Expansion of stochastic expression repertoire by tandem duplication in mouse Protocadherin-α cluster

Ryosuke Kaneko1, Manabu Abe2, Takahiro Hirabayashi3, Arikuni Uchimura3, Kenji Sakimura2, Yuchio Yanagawa1,4,5 & Takeshi Yagi3,5

1Bioresource center, Gunma University Graduate School of Medicine, 2Department of Cellular Neurobiology, Brain Research Institute, Niigata University, 3KOKORO-Biology Group, Laboratories for Integrated Biology, Graduate School of Frontier Biosciences, Osaka University, 4Department of Genetic and Behavioral Neuroscience, Gunma University Graduate School of Medicine, 5Japan Science and Technology Agency, Core Research for Evolutional Science and Technology (CREST).

Tandem duplications are concentrated within the Pcdh cluster throughout vertebrate evolution and as copy number variations (CNVs) in human populations, but the effects of tandem duplication in the Pcdh cluster remain elusive. To investigate the effects of tandem duplication in the Pcdh cluster, here we generated and analyzed a new line of the Pcdh cluster mutant mice. In the mutant allele, a 218-kb region containing the Pcdh-a2 to Pcdh-ac2 variable exons with their promoters was duplicated and the individual duplicated Pcdh isoforms can be distinguished. The individual duplicated Pcdh-α isoforms showed diverse expression level with stochastic expression manner, even though those have an identical promoter sequence. Interestingly, the 5'-located duplicated Pcdh-ac2, which is constitutively expressed in the wild-type brain, shifted to stochastic expression accompanied by increased DNA methylation. These results demonstrate that tandem duplication in the Pcdh cluster expands the stochastic expression repertoire irrespective of sequence divergence.

Genetic variations play critical roles in animal evolution and human diseases1–3. These variations involve single nucleotide polymorphisms, small insertions/deletions, and large rearrangements, including inversions, translocations, and copy number variations (CNVs). CNVs involve either a gain (duplication) or a loss (deletion) of DNA segments. Tandem duplications, which are generated by unequal crossover, are frequent within tandemly arrayed gene clusters, which are adjacent groups of paralogous genes. Such clustered genes represent about 14% of vertebrate genes4 and are involved in a variety of important physiological and biochemical functions. They include, for instance, the immunoglobulin and T-cell receptor genes5, HOX genes6, zinc finger genes7, α- and β-globin genes8,9, and olfactory receptor genes10.

Although detailed analyses of these gene clusters have revealed their sophisticated gene regulation mechanisms, for example, somatic DNA rearrangements in the immunoglobulin and T-cell receptor clusters6, collinear expression in the HOX cluster6, and monoallelic and exclusive expression in the olfactory receptor cluster10, the evolutionary events by which these sophisticated gene regulations were acquired remain unclear.

Recently, the protocadherin (Pcdh) cluster was shown to be rich in tandem duplications6 and to exhibit sophisticated gene regulation mechanisms11,12. The Pcdh cluster has been identified in wide range of vertebrate species13–15. The mammalian Pcdh cluster genes are further classified into three subfamilies: Pcdh-α, Pcdh-β, and Pcdh-γ16. The Pcdh cluster genes probably arose by the tandem duplication and sequence divergence of existing Pcdh genes at some undetermined point in vertebrate evolution17,18. In fact, several mammalian Pcdh-α isoforms (α1–α12 in mouse and α1–α13 in human) are orthologous to four coelacanth Pcdh-α isoforms (α11–α14), and numerous mammalian Pcdh-β isoforms (β1–β22 in mouse and β1–β16 in human) are orthologous to four coelacanth Pcdh-β isoforms (β1–β4), suggesting that the common ancestors of those Pcdh-α and β isoforms have become highly expanded in the mammalian lineage via tandem duplications19. The intriguing point about these expansions is that the Pcdh repertoire in mammals is more diverse than that in coelacanth. Furthermore, current human populations show a large number of CNVs in the PCDH locus5,20. However, the relevant ancestral material and human material could not be assessed, thus calling for the development of animal models of tandem duplication in the Pcdh cluster.
The diversity and sophisticated gene regulation exhibited by the Pcdh cluster genes are important for normal development of the nervous system\(^{11,12,21,22}\). The Pcdh cluster genes, which encode a group of diverse cadherin-related transmembrane proteins, are expressed mainly in the nervous system, and gene regulation mechanisms in the Pcdh clusters include both constitutive and stochastic expression in single neurons\(^{23-25}\). Gene ablation studies showed that Pcdh-\(\alpha\) and Pcdh-\(\gamma\) are required for neuronal survival, synapse formation, axonal targeting, dendritic arborization, and self-avoidance of dendrites\(^{11,12,21-25}\). These findings led to the suggestion that the Pcdh cluster genes are likely candidates for the individualization of neurons in the vertebrate brain, which would be generated through the stochastic expression of these genes\(^{11,12,21,22}\). Although these findings suggest that stochastic expression in the Pcdh cluster is important in neurodevelopment, the evolutionary origin of the stochastic expression in the Pcdh cluster has remained a mystery.

To better understand the evolutionary origin of the stochastic expression and CNVs’ effect in the Pcdh cluster, here we focused on the effect of tandem duplication in the mouse Pcdh-\(\alpha\) cluster. In the present study, we engineered a targeted tandem duplication within the mouse Pcdh-\(\alpha\) cluster, a situation somewhat comparable to that occurring during vertebrate evolution and in current human populations. The individual Pcdh-\(\alpha\) isoforms transcribed from each duplicate exon can be distinguished in the mutant mice, enabling us to determine the manner by which their expression was regulated.

The individual duplicated Pcdh-\(\alpha\) isoforms showed diverse expression level with stochastic expression manner, even though those have an identical promoter sequence. Surprisingly, the duplicated Pcdh-\(\alpha2\) isoform, which shows constitutive expression in the wild-type allele, shifted to stochastic expression accompanied by increased DNA methylation. Our results demonstrate that tandem duplication in the Pcdh cluster expands the stochastic expression repertoire irrespective of sequence divergence.

**Results**

**Targeted tandem duplication in the mouse Pcdh-\(\alpha\) cluster.** To study the consequences of tandem duplication in the Pcdh gene cluster, we generated a targeted tandem duplication in the mouse Pcdh-\(\alpha\) cluster using the inter-strain targeted meiotic recombination (iTAMERE) system\(^1\). The wild-type mouse Pcdh-\(\alpha\) cluster contains 14 large ‘variable’ exons, each of which encodes a cadherin-like type I membrane protein consisting of extracellular domains, a transmembrane domain, and a proximal cytoplasmic domain. Each variable exon is expressed from its own promoter and spliced to three short boundary (Fig. 2a). We found that the duplication did not alter the 9boundary and Slc25a2, Taf7, Diap1, and Hdac3, which have been shown to be regulated by the Pcdh-\(\alpha\) cluster. The minority of F1 pups carried the dup(2-c2) allele, in which exons Pcdh-\(\alpha9\) to Pcdh-\(\alpha2\) were duplicated (Fig. 1d), or the del(2-c2) allele, in which exons Pcdh-\(\alpha9\) to Pcdh-\(\alpha2\) were deleted (data not shown). F1 pups carrying these duplication or deletion alleles were obtained in 7% (4 of 69 pups) and 2% (1 of 69 pups), respectively. We then analyzed the tail DNA of the duplication-containing mice by PCR to detect the Cre-mediated duplication alleles, and sequenced the PCR products to confirm the presence of the predicted junction sequences generated by the Cre-mediated site-specific recombination events. Animals homozygous for the duplicated allele (Pcdh\(\alpha\)dup(2-c2)/dup(2-c2)) were obtained by crossing heterozygous (Pcdh\(\alpha\)het/dup(2-c2)/dup(2-c2) parents. The Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mice were born with the expected Mendelian distribution (Supplementary Table S1), developed normally to adulthood, and were fertile. Histochemical analysis with Nissl staining revealed an apparently normal gross anatomy of the Pcdh-\(\alpha\)dup(2-c2)/dup(2-c2) mouse brain (Fig. 1g left and Supplementary Fig. S2a). This finding was further supported by cytochrome oxidase staining showing a normal barrel structure in the Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mice (Supplementary Fig. S2a). Furthermore, neural pathway and serotonergic axon analyses by anti-neurofilament and anti-SERT staining, respectively, showed no obvious differences between the genotypes (Fig. 1g middle and Supplementary Fig. S2b). Finally, the distribution of c-fos mRNA, a well-known marker for neuronal activity, was also similar between the genotypes (Supplementary Fig. S2c). These findings suggested that the tandem duplication of exons Pcdh-\(\alpha9\) to Pcdh-\(\alpha2\) does not result in any deleterious effects on mouse development or brain morphogenesis.

**Tandem duplication maintains the expression level of neighboring genes.** Previous studies have indicated that tandem duplication may alter not only the expression of genes within the duplication boundaries but also of genes located in their genomic neighborhoods\(^3\). Prompted by these observations, we first quantified the expression levels of transcripts in the vicinity of the Pcdh-\(\alpha\) cluster in the cerebellum of 4-week-old Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mice. The analyzed non-Pcdh genes included Wdr55, Dnd1, Hars, and Acsl3. Among these neighboring genes (using Pcdh-\(\alpha\)CR probe), Pcdh-\(\alpha1\), Pcdh-\(\alpha3\), Pcdh-\(\alpha6\), Pcdh-\(\alpha7\), and Pcdh-\(\alpha9\) were constitutively expressed, and Pcdh-\(\alpha2\), Pcdh-\(\alpha10\), Pcdh-\(\alpha12\), and Pcdh-\(\alpha19\) showed constitutive expression in the wild-type (CR1) in the Pcdh-\(\alpha\) cluster to generate the “SR” allele (Fig. 1c and Supplementary Fig. S1). To insert this loxP site into the B6 allele, we used the RENKA ES cell line, which is on a pure B6 genetic background\(^6\).

To duplicate the sequence between the loxP site of the G16Neo allele and that of the SR allele using the Cre-loxP system, we obtained male mice that possessed the G16Neo and SR alleles (G16Neo/SR) and the Sycp1-Cre transgene, which elicits Cre recombinase expression specifically in the testis\(^8\). The male mice were crossed with B6 female mice, and the genotypes of the F1 pups were analyzed by PCR. The minority of F1 pups carried the dup(2-c2) allele, in which exons Pcdh-\(\alpha2\) to Pcdh-\(\alpha2\) were duplicated (Fig. 1d), or the del(2-c2) allele, in which exons Pcdh-\(\alpha2\) to Pcdh-\(\alpha2\) were deleted (data not shown). F1 pups carrying these duplication or deletion alleles were obtained at 5.8% (4 of 69 pups) and 1.4% (1 of 69 pups), respectively. We then analyzed the tail DNA of the duplication-containing mice by PCR to detect the Cre-mediated duplication alleles, and sequenced the PCR products to confirm the presence of the predicted junction sequences generated by the Cre-mediated site-specific recombination events. Animals homozygous for the duplicated allele (Pcdh\(\alpha\)dup(2-c2)/dup(2-c2)) were obtained by crossing heterozygous (Pcdh\(\alpha\)het/dup(2-c2)/dup(2-c2) parents. The Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) pups were born with the expected Mendelian distribution (Supplementary Table S1), developed normally to adulthood, and were fertile. Histochemical analysis with Nissl staining revealed an apparently normal gross anatomy of the Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mouse brain (Fig. 1g left and Supplementary Fig. S2a). This finding was further supported by cytochrome oxidase staining showing a normal barrel structure in the Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mice (Supplementary Fig. S2a). Furthermore, neural pathway and serotonergic axon analyses by anti-neurofilament and anti-SERT staining, respectively, showed no obvious differences between the genotypes (Fig. 1g middle and Supplementary Fig. S2b). Finally, the distribution of c-fos mRNA, a well-known marker for neuronal activity, was also similar between the genotypes (Supplementary Fig. S2c). These findings suggested that the tandem duplication of exons Pcdh-\(\alpha9\) to Pcdh-\(\alpha2\) does not result in any deleterious effects on mouse development or brain morphogenesis.

**Tandem duplication re-allocates the manner of Pcdh-\(\gamma\) expression.** We next examined the distribution of Pcdh transcripts in the cerebellum of 4-week-old Pcdh\(\alpha\)dup(2-c2)/dup(2-c2) mice by in situ hybridization (ISH). We analyzed the expression of all the Pcdh-\(\gamma\) genes (using Pcdh-\(\gamma\)CR probe), Pcdh-\(\gamma1\), Pcdh-\(\gamma3\), Pcdh-\(\gamma2\), Pcdh-\(\beta2\), and all the Pcdh-\(\gamma\) genes (using Pcdh-\(\gamma\)CR probe). Similar
Figure 1 | Targeted tandem duplication in the Pcdh-α cluster. (a) Genomic structure of the Pcdh-α wild-type allele. The Pcdh-α allele consists of variable region exons (1–12, c1 and c2) and constant region exons (CR1–CR3). Each variable region exon is transcribed from its own promoter in a stochastic (black arrows) or constitutive (red arrows) manner. A Pcdh-α transcript is produced from one of the variable region exons and the set of constant region exons by splicing. (b) The G16Neo allele: a loxP site was inserted between Pcdh-α1 and Pcdh-α2. The loxP site is shown as a red triangle. (c) The SR allele: a loxP site was inserted between Pcdh-α2 and CR1. (d) The dup(2-c2) allele: duplication of Pcdh-α2-Pcdh-αc2. The dup(2-c2) allele was produced by Cre-loxP-mediated trans-allelic meiotic recombination between the G16neo and SR alleles. The duplicated segments are shown under the position of the original segments. Importantly, each duplicate Pcdh-α3, Pcdh-α5, Pcdh-α6, Pcdh-α7, Pcdh-α9, Pcdh-α10, Pcdh-α12, and Pcdh-αc2 could be distinguished by SNP analysis; the 5 genes were from B6 and the 3 genes were from CBA. (e & f) Confirmation of the duplication allele by Southern blotting (e) and PCR (f). G, Histological analysis of the cerebellum of 4-week-old wild-type and Pcdh-α dup(2-c2)/dup(2-c2) mice. Nissl staining (left), immunostaining for neurofilament (middle), and in situ hybridization using a Pcdh-α CR probe (right). Cb1-10, 1st – 10th lobule of the cerebellum; Gra, granule cell layer; Mol, molecular layer; Pur, Purkinje cell layer.
positive signals were observed for all the genes examined between the wild-type and the Pcdhαdup(2-c2) cerebellum (Fig. 1g right and Supplementary Fig. S3). These results suggested that the gene regulatory mechanisms governing the spatial distribution patterns of the Pcdh transcript were maintained in the dup(2-c2) allele.

We next examined whether both the 5’- and 3’-located duplicated Pcdh-αs were expressed. Since each duplicate of Pcdh-α3, Pcdh-α5, Pcdh-α6, Pcdh-α7, Pcdh-α9, Pcdh-α10, Pcdh-α12, and Pcdh-αc2 could be distinguished by single nucleotide polymorphism (SNP) analysis, we focused on these exons and on Pcdh-α1, which was not duplicated. We amplified the cDNA fragments of these Pcdh-α genes using specific primer combinations, and the resultant amplifications were sequenced directly (Fig. 2c). The analysis detected most of the 5’- and 3’-located duplicates of Pcdh-α transcripts.

Next, the expression levels of these duplicated Pcdh-αs were quantified by qRT-PCR and cloning-mediated SNP analysis, which is highly sensitive and yields quantitative data (Fig. 2d). The expression level of the spliced CR transcripts, which are common to all the 5’-located and 3’-located duplicated Pcdh-αs, was unchanged in the Pcdhαdup(2-c2) mice. There were no significant differences in the expression levels of most of the 3’-located duplicated Pcdh-αs (Pcdh-α3, Pcdh-α6, Pcdh-α7, Pcdh-α10, Pcdh-α12, and Pcdh-αc2) compared to wild-type. In contrast, the expression levels of all the 5’-located duplicated Pcdh-αs and the 3’-located duplicated Pcdh-α5 and Pcdh-α9 genes were significantly reduced compared to wild-type. These observations revealed that the expression level of the total Pcdh-α genes was maintained, while that of individual isoforms was altered, indicating that expression re-allocation occurred in the dup(2-c2) allele.

Interestingly, the expression levels of the 5’-located duplicated Pcdh-αs were significantly lower than those of their duplicated 3’-located counterparts. This observation suggested that, despite their identical promoter sequences, each 5’-located and 3’-located duplicated Pcdh-α receives distinct gene-regulation influences. Taken together, these findings indicate that tandem duplication alters the manner of gene regulation in the Pcdh-α cluster.

Stochastic expression of duplicate Pcdh-α genes with identical promoter sequences. To investigate whether the duplicated Pcdh-αs retained their stochastic expression at the single-neuron level, we performed single-cell RT-PCR and SNP analysis on Purkinje cells of the Pcdhαdup(2-c2)F1 mice. The Pcdhαdup(2-c2)F1 mice were F1 mice from a JF1 × Pcdhαdup(2-c2) dup(2-c2) cross. To distinguish the 5’- located exons (B6) and 3’-located exons (CBA) in the dup(2-c2) allele, and exons in the wild-type allele (F1) by SNP analysis, we focused on Pcdh-α3, Pcdh-α5, Pcdh-α7 (Fig. 3).

To analyze the expression of Pcdh-α3, Pcdh-α5, and Pcdh-α7 in the dup(2-c2) allele, single Purkinje cells from 4-week-old Pcdhαdup(2-c2) mice were picked up by glass capillary. Complementary DNA of Pcdh-α3, Pcdh-α5, Pcdh-α7, and Pcp-2 (a marker for Purkinje cells) was synthesized from the single-cell samples in the same tube, and the resulting cDNA was then divided into three tubes and subjected to separate, first-round multiplex PCR analysis. The second round of PCR amplification was carried out individually for each tube and respective primer pairs.
used nested primers for the Pcdh-α3, Pcdh-α5, Pcdh-α7 genes, and for Pcp-2. Finally, each PCR product was subjected to direct sequencing to determine from which exon the transcript was derived: i.e., the wild-type (JF1) allele or the 5'-located (dup-5') or 3'-located (dup-3') exons in the dup(2-c2) allele.

Of the 163 single Purkinje cells analyzed, 45 yielded PCR amplicons of Pcdh-α3, Pcdh-α5, or Pcdh-α7 from the same exons in all three tubes, and all 45 cells were positive for Pcp-2, confirming that they were differentiated Purkinje cells. In addition to three of three specific transcripts from the same exon, some cells showed one or two of three transcripts (for example, Pcdh-α3 in cell #1-37); these findings suggested that the amounts of corresponding transcripts in these cells were low, and therefore we excluded these cells from the following analysis.

For Pcdh-α3, the transcripts were derived from the wild-type exon, 5'-located exon, and 3'-located exon in 1 cell, 2 cells, and 5 cells, respectively: the wild-type (#1-138), 5'-located (#1-76, and #1-153), and 3'-located exon (#1-26, #1-45, #1-87, #1-129, and #1-159). For Pcdh-α5, the transcripts were derived from the wild-type exon, 5'-located exon, and 3'-located exon in 14 cells, 5 cells, and 5 cells, respectively: the wild-type (#1-29, #1-34, #1-36, #1-39, #1-53, #1-63, #1-67, #1-94, #1-119, #1-126, #1-127, #1-153, #1-154, and #1-173), 5'-located (#1-6, #1-26, #1-37, #1-59, and #1-145), and 3'-located exon (#1-17, #1-53, #1-107, #1-114, and #1-116). For Pcdh-α7, the transcripts were derived from the wild-type exon, 5'-located exon, and 3'-located exon in 13 cells, 2 cells, and 4 cells, respectively: the wild-type (#1-2, #1-27, #1-37, #1-38, #1-61, #1-83, #1-106, #1-123, #1-143, #1-148, #1-150, #1-161, and #1-163), 5'-located (#1-44, and #1-127), and 3'-located exon (#1-67, #1-70, #1-112, and #1-145). Only one cell expressed both the wild-type and the 3'-located exon for the same Pcdh-α isoform (Pcdh-α5 in cell #1-53) simultaneously, indicating that Pcdh-α genes are not always expressed monoallelically. Importantly, although it cannot rule out the possibility of small number of analyzed cells and possible lower expression level of the 5'-located exons may reduce the chance of detection, no cell expressed both the 5'-located and the 3'-located exon for the same

Figure 3 | Stochastic expression of duplicated Pcdh-α3, Pcdh-α5, and Pcdh-α7 in single Purkinje cells. (a) Electrophoresis results of single-cell RT-PCR for the Pcdh-α3, Pcdh-α5, or Pcdh-α7 and Pcp-2 genes in individual Purkinje cells. The numbers #1-2 ~ #1-173 designate individual cells. 1–3, tubes into which the cDNA from an individual Purkinje cell was divided; independent PCRs were performed for each tube. This figure shows only the cells that gave PCR products for Pcdh-α3, Pcdh-α5, or Pcdh-α7. (b) SNP analysis to distinguish between Pcdh-α transcripts from the 5'-located (shown as dup-5') and 3'-located (shown as dup-3') isoforms on the dup(2-c2) allele and those from the wild-type allele (designated JF1). (c) Classification of Pcdh-α3, Pcdh-α5, or Pcdh-α7 expressed from the dup(2-c2) allele in individual Purkinje cells. Blue and purple bars indicate the number of Purkinje cells expressing the 5'-located and 3'-located isoforms on the dup(2-c2) allele, respectively. The cells expressing one or two of three transcripts were excluded from the analysis.
Pcdh-a isoform simultaneously. These results strongly suggested that tandem duplication maintained the stochastic expression of the Pcdh-a isoforms, as seen in the wild-type brain. The 5’-located Pcdh-a c2 acquired stochastic expression upon tandem duplication.

To investigate whether the 5’- and 3’-located duplicated Pcdh-a c2 retained its constitutive expression at the single-neuron level, single-cell RT-PCR and SNP analysis was carried out (Fig. 4). Of the 16 single Purkinje cells analyzed, all showed the Pcdh-a c2 transcript from the wild-type (JF1) and 3’-located duplicated exon in all three tubes. They were also all positive for Pcp-2, confirming that they were differentiated Purkinje cells. These results suggested that the Pcdh-a c2 wild-type exon (JF1) and 3’-located duplicated exon were expressed constitutively in differentiated Purkinje cells.

To our surprise, the Pcdh-a c2 transcript from the 5’-located duplicated exon was detected in three tubes for only two of the 16 cells (#2-4 and #2-13). Notably, 8 of the 16 cells (#2-2, #2-3, #2-6, #2-9, #2-10, #2-11, #2-15 and #2-16) showed no Pcdh-a c2 transcript from the 5’-located duplicated exon (Fig. 4d). For those that did express it, the quantity expressed from the 5’-located duplicated exon was comparable to that from the wild-type (JF1) and 3’-located duplicated exons (Fig. 4e). These results strongly suggested that the expression of the 5’-located duplicated Pcdh-a c2 changed from constitutive to stochastic.

The 5’-located Pcdh-a c2 acquired stochastic expression upon tandem duplication. To investigate whether the 5’- and 3’-located duplicated Pcdh-a c2 retained its constitutive expression at the single-neuron level, single-cell RT-PCR and SNP analysis was carried out (Fig. 4). Of the 16 single Purkinje cells analyzed, all showed the Pcdh-a c2 transcript from the wild-type (JF1) and 3’-located duplicated exon in all three tubes. They were also all positive for Pcp-2, confirming that they were differentiated Purkinje cells. These results suggested that the Pcdh-a c2 wild-type exon (JF1) and 3’-located duplicated exon were expressed constitutively in differentiated Purkinje cells.

To our surprise, the Pcdh-a c2 transcript from the 5’-located duplicated exon was detected in three tubes for only two of the 16 cells (#2-4 and #2-13). Notably, 8 of the 16 cells (#2-2, #2-3, #2-6, #2-9, #2-10, #2-11, #2-15 and #2-16) showed no Pcdh-a c2 transcript from the 5’-located duplicated exon (Fig. 4d). For those that did express it, the quantity expressed from the 5’-located duplicated exon was comparable to that from the wild-type (JF1) and 3’-located duplicated exons (Fig. 4e). These results strongly suggested that the expression of the 5’-located duplicated Pcdh-a c2 changed from constitutive to stochastic.

Tandem duplication masks cluster-structure dependent DNA hypomethylation. The 5’-located Pcdh-a c2 was down-regulated in the Pcdh-a dup(2-c2)/dup(2-c2) mouse cerebellum (Fig. 2) and acquired a stochastic expression pattern (Fig. 4), suggesting that the regulation was different between the 5’-located and the 3’-located Pcdh-a c2 in the dup(2-c2) allele. Since Pcdh-a c2 is extensively hypomethylated in the wild-type mouse brain, and higher mosaic DNA methylation levels are correlated with a lower transcription of stochastically expressed Pcdh genes, we first examined the DNA methylation of Pcdh-a c2 in the Pcdh-a dup(2-c2)/dup(2-c2) mouse cerebellum using bisulfite sequencing (Fig. 5a and 5b). The results showed that the expression of the 5’-located duplicated Pcdh-a c2 was comparably expressed to that from the wild-type (JF1) and 3’-located duplicated exons (Fig. 4e). These results strongly suggested that the expression of the 5’-located duplicated Pcdh-a c2 changed from constitutive to stochastic.

Because bisulfite sequencing was unable to discriminate between the two Pcdh-a c2s in the dup(2-c2) allele, we further analyzed the
DNA methylation analysis using HpaII digestion-mediated DNA methylation analysis (Fig. 5c and 5d). The HpaII-resistant fraction, containing methylated CCGG, predominantly included the 5'-located duplicated Pcdh-αc2 genomic DNA. The DNA methylation level on the 5'-located duplicated Pcdh-αc2 was higher than that on the 3'-located and wild-type Pcdh-αc2.

We further examined the DNA methylation of Pcdh-αc1 in the Pcdh-αc2; the positions of CpGs and HpaII/MspI sites are shown to scale by vertical lines. Each row represents a single clone. A primer set was designed to amplify the region corresponding to the 5' region of the exon, which has the same sequence in the 5'- and 3'-located duplicated Pcdh-αc2 isoforms. The percentage below each methylation pattern indicates the CpG methylation rate for the region. C, D, Discrimination of the DNA methylation between the 5' - and 3'-located duplicated Pcdh-αc2 isoforms by HpaII digestion-mediated analysis. (c) Electrophoresis results. D, Example chromatograms showing that the HpaII-resistant fraction contained the region of the 5'-located duplicated Pcdh-αc2 exon. Schematic representation of Pcdh-αc1; positions of CpGs are shown to scale by vertical lines. F, Results of bisulfite sequencing. A primer set was designed to amplify the region corresponding to the Pcdh-αc1 promoter, which has the same sequence in the 5'- and 3'-located Pcdh-αc1 duplicates. The percentage below each methylation pattern indicates the CpG methylation rate for the region.

Discussion

Tandem duplications are concentrated within the Pcdh cluster throughout vertebrate evolution and as CNVs in human populations18,20, but the effects of tandem duplication in the Pcdh cluster remain elusive. Here we revealed a critical role for tandem duplica-
tion in the Pcdh cluster gene regulation. The individual duplicated Pcdh-α isoforms showed diverse expression level with stochastic expression manner, even though those have an identical promoter sequence. Interestingly, the 5'3'-located duplicated Pcdh-α12, which is constitutively expressed in the wild-type brain, shifted to stochastic expression upon tandem duplication, accompanied by increased DNA methylation. These observations suggest that tandem duplication has been beneficial for the acquisition of the stochastic expression and the expansion of its repertoire through vertebrate evolution and in human populations (Fig. 7).

What are the mechanisms by which individual duplicated Pcdh-αs, each of which has an identical promoter sequence, are expressed stochastically? The present data are consistent with our previous findings, which is that the stochastic expression of Pcdh-α cluster genes is governed by enhancer and promoter DNA methylation. The enhancers for the Pcdh-α genes, named HS5-1 and HS7, are located downstream of the Pcdh-α cluster, and the present data suggest that the HS5-1 and/or HS7 enhancers in the duplicated allele are still effective at even greater distances than in the wild-type allele (distance between the HS5-1 enhancer and the most distal promoter (Pcdh-α1) in the wild-type allele was 280 kb; in the dup(2-c2) allele, it was 500 kb). However, the present finding of lower expression of the 5'3'-located duplicated Pcdh-α genes than the 3'3'-located duplicated Pcdh-α genes in the dup(2-c2) allele indicates that
gene clusters, such as the olfactory receptor MOR28 cluster, and the Pcdh cluster genes. Previous studies have suggested that Pcdh cluster evolution included successive tandem duplications and sequence divergences. Here, we propose a model in which stochastic expression of the duplicated Pcdh genes is immediately acquired after tandem duplication, which precedes sequence divergence (Fig. 7).

The human PCDH cluster is particularly rich in CNVs, including duplications and deletions (see Database of Genome Variants: http://dgv.tcag.ca/gb2/gbrowse/dgv2_hg18/?name=chr5:140050001.141050000). The present study showed that the tandem duplication resulted in healthy mice with a macroscopically normal brain. This result can be explained in part by the maintenance of the expression levels and distribution patterns of the total Pcdh-α transcript, by the re-allocation of Pcdh-α isoform expression, and by the maintenance of the expression levels of neighboring genes, upon duplication. Thus, it is likely that following various CNV events in the human PCDH cluster, the total expression level, dual gene-regulatory mechanisms, and stochastic expression of the human PCDH genes are maintained. This robustness may provide the predominant reason for the frequent CNVs in the human PCDH cluster. One study reported that there is no phenotypic link between a CNV in the human PCDH cluster, a 16.7-kb deletion affecting PCDHA8-A10, and psychiatric disorders. Recently, a de novo gene disruption in PCDHA13 was reported in autism. Furthermore, Anitha A. et al. reported strong genetic evidence of PCDHA as a potential candidate gene for autism. The PCDHA cluster is also a candidate locus for bipolar disorder. Furthermore, deletion of PCDHA1-PCDHA9 is associated with higher brain function, such as music perception. It will be interesting to investigate the effects of these genetic mutations on the PCDHA expression and neural circuit formation.

Recent human genome analyses revealed that tandem duplications contribute to human phenotypes, including many psychiatric disorders, color vision, Parkinson’s disease, and Rheumatoid arthritis. For example, duplications of 7q36.3, which contains the vasointestinal peptide receptor gene VIPR2, confer significant risk for schizophrenia, and VIPR2 mRNA levels are increased differently among duplication carriers. These findings suggest the importance of conducting detailed investigations addressing the effects of duplications on gene regulation. The iTAMERE approach enables careful analyses directed toward understanding the etiology of CNV-associated human disorders.

The vertebrate Pcdh cluster shows remarkable similarity to the Drosophila Dscam1 gene, within which tandem duplication is frequent throughout its evolutionary history. Importantly, recent studies demonstrated that the stochastic gene regulation in Pcdh and Dscam1 play important roles in neural circuit development by providing a source for cell surface diversity. These findings suggest essential roles for tandem duplications in the evolution of vertebrate and invertebrate nervous systems. Further studies aimed at dissecting fine-scale neural circuits in the Pcdh cluster contribute to brain function.
Animals. B6 mice were purchased from Charles River Japan. The wild-type mouse strain (C57BL/6J) was obtained from the National Institute for Genetics (Mishima, Shizuoka, Japan). All animals were maintained in a specific pathogen-free facility under a 12-h light/dark regimen. Experimental procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals of the Science Council of Japan and approved by the Animal Experiment Committee of Gunma University and Osaka University.

Generation of Pcdhγ44dup(2-c2) mice. By crossing SR mice (see Supplemental Experimental Procedures) with Scyp1-Cre transgenic mice27, we generated SR mice carrying the Scyp1-Cre transgene. We crossed these mice with mice bearing the G16Neo allele, then selected male offspring bearing the SR allele, G16Neo allele, and the Scyp1-Cre transgene (Pcdhγ44dup(2-c2), Scyp1-Cre). We then crossed Pcdhγ44dup(2-c2), Scyp1-Cre transgenic male and female pups, and generated the pups using genomic DNA extracted from the tail. Some of these pups carried the duplicated [dup(2-c2)] or deleted [del(2-c2)] allele as a result of TAMERE in the testsis23. To identify the Pcdhγ44dup(2-c2) mice, we performed Southern blotting and PCR analyses (Fig. 1e and f). The BamHI-digested genomic DNA from the tail was subjected to Southern blotting using probe A (the same probe used to identify the SR allele). A band at 8.6 kb indicated the wild-type allele, and a band at 12.4 kb indicated the dup(2-c2) allele. Pcdhγ44dup(2-c2) mice were obtained by crossing Pcdhγ44dup(2-c2) parents, which were backcrossed with B6 for more than three generations.

In situ hybridization. In situ hybridization (ISH) was performed essentially as described previously27. The details are provided in the Supplemental Experimental Procedures.

Expression analysis in cerebellum. The cerebellum was dissected from 4-week-old mice and immediately frozen in liquid nitrogen. The tissue was homogenized with a Polytron homogenizer. Total RNA was isolated using RNeasy (Qiagen), according to the supplier’s recommendations. To obtain cDNA, 2.5 μg of the total RNA was treated with DNase I (Takara) and reverse transcribed with SuperScriptIII reverse transcriptase (Invitrogen) using random primers in a 40-μl reaction volume.

For the SNP analyses, PCR for each Pcdh-α isoform was performed using 0.4 μl of cDNA from the cerebellum of a 4-week-old mouse as a template. The primer sequences used for the SNP analyses are shown in Supplementary Table S2. For direct SNP analyses, the PCR products were sequenced using a standard method. For the cloning-mediated SNP analysis, the PCR products were cloned into pT7-Blue SR/G16Neo, Sycp1-Cre). We then crossed Pcdhγ44dup(2-c2), Scyp1-Cre transgenic male and female pups, and generated the pups using genomic DNA extracted from the tail. Some of these pups carried the duplicated [dup(2-c2)] or deleted [del(2-c2)] allele as a result of TAMERE in the testis23. To identify the Pcdhγ44dup(2-c2) mice, we performed Southern blotting and PCR analyses (Fig. 1e and f). The BamHI-digested genomic DNA from the tail was subjected to Southern blotting using probe A (the same probe used to identify the SR allele). A band at 8.6 kb indicated the wild-type allele, and a band at 12.4 kb indicated the dup(2-c2) allele. Pcdhγ44dup(2-c2) mice were obtained by crossing Pcdhγ44dup(2-c2) parents, which were backcrossed with B6 for more than three generations.

Split single-cell RT-PCR. Single-cell RT-PCR was performed essentially as described previously, with small modifications28. The details are provided in the Supplemental Experimental Procedures.

DNA methylation analysis by bisulfite sequencing and HpaII digestion. The genomic DNA was prepared with the QiAamp DNA Micro Kit (Qiagen) from the cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm, or with the EpiTect Plus LyseAll Kit (Qiagen) from the brain of an E12.5 embryo, head of an E9.5 embryo, cerebellum of a 4-week-old or 3-month-old mouse, or from sperm.
32. Mishina, M. & Sakimura, K. Conditional gene targeting on the pure C57BL/6 genetic background. *Neurosci Res* 58, 105–112 (2007).
33. Henrichsen, C. N. et al. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41, 424–429 (2009).
34. Dallosso, A. R. et al. Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 9q31 in Wilms' tumor. *PLoS Genet* 5, e1000745 (2009).
35. Kaneko, R., Kawaguchi, M., Toyama, T., Taguchi, Y. & Yagi, T. Expression levels of Protocadherin-alpha transcripts are decreased by nonsense-mediated mRNA decay with frameshift mutations and by high DNA methylation in their promoter regions. *Gene* 430, 86–94 (2009).
36. Kawaguchi, M. et al. Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster. *J Biol Chem* 283, 12064–12075 (2008).
37. Tasic, B. et al. Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol Cell* 10, 21–33 (2002).
38. Toyoda, S. et al. Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* 82, 94–108 (2014).
39. Ribisch, S., Tasic, B. & Maniatis, T. Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc Natl Acad Sci U S A* 103, 19719–19724 (2006).
40. Kehayova, P., Monahan, K., Chen, W. & Maniatis, T. Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc Natl Acad Sci U S A* 108, 17195–17200 (2011).
41. Yokota, S. et al. Identification of the cluster control region for the Protocadherin-(beta) genes located beyond the Protocadherin-(gamma) cluster. *J Biol Chem* (2011).
42. Fuss, S. H., Omura, M. & Mombaerts, P. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* 130, 373–384 (2007).
43. Wang, Y. et al. Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors. *Proc Natl Acad Sci U S A* 96, 5251–5256 (1999).
44. Montavon, T., Le Garrec, J. F., Kerszberg, M. & Duboule, D. Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev* 22, 346–359 (2008).
45. Tsujimura, T., Hosoya, T. & Kawamura, S. A single enhancer regulating the differential expression of duplicated red-sensitive opsin genes in zebrafish. *PLoS Genet* 6, e1001245 (2010).
46. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. & Yagi, T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep* 2, 345–357 (2012).
47. Lachman, H. M. et al. Analysis of protocadherin alpha gene deletion variant in bipolar disorder and schizophrenia. *Psychiatr Genet* 18, 110–115 (2008).
48. Iossifov, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285–299 (2012).
49. Anitha, A. et al. Protocadherin alpha (PCDHA) as a novel susceptibility gene for autism. *Journal of psychiatry & neuroscience: JPN* 38, 192–198 (2013).
50. Pedrosa, E. et al. Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia. *Schizophr Res* 102, 210–219 (2008).
51. Ukkola-Vuo, L. et al. Genome-wide copy number variation analysis in extended families and unrelated individuals characterized for musical aptitude and creativity in music. *PLoS One* 8, e56356 (2013).
52. Fanciulli, M., Petretto, E. & Altman, T. J. Gene copy number variation and common human disease. *Clin Genet* 77, 201–213 (2010).
53. Vacic, V. et al. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 471, 499–503 (2011).
54. Crayton, M. E., 3rd, Powell, B. C., Vision, T. J. & Giddings, M. C. Tracking the evolution of alternatively spliced exons within the Dscam family. *BMC evolutionary biology* 6, 16 (2006).
55. Esumi, S., Kaneko, R., Kawamura, Y. & Yagi, T. Split single-cell RT-PCR analysis of Purkinje cells. *Nat Protoc* 1, 2143–2151 (2006).
56. Li, L. C. & Dahlia, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427–1431 (2002).
57. Kumaki, Y., Oda, M. & Okano, M. QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* 36, W170–175 (2008).

**Acknowledgments**

We are grateful to Dr. T. Shiroishi (The National Institute of Genetics, Japan) for providing the JF1 mouse strain, Dr. M. Mishina (University of Tokyo, Japan) for providing the FLP66 mouse strain, and Dr. J. Takeda (Osaka University, Japan) for providing the pPE7.neoW-PZLF plasmid. We thank members of the Yagi laboratory, Yanagawa laboratory, and Drs. A. Oue, H. Hata for suggestions and discussion during the course of this work. We appreciate Mses. R. Natsume, C. Gaoji, A. Nishigao, Y. Hidaka, T. Kakinuma, S. Sato, and A. Morita for their technical support and Ms. J. Shuto for their excellent secretarial assistance. This work was supported by Grant-in-Aid for Young Scientists (B) (ISPS) (No. 19770147, 22700326), Grant-in-Aid for Scientific Research (C) (ISPS) (No. 24500375), the Life Science Foundation of Japan, and the Takeda Science Foundation (Ryosuke Kaneko), and Grant-in-Aid for Scientific Research (S) (ISPS) (No. 19100006), Innovative Areas “Mesoscopic Neurocircuity” (No. 23115155 and No. 25115720) and (Comprehensive Brain Science Network) from the Ministry of Education, Science, Sports, and Culture of Japan (MEXT), and IST-CREST (Takeshi Yagi).

**Author contributions**

R.K. and T.Y. conceived and designed the experiments, analyzed the data and wrote the manuscript. R.K. and M.A. performed research; R.K., M.A., T.H., A.U., K.S. and Y.Y. contributed unpublished reagents/analytic tools.

**Additional information**

**Supplementary information** accompanies this paper at http://www.nature.com/scientificreports

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Kaneko, R. et al. Expansion of stochastic expression repertoire by tandem duplication in mouse Protocadherin-α cluster. *Sci. Rep.* 4, 6263; DOI:10.1038/srep06263 (2014).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/