO hereafter). Homozygous C4KO mutants exhibited the same hunched posture, tremor, and inability to nurse or right themselves observed in *Pcdhg*$^{del/del}$ and 13R1 mutants and died shortly after birth. Sanger sequencing confirmed that the analogous regions of the other *Pcdhg* V exons were not disrupted in this line (Figure 7-figure supplement 2B), while Western blot analyses verified that the other isoforms were produced as expected (Figure 7C). Together, these two additional mouse lines show that mutation of *Pcdhgc4* alone recapitulated the overt phenotype of losing the entire cluster, whilst expression of this single γ-Pcdh isoform was sufficient to rescue viability, even when all 21 other *Pcdhg* isoforms were absent.

Pcdhgc4 is the crucial isoform for neuronal survival

To ask if these overt phenotypes of death *vs.* survival in C4KO and 1R1 mutants extended to neuronal cell survival, we again analyzed spinal cords from P0 neonatal mutants. As expected from their outward appearance, 1R1 homozygous spinal cords appeared grossly normal in overall size and neuronal density (NeuN+ cells), with little if any reactive gliosis (GFAP+; Figure 8B,D). In contrast, C4KO mutant spinal cords were grossly smaller and exhibited clear ventral interneuron loss (Figure 8A) accompanied by reactive gliosis (Figure 8C) that was essentially identical to that observed in complete *Pcdhg* null mutants (Prasad et al., 2008; Wang et al., 2002b) and 13R1 mutants (Figure 4). Analysis of individual cell types revealed significant loss of FoxP2-labeled cells and Pax2-positive neurons in the ventral spinal cord of C4KO mutants (Figure 8E,G,K,L). Significant reductions in 1R1 mutant cell density were observed, though they were much more modest (Figure 8F,H,K,L). Consistent with this, there were significantly more CC3-positive neurons in C4KO mutants than in 1R1 or WT neonates (Figure 8I,J,M). As expected, these patterns of spinal interneuron survival corresponded to the presence (C4KO; Figure 8-figure supplement 1A) or absence (1R1; Figure 8-figure supplement 1B) of aggregated Ia afferent terminals around motor neurons in the ventral horn.

As 1R1 mutants survive into adulthood, we analyzed neuronal survival in the retina at 14 days of age and in adult. There was no indication of retinal thinning at either age. As above, cell densities were calculated (in cells per mm^2) for four cell types: dopaminergic amacrine cells (TH+), VGLUT3+ amacrine cells, ipRGCs (Melanopsin+), and Brn3a+ RGCs (Figure 8O-R). The means of the cell densities were compared across genotype and age with a two-way ANOVA followed by pairwise comparisons with a Tukey analysis. There was a significant genotype effect in Brn3a+ cell density ($p = 0.034$), and a significant pairwise difference between wild type animals at 2 weeks and 1R1 mutants in adult ($p = 0.046$). No other differences reached statistical significance. Therefore, neuronal survival was largely normal in 1R1 mutants, with any cell loss being very modest compared to mutants lacking γC4 (13R1/cRKO: Figure 5, Figure 6; *Pcdhg* nulls: Lefebvre et al., 2008).

PCDHGC4 *is constrained in humans*

With the number of human genomes and exomes that have been sequenced, it is possible to test if variation occurs at the rate expected from random mutation. If predicted loss of function (LOF) mutations (e.g., frame-shifts and premature stop codon insertions) within a given gene are observed at lower rates than expected, it indicates that that gene is essential, and that loss of function is not tolerated (referred to as "constraint"). To ask if specific cPcdh isoforms are constrained in the human population, we queried the Genome Aggregation Database (gnomAD, Broad Institute), an aggregation of genomic variation across 141,456 human exomes and whole genomes (Karczewski et al., 2019). For each cPcdh isoform, the ratio of observed to expected (o/e) LOF mutations was reported along with its 90% confidence interval (CI, vertical lines, Figure 9A-B). A gene is considered constrained if the upper bound of the 90% CI falls below 0.35 (red line, Figure 9A-B). Within the *PCDHG* locus, only *PCDHGC4* met this criterion (o/e = 0.14, 90% CI = 0.07-0.31, Figure 9A). Amongst the *PCDHA* and *PCDHB* isoforms, only *PCDHAC2*, the specific isoform required in serotonergic neurons, was constrained (o/e = 0.16, 90% CI = 0.09-0.34). Thus, consistent with our analysis of the orthologous gene in our allelic series of mice, *PCDHGC4* is likely also essential in humans.

DISCUSSION

The 22 γ-Pcdhs comprise a diverse family of cadherin superfamily adhesion molecules with multiple distinct functions in distinct cell types. Extant data suggest three possible models for the role of isoform diversity: (1) a model of *diversity* where many isoforms are required for a given function; (2) a model of *redundancy* in which any single isoform (or small subset of isoforms) can serve a given function; and (3) a model of isoform *specificity* where one specific isoform (or small subset of isoforms) is strictly required for a given function. Here, we used a CRISPR/Cas9 strategy to reduce γ-Pcdh isoform diversity in an unbiased fashion, creating a new allelic series of mouse strains. In the course of analyzing

several new *Pcdhg* alleles, we discovered that the control of neuronal survival and postnatal viability is best described by the third model. We found–surprisingly, given its inability to reach the cell surface and mediate homophilic adhesion without other cPcdh *cis*-interaction partners–that γC4 is the sole necessary and sufficient isoform.

Our strategy was to inject Cas9 mRNA along with 20 individual sgRNAs targeting each of the *Pcdhg* V exons simultaneously. By screening G1 offspring from many founders, we collected an array of mutants with distinct V exon mutation patterns, ranging from a single isoform disrupted (21 intact) to 21 isoforms disrupted (1 intact), all of which were cryopreserved. Each isoform was disrupted in at least one mutant line (Table 1). Furthermore, there was evidence of double-stranded breaks at each sgRNA site identifiable as indels or junctions between guide sites. Our G1 screening was done by next generation sequencing of a custom amplicon array. This technique was effective for identifying indels within each individual amplicon (e.g., smaller indels at each guide site), but was not exhaustive in identifying rearrangements between guide sites, even when both sides of the junction were predicted to be recognized by the amplicon array primers. This could be due to the new DNA sequence: 1) being difficult to amplify within the array, and therefore not sequenced; 2) not being accurately aligned to the reference genome; or 3) not being identified as a new junction by the bioinformatic algorithm BreaKmer. Indeed, some junctions that were not recognized by BreaKmer were clearly identified by visual inspection of the sequence alignments within IGV as paired reads mapping to two different exons (Table 1 – Table supplement 1). Those identified in two different mutant lines – 3R1 and 1R1 – were confirmed by whole genome sequencing. It is thus likely that many of the cryopreserved lines not yet recovered and made homozygous harbor additional mutations to those described here from the initial analysis (Table 1, Table 1 – Table supplement 1). We suggest that future studies targeting multiple sites within a relatively small region should use other methods to identify rearrangements between guide sites, such as targeted long-read sequencing (Bennett-Baker and Mueller, 2017).

One concern when using a pooled sgRNA approach such as ours is the compounding of potential off-target mutations. To assess this, we sequenced the top 95 predicted off-target sites in G1 offspring by amplicon sequencing. We did not identify indels attributable to Cas9 activity at any of these sites. While not exhaustive, this analysis is consistent with recent studies suggesting that off-target mutations from CRISPR/Cas9 are relatively rare in mouse model production (Iyer et al., 2015; Mianné et al., 2016; Nakajima et al., 2016). To further confirm the specificity of our CRISPR/Cas9 targeting, we performed whole genome sequencing on homozygous mutants from four lines at least three generations after G1. Here, we were particularly concerned with the closely linked and highly similar protocadherins encoded by the two adjacent gene clusters (*Pcdha* and *Pcdhb*) as well as *Pcdh1* and *Pcdh12*, non-clustered protocadherins located within 450 kb of the *Pcdhg* cluster. There were no indels or rearrangements identified within any of these genes. We conclude that, at the resolution assayed here, off-target mutations do not contribute to our results. This conclusion is bolstered by the segregation of postnatal and neuronal survival phenotypes observed across multiple mouse lines: the two lines harboring frame-shifting mutations in *Pcdhgc4* (13R1 and C4KO, generated separately with distinct methodologies) died and exhibited similarly exacerbated neuronal apoptosis, while the two mouse lines in which *Pcdhgc4* was untouched (3R1) or harbored a small in-frame deletion (3R2) survived and exhibited normal neuronal survival. The fact that mice harboring an additional allele (1R1) in which *Pcdhgc4* is the only remaining functional isoform are viable and exhibit only mild neuronal survival alterations further supports the specificity of CRISPR targeting.

Isoform choice in the Pcdh gene clusters is regulated by CCCTC-binding factor (CTCF) and cohesin, which bind directly to a conserved sequence element (CSE) in each individual V exon promoter (Golan-Mashiach et al., 2012; Kehayova et al., 2011; Monahan et al., 2012). These proteins organize DNA loops, bringing the promoters of expressed isoforms into proximity with an enhancer region 3' to the constant exons (Jiang et al., 2017). CTCF/cohesin complex binding is restricted by promoter methylation

laid down during embryogenesis by the methyltransferase Dnmt3b. Indeed, in the absence of *Dnmt3b*, many more cPcdh isoforms were expressed by each neuron (Toyoda et al., 2014); conversely, in CTCF knockout forebrain neurons, expression of nearly all cPcdh genes was markedly reduced (Hirayama et al., 2012). Here, we analyzed the isoform expression in cerebral cortex from four new *Pcdhg* mutant strains using qPCR. In the *Pcdhg* cluster, indels or other genomic DNA junctions that allowed splicing from V exons to the constant exons resulted in expressed transcripts, while expression from isoforms deleted from the genome or inverted were undetectable, as expected (Figure 3, Figure 7-figure supplement 3). The lone exception was the 3'-most isoform, γC5, which was significantly reduced in 1R1 mutants which harbored a single base pair insertion in the *Pcdhgc5* variable exon (Figure 7-figure supplement 3). We also analyzed select isoforms from the *Pcdha* and *Pcdhb* clusters. Here, large deletions in the 5' end of the *Pcdhg* cluster resulted in significantly higher expression from the 3' *Pcdhb* isoforms (β11, β15, and β22, Figure 3-figure supplement 1, Figure 7-figure supplement 3). This was most pronounced in 1R1 mutants, which harbor the largest deletion in the *Pcdhg* cluster, but was not detected in 13R1 mutants, which have the smallest deletion of the mutants assayed. This is likely due to regulatory elements within DNaseI hypersensitive sites HS16-20, located 3' to the *Pcdhg* cluster, that are essential for expression from the *Pcdhb* cluster (Yokota et al., 2011). The movement of these elements closer to the *Pcdhb* cluster in the 1R1, 3R1, and 3R2 alleles is likely responsible for the increased expression observed. Consistent with this interpretation, it has previously been reported that moving *Pcdha* isoforms closer to regulatory elements located between the *Pcdha* and *Pcdhb* cluster (HS7 or HS5-1) by deleting intervening exons resulted in their increased expression (Noguchi et al., 2009). Importantly, this increased expression of *Pcdhb* cluster genes cannot be responsible for the viability of the 1R1, 3R1, or 3R2 mouse lines: similar *Pcdhb* overexpression was reported previously in $Pcdhg^{del/del}$ animals, which nevertheless die at birth (Chen et al., 2012).

Previous interrogations of *Pcdhg* isoform diversity either deleted two sets of three V exons (A1-A3 or C3-C5) (Chen et al., 2012) or drove overexpression of a single isoform from a transgene (Lefebvre et al., 2012; Molumby et al., 2017, 2016). Our new series of mutants, including both those initially analyzed and the many that remain cryopreserved, will allow for a finer dissection of the role of isoform diversity and the identification of potential isoform-specific functions. Chen et al. (2012) showed that mice lacking γC3-C5 exhibited neonatal lethality and exacerbated spinal interneuron apoptosis similar to that of *Pcdhg* null mutants; in contrast, mice lacking A1-A3 were viable and outwardly normal. Our new results indicate that γC4 is, in fact, the sole essential isoform for neuronal survival and postnatal viability, indicating that loss of this isoform was responsible for the lethality observed in ΔC3-5 mice (Chen et al., 2012). This is particularly surprising, as the γC4 isoform is peculiar in several ways. Cell aggregation experiments in the K562 suspension cell line indicate that γC4, uniquely among γ-Pcdhs, cannot mediate *trans* homophilic binding on its own. Like the α-Pcdhs, it requires interaction with carrier isoforms (either other γ- or β-Pcdhs) to reach the cell membrane (Thu et al., 2014), possibly because EC6 of γC4 inhibits surface delivery and the formation of *cis*-homodimers (Rubinstein et al., 2015; Thu et al., 2014). This raises the novel possibility that the postnatal viability and largely (though not entirely) retained neuronal survival observed in 1R1 mutants, in which γC4 is the only functional γ-Pcdh isoform, may reflect functions for this protein in cellular compartments other than the plasma membrane. Alternatively, γC4 may reach the membrane, escorted by β-Pcdhs (several of which are upregulated in 1R1 mutants) and participate in homophilic interactions that trigger intracellular signaling pathways specific to the γC4 VCD, which remain to be identified. Precedence for this latter possibility comes from our recent finding that the VCD of γC3 can, uniquely amongst γ-Pcdhs, bind to and sequester Axin1 at the membrane, which leads to suppression of some components of the Wnt signaling pathway (Mah et al., 2016).

Thus, while the downstream mechanisms through which γ-Pcdhs promote neuronal survival remain unknown, our findings suggest that either protein interactions specifically mediated by γC4, or a unique localization for this isoform, will be involved. In 1R1 mutants, γC4 alone was not entirely sufficient for normal neuron number in the spinal cord or retina (Figure 8). This could be explained potentially by the significantly reduced expression levels of γC4 (and thus of total γ-Pcdh proteins) in these mutants (Figure 7C, Figure 7-figure supplement 1), or it could indicate that other cPcdh isoforms contribute to survival of some neuronal subsets in a collaborative, if not strictly essential, manner. Consistent with this second possibility, cell death phenotypes observed in *Pcdhg* mutants were made more severe by additional disruption of the *Pcdha* and *Pcdhb* clusters, despite the fact that *Pcdha* or *Pcdhb* disruption did not increase cell death when *Pcdhg* remained intact (Hasegawa et al., 2016; Ing-Esteves et al., 2018). In any case, our genetic interrogation of the *Pcdhg* gene cluster clearly demonstrates that γC4 is the only γ-Pcdh isoform strictly necessary for neuronal survival and postnatal viability. Consistent with this, the orthologous *PCDHGC4* is the only *PCDHG* gene (and, along with *PCDHAC2*, one of only 2 clustered Pcdh genes overall) that is constrained in humans, thus indicating an essential role conserved throughout mammalian evolution that is of potential clinical relevance. In this respect, the critical next step will be to elucidate the unique protein-protein interactions and intracellular signaling pathways in which γC4 participates during the development of the nervous system.

MATERIALS AND METHODS

Mouse strains

All animals were housed in the research animal facility either at The Jackson Laboratory, The University of Iowa, or Wayne State University under standard housing conditions with a 12 h/12 h light/dark cycle and food and water ad libitum. All procedures using animals were performed in accordance

491 with The Guide for the Care and Use of Laboratory Animals and were reviewed and approved by the
492 Institutional Animal Care and Use Committee at each respective institution. All experiments included a
493 mix of male and female animals. Previously described animals include *Pcdhg*^{fcon3} (Prasad et al., 2008) and
494 *Pax6α-Cre* (Marquardt et al., 2001). The newly generated reduced *Pcdhg* diversity mutants are
495 cryopreserved at The Jackson Laboratory (Stock numbers listed in Table 2).
496 *Generation of Pcdhg reduced diversity mutants*
497 Guide RNA (sgRNA) sequences were designed to target the near 5' regions to the start codons of
498 each variable exon (Table 3). Guides were designed using the tool at crispr.mit.edu and were analyzed
499 using RGEN tools to minimize off target sites and to maximize the likelihood of frameshifting mutations
500 (Bae et al., 2014a, 2014b; Park et al., 2015). 20 total guides were synthesized (IDT), as exons B4 and B5
501 had a high level of 5' homology, as did exons B6 and B7. Guides were received lyophilized, resuspended
502 in water, mixed in equal parts, then lyophilized again. This mixture was diluted and resuspended with *S.*
503 *pyogenes* Cas9 mRNA to generate three guide concentrations: 50 ng/μl for each guide, 10 ng/μl for each
504 guide, or 5 ng/μl for each guide. In each dilution, Cas9 mRNA was present at a concentration of 100
505 ng/μl. These mixtures were microinjected into C57BL/6J zygotes, which were subsequently implanted
506 into pseudopregnant female C57BL/6J mice.
507 The resulting live founders were screened by PCR and Sanger sequencing (primers in Table 4).
508 There were three iterations of screening. First, we chose 7 exons distributed across the locus for PCR
509 amplification and Sanger sequencing in 100 live founders. Heterozygous and homozygous indels were
510 identified by analyzing the sequence traces, and 15 founders were found to harbor some mutation. 50
511 of the 100 founders were from the lowest concentration of sgRNA injection (5 ng/μl). No mutations
512 were found in the first round of screening in these animals, and they were not analyzed further. In the
513 second round of screening, the 35 remaining founders were analyzed for indels at the other 15 exons. In
514 the third round, they were screened for rearrangements between exons by PCR with single forward

primers mixed with pooled reverse primers. PCR products detected above background level were purified and sequenced using the forward primer. Mice carrying any mutation detected in any round of screening were bred for one generation with C57BL/6J animals. The resulting G1 offspring were screened by amplicon sequencing. Males were prioritized when available for ease of sperm cryopreservation.

To generate *Pcdhg*^{C4KO} mice, only sgRNA complimentary to exon C4 was microinjected (50 ng/μl, Table 3) along with Cas9 mRNA (100 ng/μl). Of the resulting 16 live founders, 9 contained mosaic indels. 3 founders were crossed with C57BL/6J, one of which transmitted a 13 bp deletion. These mice were crossed to homozygosity, and all 22 variable exons were screened by PCR and Sanger sequencing to verify that only the *Pcdhgc4* exon was disrupted.

Screening by amplicon sequencing

A TruSeq Custom Amplicon 1.5 assay was designed with Illumina DesignStudio. The amplicon length was 250 bp, 201 bp of which corresponded to target sequence. 153 targets were covered by 225 amplicons for a cumulative target length of 31,721 bp. Target regions were chosen to be centered at the sgRNA sites, the analogous sites in the variable exons in *Pcdha* and *Pcdhb,* and 95 predicted potential off-target sites. The target coordinates are listed in Supplementary File 1. Genomic DNA was extracted using the DNeasy 96 Blood & Tissue Kit (Qiagen) from 94 G1 animals and processed along with a C57BL/6J wild type control and a negative control. Libraries were constructed using the TruSeq Custom Amplicon (Illumina) Library prep kit. Briefly, custom probes were hybridized to the flanking region of interest in gDNA, DNA polymerase extended across the target region, and unique barcodes and sequencing primers are added by PCR. Libraries were pooled and sequenced on a MiSeq instrument from Illumina with paired-end reads, 150 bp long.

Sequences were matched to samples by barcode and aligned to the target regions (BaseSpace, Illumina). Variants were called using Genome Analysis Toolkit (GATK) (McKenna et al., 2010). As

539 CRISPR/Cas9 mediated NHEJ events result in small to medium sized indels, but not SNPs (Allen et al.,
540 2018), we focused on indels that exceeded the following thresholds: QUAL ≥ 850; DP ≥ 90; QD ≥ 5. We
541 also filtered out indels detected in the wild type C57BL/6J control, as these reflect discrepancies from
542 the reference genome. Subsequent whole genome sequencing confirmed these mutations in the *Pcdhg*
543 locus, but not potential off-target mutations expected to be linked (e.g., *Pcdhb7*). Furthermore, one
544 mutation that did not exceed our threshold was detected by whole genome sequencing (A12 in
545 *Pcdhg*^{em35}). Therefore, Table 1 – Table supplement 1 summarizes mutations detected in the *Pcdhg* locus
546 above and below the threshold, as indicated.

547 Because of the pooled amplicon reaction, we reasoned that large rearrangements between
548 guide sites could be detected by amplicon sequencing if enough sequence remained on either side of
549 the junction. To find these junctions we used BreaKmer (Abo et al., 2015). We were able to detect some
550 of these junctions with BreaKmer, but subsequent whole genome sequencing revealed rearrangements
551 missed by BreaKmer in each sequenced mouse line. In some cases, this could be because deletions
552 extended beyond the amplicon site on one or both sides of the junction. However, many junctions were
553 flanked by enough amplicon sequence to expect them to be detectible in this data set. Indeed, by
554 visually analyzing the read alignments in the Integrative Genomics Viewer (IGV, Broad Institute) we were
555 able to find some of these junctions when mating paired-end reads mapped to different exons (Figure 1-
556 figure supplement 1B). We used this approach to identify additional junctions in the cryopreserved
557 mouse lines as indicated in Table 1 – Table supplement 1.

558 *Whole genome sequencing*

559 Genomic DNA was isolated from homozygous mutants using the DNA Extraction from Fresh Frozen
560 Tissue protocol (10X Genomics). Briefly, nuclei were isolated from tissue, lysed using proteinase K, then
561 DNA purified using magnetic beads. Linked-read whole genome libraries were constructed using the
562 Genome Chip Kit v2 (10X Genomics). Briefly, high molecular weight DNA was partitioned into Gel Bead-

In-EMulsions (GEMs) where unique barcoded primers were added to individual molecules of DNA. After the GEMs were dissolved, Illumina specific sequencing primers and barcodes were added by isothermal amplification, then library construction was completed via end repair, a-tailing, adapter ligation, and amplification. Libraries were sequenced on a HiSeq X (Illumina) by Novogene (Sacramenta, CA) with 150 bp paired-end reads. Sequences were aligned using the Long Ranger analysis pipeline from 10X Chromium and visualized in Loupe (10X Genomics) or IGV (Broad Institute). Paired-end reads were used to find NHEJ junctions between guide sites, while reads spanning the junctions were analyzed to uncover the specific sequences of these junctions.

Data deposition

All Illumina sequencing data is available at the Gene Expression Omnibus (GEO) with the accession number XXXXX (in progress) or the Sequence Read Archive (SRA) with number XXXX.

Quantitative RT-PCR

RNA was isolated from cerebral cortex of animals at P0 (13R1 mutants and littermate controls) or 2-12 weeks of age (other mutants and littermate controls) using Trizol reagent. Five μg of RNA per sample was used to make cDNA using Superscript III (Invitrogen) according to the manufacturers protocol. qPCR was performed in triplicate (technical replicates) using primers listed in Figure 3– Figure supplement 2 with SYBR Green PCR Master Mix. The relative abundance of each transcript was calculated using the ΔΔCt method, normalized to *GAPDH* and littermate controls. The relative levels of each transcript in controls, 3R1, and 3R2 mutants were compared using an ANOVA with Tukey post-hoc tests. 13R1 and 1R1 mutants were compared with littermate controls for each transcript using a t-test.

Western blotting

Whole brains at P0 or adult ages (3-8 months) were homogenized in RIPA buffer (0.1% SDS, 0.25% sodium deoxycholate, 1% NP-40, 150 mM NaCl, 50 mM Tris-HCl, pH 7.4, 5 mM NaF) plus protease inhibitors (Roche Mini cOmplete) using a Dounce homogenizer and a Wheaton overhead stirrer. The

lysate was centrifuged at 16,000 X g for 15 min at 4°C to remove cell debris and proteins were quantified using a BCA assay kit (Pierce/Thermo Scientific). Thirty or 40 µg of protein was resolved on Mini-PROTEAN TGX precast 12% SDS-PAGE gels (Bio-Rad) and proteins were transferred to a nitrocellulose membrane using the Bio-Rad Trans-Blot Turbo Transfer System. Membranes were blocked using 10% skim milk for 1 hour and incubated with the primary antibody in buffer (2.5% BSA in TBS-T (Tris buffered saline with 0.1% Tween-20) overnight at 4°C. The following day, membranes were washed using TBS-T and incubated in HRP-conjugated secondary antibody for 1 hour. Membranes were washed, developed using the SuperSignal West Pico ECL reagents (Thermo Scientific), and imaged using a Li-Cor Odyssey Fc Imaging system.

Immunofluorescence

Tissues were processed and stained as described previously (Garrett et al., 2016; Prasad et al., 2008). Briefly, neonatal spinal columns were removed and fixed by immersion in 4% paraformaldehyde (PFA) for four hours at 4° C, followed by extensive washing in PBS and cryopreservation in 30% sucrose. Eyes were enucleated and dissected to remove the cornea and lens, then fixed overnight by immersion in 4% PFA at 4° C followed by cryopreservation in 30% sucrose. Tissue was embedded in OCT (Sakura-Finetek) and sectioned with a cryostat onto positively charged Superfrost Plus slides (Fisher Scientific). After blocking in 2.5% bovine serum albumin with 0.1% Triton-X-100 in PBS, primary antibodies were incubated on the slides overnight at 4° C in a humidified chamber, followed by secondary antibodies for 1 hour at room temperature. Whole mount retinas were stained free floating in primary antibody diluted in blocking solution with 0.5% Triton-X-100 for 48-72 hours, and in secondary antibody for 24 hours. Sections were counter-stained with DAPI (4',6-diamidino-2-phenylindole), prior to mounting with Fluoro-Gel mounting media (Electron Microscopy Services #17985-11).

Antibodies

Primary antibodies used included the following: Guinea pig anti-VGLUT3 (1:10,000, Millipore), sheep anti-tyrosine hydroxylase (1:500, Millipore), rabbit anti-melanopsin (1:10,000, ATS), mouse anti-Brn3a (1:500, Millipore), goat anti-choline acetyltransferase (1:500, Millipore), mouse anti-NeuN (1:500, Millipore), mouse anti-GFAP (1:500, Sigma), rabbit anti-FoxP2 (1:4000, Abcam), rabbit anti-Pax2 (1:200, Zymed), rabbit anti-cleaved caspase 3 (1:100, Cell Signaling Technologies), mouse anti-parvalbumin (1:500, Sigma), and mouse anti-GAPDH (1:500, Abcam). Mouse monoclonal antibodies against γ-Pcdh proteins used for western blots (1:500-1:1000) were generated by NeuroMab in collaboration with the Weiner laboratory (Lobas et al., 2012) and obtained from Antibodies, Inc.: N159/5 (detecting an epitope in constant exon 1 or 2 and thus all 22 γ-Pcdh isoforms); N144/32 (detecting all γA subfamily isoforms); N148/30 (specific for γB2); N174B/27 (specific for γC3). A rabbit polyclonal antibody raised at Affinity BioReagents against the peptide sequence VAGEVNQRHFRVDLD (within EC1) from murine γC4 was also used for western blotting (1:1000). Secondary antibodies were conjugated with Alexa-488, -568, or -647 (1:500, Invitrogen) or HRP (1:1000-1:5000, Jackson Immunoresearch).

Image quantification

Control and mutant spinal cords were imaged using epifluorescence at equivalent thoraco-lumbar locations using a Leica SPE TCS confocal microscope and captured with Leica Application Suite software. From the resulting images, cell counts were performed using Cell Counter plugin in Fiji image analysis software (Schindelin et al., 2012). For FoxP2 and CC3 analysis, all immunoreactive cells were counted per hemicord from 12 hemicords per animal and at least 3 animals per genotype. For Pax2 analysis, all interneurons ventral to the central canal were quantified and compared. Values for wild type and mutant genotypes were compared using One-way ANOVA with Dunnett's multiple comparison test. Sample size was based on prior studies, as effective group sizes were known (Prasad et al., 2008).

Whole mount retinas were imaged by confocal microscopy. In the resulting image stacks, cell density was measured using the Cell Counter plugin in Fiji image analysis software (Schindelin et al.,

2012). Values from at least two fields per retina, sampled from different regions midway between the center and periphery (technical replicates), were averaged. These averaged values from at least six retinas per genotype (biological replicates) were compared using an ANOVA with Tukey post-hoc tests. Sample size was based on prior studies, as effective group sizes were known. Retinas were analyzed at 2 weeks of age, with the exception of 1R1 mutants and littermate controls, which were analyzed at 2 weeks and adult (3-6 months of age). Here, values were compared using a two-way ANOVA with age and genotype as independent variables, followed by Tukey pairwise comparisons.

All p-values are listed in Supplementary File 3

gnomAD database analysis

For each cPcdh isoform in human, the gnomAD browser was queried (gnomAD.broadinstitute.org; (Karczewski et al., 2019)) and the ratio of observed to expected (o/e) loss of function variants was recorded along with its 90% confidence interval.

ACKNOWLEDGEMENTS

We would like to thank the scientific services at the Jackson Laboratory for assistance throughout this project, including Genetic Engineering Technologies, Microinjection, and Reproductive Sciences services for the production and preservation of new mutants, the Genome Technologies service for sequencing, and the Bioinformatics service for data analysis. We would also like to thank Kate Miers for assistance with the mouse colony. This work was supported by NIH Grant NS090030 to R.W.B. and J.A.W. and NIH Grant NS055272 to J.A.W. The Scientific Services at The Jackson Laboratory are supported by NIH Grant CA034196.

COMPETING INTERESTS

The authors declare no financial or non-financial competing interests.

658

659 **REFERENCES**

660

661 Abo RP, Ducar M, Garcia EP, Thorner AR, Rojas-Rudilla V, Lin L, Sholl LM, Hahn WC, Meyerson M,
662 Lindeman NI, Van Hummelen P, MacConaill LE. 2015. BreaKmer: detection of structural variation in
663 targeted massively parallel sequencing data using kmers. *Nucleic Acids Res* **43**:e19–e19.
664 doi:10.1093/nar/gku1211

665

666 Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, Angeli P, Páleníková P, Khodak A, Kiselev V,
667 Kosicki M, Bassett AR, Harding H, Galanty Y, Muñoz-Martínez F, Metzakopian E, Jackson SP, Parts L.
668 2018. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat*
669 *Biotechnol* **37**:64. doi:10.1038/nbt.4317

670

671 Bae S, Kweon J, Kim HS, Kim J-SS. 2014a. Microhomology-based choice of Cas9 nuclease target sites.
672 *Nature methods* **11**:705–6. doi:10.1038/nmeth.3015

673

674 Bae S, Park J, Kim J-SS. 2014b. Cas-OFFinder: a fast and versatile algorithm that searches for potential
675 off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics (Oxford, England)* **30**:1473–5.
676 doi:10.1093/bioinformatics/btu048

677

678 Bennett-Baker PE, Mueller JL. 2017. CRISPR-mediated isolation of specific megabase segments of
679 genomic DNA. *Nucleic acids research* **45**:e165. doi:10.1093/nar/gkx749

680

681 Brasch J, Goodman KM, Noble AJ, Rapp M, Mannepalli S, Bahna F, Dandey VP, Bepler T, Berger B,

682 Maniatis T, Potter CS, Carragher B, Honig B, Shapiro L. 2019. Visualization of clustered protocadherin
683 neuronal self-recognition complexes. *Nature*. doi:10.1038/s41586-019-1089-3
684
685 Chen WV, Alvarez FJ, Lefebvre JL, Friedman B, Nwakeze C, Geiman E, Smith C, Thu CA, Tapia JC, Tasic B,
686 Sanes JR, Maniatis T. 2012. Functional significance of isoform diversification in the protocadherin
687 gamma gene cluster. *Neuron* **75**:402–9. doi:10.1016/j.neuron.2012.06.039
688
689 Chen WV, Nwakeze CL, Denny CA, O'Keeffe S, Rieger MA, Mountoufaris G, Kirner A, Dougherty JD, Hen
690 R, Wu Q, Maniatis T. 2017. Pcdhαc2 is required for axonal tiling and assembly of serotonergic circuitries
691 in mice. *Science* **356**:406–411. doi:10.1126/science.aal3231
692
693 Cong L, Ran F, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013.
694 Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, NY)* **339**:819–23.
695 doi:10.1126/science.1231143
696
697 Emond MR, Jontes JD. 2008. Inhibition of protocadherin-α function results in neuronal death in the
698 developing zebrafish. *Developmental Biology* **321**:175–187. doi:10.1016/j.ydbio.2008.06.011
699
700 Frank M, Ebert M, Shan W, Phillips GR, Arndt K, Colman DR, Kemler R. 2005. Differential expression of
701 individual gamma-protocadherins during mouse brain development. *Molecular and Cellular*
702 *Neuroscience* **29**:603–616. doi:10.1016/j.mcn.2005.05.001
703
704 Fuerst PG, Bruce F, Tian M, Wei W, Elstrott J, Feller MB, Erskine L, Singer JH, Burgess RW. 2009. DSCAM
705 and DSCAML1 Function in Self-Avoidance in Multiple Cell Types in the Developing Mouse Retina. *Neuron*

706 **64**:484–497. doi:10.1016/j.neuron.2009.09.027

707

708 Fuerst PG, Koizumi A, Masland RH, Burgess RW. 2008. Neurite arborization and mosaic spacing in the

709 mouse retina require DSCAM. *Nature* **451**:470. doi:10.1038/nature06514

710

711 Garrett AM, Khalil A, Walton DO, Burgess RW. 2018. DSCAM promotes self-avoidance in the developing

712 mouse retina by masking the functions of cadherin superfamily members. *Proc Natl Acad Sci USA*

713 **115**:E10216. doi:10.1073/pnas.1809430115

714

715 Garrett AM, Schreiner D, Lobas MA, Weiner JA. 2012. γ-Protocadherins Control Cortical Dendrite

716 Arborization by Regulating the Activity of a FAK/PKC/MARCKS Signaling Pathway. *Neuron* **74**:269–276.

717 doi:10.1016/j.neuron.2012.01.028

718

719 Garrett AM, Tadenev AL, Hammond YT, Fuerst PG, Burgess RW. 2016. Replacing the PDZ-interacting C-

720 termini of DSCAM and DSCAML1 with epitope tags causes different phenotypic severity in different cell

721 populations. *eLife* **5**:e16144. doi:10.7554/eLife.16144

722

723 Garrett AM, Weiner JA. 2009. Control of CNS Synapse Development by γ-Protocadherin-Mediated

724 Astrocyte–Neuron Contact. *The Journal of Neuroscience* **29**:11723–11731.

725 doi:10.1523/JNEUROSCI.2818-09.2009

726

727 Golan-Mashiach M, Grunspan M, Emmanuel R, Gibbs-Bar L, Dikstein R, Shapiro E. 2012. Identification of

728 CTCF as a master regulator of the clustered protocadherin genes. *Nucleic acids research* **40**:3378–91.

729 doi:10.1093/nar/gkr1260

730

731 Goodman KM, Rubinstein R, Thu CA, Bahna F, Mannepalli S, Ahlsén G, Rittenhouse C, Maniatis T, Honig
732 B, Shapiro L. 2016a. Structural Basis of Diverse Homophilic Recognition by Clustered α- and β-
733 Protocadherins. *Neuron* **90**:709–23. doi:10.1016/j.neuron.2016.04.004

734

735 Goodman KM, Rubinstein R, Thu CA, Mannepalli S, Bahna F, Ahlsén G, Rittenhouse C, Maniatis T, Honig
736 B, Shapiro L. 2016b. γ-Protocadherin structural diversity and functional implications. *eLife* **5**.
737 doi:10.7554/eLife.20930

738

739 Hasegawa S, Kobayashi H, Kumagai M, Nishimaru H, Tarusawa E, Kanda H, Sanbo M, Yoshimura Y,
740 Hirabayashi M, Hirabayashi T, Yagi T. 2017. Clustered Protocadherins Are Required for Building
741 Functional Neural Circuits. *Frontiers in Molecular Neuroscience* **10**:114. doi:10.3389/fnmol.2017.00114

742

743 Hasegawa S, Kumagai M, Hagihara M, Nishimaru H, Hirano K, Kaneko R, Okayama A, Hirayama T, Sanbo
744 M, Hirabayashi M, Watanabe M, Hirabayashi T, Yagi T. 2016. Distinct and Cooperative Functions for the
745 Protocadherin-α, -β and -γ Clusters in Neuronal Survival and Axon Targeting. *Frontiers in Molecular*
746 *Neuroscience* **9**:155. doi:10.3389/fnmol.2016.00155

747

748 Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T. 2012. CTCF Is Required for Neural Development
749 and Stochastic Expression of Clustered Pcdh Genes in Neurons. *Cell Reports* **2**:345–357.
750 doi:10.1016/j.celrep.2012.06.014

751

752 Ing-Esteves S, Kostadinov D, Marocha J, Sing AD, Joseph KS, Laboulaye MA, Sanes JR, Lefebvre JL. 2018.
753 Combinatorial Effects of Alpha- and Gamma-Protocadherins on Neuronal Survival and Dendritic Self-

754 Avoidance. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **38**:2713–
755 2729. doi:10.1523/JNEUROSCI.3035-17.2018
756
757 Iyer V, Shen B, Zhang W, Hodgkins A, Keane T, Huang X, Skarnes WC. 2015. Off-target mutations are rare
758 in Cas9-modified mice. *Nat Methods* **12**:479–479. doi:10.1038/nmeth.3408
759
760 Jiang Y, Loh Y-HE, Rajarajan P, Hirayama T, Liao W, Kassim BS, Javidfar B, Hartley BJ, Kleofas L, Park RB,
761 Labonte B, Ho S-MM, Chandrasekaran S, Do C, Ramirez BR, Peter CJ, Julia T, faie B, Morishita H, Roussos
762 P, Nestler EJ, Schaefer A, Tycko B, Brennand KJ, Yagi T, Shen L, Akbarian S. 2017. The methyltransferase
763 SETDB1 regulates a large neuron-specific topological chromatin domain. *Nature genetics* **49**:1239–1250.
764 doi:10.1038/ng.3906
765
766 Kaneko R, Kato H, Kawamura Y, Esumi S, Hirayama T, Hirabayashi T, Yagi T. 2006. Allelic Gene Regulation
767 of Pcdh-α and Pcdh-γ Clusters Involving Both Monoallelic and Biallelic Expression in Single Purkinje Cells.
768 *Journal of Biological Chemistry* **281**:30551–30560. doi:10.1074/jbc.M605677200
769
770 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A,
771 Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG,
772 Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX,
773 Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel E, Weisburd B, Lek M, Ware JS,
774 Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel
775 S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-
776 Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Consortium T,
777 Neale BM, Daly MJ, MacArthur DG. 2019. Variation across 141,456 human exomes and genomes reveals

778 the spectrum of loss-of-function intolerance across human protein-coding genes. *Biorxiv* 531210.
779 doi:10.1101/531210
780
781 Katori S, Hamada S, Noguchi Y, Fukuda E, Yamamoto T, Yamamoto H, Hasegawa S, Yagi T. 2009.
782 Protocadherin-α Family Is Required for Serotonergic Projections to Appropriately Innervate Target Brain
783 Areas. *The Journal of Neuroscience* **29**:9137–9147. doi:10.1523/JNEUROSCI.5478-08.2009
784
785 Keeler AB, Schreiner D, Weiner JA. 2015. Protein Kinase C Phosphorylation of a γ-Protocadherin C-
786 terminal Lipid Binding Domain Regulates Focal Adhesion Kinase Inhibition and Dendrite Arborization.
787 *Journal of Biological Chemistry* **290**:20674–20686. doi:10.1074/jbc.M115.642306
788
789 Kehayova P, Monahan K, Chen W, Maniatis T. 2011. Regulatory elements required for the activation and
790 repression of the protocadherin-alpha gene cluster. *Proceedings of the National Academy of Sciences of*
791 *the United States of America* **108**:17195–200. doi:10.1073/pnas.1114357108
792
793 Kostadinov D, Sanes JR. 2015. Protocadherin-dependent dendritic self-avoidance regulates neural
794 connectivity and circuit function. *eLife* **4**:e08964. doi:10.7554/eLife.08964
795
796 Lefebvre JL, Kostadinov D, Chen WV, Maniatis T, Sanes JR. 2012. Protocadherins mediate dendritic self-
797 avoidance in the mammalian nervous system. *Nature* **488**:517. doi:10.1038/nature11305
798
799 Lefebvre JL, Zhang Y, Meister M, Wang X, Sanes JR. 2008. γ-Protocadherins regulate neuronal survival
800 but are dispensable for circuit formation in retina. *Development* **135**:4141–4151.
801 doi:10.1242/dev.027912

Lewis KE. 2006. How do genes regulate simple behaviours? Understanding how different neurons in the vertebrate spinal cord are genetically specified. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **361**:45–66. doi:10.1098/rstb.2005.1778

Lobas MA, Helsper L, Vernon CG, Schreiner D, Zhang Y, Holtzman MJ, Thedens DR, Weiner JA. 2012. Molecular heterogeneity in the choroid plexus epithelium: the 22-member γ-protocadherin family is differentially expressed, apically localized, and implicated in CSF regulation. *Journal of neurochemistry* **120**:913–27. doi:10.1111/j.1471-4159.2011.07587.x

Mah KM, Houston DW, Weiner JA. 2016. The γ-Protocadherin-C3 isoform inhibits canonical Wnt signalling by binding to and stabilizing Axin1 at the membrane. *Scientific reports* **6**:31665. doi:10.1038/srep31665

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science (New York, NY)* **339**:823–6. doi:10.1126/science.1232033

Marquardt T, Ashery-Padan R, Andrejewski N, Scardigli R, Guillemot F, Gruss P. 2001. Pax6 Is Required for the Multipotent State of Retinal Progenitor Cells. *Cell* **105**:43–55. doi:10.1016/S0092-8674(01)00295-1

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**:1297–303. doi:10.1101/gr.107524.110

826

827 Meguro R, Hishida R, Tsukano H, Yoshitake K, Imamura R, Tohmi M, Kitsukawa T, Hirabayashi T, Yagi T,
828 Takebayashi H, Shibuki K. 2015. Impaired clustered protocadherin-α leads to aggregated
829 retinogeniculate terminals and impaired visual acuity in mice. *Journal of neurochemistry* **133**:66–72.
830 doi:10.1111/jnc.13053

831

832 Mianné J, Chessum L, Kumar S, Aguilar C, Codner G, Hutchison M, Parker A, Mallon A-M, Wells S, Simon
833 MM, Teboul L, Brown SD, Bowl MR. 2016. Correction of the auditory phenotype in C57BL/6N mice via
834 CRISPR/Cas9-mediated homology directed repair. *Genome Med* **8**:16. doi:10.1186/s13073-016-0273-4

835

836 Molumby MJ, Anderson RM, Newbold DJ, Koblesky NK, Garrett AM, Schreiner D, Radley JJ, Weiner JA.
837 2017. γ-Protocadherins Interact with Neuroligin-1 and Negatively Regulate Dendritic Spine
838 Morphogenesis. *Cell Reports* **18**:2702–2714. doi:10.1016/j.celrep.2017.02.060

839

840 Molumby MJ, Keeler AB, Weiner JA. 2016. Homophilic Protocadherin Cell-Cell Interactions Promote
841 Dendrite Complexity. *Cell Reports* **15**:1037–1050. doi:10.1016/j.celrep.2016.03.093

842

843 Monahan K, Rudnick ND, Kehayova PD, Pauli F, Newberry KM, Myers RM, Maniatis T. 2012. Role of
844 CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-α
845 gene expression. *Proceedings of the National Academy of Sciences of the United States of America*
846 **109**:9125–30. doi:10.1073/pnas.1205074109

847

848 Mountoufaris G, Chen WV, Hirabayashi Y, O'Keeffe S, Chevee M, Nwakeze CL, Polleux F, Maniatis T.
849 2017. Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science (New*

York, NY) **356**:411–414. doi:10.1126/science.aai8801

Nakajima K, Kazuno A, Kelsoe J, Nakanishi M, Takumi T, Kato T. 2016. Exome sequencing in the knockin mice generated using the CRISPR/Cas system. *Sci Rep-uk* **6**:34703. doi:10.1038/srep34703

Noguchi Y, Hirabayashi T, Katori S, Kawamura Y, Sanbo M, Hirabayashi M, Kiyonari H, Nakao K, Uchimura A, Yagi T. 2009. Total Expression and Dual Gene-regulatory Mechanisms Maintained in Deletions and Duplications of the Pcdha Cluster. *Journal of Biological Chemistry* **284**:32002–32014. doi:10.1074/jbc.M109.046938

Ono R, Ishii M, Fujihara Y, Kitazawa M, Usami T, Kaneko-Ishino T, Kanno J, Ikawa M, Ishino F. 2015. Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes. *Scientific Reports* **5**:srep12281. doi:10.1038/srep12281

Park J, Bae S, Kim J-SS. 2015. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics (Oxford, England)* **31**:4014–6. doi:10.1093/bioinformatics/btv537

Peek SL, Mah K, Weiner JA. 2017. Regulation of neural circuit formation by protocadherins. *Cellular and Molecular Life Sciences* **74**:4133–4157. doi:10.1007/s00018-017-2572-3

Prasad T, Wang X, Gray PA, Weiner JA. 2008. A differential developmental pattern of spinal interneuron apoptosis during synaptogenesis: insights from genetic analyses of the protocadherin-γ gene cluster. *Development* **135**:4153–4164. doi:10.1242/dev.026807

874 Prasad T, Weiner JA. 2011. Direct and Indirect Regulation of Spinal Cord Ia Afferent Terminal Formation
875 by the γ-Protocadherins. *Frontiers in Molecular Neuroscience* **4**:54. doi:10.3389/fnmol.2011.00054
876
877 Rubinstein R, Goodman KM, Maniatis T, Shapiro L, Honig B. 2017. Structural origins of clustered
878 protocadherin-mediated neuronal barcoding. *Seminars in cell & developmental biology* **69**:140–150.
879 doi:10.1016/j.semcdb.2017.07.023
880
881 Rubinstein R, Thu CA, Goodman KM, Wolcott HN, Bahna F, Mannepalli S, Ahlsen G, Chevee M, Halim A,
882 Clausen H, Maniatis T, Shapiro L, Honig B. 2015. Molecular logic of neuronal self-recognition through
883 protocadherin domain interactions. *Cell* **163**:629–42. doi:10.1016/j.cell.2015.09.026
884
885 Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila
886 Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell* **101**:671–684.
887 doi:10.1016/S0092-8674(00)80878-8
888
889 Schreiner D, Weiner JA. 2010. Combinatorial homophilic interaction between γ-protocadherin multimers
890 greatly expands the molecular diversity of cell adhesion. *Proceedings of the National Academy of*
891 *Sciences* **107**:14893–14898. doi:10.1073/pnas.1004526107
892
893 Su H, Marcheva B, Meng S, Liang FA, Kohsaka A, Kobayashi Y, Xu AW, Bass J, Wang X. 2010. Gamma-
894 protocadherins regulate the functional integrity of hypothalamic feeding circuitry in mice.
895 *Developmental Biology* **339**:38–50. doi:10.1016/j.ydbio.2009.12.010
896
897 Suo L, Lu H, Ying G, Capecchi MR, Wu Q. 2012. Protocadherin clusters and cell adhesion kinase regulate

dendrite complexity through Rho GTPase. *Journal of Molecular Cell Biology* **4**:362–376. doi:10.1093/jmcb/mjs034

Tasic B, Nabholz CE, Baldwin KK, Kim Y, Rueckert EH, Ribich SA, Cramer P, Wu Q, Axel R, Maniatis T. 2002. Promoter Choice Determines Splice Site Selection in Protocadherin α and γ Pre-mRNA Splicing. *Molecular Cell* **10**:21–33. doi:10.1016/S1097-2765(02)00578-6

Thu CA, Chen WV, Rubinstein R, Chevee M, Wolcott HN, Felsovalyi KO, Tapia JC, Shapiro L, Honig B, Maniatis T. 2014. Single-cell identity generated by combinatorial homophilic interactions between α, β, and γ protocadherins. *Cell* **158**:1045–1059. doi:10.1016/j.cell.2014.07.012

Toyoda S, Kawaguchi M, Kobayashi T, Tarusawa E, Toyama T, Okano M, Oda M, Nakauchi H, Yoshimura Y, Sanbo M, Hirabayashi M, Hirayama T, Hirabayashi T, Yagi T. 2014. Developmental Epigenetic Modification Regulates Stochastic Expression of Clustered Protocadherin Genes, Generating Single Neuron Diversity. *Neuron* **82**:94–108. doi:10.1016/j.neuron.2014.02.005

Wang X, Su H, Bradley A. 2002a. Molecular mechanisms governing Pcdh-γ gene expression: Evidence for a multiple promoter and cis-alternative splicing model. *Genes & Development* **16**:1890–1905. doi:10.1101/gad.1004802

Wang X, Weiner JA, Levi S, Craig A, Bradley A, Sanes JR. 2002b. Gamma Protocadherins Are Required for Survival of Spinal Interneurons. *Neuron* **36**:843–854. doi:10.1016/S0896-6273(02)01090-5

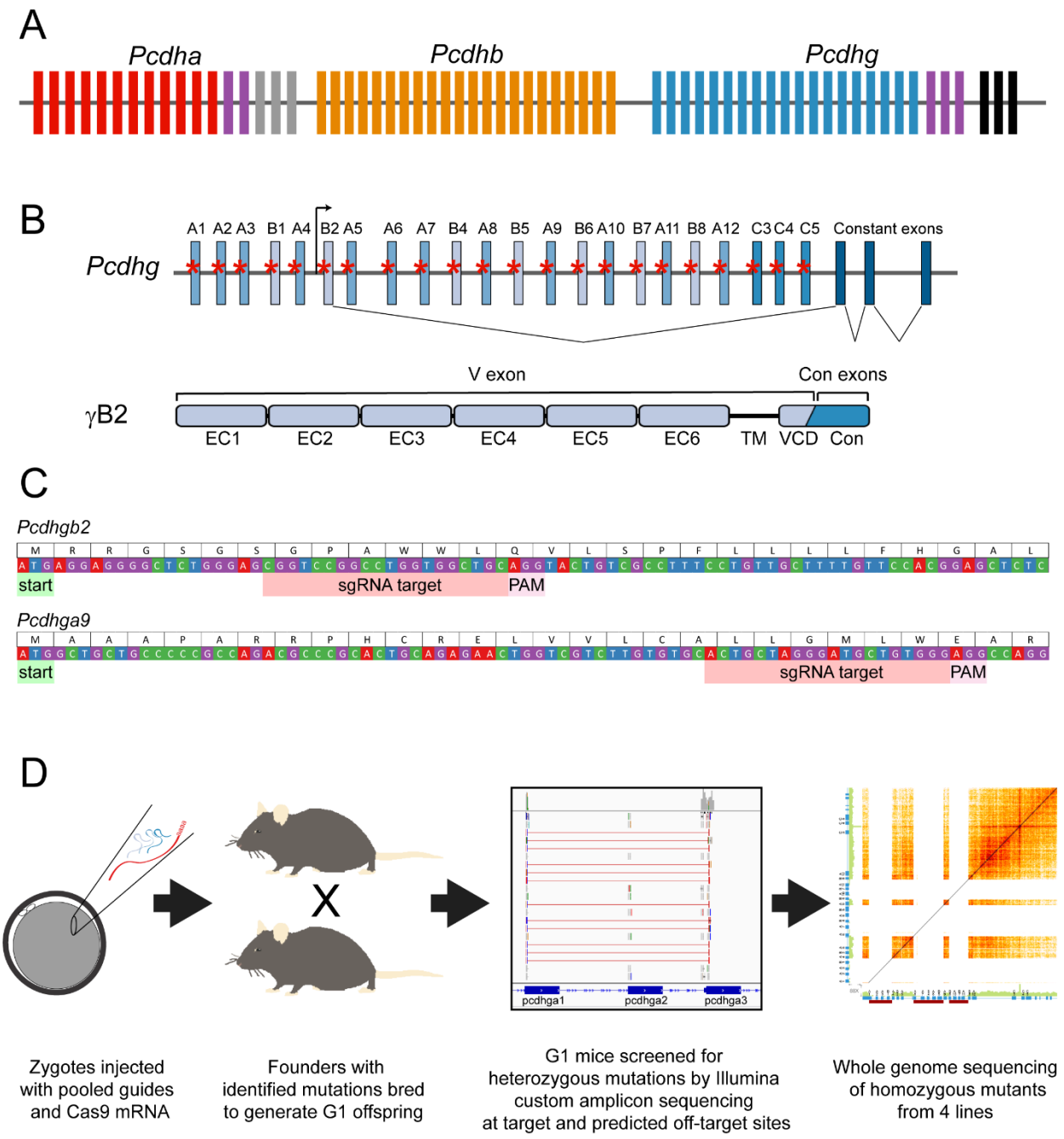
Weiner JA, Wang X, Tapia J, Sanes JR. 2005. Gamma protocadherins are required for synaptic

development in the spinal cord. *Proceedings of the National Academy of Sciences of the United States of America* **102**:8–14. doi:10.1073/pnas.0407931101

Yokota S, Hirayama T, Hirano K, Kaneko R, Toyoda S, Kawamura Y, Hirabayashi M, Hirabayashi T, Yagi T. 2011. Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *The Journal of biological chemistry* **286**:31885–95. doi:10.1074/jbc.M111.245605

Zipursky LS, Sanes JR. 2010. Chemoaffinity Revisited: Dscams, Protocadherins, and Neural Circuit Assembly. *Cell* **143**:343–353. doi:10.1016/j.cell.2010.10.009

935 **FIGURES**



936

937 **Figure 1: CRISPR/Cas9 strategy to reduce *Pcdhg* isoform diversity. A**) A schematic of the clustered Pcdh

938 loci with *Pcdha*, *Pcdhb* and *Pcdhg* arrayed in tandem. Constant exons for *Pcdha* are in gray and those for

939 *Pcdhg* in black. The homologous C variable (V) exons are indicated by purple boxes. **B**) The *Pcdhg* cluster

940 is comprised of 22 V exons and 3 constant exons. The V exons are subdivided into γA, γB and γC groups

based on sequence homology. sgRNA were designed to target each V exon (red asterisks). As shown for *Pcdhgb2* as an example, each V exon has its own promoter, and upon transcription, is spliced to the three constant exons. In the resulting protein, the extracellular domain (cadherin repeats EC1-EC6), the transmembrane domain (TM) and variable cytoplasmic domain (VCD) are encoded by the single V exon, while the C-terminal constant domain (Con) is encoded by the three constant exons. **C**) Examples for *Pcdhgb2* and *Pcdhga9* indicate the general design strategy for sgRNA, targeting downstream of and proximal to the start codon. **D**) The workflow for generating an allelic series of mutants with reduced *Pcdhg* isoform diversity. See also Figure 1-figure supplement 1.

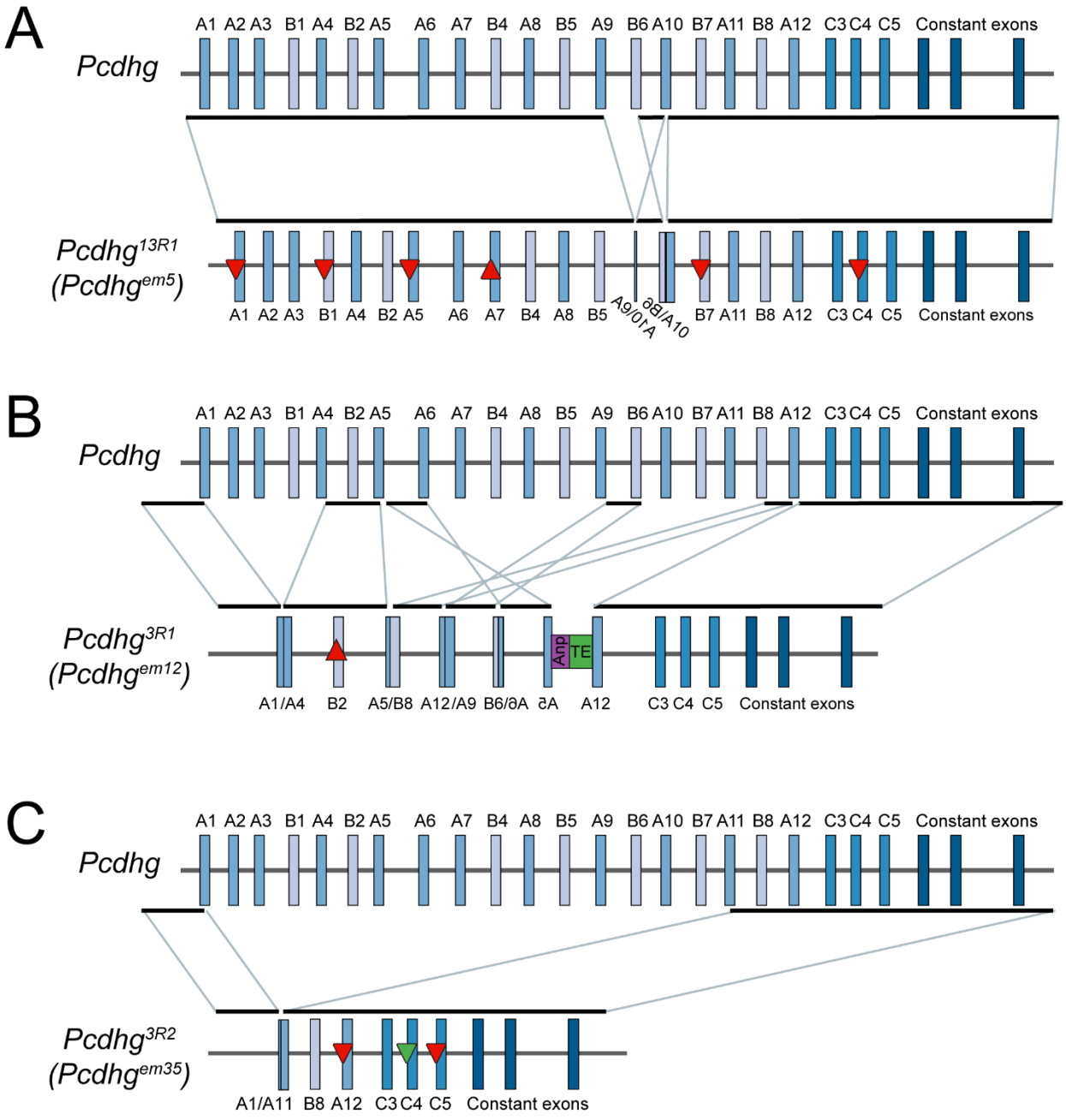


Figure 2: Three new *Pcdhg* alleles with reduced isoform diversity *via* CRISPR/Cas9 genome editing. Schematics of the mutated *Pcdhg* alleles verified or identified by whole genome sequencing in three strains illustrate the range of mutations induced by CRISPR/Cas9 targeting. **A**) In 13R1 mutants, frame-shifting indels disrupted 6 exons (upward triangle indicates insertion, downward indicates deletion, red

indicates frame-shift) and an inversion and deletion disrupted exons A9, B6, and A10. **B**) In 3R1 mutants, there was a small frame-shifting insertion in exon B2 along with multiple deletions and rearrangements between exons. For example, genomic DNA between the exon A5 and exon A6 guide sites was inverted and inserted at the B6 guide site, followed by coding sequence from the gene Anp32a (purple box) and a transposable element (green box). Only exons C3, C4, and C5 remain intact. **C**) In 3R2 mutants, a single large deletion resulted in a fusion between guide sites at exon A1 and exon A11. This was accompanied by small, frame-shifting deletions in exons A12 and C5 and a small in-frame deletion in exon C4 (downward pointing green triangle). See also Figure 2-figure supplement 1.

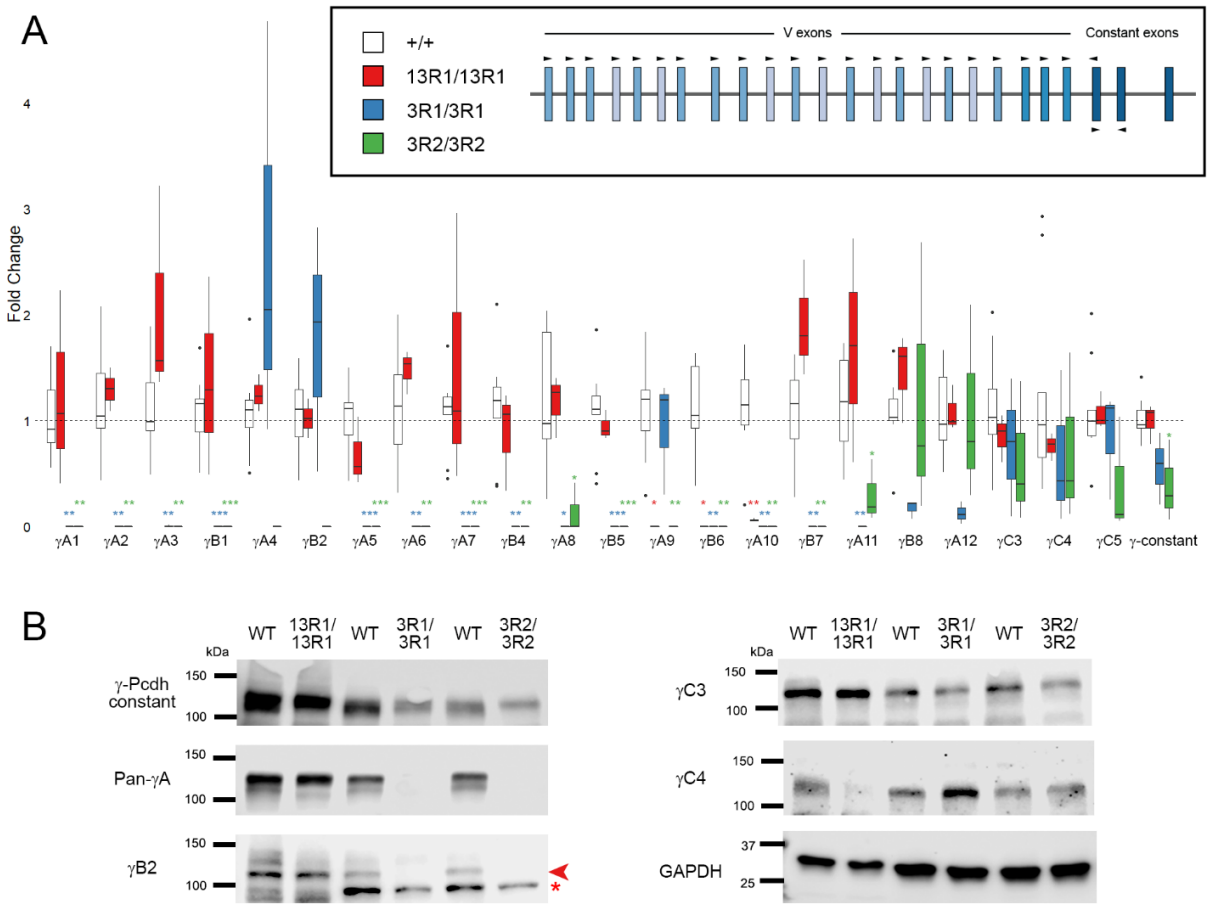
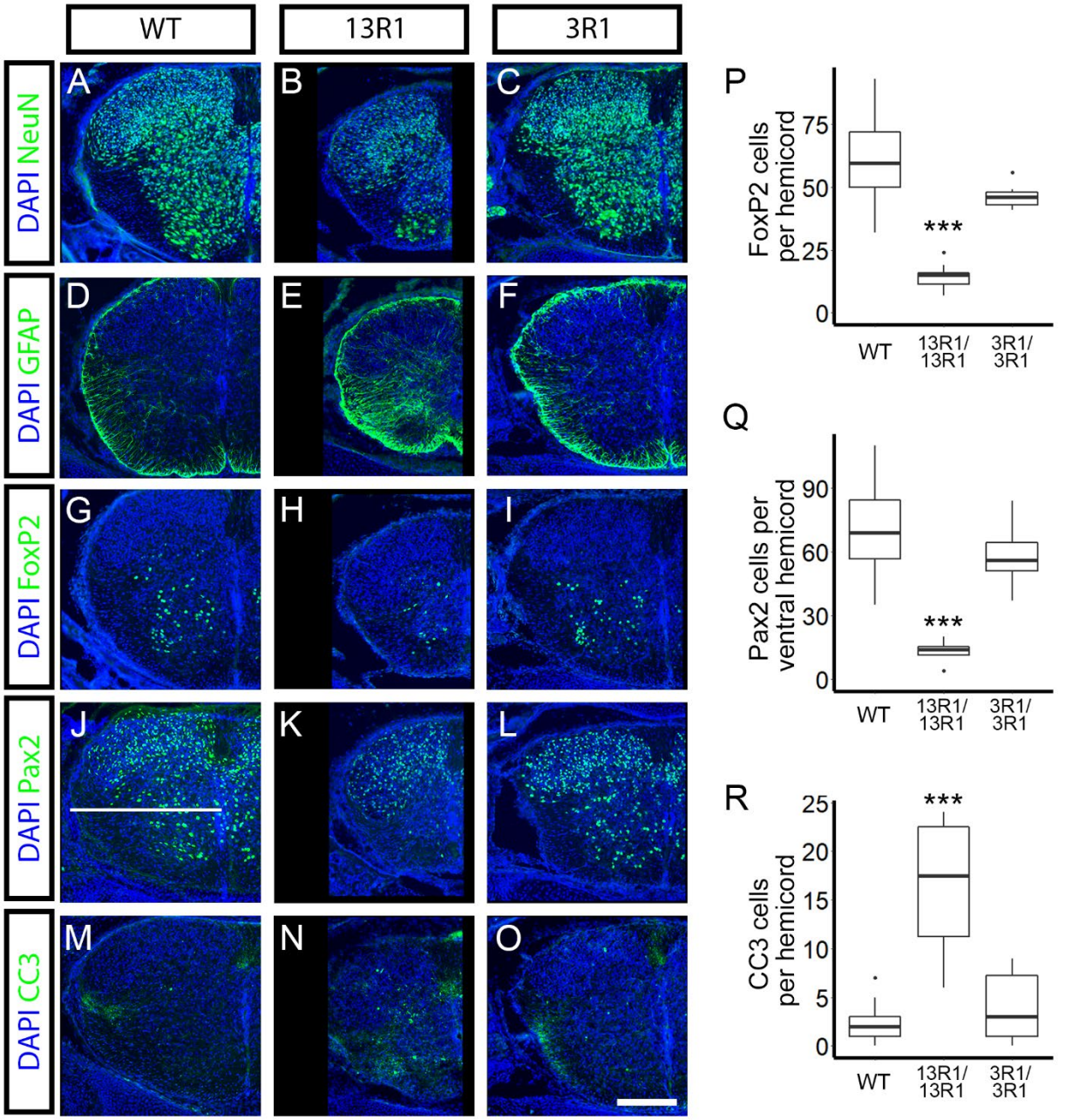


Figure 3: *Pcdhg* isoform expression reflects genomic DNA mutations. A) Quantitative RT-PCR analysis of cerebral cortex cDNA from 13R1 mutants (red), 3R1 mutants (blue) and 3R2 mutants (green) verified that intact isoforms were expressed at expected levels. Expression of constant exons was significantly reduced in 3R2 mutants only. 13R1 mutants were analyzed at P0, while 3R1 and 3R2 animals were analyzed at P14, each with littermate controls (white). * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ by Tukey post-hoc test comparing the indicated genotype with wild type. n=3-9 animals per genotype. Box plots represent the median, first and third quartile, range, and outliers. **B**) Western blots of brain lysates from 13R1 (at P0), 3R1, and 3R2 mutants (at 1 month of age) with littermate controls confirm the protein isoform expression predicted by allele sequencing. Arrowhead indicates the specific γB2 band at the predicted size; asterisk indicates non-specific background band of incorrect size present in adult samples only. See also Figure 3-figure supplement 1, and Figure 3-figure supplement 2.



977

978 **Figure 4: Spinal interneurons undergo excessive apoptosis in 13R1, but not 3R1 mutants.** Cryosections

979 from P0 spinal cords were immunostained for the indicated markers (green; DAPI counterstain for

980 nuclei, blue). NeuN staining in cryosections from P0 spinal cords of (**A**) wild type, (**B**) 13R1 mutants, and

981 (**C**) 3R1 mutants revealed that 13R1 mutant cords were smaller with substantial loss of ventral

982 interneurons. (**D-F**) Comparison of GFAP-positive astrocytes within the ventral spinal cord showed

983 reactive gliosis in (**E**) 13R1, but not in (**D**) wild type or (**F**) 3R1. Specific interneuron populations were
984 quantified, including (**G-I,P**) FoxP2+ and (**J-L,Q**) Pax2+ ventral interneurons. Both populations were
985 significantly reduced in 13R1 mutants compared to wild type or 3R1 animals (**P**, **Q**). The horizontal white
986 line in **J** indicates the divide used to quantify ventral interneurons. **M-O**) Staining for cleaved caspase 3
987 (CC3) reveals more apoptotic cells in (**N**) 13R1 than in (**M**) wild type or (**O**) 3R1 mutants (quantified in **R**).
988 Scale bar is 200 μm. ** = $p < 0.01$; *** = $p < 0.001$ by Dunnett's multiple comparison test comparing the
989 indicated genotype with wild type. n=32-48 hemicords total from 3 animals per genotype. Box plots
990 represent the median, first and third quartile, range, and outliers. See also Figure 4-figure supplement 1.
991

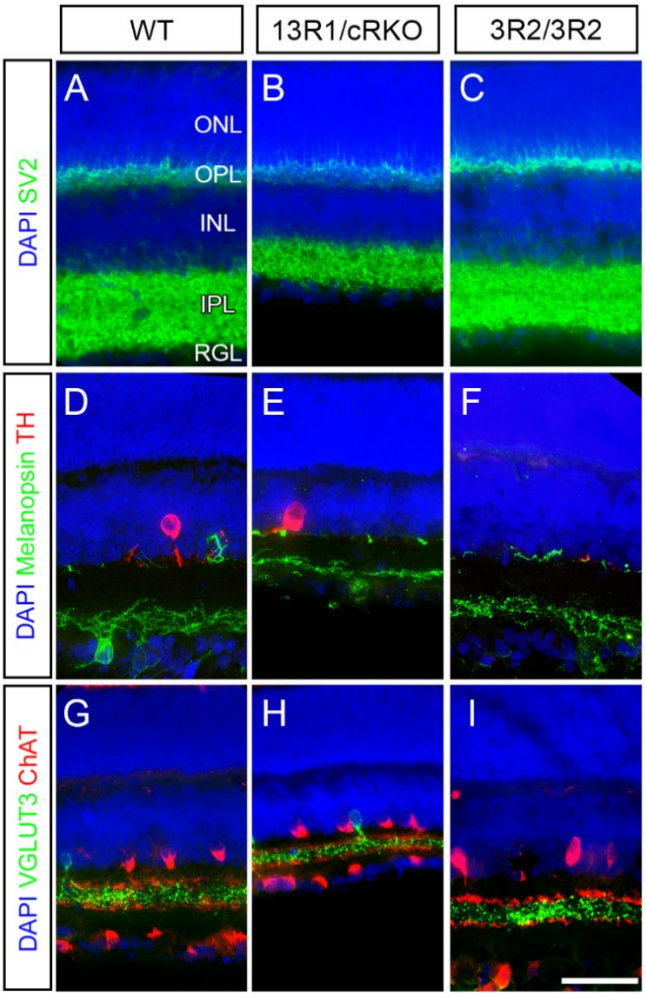


Figure 5: Substantial loss of retinal thickness without layer disorganization in 13R1 mutants. A-C) Cryosections taken in cross section through retinas of (**A**) wild type, (**B**) 13R1/cRKO, and (**C**) 3R2 mutants at two weeks of age were immunostained for SV2 to label the synaptic layers (outer and inner plexiform layers; OPL, IPL) and DAPI to label the inner and outer nuclear layers (ONL, INL) and retinal ganglion cell layer (RGL). The IPL and INL were both notably thinner in 13R1/cRKO than in wild type or 3R2 mutants. This was not accompanied by disorganization of neuronal subtype stratification within the IPL, as revealed by (**D-F**) Melanopsin and TH immunolabeling of ipRGCs and dopaminergic amacrine cells, respectively, and (**G-I**) VGLUT3 and ChAT immunolabeling of glutamatergic amacrine cells and starburst amacrine cells, respectively. Scale bar is 50 μm. Images are representative of at least 6 retinas per genotype analyzed.

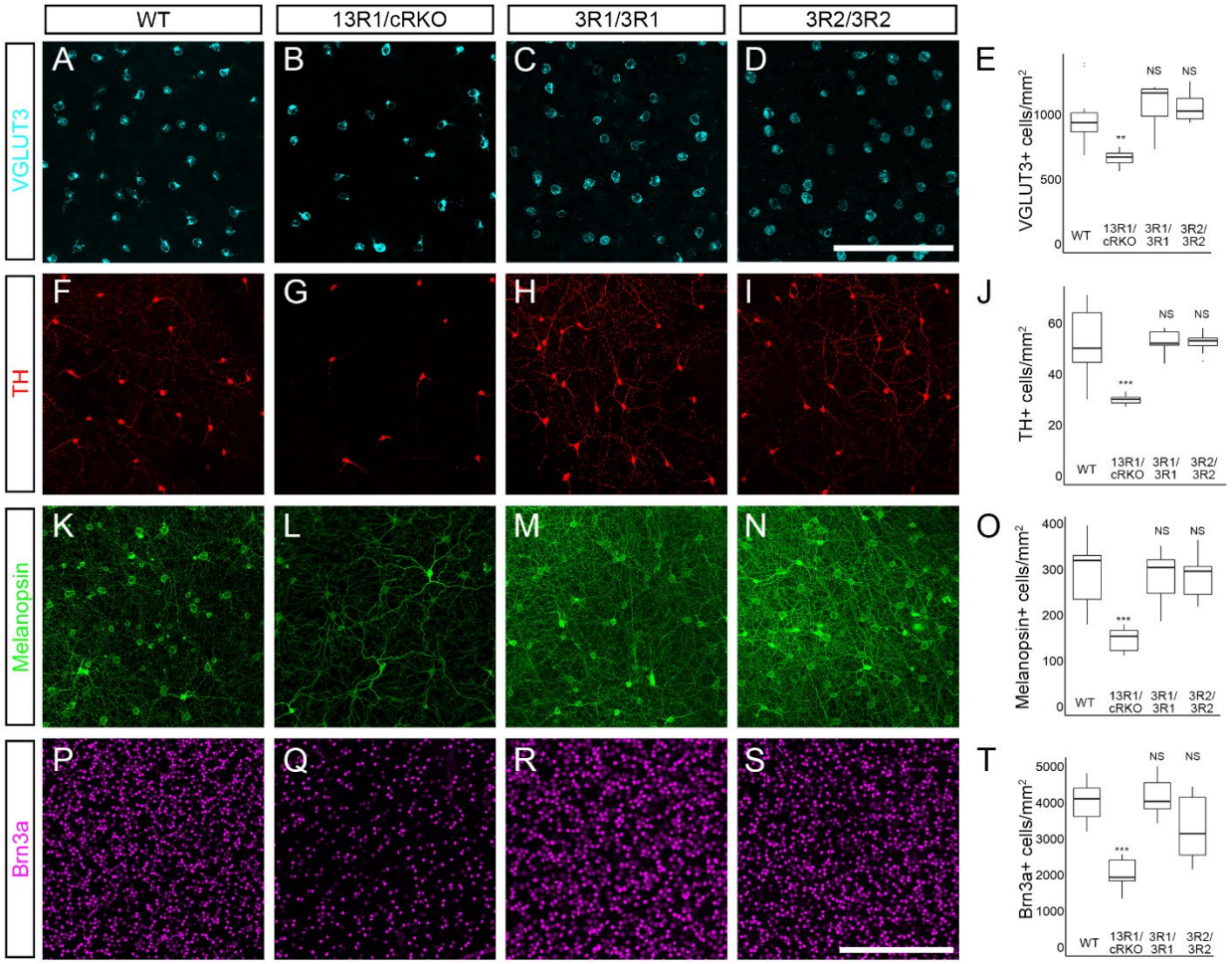


Figure 6: Reduced retinal neuron numbers in 13R1 mutants. Whole mount retinas from two-week-old mice were immunostained to label neuronal subtypes, imaged *en face* by confocal microscopy, and cell densities quantified. Analyzed cell types include (**A-E**) VGLUT3+ amacrine cells, (**F-J**) TH+ dopaminergic amacrine cells, (**K-O**) Melanopsin+ ipRGCs, and (**P-T**) Brn3a+ RGCs. For all four of the subtypes assayed, densities were significantly reduced in 13R1/cRKO retinas compared to wild-type. Densities were not reduced in either 3R1 or 3R2. Scale bar in **D** is 100 μm (for **A-D**); scale bar in **S** is 300 μm (for **F-S**). ** = $p < 0.01$; *** = $p < 0.001$ by Tukey post-hoc test comparing the indicated genotype with wild type. n=6 retinas per genotype. Box plots represent the median, first and third quartile, range, and outliers.

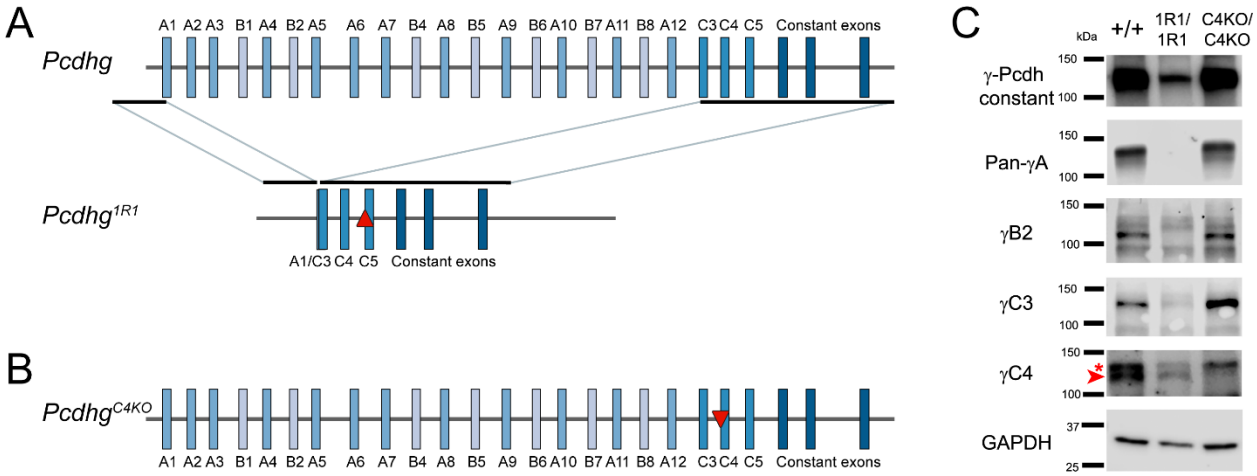


Figure 7: Two further mouse alleles that indicate γC4 is the sole γ-Pcdh isoform necessary and sufficient for postnatal survival. A) A schematic of the 1R1 allele illustrates a large deletion from the sgRNA guide site in exon A1 to that in exon C3, as well as a frame-shifting insertion in exon C5 (red upward triangle). Only exon C4 remained intact. All mutations were verified by whole-genome sequencing. **B**) C4KO mutants were made by targeting exon C4 only for CRISPR/Cas9 genome editing, resulting in a 13 base pair deletion (red downward triangle). **C**) Western blot analyses on P0 cortex confirmed the protein expression expected from sequence analysis. Particularly, γC4 was present in 1R1 mutants (arrowhead, specific lower band of expected size; asterisk indicates larger non-specific band) but absent in C4KO animals, while all other isoforms analyzed were absent in 1R1 but present in C4KO. See also Figure 7-figure supplement 1, Figure 7-figure supplement 2, and Figure 7-figure supplement 3.

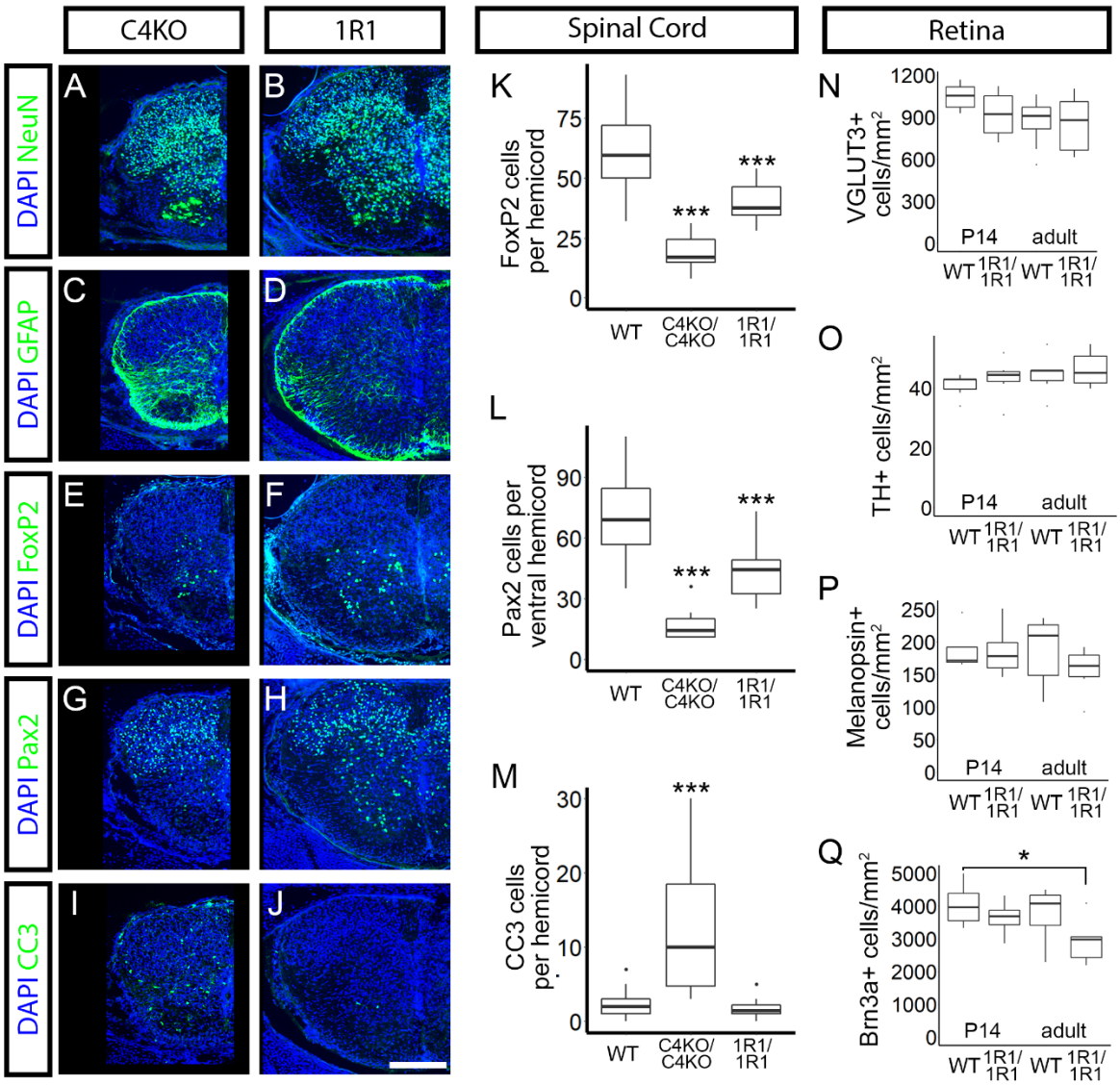


Figure 8: The isoform γC4 is necessary for normal neuronal survival. Cryosections from P0 spinal cords were immunostained for the indicated markers (green; DAPI counterstain for nuclei, blue) **A-B**) Immunostaing for NeuN revealed grossly smaller spinal cords and neuronal loss in C4KO mutants, but not 1R1 animals. **C-D**) This was accompanied by reactive gliosis in C4KO. **E-H**) Interneuron subtypes were analyzed as in Figure 4 (**E-F**, FoxP2 and **G-H**, Pax2). **K-L**) Ventral interneurons were drastically reduced in C4KO mutants; decreases in 1R1 mutants were statistically significant, but substantially more modest. **I-J**) Apoptosis was increased in C4KO animals, but not in 1R1 mutants, as demonstrated by CC3

immunolabeling (quantified in **M**). **N-Q**) Whole mount retinas from 1R1 mutants were assayed for neuronal loss at 2 weeks of age and in adult as in Figure 6. There were no reductions in (**N**) VGLUT3+ amacrine cells, (**O**) TH+ dopaminergic amacrine cells, or (**P**) Melanopsin+ ipRGCs. **Q**) At 2 weeks of age, Brn3b+ RGCs were also not reduced, but their density was significantly lower in adult retinas. Scale bar is 200 μm. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ by Dunnett (spinal cord) or Tukey (retina) test comparing the indicated genotype with wild type. Wild type values in K-M are represented from Figure 4. n=36 hemicords total from 3 animals per genotype or n=6 retinas per genotype. Box plots represent the median, first and third quartile, range, and outliers. See also Figure 8-figure supplement 1.

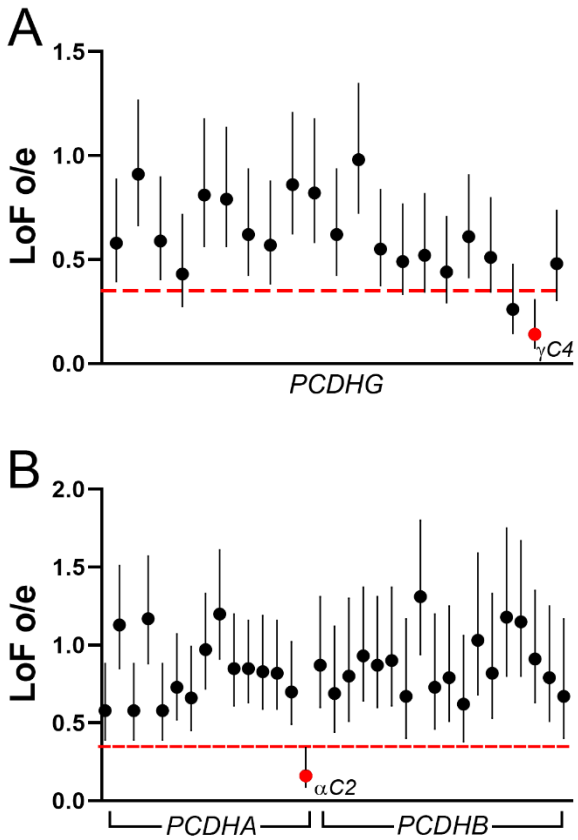


Figure 9: *PCDHGC4* is constrained in humans. The gnomAD database (Karczewski et al., 2019; https://gnomad.broadinstitute.org/) was queried for each cPcdh isoform to determine the extent to which loss-of-function (LoF) mutations were tolerated within the human population. **A)** The ratio of observed to expected (o/e) LoF mutations (y-axis) surrounded by its 90% CI is plotted for each *PCDHG* isoform (x-axis). *PCDHGC4* (red circle) is the only isoform in which the upper bound of the 90% CI falls below 0.35 (red line), indicating constraint. **B)** LoF o/e (y-axis) is plotted for each isoform from the *PCDHA* and *PCDHB* cluster (x-axis). *PCDHAC2* is the only constrained isoform (red circle).

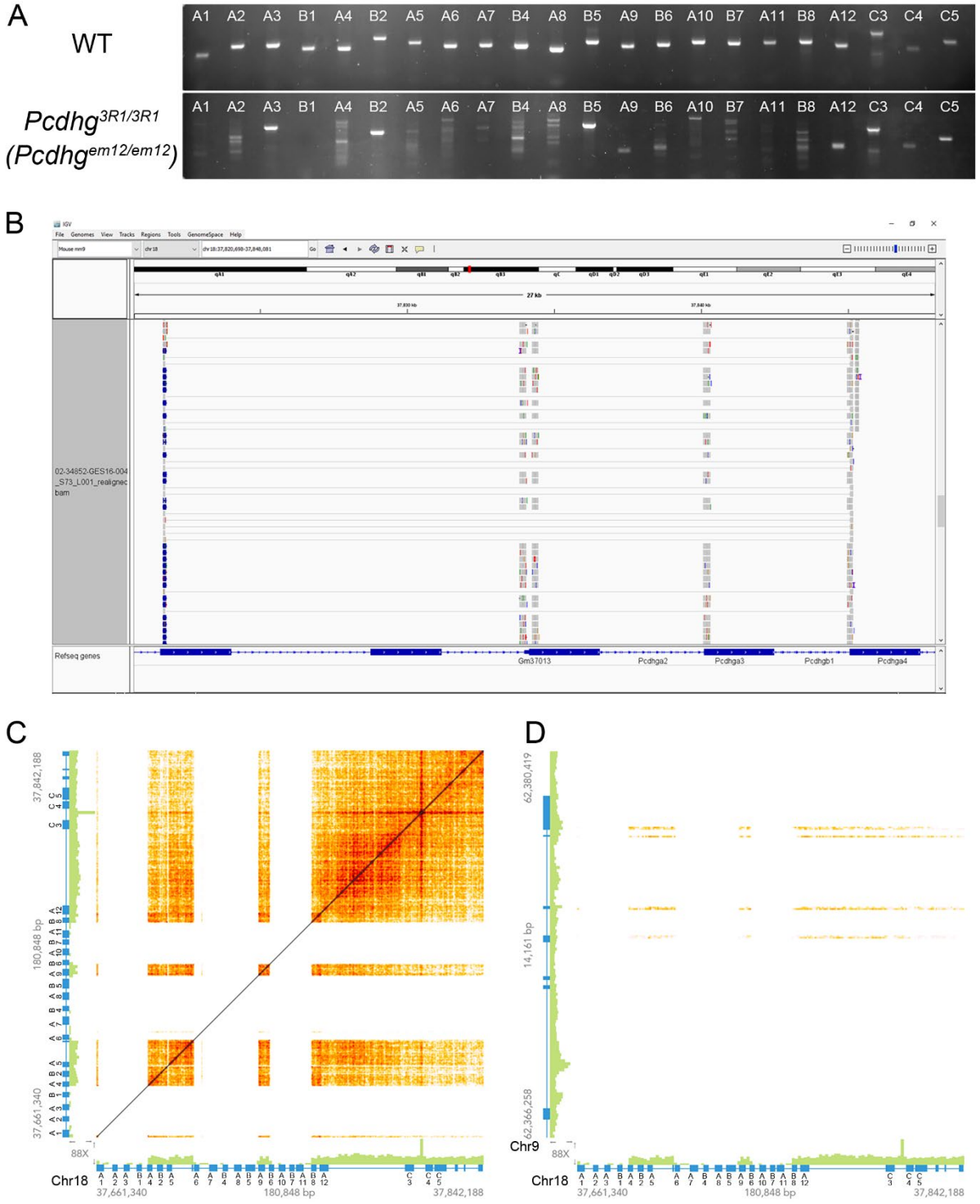


Figure 1-figure supplement 1: Workflow for identifying mutations with 3R1 as an example. A) PCR from genomic DNA in 3R1 homozygous mutants compared with wild type, using primer pairs spanning

each sgRNA target site. PCR products were either absent or abnormal in size from more exons than predicted by the custom amplicon sequencing analysis. **B**) Amplicon sequencing reads from a heterozygous 1R1 mutant were aligned to the mouse genome and visualized in the Integrated Genome Viewer (IGV) with paired-end reads highlighted. In this example, multiple read pairs map with one end in exon A1 and the other in exon A4, indicating a junction between A1 and A4 that was not detected by BreaKmer. **C**) Linked-read whole genome sequencing was performed on DNA from a homozygous 3R1 mutant. The matrix represents the locations of mapped reads with common barcodes, indicating that they originated from common DNA fragments within a microfluidic droplet. The *Pcdhg* locus is presented on the X and Y axes, revealing three large deletions including the deletion resulting in the A1-A4 fusion found in **B**. **D**). Linked-read sequencing also revealed that *Anp32a* coding sequence incorporated into the Pcdhg locus, as reads mapping to its exons (but not introns) had barcodes in common with those mapping to *Pcdhg* (*Anp32a* is on the Y axis, *Pcdhg* on the X axis).

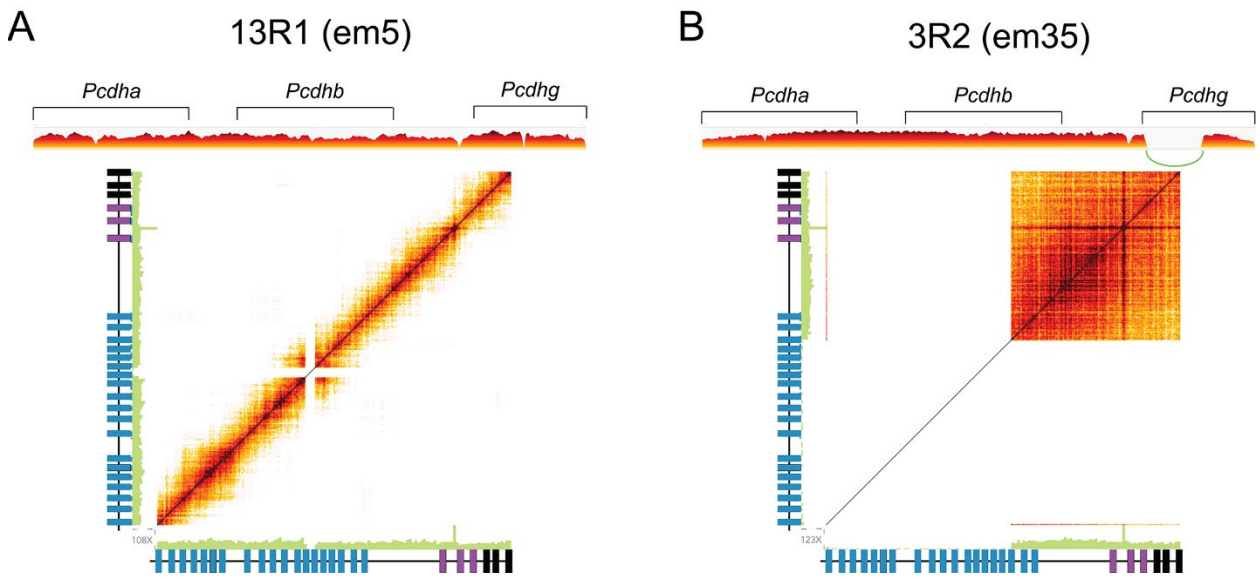


Figure 2-figure supplement 1: Linked-read whole genome sequencing of two further *Pcdhg* mutant lines. 10X Chromium linked-read sequencing results from (**A**) 13R1 and (**B**) 3R2 mutants demonstrate normal coverage through the *Pcdha* and *Pcdhb* clusters (upper panel), but coverage gaps where sequence was deleted in the *Pcdhg* locus (upper and lower panels). Short reads with the same barcode (i.e., from the same initial larger fragment) are connected on the matrix in the lower panel. Actual read sequences were used to identify junctions.

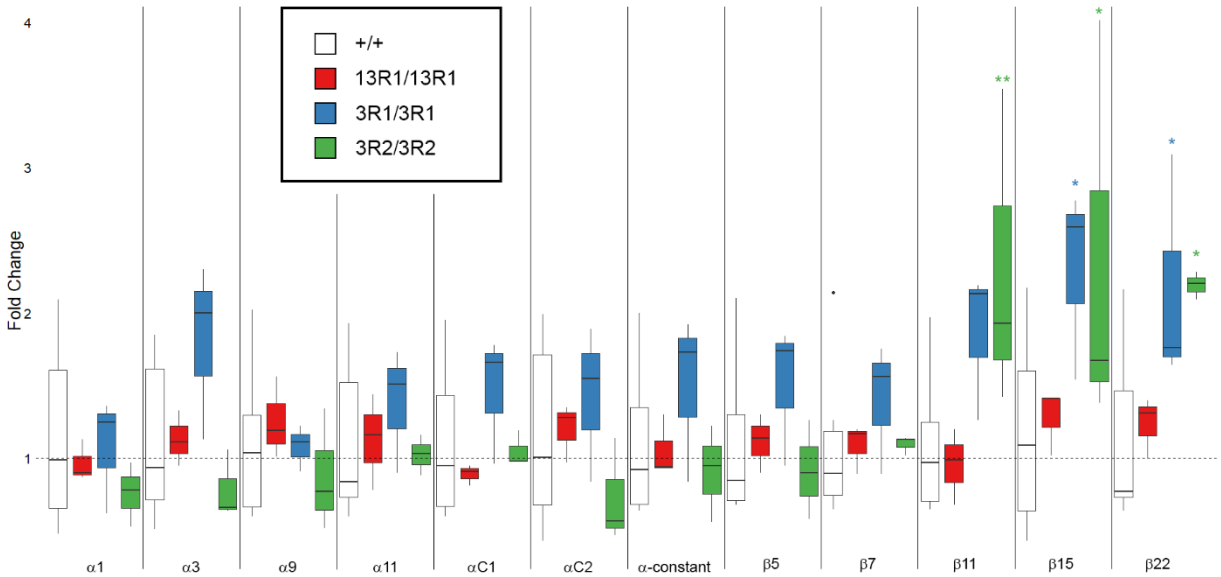


Figure 3-figure supplement 1: Large deletions in *Pcdhg* increased mRNA expression from the 3' end of *Pcdhb*. Quantitative real-time PCR of cDNA reverse-transcribed from RNA isolated form cerebral cortex from 13R1 mutants (red), 3R1 mutants (blue) and 3R2 mutants (green) demonstrated no change in isoform expression from the *Pcdha* cluster genes analyzed. *Pcdhb* isoforms at the 3' end of the cluster were increased in mutants with large deletions in *Pcdhg* (i.e., in 3R1 and 3R2 mutants, but not in 13R1 mutants). * = $p < 0.05$; ** = $p < 0.01$ by Tukey post-hoc test comparing the indicated genotype with wild type. n=3-9 animals per genotype. Box plots represent the median, first and third quartile, range, and outliers.

Name	sequence	note
Pcdha1QF	CCCAGGTTTGAACATAGC	Used with PcdhaQR
Pcdha3QF	GACAAACTGGTTGGAGACAT	Used with PcdhaQR
Pcdha9QF	TACACCTTGCCCAGTGG	Used with PcdhaQR
Pcdha11QF	CTGGGTAGAGATGAAAGGG	Used with PcdhaQR
PcdhaC1QF	GGGGATCATTCAAATGTGGA	Used with PcdhaQR
PcdhaC2QF	CAACAGGCAACTCACCG	Used with PcdhaQR
PcdhaQR	CGAGGCAGAGTAGCGCC	
PcdhaconQF	CGCTACTCTGCCTCGCTAA	
PcdhaconQR	GCTGTTGCTGTTGACACCG	
Pcdhb5QF	GGTGATAGAGGTGTTGGATGTG	
Pcdhb5QR	GATGATAGCGACCACAGTTTCT	
Pcdhb7QF	CTGGGAAAGGCTTGGTAATAGT	
Pcdhb7QR	GAGCTGGTCAGTGAAGATATGG	
Pcdhb11QF	AGCTGGATTTCGAGGGAATTAG	
Pcdhb11QR	CTCTGGAGCGTTGTCATTCA	
Pcdhb15QF	TGTCCCAGAACTACCAGTATGA	
Pcdhb15QR	GTGCTTTCGGTTAGGAAGTTAGA	
Pcdhb22QF	CTGGAACCTTCCGTTGAGAA	
Pcdhb22QR	GTCGGTGACGGAGATAGTAATG	
PcdhgA1QF	AGGAATTTTTGTCAGCACCCC	Used with PcdhgQR1
PcdhgA2QF	GATTTCCTCTCAGCACCTCAG	Used with PcdhgQR1
PcdhgA3QF	GCGAGCCTCTTATAATACCTCAAG	Used with PcdhgQR1
PcdhgA4QF	AGTGATCCTCTCCTGGTATCTC	Used with PcdhgQR1
PcdhgA5QF	ACACACAAAGAAGAGCCCG	Used with PcdhgQR2
PcdhgA6QF	TGCAAAGAGGAAGACTCTCTTGA	Used with PcdhgQR2
PcdhgA7QF	GAAGCCCCAAGTTCCCAG	Used with PcdhgQR2
PcdhgA8QF	GAATAAGGATGAAGATGCTTGCG	Used with PcdhgQR2
PcdhgA9QF	ACACTCCTTTGGTTCCTCAAG	Used with PcdhgQR2
PcdhgA10QF	TGATTCCAAGTGTCCTGTAGAAG	Used with PcdhgQR2
PcdhgA11QF	ACCGACAGATATTCAGCAAGC	Used with PcdhgQR2
PcdhgA12QF	CATGCATTAAATCAGCAAGCCC	Used with PcdhgQR2
PcdhgB1QF	TCCTTCTGTGGTTGTATGTGG	Used with PcdhgQR1
PcdhgB2QF	TGTCTCGGATTCTATCTCAAAGC	Used with PcdhgQR2
PcdhgB4QF	AATGAGTCAACCTCCCATCAG	Used with PcdhgQR2
PcdhgB5QF	TTCCCACCCCGAACCTCTAA	Used with PcdhgQR2
PcdhgB6QF	TTCCGGTAGTTCTCCTGGGG	Used with PcdhgQR1
PcdhgB7QF	TCCAGCCGCACAAGATATTC	Used with PcdhgQR1
PcdhgB8QF	CGAGACCTTTGTACGGAAGC	Used with PcdhgQR1
PcdhgC3QF	CTATAGACAGGTGTTGGGTGC	Used with PcdhgQR2
PcdhgC4QF	GTCCACCCTCTGATCTTCTCTA	Used with PcdhgQR2
PcdhgC5QF	CTTCACTTTCCTCAGGCCC	Used with PcdhgQR1
PcdhgQR1	TTGAGAGAAACGCCAGTCAG	
PcdhgQR2	GCATCTCTGTATCAAACTGGTTG	
PcdhgconQF	CCAACACTGACTGGCGTTTC	
PcdhgconQR	ATGGCTTGCAGCATCTCTGT	

Figure 3 – figure supplement 2: Primers used for quantitative RT-PCR

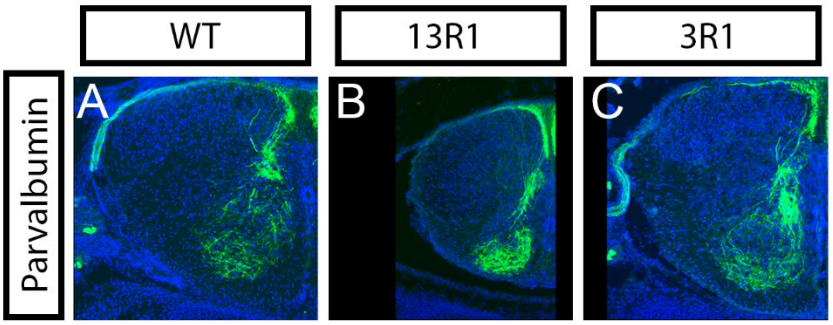
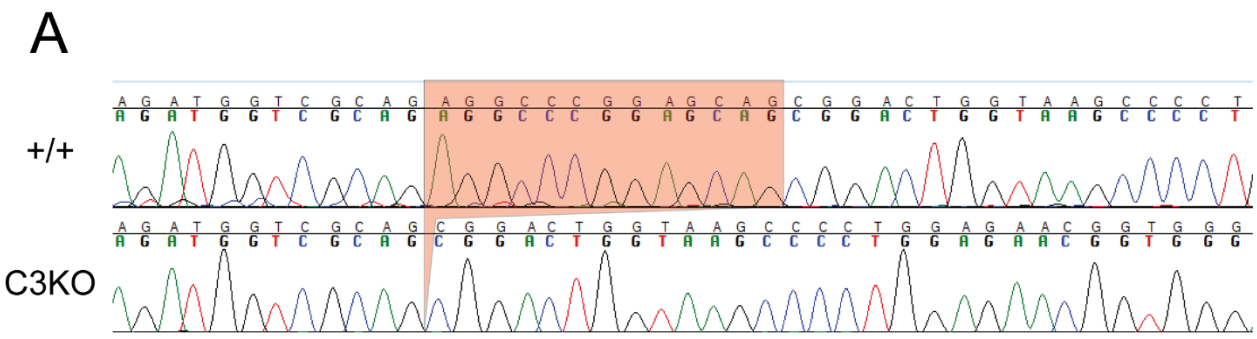
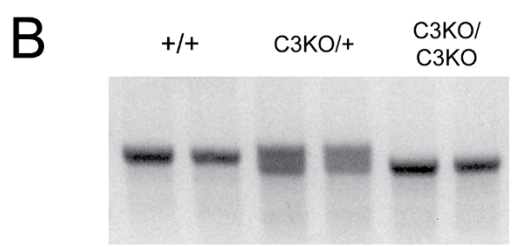


Figure 4-figure supplement 1: Aggregation of Ia afferent axon terminals in mutants with interneuron apoptosis. Parvalbumin staining of proprioceptive Ia afferent axons in spinal cord sections from P0 **A**) wild type, **B**) 13R1 homozygous mutants, and **C**) 3R1 homozygous mutants reveals axon terminal clumping in 13R1, but not 3R1 or wild type animals.





C

Genotype	Observed	Expected
+/+	32	29.5
C3KO/+	61	59
C3KO/C3KO	25	29.5
Total	118	118

$\chi^2 = 0.62$

1093

1094 **Figure 7-figure supplement 1: *Pcdhgc3* homozygous mutants survive at expected ratios. A**)

1095 CRISPR/Cas9 targeting of *Pcdhgc3* only resulted in a 13 bp deletion in the *Pcdhg*^{C3KO} allele (referred to as

1096 C3KO here). **B**) Tail genotyping PCR spanning the deletion was used to identify wild type, heterozygous,

1097 and homozygous mutants. **C**) Homozygous C3KO mutants survive in numbers not significantly different

1098 from the expected Mendelian ratio (n=118 offspring from 18 litters).

1099

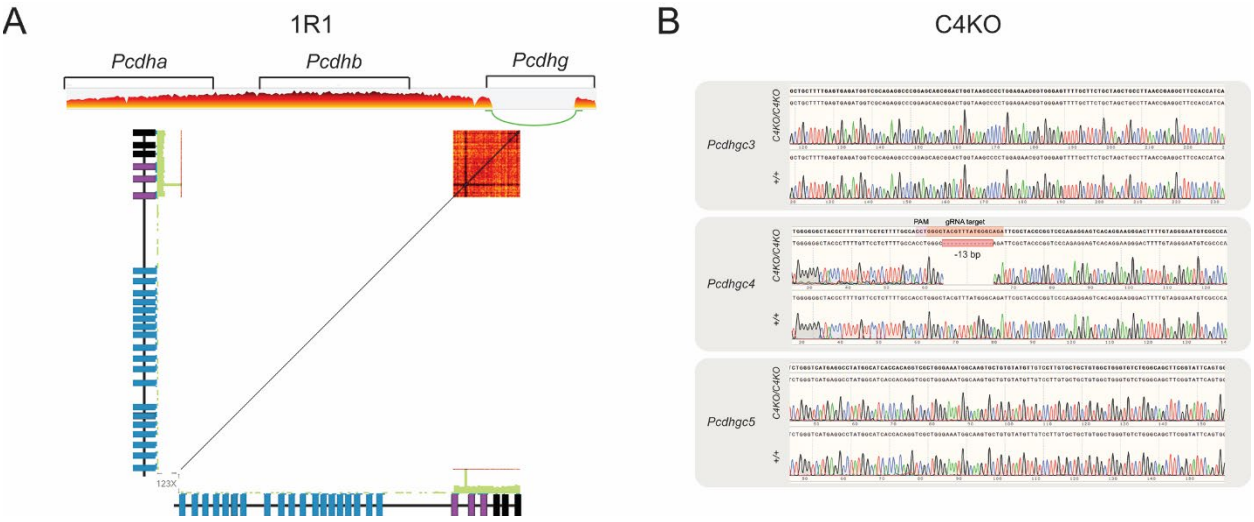


Figure 7-figure supplement 2: Confirmation of 1R1 and C4KO mutations. A) 10X Chromium linked-read sequencing results from 1R1 homozygous mutants demonstrates normal coverage through the *Pcdha* and *Pcdhb* clusters (upper panel), but a large gap between *Pcdhga1* and *Pcdhgc3* (upper and lower panels). Short reads with the same barcode (i.e., from the same initial larger fragment) are connected on the matrix in the lower panel. Actual read sequences were used to identify the junction. **B**) Sanger sequencing was performed on PCR from genomic DNA from C4KO homozygous mutants. A frame-shifting 13 bp deletion was identified at the guide site in *Pcdhgc4*, but no mutations were found in any of the other isoforms (*Pcdhgc3* and *Pcdhgc5* are shown here, all other isoforms were also sequenced).

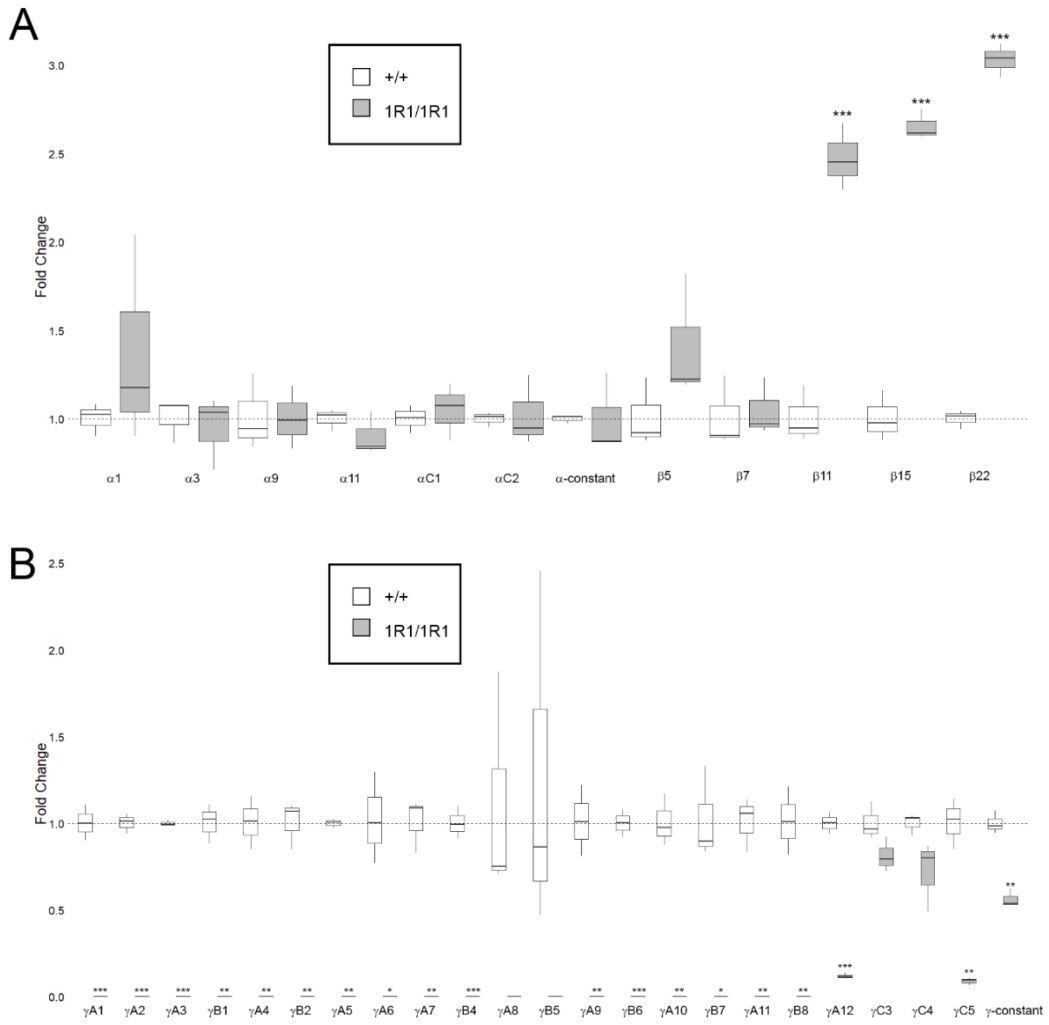


Figure 7-figure supplement 3: Quantitative RT-PCR from 1R1 mutants. A) Quantitative real-time PCR of cDNA reverse-transcribed from RNA isolated form cerebral cortex from 1R1 mutants (gray) compared to control (white) demonstrated no change in isoform expression from the *Pcdha* cluster. Expression of *Pcdhb* isoforms at the 3' end of the cluster was increased in 1R1 mutants, consistent with the effect from large deletions in 3R1 and 3R2 mutants. **B**) Expression of the *Pcdhg* cluster reflected genomic mutations, including expression from the γA1-γC3 fusion. γC5 isoform expression was significantly reduced (disrupted by 1 base pair insertion), and total γ-constant expression was reduced by half. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ by student's t-test. n=3 animals per genotype. Box plots represent the median, first and third quartile, range, and outliers.

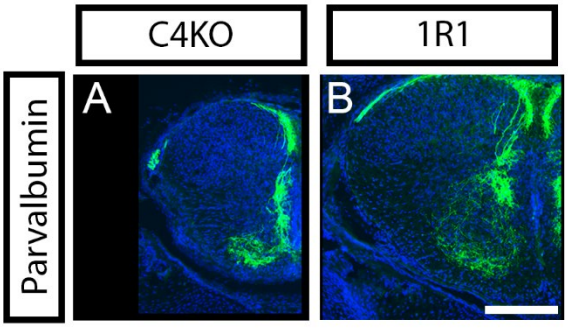


Figure 8-figure supplement 1: γC4 is necessary and sufficient among γ-Pcdh isoforms for normal Ia afferent terminal targeting. A) Parvalbumin staining in spinal cord sections from P0 C4KO mutants revealed aggregation of proprioceptive Ia afferent axons comparable to that observed in null animals or in 13R1. **B**) Conversely, 1R1 homozygous mutants exhibited normal terminal morphology similar to wild type, 3R1, and 3R2 pups.

Allele	Name here	A1	A2	A3	B1	A4	B2	A5	A6	A7	B4	A8	B5	A9	B6	A10	B7	A11	B8	A12	C3	C4	C5	Intact exons
Pcdhg^{em4}				X	X			X																19
Pcdhg^{em5}	13R1	X			X			X		X				[--	---	--]	X					X		13
Pcdhg^{em6}		[--	---	--]				IF								X	IF	X	IF					17
Pcdhg^{em7}				X												X						X		19
Pcdhg^{em8}	1R1	[--	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--]		X	1
Pcdhg^{em9}		[--	---	---	---	---	---	--]	IF				X		IF					X				13
Pcdhg^{em10}								IF								X					IF	IF	X	20
Pcdhg^{em11}		IF						X																21
Pcdhg^{em12}	3R1	[--	---	---	---	--]	X	[--	---	---	---	---	---	---	---	---	---	---	---	--]				3
Pcdhg^{em13}		X	X	X									X											18
Pcdhg^{em14}		X		X				X				X												18
Pcdhg^{em16}													X	X							X			19
Pcdhg^{em17}		[--	---	---	---	---	---	--]							IF					X	X			13
Pcdhg^{em19}					[--	---	---	---	---	---	---	---	---	---	---	---	---	---	--]	X				6
Pcdhg^{em23}		[--	---	---	---	---	---	---	---	--]														13
Pcdhg^{em24}																			X					21
Pcdhg^{em25}			X	X												X		[--	---	--]				16
Pcdhg^{em27}		[--	---	---	---	---	---	---	---	---	--]													12
Pcdhg^{em28}		X		X													X							19
Pcdhg^{em31}						X														X	IF	IF	X	19
Pcdhg^{em32}				IF	[--	---	---	---	--]	X														16
Pcdhg^{em33}		X	X	[--	---	---	---	--]			[--	---	---	---	---	---	---	---	--]	X				5
Pcdhg^{em34}		[--	---	---	---	---	---	--]					IF		IF	X								14
Pcdhg^{em35}	3R2	[--	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--]		X		IF	X	3
Pcdhg^{em42}		X											IF											21
Pcdhg^{em44}																					X			21

IF	= in-frame indel
X	= frame-shifting indel
[-- --- --]	= rearrangement between breakpoints

Table 1: Summary of *Pcdhg* exons disrupted in CRISPR/Cas9 targeted strains. Individual mouse lines are summarized with the allele name along with the name used to describe the lines analyzed here. Red boxes indicate disrupted exons – "X" for discrete frame-shifting indel and dashed lines for large scale rearrangements or deletions between the indicated exons. "IF" indicates an in-frame indel predicted to result in an expressed protein.

F	sgRNA conc.	Allele	A1	A2	A3	B1	A4	B2	A5	A6	A7	B4	A8	B5	A9	B6	A10	B7	A11	B8	A12	C3	C4	C5
55-2	50 ng/µl	*Pcdhg^{em4}*			-2,-3	-5			-5,1,-1															
54-2	50 ng/µl	*Pcdhg^{em5}*	-2			-5			-34		2				[---	---	---]	-5					-5	
53-5	50 ng/µl	*Pcdhg^{em6}*	[---	---	------]				-6								-5	-6	-7	-9				
53-5	50 ng/µl	*Pcdhg^{em7}*			-4,-6												-70						-5	
52-4	50 ng/µl	*Pcdhg^{em8}*	[---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---]		1
52-4	50 ng/µl	*Pcdhg^{em9}*	[---	---	---	---	---	---	---]	3				2,7		-9					-10			
55-3	50 ng/µl	*Pcdhg^{em10}*							3								-70					-9	-9	-7
55-3	50 ng/µl	*Pcdhg^{em11}*	-12						-11															
54-3	50 ng/µl	*Pcdhg^{em12}*	[---	---	---	---	---]	52	[---	---	---	---	---	---	---	---	---	---	---	---]	---]			
43-4	10 ng/µl	*Pcdhg^{em13}*	-1	-76	-6,2									-1										
42-5	10 ng/µl	*Pcdhg^{em14}*	1		-2,-3				-3,-1,-6				-2,3,-1											
37-2	10 ng/µl	*Pcdhg^{em16}*												-19	-10							-2		
37-2	10 ng/µl	*Pcdhg^{em17}*	[---	---	---	---	---	---	---]							3					-10	-2		
37-2	10 ng/µl	*Pcdhg^{em19}*				[---	---	---	---	---	---	---	---	---	---	---	---	---	---	---]	-10			
54-3	50 ng/µl	*Pcdhg^{em23}*	[---	---	---	---	---	---	---	---	---]													
54-3	50 ng/µl	*Pcdhg^{em24}*																		-1				
54-1	50 ng/µl	*Pcdhg^{em25}*		-1	-1												-71		[---	---	---]			
55-1	50 ng/µl	*Pcdhg^{em27}*	[---	---	---	[---	---	---	---	---]	---	---]												
54-1	50 ng/µl	*Pcdhg^{em28}*	1		-7													-10						
55-3	50 ng/µl	*Pcdhg^{em31}*					3,-1														-10	-9	-9	-7
52-4	50 ng/µl	*Pcdhg^{em32}*			-3	[---	---	---	-6	---]	16													
43-4	10 ng/µl	*Pcdhg^{em33}*	1	-16	[---	---	---	---	------]			[---	---	---	---	---	---	---	---	---	---]	-10		
52-4	50 ng/µl	*Pcdhg^{em34}*	[---	---	---	---	---	---	------]					9		-9	-71							
55-3	50 ng/µl	*Pcdhg^{em35}*	[---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---]		-10		-9	-7
52-1	50 ng/µl	*Pcdhg^{em42}*	7											-6										
55-2	50 ng/µl	*Pcdhg^{em44}*																				-2		

N	= Custom amplicon analysis (GATK, above threshold)
N	= Custom amplicon analysis (GATK, below threshold)
N	= Custom amplicon analysis (Breakmer)
N	= Custom amplicon analysis (visual inspection of alignment)
N	= Whole genome sequencing

Table 1 – Table supplement 1: Methods used to identify mutations. Mutations were identified by analyzing sequence results from the custom amplicon using GATK (blue or purple), Breakmer (orange), or visual inspection of aligned reads (yellow), or by analyzing whole genome sequencing (green). Numbers indicate the size of insertion (positive numbers) or deletion (negative numbers). Dashed lines indicate large scale rearrangements or deletions between the indicated exons.

1138

exon	sgRNA	PAM
A1	GGGATCCGCATCGTTTCCAG	AGG
A2	GTGCCGGAGGAGATCGACAA	AGG
A3	TTTCAGGCAGATCCGCTACT	CGG
A4	CTCCTTACAAATCAGACCGC	AGG
A5	ACGCTGTGCGCGCCAGGAAA	GGG
A6	GCTCCGCTACTCTATTCCCG	AGG
A7	CATTCTCCTGGGGATGCGGT	GGG
A8	CCCCGCGAGCTGGCGGAGCG	CGG
A9	ACTGCTAGGGATGCTGTGGG	AGG
A10	CCCAGGGCATCCATAAGAAA	AGG
A11	ACAGCGGGATCAGCGCGCGG	CGG
A12	AACGGCGGGAGGACTGCAAA	GGG
B1	GCCCATCCACTACTCCATTC	CGG
B2	CGGTCCGGCCTGGTGGCTGC	AGG
B4/B5	GGCACTGCTGAGAGGGGCCG	GGG
B6/B7	GCGGACCGGCTGGCCCAGCG	CGG
B8	GCAAGTCCGGTACTCTATTC	CGG
C3	GGTAAGCCCCTGGAGAACGG	TGG
C4	TCTGCCCATAAACGTAGCCC	AGG
C5	TGGCATCACCACAGGTCGCT	GGG

1139

1140 **Table 2: sgRNA sequences**

1141

Allele name here	Strain name	Stock number
Pcdhg^{em4}	C57BL/6J-Pcdhg<em4Rwb>/Rwb	29164
Pcdhg^{em5} (Pcdhg^{13R1})	C57BL/6J-Pcdhg<em5Rwb>/Rwb	29165
Pcdhg^{em6}	C57BL/6J-Pcdhg<em6Rwb>/Rwb	29166
Pcdhg^{em7}	C57BL/6J-Pcdhg<em7Rwb>/Rwb	29167
Pcdhg^{em8} (Pcdhg^{1R1})	C57BL/6J-Pcdhg<em8Rwb>/Rwb	29168
Pcdhg^{em9}	C57BL/6J-Pcdhg<em9Rwb>/Rwb	29169
Pcdhg^{em10}	C57BL/6J-Pcdhg<em10Rwb>/Rwb	29170
Pcdhg^{em11}	C57BL/6J-Pcdhg<em11Rwb>/Rwb	29171
Pcdhg^{em12} (Pcdhg^{3R1})	C57BL/6J-Pcdhg<em12Rwb>/Rwb	29172
Pcdhg^{em13}	C57BL/6J-Pcdhg<em13Rwb>/Rwb	29173
Pcdhg^{em14}	C57BL/6J-Pcdhg<em14Rwb>/Rwb	29174
Pcdhg^{em16}	C57BL/6J-Pcdhg<em16Rwb>/Rwb	29176
Pcdhg^{em17}	C57BL/6J-Pcdhg<em17Rwb>/Rwb	29177
Pcdhg^{em19}	C57BL/6J-Pcdhg<em19Rwb>/Rwb	29179
Pcdhg^{em23}	C57BL/6J-Pcdhg<em23Rwb>/Rwb	29183
Pcdhg^{em24}	C57BL/6J-Pcdhg<em24Rwb>/Rwb	29184
Pcdhg^{em25}	C57BL/6J-Pcdhg<em25Rwb>/Rwb	29185
Pcdhg^{em27}	C57BL/6J-Pcdhg<em27Rwb>/Rwb	29187
Pcdhg^{em28}	C57BL/6J-Pcdhg<em28Rwb>/Rwb	29188
Pcdhg^{em31}	C57BL/6J-Pcdhg<em31Rwb>/Rwb	29191
Pcdhg^{em32}	C57BL/6J-Pcdhg<em32Rwb>/Rwb	29192
Pcdhg^{em33}	C57BL/6J-Pcdhg<em33Rwb>/Rwb	29193
Pcdhg^{em34}	C57BL/6J-Pcdhg<em34Rwb>/Rwb	29194
Pcdhg^{em35} (Pcdhg^{3R2})	C57BL/6J-Pcdhg<em35Rwb>/Rwb	29195
Pcdhg^{em42}	C57BL/6J-Pcdhg<em42Rwb>/Rwb	29161
Pcdhg^{em44}	C57BL/6J-Pcdhg<em44Rwb>/Rwb	29163
Pcdhg^{em3} (Pcdhg^{C3KO})	C57BL/6J-Pcdhg<em3Rwb>/Rwb	27939
Pcdhg^{em41} (Pcdhg^{C4KO})	C57BL/6J-Pcdhg<em41Rwb>/Rwb	33473

Table 3: Summary of new alleles created here.

Name	sequence
PCDHGA1KF	TCTCTGGAGCTACTGCTGGA
PCDHGA1KR	AGCTCTTCCCGGTCTATCCT
PCDHGA2KF	CCCCACTGTGCTGAAAGCTA
PCDHGA2KR	CCTCCCTGTCTATCCTGCCT
PCDHGA3KF	CCATCGTGGGAATCAGAGGG
PCDHGA3KR	ACACAGCTCTTCCCGGTCTA
PCDHGA4KF	CACCAGCGTCCAACAGTACA
PCDHGA4KR	GATTACCCACCACAGAGCCC
PCDHGA5KF	TCTTGGGGAAAGCCACCATC
PCDHGA5KR	CTGGAGATAATGCGGACCCC
PCDHGA6KF	TGGATTACCAAGTGGGCTGG
PCDHGA6KR	TGGAGATGATGCGAATCCCG
PCDHGA7KF	CGGACACAACTGGGTTCTGA
PCDHGA7KR	ACCTTTGGTGATGATGCGGA
PCDHGA8KF	CAGCGTAGGACAGATTCGCT
PCDHGA8KR	CAAGACAAGGCGTGCTCTGA
PCDHGA9KF	GCTCCGGGACTACAGCAAAA
PCDHGA9KR	GGCACGGAATAGCGGATTTG
PCDHGA10KF	CTTTAAGGTGCTTCCGCTGC
PCDHGA10KR	CCTCTGGAGATGATGCGGAC
PCDHGA11KF	TTTGGATGCGGGATACCTGG
PCDHGA11KR	CCTGGAGATGATGCGGACTC
PCDHGA12KF	CCCTCTAGGAGCAAACTGGC
PCDHGA12KR	AGCCCCAGATCCTTGGAGAT
PCDHGB1KF	TACCTGGCAAGCTCAGAACG
PCDHGB1KR	CCCCACTCTCTGCACTAACG
PCDHGB2KF	AGACGCGTTAGGAAACTGGG
PCDHGB2KR	AATCCTTCTCTGCGCTGACC
PCDHGB4KF	GAGCCCGGAATATCCACACC
PCDHGB4KR	CTTTCGCGTCGGTAACTCCT
PCDHGB5KF	GCAAGGAAAGGGGAGCAAGA
PCDHGB5KR	CACCAGCTTCTTCCCGCATA
PCDHGB6KF	CCTCGTACCCTGCTGTGAAG
PCDHGB6KR	CTCCGAATCCACGCTGAAGT
PCDHGB7KF	ACCCTCGTAGGCTGAACTCT
PCDHGB7KR	CTCCGAATCCACGCTGAAGT
PCDHGB8KF	GAAGCCTCCTCCCTTCACAC
PCDHGB8KR	CAGAGGGCTTTGGAAGCAGA
PCDHGC3KF	CCGGGATGAGGCAGAGACTGAA
PCDHGC3KR	TTACAGTGCAGGAGGGCAGCGT
PCDHGC4KF	AGAATTAGCGGATGGCAGCA
PCDHGC4KR	TTTGGTTCACCTCTCCAGCG
PCDHGC5KF	TTCAGCTTCTGCACTCCAGG
PCDHGC5KR	CAAAGCTCCGCTCACCAAAC

1145 **Table 4: Primers used to amplify target regions of *Pcdhg* variable exons.**

SUPPLEMENTARY INFORMATION

Supplemental File 1: Genomic coordinates of custom amplicon sequencing targets

Supplemental File 2: Primary protein sequences resulting from CRISPR/Cas9-mediated mutations

Supplemental File 3: p-values from all analyses