Clustered protocadherins
Weisheng V. Chen and Tom Maniatis*

Summary
The majority of vertebrate protocadherin (Pcdh) genes are clustered in a single genomic locus, and this remarkable genomic organization is highly conserved from teleosts to humans. These clustered Pcdhs are differentially expressed in individual neurons, they engage in homophilic trans-interactions as multimers and they are required for diverse neurodevelopmental processes, including neurite self-avoidance. Here, we provide a concise overview of the molecular and cellular biology of clustered Pcdhs, highlighting how they generate single cell diversity in the vertebrate nervous system and how such diversity may be used in neural circuit assembly.

Key words: CTCF, Cohesin, Promoter choice, Single cell diversity, Homophilic interaction, Self-avoidance

Introduction
Protocadherins (Pcdhs), which are predominantly expressed in the nervous system, constitute the largest subfamily of the cadherin superfamily of cell-adhesion molecules. The founding members of the Pcdh gene family were discovered by Suzuki and co-workers in an effort to isolate novel cadherin repeat-containing genes (Sano et al., 1993). Distinct Pcdh members were subsequently cloned, and these included a set of eight cDNAs encoding cadherin-related neuronal receptors (CNRs) (Kohmura et al., 1998). A striking characteristic of the corresponding CNR mRNAs is that their 5’ ends are distinct, whereas their 3’ ends are all identical, suggesting that they are generated by alternative pre-mRNA splicing. Characterization of the human genomic DNA encoding the CNR mRNAs revealed that they are encoded in a large Pcdh gene cluster designated α (Pcdha), which is located immediately upstream of two additional Pcdh gene clusters designated β (Pcdhb) and γ (Pcdhg) (Wu and Maniatis, 1999). Remarkably, the genomic organization of the Pcdh gene clusters resembles that of the immunoglobulin and T-cell receptor genes, both of which generate enormous diversity in the immune system through a mechanism that...
involves somatic cell DNA rearrangement. Subsequent studies confirmed the possibility that the clustered Pcdhs serve as a source of molecular diversity in the nervous system, albeit through a different mechanism. The enormous cell surface diversity resulting from the combinatorial expression of Pcdh isoforms, together with the extraordinary specificity afforded by their homophilic interactions, have led to the speculation that clustered Pcdhs are functional counterparts of the Drosophila Dscam1 proteins, which play a central role in neural circuit assembly in the invertebrate nervous system (Zipursky and Sanes, 2010). Indeed, recent studies have demonstrated that the Pcdhg gene cluster is required for neurite self-avoidance in the mouse, in a manner similar to that of the Dscam1 gene in the fly (Lefebvre et al., 2012). Thus, it appears that clustered Pcdhs may function as molecular barcodes for self-recognition by individual neurons in the vertebrate nervous system. Here, we briefly review studies that led to this hypothesis, as well as those that have provided insights into the multifaceted roles of clustered Pcdhs in neural development.

**Protocadherin gene clusters**

In mammals, two types of Pcdh genes have been defined: the non-clustered Pcdhs, which are scattered throughout the genome (Kim et al., 2011); and the clustered Pcdhs, which are organized in three closely linked gene clusters designated α, β and γ (Wu and Maniatis, 1999). The mouse Pcdha, Pcdhb and Pcdhc gene clusters comprise 14, 22 and 22, respectively, large ‘variable’ exons arrayed in tandem, each encoding the entire extracellular domain of the protein, a transmembrane domain and a variable intracellular domain. In the Pcdha and Pcdhc clusters (but not in the Pcdhb cluster), multiple variable exons are located upstream of three ‘constant’ exons, which encode a common distal intracellular domain of the corresponding Pcdh or Pcdhg proteins. Phylogenetic analyses indicate that the Pcdh isoforms encoded by each gene cluster are highly homologous to each other. However, a salient exception is a set of five Pcdh isoforms encoded by the Pcdha gene cluster (Pcdha1 and Pcdha2) and the Pcdhc gene cluster (Pcdhgc3, Pcdhgc4 and Pcdhgc5), which are evolutionarily divergent from other members of their respective cluster but are instead more similar to each other. These divergent Pcdh members have been designated C-type isoforms. The unique genomic organization of Pcdh gene clusters is highly conserved in vertebrates, although the Pcdhb cluster is lacking in some lower species, such as the pufferfish Fugu and zebrafish (Yu et al., 2007).

The genomic organization of clustered Pcdhs suggests an elegant mechanism for generating single cell diversity in the brain. Although the genomic organization of Pcdh gene clusters is remarkably similar to that of the immunoglobulin and T-cell receptor gene clusters, early studies ruled out somatic cell DNA rearrangement (Tasic et al., 2002). In addition, although pre-mRNA trans-splicing was detected between variable and constant exons in the three gene clusters, the low level of these events, together with other evidence, ruled out the possibility that functional Pcdh diversity is generated through a trans-splicing mechanism (Tasic et al., 2002). Rather, detailed characterization of Pcdha (Tasic et al., 2002) and Pcdhg (Wang et al., 2002a) gene expression revealed that each variable exon is preceded by a promoter, and the selective transcription of individual Pcdh isoforms is achieved by promoter choice followed by alternative pre-mRNA cis-splicing. Direct evidence for single-cell diversity of Pcdh gene expression was provided by single-cell RT-PCR studies of individual Purkinje neurons from interspecies F1 hybrid mice (Esumi et al., 2005; Kaneko et al., 2006; Hirano et al., 2012). Based on these studies, a model was proposed in which several alternate Pcdh isoforms in each of the three clusters are stochastically and independently transcribed on the two homologous chromosomes, whereas all five C-type isoforms are constitutively expressed from both chromosomes in every neuron. Although the generality of this Purkinje cell-based model remains to be established, the proposed combinatorial expression of Pcdh isoforms from all three gene clusters can generate enormous cell surface diversity.

**Mechanisms for generating single cell diversity**

How does such a complex pattern of promoter choice occur in individual neurons? This question was first addressed by identifying and characterizing transcriptional regulatory elements and DNA-binding proteins that function in the Pcdha gene cluster. A conserved sequence element (CSE) was identified immediately upstream of the transcriptional start site of most variable exons within all three Pcdh gene clusters (Wu et al., 2001), and this element was shown to function as a crucial promoter motif (Tasic et al., 2002; Monahan et al., 2012). Essential transcriptional enhancer elements were later identified through a search for conserved intergenic sequences, detection of DNase I hypersensitive (HS) sites and validation through the analysis of reporter gene expression in transgenic mice (Ribich et al., 2006). Two enhancer elements were identified in the Pcdha cluster: HS5-1, which is located downstream of the last constant exon; and HS7, which is located in the intron between last two constant exons. These two enhancers were individually deleted in mice, and both were shown to be required for maximal levels of Pcdha gene expression (Ribich et al., 2006; Kehayova et al., 2011). Remarkably, the enhancer activity of HS5-1 correlates with binding of the insulator protein CCCTC-binding factor (CTCF) to target promoters (Kehayova et al., 2011). Interestingly, HS5-1 was also shown to function as a transcriptional silencer of Pcdha gene expression in non-neuronal cells, and this silencing activity requires binding of the RE1 silencing transcription factor (REST, also known as NRSF) repressor complex (Kehayova et al., 2011). Thus, the HS5-1 element functions as an on/off switch, turning expression of the Pcdha genes on in neurons and off in non-neuronal cells.

Detailed characterization of both mouse (Monahan et al., 2012) and human (Guo et al., 2012) neuroblastoma cell lines revealed a direct correlation between transcriptional activity and binding of the CTCF/cohesin complex to the active promoters and enhancers of Pcdha genes. Further analyses using the chromatin immunoprecipitation followed by sequencing (ChIP-Seq) method revealed that, for each actively transcribed alternate Pcdha gene, the CTCF/cohesin complex binds to the CSE in its promoter, as well as to a second conserved sequence within the variable exon. Similarly, the CTCF/cohesin complex also binds to two sites in the HS5-1 enhancer (HS5-1a and HS5-1b). Of the two constitutively expressed C-type genes, only Pcdha1 contains a CSE site in its promoter, which also binds to the CTCF/cohesin complex, but lacks the exonic CTCF-bonding site. By contrast, the Pcdha2 promoter does not contain the CSE motif, and binds to cohesin but not to CTCF. Similarly, the HS7 enhancer lacks a CTCF-binding site and binds only to cohesin. Chromosome conformation capture (3C) assays showed that both the HS5-1 and HS7 enhancers directly interact with active promoters through long-range DNA looping, with the former (mediated by HS5-1) being CTCF and cohesin dependent, and the latter (mediated by HS7) being CTCF independent but cohesin dependent (Guo et al., 2012).

On the basis of these observations, it was proposed that CTCF and/or cohesin mediate DNA-looping interactions in an active
‘transcriptional hub’, in which several Pcdha promoters simultaneously interact with both enhancers (Guo et al., 2012). This arrangement provides a possible mechanism for the stochastic, monoallelic expression of alternate Pcdh isoforms, as well as the constitutive, biallelic expression of C-type isoforms in the same cell. Putative enhancer elements with similar characteristics (HS5-1aL, HS5-1bL and HS7L) were identified at similar locations in the Pcdhg cluster (Guo et al., 2012). In addition, an enhancer element (HS16-20) was identified downstream of the Pcdhg cluster (Yokota et al., 2011), which appears to act on the Pcdhb cluster and also binds to the CTCF/cohesin complex (Monahan et al., 2012). These observations suggest that similar mechanisms are likely to be involved in the expression of all three Pcdh gene clusters. The dependence on, as well as the specificity of, CTCF and cohesin binding in the regulation of Pcdh gene expression were shown by Ctcf and cohesin knockdown experiments in neuroblastoma cells (Golan-Mashiach et al., 2012; Guo et al., 2012; Monahan et al., 2012), as well as in Ctcf and cohesin subunit conditional knockout mice (Kawauchi et al., 2009; Hirayama et al., 2012; Remeseiro et al., 2012).

**Homophilic interactions and intracellular signaling**

All three classes of clustered Pcdh proteins are detected throughout the neuronal soma, dendrites and axons, and are observed in synapses and growth cones (Kohmura et al., 1998; Wang et al., 2002b; Kallenbach et al., 2003; Phillips et al., 2003; Junghans et al., 2008). The presence of Pcdh proteins at the synapse led to speculation that these cell-surface proteins may provide a synaptic adhesive code for specifying neuronal connections (Shapiro and Colman, 1999; Takeichi, 2007). However, direct evidence for this is lacking. Classical cadherins mediate cell-cell adhesion via trans-homodimerization between their extracellular domains extended from apposing cell membranes, and they assemble intercellular adherens junctions through cis-clustering (Brusch et al., 2012). Like classical cadherins, members of the Pcdh family were also shown to mediate cell-cell adhesion in cell-based assays (Obata et al., 1995; Frank et al., 2005; Reiss et al., 2006), and multiple Pcdhg proteins were shown to engage in homophilic trans-interactions (Fernández-Monreal et al., 2009; Schreiner and Weiner, 2010). Unlike classical cadherins, however, the binding specificity of Pcdhs is not determined by the EC1 domain, but instead is mediated by the EC2 and EC3 domains (Schreiner and Weiner, 2010), which are highly divergent between distinct Pcdh isoforms (Wu, 2005). It has been proposed that cis-tetramers, rather than monomers (as in the cases of classical cadherins or Dscams), represent the unit of Pcdh homophilic trans-interactions (Schreiner and Weiner, 2010). As Pcdha and Pcdhg proteins (and most likely Pcdhb proteins) form multimeric complexes in cis (Murata et al., 2004; Bonn et al., 2007; Han et al., 2010; Schalm et al., 2010), multimerization of stochastically expressed Pcdh isoforms of all three classes in single neurons could produce a large number of distinct homophilic interaction units, significantly expanding the cell-surface diversity afforded by stochastic gene expression alone. The enormous diversity and specificity mediated by homophilic trans-interactions between Pcdh cis-multimers may thus provide the molecular basis for contact-dependent neurite self-avoidance such as that observed in the fly Dscam1 system (Hattori et al., 2008).

In addition to mediating cell-cell adhesion (or contact-dependent neurite repulsion) via homophilic interactions, clustered Pcdh proteins also mediate intracellular signaling. All three classes of clustered Pcdh proteins have distinct cytoplasmic regions, which do not have recognizable motifs and lack the catenin-binding domains that are found in classical cadherins. Although the intracellular domains encoded by the constant exons of the Pcdha and Pcdhg clusters differ significantly from each other, they are strictly conserved in vertebrate evolution, suggesting a conserved cellular function. Both Pcdha and Pcdhg proteins bind the receptor tyrosine kinase Ret, which is required for the stabilization and phosphorylation of their intracellular domains (Schalm et al., 2010). Gliarial cell line-derived neurotrophic factor (GDNF), a ligand of the Ret receptor, induces the phosphorylation of Pcdhs in cultured motoneurons and sympathetic neurons. The Pcdhs in turn are required for the stabilization of activated Ret, indicating that Pcdhs and Ret are functional components of a phosphorylation-dependent signaling complex (Schalm et al., 2010). Both Pcdha and Pcdhg intracellular domains also interact with the cell-adhesion kinases proline-rich tyrosin kinase 2 (Pyk2) and focal adhesion kinase (Fak), resulting in inhibition of kinase activity (Chen et al., 2009). These interactions have been implicated in the defects in cell survival and dendritic patterning observed in Pcdha- or Pcdhg-deficient neurons (Chen et al., 2009; Garrett et al., 2012; Suo et al., 2012). A large number of other potential interacting proteins, including phosphatases, kinases, adhesion molecules and synaptic proteins, have been reported (Han et al., 2010; Schalm et al., 2010), suggesting that clustered Pcdhs may form large heteromeric complexes with broad functions yet to be elucidated.

The role of clustered Pcdhs in intracellular signaling is further supported by the finding that they are proteolytically processed by the γ-secretase complex, which releases soluble intracellular fragments into the cytoplasm (Haas et al., 2005; Hambsch et al., 2005; Reiss et al., 2006; Bonn et al., 2007; Buchanan et al., 2010). This process, which requires endocytosis, is developmentally regulated, and a decrease in cleavage products is observed upon neuronal differentiation (Buchanan et al., 2010). The released intracellular fragments may function locally in the cytoplasm and/or enter the nucleus and regulate gene expression, as is the case for other cell-surface proteins such as Notch and N-cadherin (Rajagopal et al., 2012).

**Roles in the nervous system**

Genetic manipulations of Pcdha and Pcdhg clusters in mice have revealed multiple roles for clustered Pcdhs in neural development. The first role identified for Pcdh family members is in neuronal survival, as evidenced by the massive apoptosis and eventual loss of spinal interneurons and retinal cells in Pcdhg-deficient mice (Wang et al., 2002b; Weiner et al., 2005; Lefebvre et al., 2008; Prasad et al., 2008). In both the spinal cord and retina, the neuronal loss is accompanied by reduced numbers of synapses, raising the possibility that the loss of Pcdh proteins may lead to synaptic loss, which in turn compromises neuronal survival. This possibility is supported by the observation that synaptic density in Pcdhg mutant spinal cords remains reduced in the BCL2-associated X protein-null (Bax−/−) background where neuronal apoptosis is genetically blocked (Weiner et al., 2005). Intriguingly, however, an essentially complete rescue of synaptic loss was observed in the retina in the absence of neuronal apoptosis (Lefebvre et al., 2008), suggesting that alternative mechanisms may account for the observed neuronal cell death in Pcdhg-deficient mice.

This dilemma was resolved in a recent study, in which the three C-type genes (Pcdhgc3, Pcdhgc4 and Pcdhgc5) were selectively deleted from the mouse Pcdhg cluster, leaving the alternate isoforms intact (Chen et al., 2012). The resultant triple C-type isomorph knockout mice are phenotypically indistinguishable from the full
Pcdhg cluster deletion mice, and both knockouts die shortly after birth and exhibit similar levels and patterns of neuronal cell loss and synaptic alterations in the spinal cord and retina. Remarkably, genetically blocking apoptosis rescues the neonatal lethality of the C-type knockouts, but not that of the full Pcdhg cluster deletion mice. Moreover, the severe defects observed with specific types of synapses on motoneurons (Prasad and Weiner, 2011; Chen et al., 2012) were also rescued in both types of mutants, indicating that these defects are not the cause but the result of neuronal cell death, although additional synaptic defects may occur independently (Weiner et al., 2005; Garrett and Weiner, 2009). These observations suggest that the role of the Pcdhg cluster in neuronal survival is primarily, if not entirely, mediated by one or more of the C-type isoforms, which are phylogenetically divergent among Pcdhs, and exhibit different expression patterns from the alternate isoforms. However, as the full cluster deletion mutants still die as neonates in the absence of apoptosis, the Pcdhg proteins must have an additional independent role that is essential for postnatal development, most likely in neural circuit assembly.

Such a role was dramatically revealed in a detailed genetic study, where Pcdhg proteins were shown to mediate neurite self-avoidance in retinal starburst amacrine cells (SACs), as well as in cerebellar Purkinje cells (Lefebvre et al., 2012). Self-avoidance is a fundamental principle in neural circuit assembly, whereby sister axons and dendrites of the same neuron recognize and repel each other, while allowing multiple neurons to share the same receptive or projection field (Kramer and Kuwada, 1983). In contrast to the radially symmetric pattern of the wild-type SACs, dendrites of Pcdhg-deficient SACs recurrently cross each other and form bundles, reminiscent of those seen in Dscam1-deficient dendritic arborization (da) neurons in the fly (Hattori et al., 2007; Matthews et al., 2007). This dendritic self-avoidance phenotype is cell-autonomous and independent of neuronal apoptosis. In addition, dendritic morphology is normal in SACs lacking small subsets of Pcdhg genes, indicating that the complete spectrum of Pcdhg isoforms is not required. Remarkably, expression of a single, arbitrarily chosen Pcdhg isoform in the Pcdhg-deficient SACs rescued the self-avoidance phenotype, and promoted repulsion between neurites from neighboring neurons expressing the same isoform (Lefebvre et al., 2012). These observations parallel those in fly Dscam1 subset deletion and single isoform mutants (Hattori et al., 2007; Matthews et al., 2007; Hattori et al., 2009), strongly suggesting that the clustered Pcdhg proteins function similarly to invertebrate Dscam1 isoforms in regulating neurite self-avoidance in vertebrates.

In addition to mediating neurite self-avoidance, Pcdhg proteins are also required for dendritic patterning, as shown by conditional Pcdhg knockout mice that display defects in dendritic branching and arborization in cortical neurons (Garrett et al., 2012). Pcdhg proteins were proposed to function in this process by inhibiting the kinase activities of associated cell-adhesion kinases Fak, which downregulates the associated cell-adhesion kinase Fak, which downregulates the activity of the protein kinase C/phospholipase C/miystooylated alanine-rich C-kinase substrate (PKC/PLC/MARCKS) pathway, which in turn promotes dendritic arborization. Although synaptic defects were observed in Pcdhg mutants (Wang et al., 2002b; Weiner et al., 2005; Garrett and Weiner, 2009), the issue of whether Pcdhg proteins play a direct role in synaptic interactions remains to be elucidated. Suggestive evidence for this idea was provided by the observation that Pcdhc5 specifically interacts with the γ-amino-butyric acid A (GABA_A) receptor, which is required for stabilization and maintenance of GABAergic synapses in cultured hippocampal neurons (Li et al., 2012).

In contrast to Pcdhg-deficient mice, which die as neonates, Pcdha knockouts and severe hypomorphs are viable and fertile with no obvious defects (Hasegawa et al., 2008; Katori et al., 2009; Suo et al., 2012). However, abnormal axonal projections of olfactory sensory neurons and serotonergic neurons were observed in Pcdha mutants that lack the constant region (Hasegawa et al., 2008; Katori et al., 2009). Interestingly, a single constitutively expressed Pcdha isoform (Pcdha1, generated by deleting Pcdha2 through Pcdha8 in the cluster) was shown to be sufficient to rescue the axonal coalescence defects of olfactory sensory neurons (Hasegawa et al., 2008), raising the possibility that intracellular Pdha function, rather than isoform diversity, is required for this process. Similar to Pcdhg mutants, Pcdha knockouts also exhibit defects in dendritic branching, as well as in spine morphogenesis. It has been proposed that Pdha proteins regulate dendritic and spine development by inhibiting the kinase activities of associated cell-adhesion kinases Fak/Pyk2, which in turn activates small GTPases such as Ras homolog gene family member A (RhoA) and Ras-related C3 botulinum toxin substrate 1 (Rac1), leading to cytoskeleton reorganization (Suo et al., 2012).

Functional studies of mice lacking the Pcdhb gene cluster have not been reported. Because there are three Pcdh gene clusters encoding 58 isoforms that are expressed broadly in overlapping brain regions, their roles in specific cellular context may not be revealed with single cluster knockouts due to functional compensation from members of the other two clusters. This possibility may explain the fact that certain phenotypes (e.g. neuronal apoptosis or dendritic self-avoidance) are observed only in specific cell types in mice lacking one Pcdg gene cluster (e.g. Pcdhg knockouts).

Conclusions and perspectives

Since the discovery of the Pdhd gene clusters over a decade ago, significant progress has been made in understanding how the clustered Pdhd genes generate single cell diversity in the nervous system, and how this diversity may function in neural circuit assembly. Nevertheless, our knowledge of gene regulation, cellular function and the in vivo roles of the clustered Pdhd proteins is far from complete, and many fundamental issues remain to be resolved. For example, the stochastic/monoallelic and constitutive/biallelic transcription model, based on single cell RT-PCR study of one cell type (the Purkinje neuron) at one developmental stage (P21), must be tested in other types of neurons and at different developmental stages to determine whether this is a general mechanism. In addition, there is much to be learned regarding the diversity and complexity of trans- and cis-interactions of Pdhd multimers at the cell surface, the structure of the interacting complexes, and their physical and functional association with signaling pathway components. We have very limited knowledge regarding the mechanism(s) by which the Pdhd genes are cleaved at the cell surface, as well as the function of the released protein fragments. How C-type Pdhd isoforms are involved in neuronal apoptosis, and whether C-type Pdha isoforms also have distinct functions are yet to be addressed. Finally, although Pdhd proteins are required for neurite self-avoidance, the issue remains of whether this is a general role for clustered Pdhd in a broad array of neuronal cell types, and whether the clustered Pdhd are the true functional counterparts of the fly Dscam1 proteins. The highly conserved genomic structure of clustered Pdhd proteins, the complex mechanisms that govern their single cell expression and the nature of their homophilic interactions strongly suggest that isoform diversity is central to understanding their functions. However, the functional significance of the


