Generation of single cell identity by homophilic interactions between combinations of α, β and γ protocadherins

Chan Aye Thu[1], Weisheng V. Chen[1], Rotem Rubinstein[2,3,4,5], Maxime Chevee[1], Holly N Wolcott[2], Klara O Felsovalyi[2,3,4,5], Juan Carlos Tapia[6], Lawrence Shapiro[2], Barry Honig[2,3,4,5]*, and Tom Maniatis[1]*

[1]Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 701 W 168th Street, Room 601, New York, NY 10032

[2]Department of Biochemistry and Molecular Biophysics, Columbia University, 1150 St. Nicholas Avenue, Russ Berrie Pavilion, Room 403, New York, NY 10032 USA

[3]Center for Computational Biology and Bioinformatics, Columbia University, 1130 St. Nicholas Avenue, Room 815, New York, NY 10032 USA

[4]Department of System Biology, Columbia University, 1130 St. Nicholas Avenue, Room 815, New York, NY 10032 USA

[5]Howard Hughes Medical Institute

[6]Department of Neuroscience, Columbia University, 1051 Riverside Drive, New York, NY 10032

*Corresponding authors: tm2472@cumc.columbia.edu; bh6@columbia.edu

Running title

Homophilic interactions of α, β and γ -Pcdhs

Summary

Individual mammalian neurons express distinct repertoires of protocadherin (Pcdh) -α, -β and -γ proteins that function in neural circuit assembly. Here we show that all three types of Pcdhs can engage in specific homophilic interactions, that cell surface delivery of alternate Pcdhα isoforms requires *cis* interactions with other Pcdh isoforms, and that the extracellular cadherin domain EC6 plays a critical role in this process. Analysis of specific combinations of up to five Pcdh isoforms showed that Pcdh homophilic recognition specificities strictly depend on the identity of all of the expressed isoforms, such that mismatched isoforms interfere with cell-cell interactions. We present a theoretical analysis showing that the assembly of Pcdh-α, –β and –γ isoforms into multimeric recognition units, and the observed tolerance for mismatched isoforms can generate the cell surface diversity necessary for single-cell identity. However, competing demands of non-self discrimination and self-recognition place limitations on the mechanisms by which recognition units can function.

Highlights

- Pcdh–α,-β and -γ isoforms mediate specific homophilic interactions
- EC6 domains of Pcdhα and PcdhγC4 isoforms inhibit cell surface delivery
- Homophilic specificity of co-expressed Pcdh isoforms provides single cell identity
- Conflicting requirements of self-recognition and non-self discrimination

Introduction

An essential feature of neural circuit assembly is that the cellular processes (axons and dendrites) of the same neuron do not contact one another, but do interact with processes of other neurons. This feature requires “self-avoidance” between sister neurites of the same cell, a phenomenon that is highly conserved in evolution. Self-avoidance, in turn, requires a mechanism by which individual neurons distinguish self from non-self (Zipursky and Grueber, 2013).

A model for self-recognition, based on studies of the *Drosophila Dscam1* gene (Schmucker et al., 2000), posits that individual neurons stochastically express unique combinations of distinct Dscam1 protein isoforms that are capable of engaging in highly specific homophilic *trans* interactions between proteins on apposing cell surfaces (Hattori et al., 2008). If neurites of the same neuron contact each other, the identical Dscam1 protein repertoire on their cell surfaces will result in homophilic interactions, which in turn leads to contact-dependent repulsion and neurite self-avoidance. In contrast, neurites from different neurons display distinct combinations of Dscam1 isoforms that do not engage in homophilic interactions, and thus not repel one another (Hattori et al., 2008).

The generation of extraordinary Dscam1 isoform diversity is a consequence of the unique structure of the *Drosophila Dscam1* gene, and stochastic alternative splicing of *Dscam1* pre-mRNAs (Miura et al., 2013; Neves et al., 2004; Sun et al., 2013; Zhan et al., 2004). In *Drosophila* this leads to the generation of 19,008 Dscam1 protein isoforms with distinct ectodomains, the vast majority of which can engage in highly specific homophilic interactions, apparently as monomers (Wojtowicz et al., 2004; Wojtowicz et al., 2007; Yagi, 2013) Genetic studies have shown that thousands of Dscam1 isoforms

are required for neurite self-avoidance and non-self discrimination (Hattori et al., 2009). By contrast to *Drosophila Dscam1*, vertebrate *Dscam* genes do not generate significant cell surface diversity (Schmucker and Chen, 2009), suggesting that other genes may serve this function in vertebrates. The most promising candidates are the clustered protocadherin (*Pcdh*) genes (for recent reviews see (Chen and Maniatis, 2013; Yagi, 2012; Zipursky and Grueber, 2013; Zipursky and Sanes, 2010)).

In the mouse 58 Pcdh proteins are encoded by the *Pcdha*, *Pcdhb*, and *Pcdhg* gene clusters, which are arranged in tandem (Figure 1A) (Wu and Maniatis, 1999; Wu et al., 2001). Each of the *Pcdh* gene clusters contains multiple variable exons that encode the entire ectodomain composed of six extracellular cadherin domains (EC1-6), a transmembrane region (TM), and a short cytoplasmic extension. The *Pcdh*α *and* *Pcdh*γ gene clusters also contain three cluster-specific "constant" exons that encode a common intracellular domain (ICD). The last two variable exons in the *Pcdha* gene cluster, and the last three variable exons of the *Pcdhg* gene cluster are divergent from other Pcdh "alternate" isoforms and are referred to as "C-type" Pcdhs (Wu and Maniatis, 1999; Wu et al., 2001). Each of the variable exons is preceded by a promoter, and Pcdh expression occurs through promoter choice (Ribich et al., 2006; Tasic et al., 2002; Wang et al., 2002). Single cell RT-PCR studies in cerebellar Purkinje cells indicate that promoter choice of alternate isoforms is stochastic and independent on the two allelic chromosomes, whereas C-type Pcdhs are constitutively and biallelically expressed (Esumi et al., 2005; Hirano et al., 2012; Kaneko et al., 2006,). As a result, each neuron expresses approximately 15 Pcdh isoforms, including a random repertoire of 10 alternate α, β and γ isoforms and all 5 C-type isoforms (Yagi, 2012).

A critical functional connection between *Drosophila* Dscam1 isoforms and vertebrate clustered Pcdhs was made by the observation that conditional deletion of the mouse *Pcdhg* gene cluster in retinal starburst amacrine cells or in Purkinje cells results in defective dendritic self-avoidance (Lefebvre et al., 2012). This observation, in conjunction with the stochastic promoter choice mechanism, suggests that clustered Pcdhs may also mediate neurite self-avoidance by specifying single cell identity. Consistent with this suggestion, previous studies showed that a subset of Pcdh-γ isoforms can engage in specific homophilic interactions (Reiss et al., 2006; Schreiner and Weiner, 2010), suggesting that Pcdhs mediate contact-dependent repulsion in a manner similar to that of invertebrate Dscam1 proteins. However, the question of whether all Pcdh-α, –β, -γ and C-type isoforms engage in homophilic interactions, which would be required to generate sufficient diversity, has yet to be answered. Paradoxically, there are only 58 distinct clustered Pcdh isoforms in the mouse as compared to 19,008 Dscam1 isoforms with distinct ectodomains in *Drosophila*, raising the question of whether the molecular diversity provided by clustered Pcdhs is sufficient for discrimination between self and non-self. A possible answer to this question was proposed in a previous study that suggested that Pcdhγs can associate promiscuously as *cis* (same cell) tetramers that bind with homophilic specificity in *trans* (different cells)(Schreiner and Weiner, 2010). The large number of possible Pcdh tetramers would then dramatically increase cell surface diversity (Schreiner and Weiner, 2010; Yagi, 2012). However, In order to reach the level of diversity predicted by this model, and to determine whether alternate models are possible or likely, it is necessary to establish

the binding behavior of all the clustered Pcdhs, including that of the Pcdh α and β isoforms, which were not previously tested.

Here we provide direct evidence that all but one of the 58 clustered Pcdh isoforms mediate highly specific homophilic *trans* interactions. We show that the EC6 domains of alternate Pcdhαs and PcdhγC4 inhibit cell surface delivery and that *cis* interactions involving the membrane proximal EC domains (EC5-EC6) of other Pcdh isoforms can relieve this inhibition. Furthermore, when multiple clustered Pcdh isoforms representing all three clusters are co-expressed, strict homophilic cell-cell recognition is observed. Remarkably, cells expressing as many as five different Pcdh isoforms display specific homophilic interactions in cell aggregation assays. However, aggregation is prevented by the expression of a single mismatched Pcdh isoform. In contrast, when the mismatch is generated by co-expression of classical N-cadherin (N-cad), there is no effect on homophilic recognition mediated by the Pcdhs, revealing a fundamental difference between the behaviors of classical cadherins and Pcdhs. Based on these findings we present a theoretical analysis of the dependence of Pcdh diversity on the number of subunits in putative *cis*-multimeric recognition units, and on the number of common isoforms that can be tolerated between two contacting cells without resulting in incorrect self-recognition. We discuss the competing requirements of self-recognition and non-self discrimination, and argue that these requirements raise questions concerning the validity of a current model in which the basic Pcdh recognition unit is a random assembly of Pcdh tetramers.

Results

Cluster-wide analysis of Pcdh-mediated homophilic interactions

The mouse *Pcdh* gene cluster encodes diverse subfamilies of cell surface proteins: 12 alternate Pcdhα, 22 Pcdhβ, 19 Pcdhγ isoforms and 5 C-type Pcdh-α or -γ isoforms (Figures 1A and 1B). We examined the ability of each Pcdh isoform to mediate homophilic recognition using a K562 cell aggregation assay. K562 cells are non-adherent in culture with no endogenous Pcdh expression, and thus provide an assay for homophilic interactions mediated by transfected clustered Pcdh cDNAs (Reiss et al., 2006; Schreiner and Weiner, 2010). It is important to note that while this aggregation assay provides an excellent system for studying homophilic interactions between Pcdh proteins on the cell surface, it cannot provide information regarding the self-avoidance (neurite repulsion) function of Pcdhs in the nervous system.

We carried out a systematic analysis of the homophilic interactions of all 58 Pcdh proteins (α, β, γ and C type Pcdhs) by transfecting cDNA plasmids encoding individual Pcdh C-terminal mCherry fusion proteins into K562 cells and visualizing cell aggregation. We found that all 22 Pcdhβs, 19 alternate Pcdhγs, and the C-type Pcdhs -αC2, -γC3 and -γC5 form aggregates when assayed individually (Figure 1D), We note that the size of the aggregates observed varies significantly (Figure S1B), which is likely the consequence of differences in expression, cell surface delivery, or intrinsic trans-binding affinities of individual Pcdh isoforms. By contrast, none of the alternate Pcdhα

isoforms nor Pcdh-αC1 or -γC4 form aggregates (Figure 1D), presumably due to the lack of membrane localization (Bonn et al., 2007; Murata et al., 2004).

Pcdhβs, Pcdhγs, and a subset of C-type Pcdhs display highly specific homophilic interactions

The EC2 and EC3 domains, which display the highest level of amino acid sequence diversity among the EC domains (Figure S2A) (Wu, 2005) were previously shown to comprise the specificity-determining region for a subset of Pcdhγ isoforms (Schreiner and Weiner, 2010). In order to determine the stringency of recognition specificity, we generated pairwise sequence identity heat maps of the EC2-EC3 domains (Figures 2B and S2B). Using these heat maps, we identified Pcdh pairs with greater than 80% pairwise sequence identity in their EC2 and EC3 domains. If the most closely related Pcdhs within the same cluster fail to recognize each other thru heterophillic interactions, it is unlikely that the more distantly related Pcdhs would interact. Notably, among the closely related Pcdh pairs, Pcdhβ6-Pcdhβ8 and PcdhγA8-PcdhγA9 both share more than 90% sequence identity within their EC2-EC3 domains. Eight of the closely related Pcdhs were tested along with twelve more distantly related Pcdhs. In total, we tested 89 unique pairs of Pcdhs with sequence identity for non-self pairs ranging from 50-95% in their EC2-EC3 domains.

Each protein was expressed with mCherry or mVenus fused to the C-terminus and tested for binding specificity (Figure 2A). Pairwise Pcdh isoform combinations were tested within each Pcdh subtype (Figures 2C-2E and S2C) and between different subtypes (Figure S2D). Only self-pairs on the matrix diagonals displayed intermixing of

mCherry- and mVenus-expressing cells, while all non-self pairs exclusively segregate into red and green cell homophilic aggregates. Despite their high level of sequence identity, even the Pcdhβ6-Pcdhβ8 and PcdhγA8-PcdhγA9 pairs form separate, non-interacting homophilic cell aggregates. Thus, all of the Pcdh γ and β proteins tested display strict *trans*homophilic specificity.

Pcdhαs mediate homophilic recognition when delivered to the cell surface

As mentioned above, Pcdhα isoforms are not delivered to the plasma membrane when expressed alone (Bonn et al., 2007; Murata et al., 2004), so this likely explains why all of the Pcdhα isoforms fail to engage in homophilic interactions in the K562 assay (Figure 1D). We therefore used Pcdh constructs bearing an extracellular c-Myc tag to visualize cell surface localization by immunofluorescence in K562 cells. We first showed that Pcdh-β17, -γB6, and the C-type -αC2 and -γC3 isoforms, all of which engage in homophilic interactions (Figure 1D), can be detected on the cell surface (Figure 3A (ii-v)). By contrast, neither the wild type nor intracellular domain (ICD) deleted Pcdhα4 can be detected on the cell surface (Figure 3A (i and vi)). This observation is consistent with the idea that failure to detect homophilic interactions of Pcdhαs in the cell aggregation assay is due to failure of Pcdhαs to localize to the plasma membrane.

Previous studies have shown that Pcdhγs can facilitate membrane delivery of Pcdhαs (Murata et al., 2004). We confirmed this finding with PcdhγB6 (Figure 3A (ix)) and in addition, found that Pcdhβ17 (Figure 3A (viii)) and the C-type Pcdh-αC2 and -γC3 isoforms (Figure 3A (vii and x)) could also facilitate membrane delivery of Pcdhα4.

Deletion of Pcdh EC1 domains was previously shown to abrogate Pcdh homophilic interactions (Figure 3C (i-iv)) (Schreiner and Weiner, 2010) but not their cell surface delivery (Figure S3A (ii)). In order to determine whether Pcdhαs can directly mediate homophilic interactions, we co-expressed Pcdhα isoforms with ΔEC1-Pcdh isoforms, reasoning that *cis* interactions with these ΔEC1-Pcdh constructs would assist in cell surface delivery but would not participate in *trans* binding. Thus, the EC1-deleted "carrier" proteins should not affect the recognition specificity (see e.g. Figure 3C (ix – xii). We confirmed that all of the ΔEC1-Pcdh proteins tested can deliver Pcdhα4 to the cell surface (Figure 3A (xi-xiv)) and facilitate cell aggregation (Figure 3C (v-viii)). Consistent with these observations, Flag-tagged Pcdhα4 co-immunoprecipitates with mCherry fusions of each of the ΔEC1-Pcdhs or wild-type Pcdh carrier proteins tested (Figure S3B). Using carrier Pcdhs for membrane delivery, we found that all 12 alternate Pcdhαs mediate strict homophilic interactions (Figures 3B, S3D and S3E). Similar to the Pcdh -βs and –γs, EC1 deletion in Pcdhα4 abolished its homophilic binding activity when co-expressed with a carrier protein (Figure 3B (vii) and (xiv)).

In addition to the Pcdhα isoforms, PcdhγC4 and PcdhαC1 did not mediate homophilic interactions when transfected alone (Figure 1D). PcdhγC4 exhibits behavior similar to that of the Pcdhαs: its membrane delivery and homophilic interactions are promoted by co-transfection with carrier Pcdhs (Figure S3F, second row). By contrast, we found that PcdhαC1 homophilic interactions could not be rescued by co-expression with any of the carrier Pcdhs (Figure S3F, third row). To determine whether the co-transfected Pcdhα isoform defines binding specificity, we selected the closely related Pcdhα pairs, Pcdh-α8 and -α7 (97% identity in the EC2-EC3 domains) and Pcdh-α8

and -α4 (74% identity) for testing in cell aggregation assay (Figure 3D). Cells expressing the same Pcdhα isoform showed homophilic interactions (Figure 3E (ii)), whereas those expressing different Pcdhα isoforms did not interact (Figure 3E (i and iii)). Conversely, when the Pcdhα isoform is the same for all transfectants, but the carrier Pcdhs are varied (Figure 3F), intermixing of the red and green cells is observed between all transfectants, irrespective of the identity of the carrier Pcdh. These results demonstrate that the recognition specificity between cells co-transfected with an alternate Pcdhα and a carrier Pcdh depends only on the identity of the Pcdhα isoform.

Role of the membrane-proximal EC6 domain in cell surface localization

To identify the regions of Pcdhβ/γ proteins responsible for the carrier function, we produced an EC-domain deletion series of PcdhγB6 in which EC domains were successively deleted starting with EC1. Each of these constructs failed to mediate homophilic interactions (Figures 4A (i-vi) and S4G). We then co-transfected Pcdhα4 with each of the PcdhγB6 ΔEC constructs and assayed for cell aggregation. When co-transfected with Pcdhα4, aggregation was observed when up to four EC domains were deleted from PcdhγB6 (Figures 4A (vii-x) and S4G). Cell aggregation was not observed in co-transfectants in which the first five or all six EC domains were deleted from PcdhγB6 (Figures 4A (xi and xii) and S4G). When co-transfected with Pcdhα4, the PcdhγB6ΔEC1-4 mediates efficient membrane delivery of Pcdhα4 (Figure 4A (xv)). PcdhγB6ΔEC1-5 localizes to the cell surface when transfected alone (Figure 4A (xiv)), yet does not deliver Pcdhα4 to the cell surface (Figure 4A (xvi)). Similarly,

Pcdhβ17ΔEC1-4 also mediates efficient membrane delivery of Pcdhα4 (Figure S4A). We conclude that the EC5 and EC6 domains of Pcdhβ or γ are necessary to deliver the Pcdhα isoform to the cell membrane.

To determine which Pcdh domains regulate membrane delivery we performed experiments in which domains were shuffled between Pcdhα4, which does not localize to the membrane (Figure 3A (i)), and PcdhγC3, which does (Figure 3A (x)). Constructs in which EC domains or the ICD of Pcdhα4 were replaced with the corresponding domains of PcdhγC3, or vice versa, were produced and tested for cell aggregation activity (Figure 4B and S4H), a proxy for membrane delivery. Chimeric constructs bearing the EC6 domain of PcdhγC3 mediate homophilic interactions (Figures 4B (i-iv, vii, xiv) and S4B (ii))) and are delivered to the cell surface (Figure S4B (iv)). By contrast, chimeric constructs that include the EC6 domain from Pcdhα4 showed no cell aggregation activity (Figures 4B (vi, ix-xiii, xv) and S4B (i)) due to the failure to localize to the plasma membrane (Figure S4B (iiii)). To address the possibility that the domain substitutions affect properties other than cell surface delivery, we co-transfected all Pcdhγ-Pcdhα chimera constructs containing the EC6 domain of Pcdhα4 (Figure S4C (i-vi)) with the carrier PcdhγB6ΔEC1. We found that these co-transfectants mediate homophilic interactions (Figure S4C (vii-xii)), demonstrating that the chimeric proteins are functional. Similar domain shuffling experiments were performed for other alternate Pcdhα isoforms (Figure S4D) and C type isoforms (Figures S4E and S4F). We conclude that the EC6 domain of alternate Pcdhα isoforms and of the PcdhγC4 isoform inhibit membrane delivery.

We next determined whether deletion of the EC6 domain in Pcdhα isoforms can rescue membrane delivery and homophilic binding. We found that Pcdhα4ΔEC6 is, in fact, efficiently delivered to the cell surface (Figure 4C (iii)) and mediates cell aggregation (Figures 4C (i) and S4I). These results, together with the domain swapping experiments (Figures 4B and S4D-S4F), show that the EC6 domain regulates Pcdh cell surface delivery but is not required for homophilic *trans* interactions.

Co-expression of multiple Pcdh isoforms generates new homophilic specificities

Previous studies suggested that multiple Pcdhγ isoforms form *cis* tetramers capable of mediating homophilic interactions (Schreiner and Weiner, 2010). Since all Pcdh-α, β and γ isoforms except PcdhαC1, mediate homophilic interactions (Figures 1D, 3B and S3D), and appear to associate with each other in *cis* (Figure S3B)(Han et al., 2010; Schalm et al., 2010), we tested the possibility that recognition specificity is diversified by co-expression of multiple Pcdh isoforms from all three subfamilies. Cells co-expressing Pcdhα4 and Pcdhβ4 were mixed with cells expressing both of these isoforms, or only one (Figure 5A (i-iii)). Cells expressing two distinct isoforms failed to co-aggregate with cells expressing either isoform alone. However, robust aggregation was observed with cells that co-express both isoforms. Similar results were observed for each of the Pcdh pairs shown in Figures 5A, S5A and S5C. These results suggest that the presence of one non-matching isoform can interfere with co-aggregation.

To test whether this type of interference is unique to Pcdhs, we carried out experiments similar to those reported in Figures 5A, but using cells co-transfected with N-cad and different Pcdhs. Figure 5B shows the results of aggregation assays with

cells expressing various combinations of N-cad and Pcdhβ18 or PcdhγB6. Three types of aggregation behavior are observed. These three behaviors can be described as (1) formation of completely separate red and green aggregates, (2) complete intermixing between cell populations, and (3) formation of separate red and green aggregates that adhere to one another. Two of these aggregation phenotypes are seen in the top panels of Figure 5B, where red cells expressing N-cad form separate aggregates from green cells expressing any of the two Pcdhs (i and ii), but form a completely mixed aggregate with green cells expressing N-cad (iii). For these two cases, the different aggregation behaviors reflect the fact that N-cad does not bind to these Pcdhs, but binds strongly to itself. Figures 5B (iv-ix), S5D, S5F and S5G depicts the behavior of cells co-expressing N-cad and one Pcdh when they are allowed to mix with either N-cad or Pcdh expressors. In each case, these green cells form completely intermixed aggregates with red cells expressing N-cad alone, or with red cells expressing N-cad and the identical Pcdh or non-matching Pcdh, reflecting strong homophilic N-cad interactions with which Pcdhs do not interfere. The third type of aggregation behavior is observed when the red cells express both N-cad and a Pcdh isoform, and the green cells express only the identical Pcdh isoform (Figure 5B vi and ix). In this case, separate green and red homophilic aggregates are formed, but importantly they now adhere to one another. Similarly, all three types of behavior are observed for cells co-expressing N-cad and two Pcdhs (Figures S6D-E and S6G). The behavior of N-cad/Pcdh co-transfectants is strikingly different from that observed for Pcdh co-transfectants with mismatches, in which all homophilic aggregates remain completely separate.

These results strongly suggest that Pcdhs interact in *cis* so as to create new homophilic specificities that differ from the specificities of the individual Pcdh isoforms. In contrast, N-cad/Pcdh co-transfectants behave in a way that can be explained by a summation of the properties of the individual proteins, showing no evidence of *cis* interaction between them (Figure 5H). Thus, interference appears to be a property that is unique to Pcdhs. We note that co-IP experiments are consistent with *cis* interactions between Pcdhs, and with their absence between Pcdhs and N-cad (Figures S5E-S5G, and S6F).

To further characterize the Pcdh interference phenomenon, we assessed the ability of cells co-transfected with up to five Pcdh isoforms to co-aggregate with cells containing various numbers of mismatches (Figures 5C-G, 6A-C, S5B, S5G, S6A-C and S6F). In all cases, mixed aggregates were observed only for cells expressing identical isoforms whereas cells expressing mismatched isoforms formed separate non-adhering aggregates (Figure 5 and 6). Remarkably, even cells co-expressing distinct sets of four or five isoforms with even a single non-matching isoform resulted in the formation of large non-contacting homophilic aggregates with no contacts between them (Figure 6C).

Discussion

The stochastic single-cell expression of clustered Pcdhs, the diversity of Pcdh extracellular domains, and the demonstration that the *Pcdhg* gene cluster is required for dendritic self-avoidance in starburst amacrine and Purkinjie cells support the hypothesis that the clustered Pcdhs provide single cell identity necessary for self-recognition in vertebrate nervous systems (Chen and Maniatis, 2013; Yagi, 2012; Zipursky and Grueber, 2013; Zipursky and Sanes, 2010). Here we provide evidence that different combinations of Pcdh α, β and γ isoforms interact in *cis* to generate combinatorial *trans* recognition specificities. The importance of Pcdh *cis* interactions is demonstrated by their role in delivering Pcdhα isoforms to the membrane. Below, we summarize evidence supporting these conclusions, we provide a theoretical analysis of Pcdh single cell diversity and we discuss the implications of this analysis on a prevailing model based on tetrameric *cis* recognition units (Schreiner and Weiner, 2010; Yagi, 2012). We conclude that, although recognition involving coupled *cis* and *trans* interactions (Wu et al., 2011; Wu et al., 2010) lies at the core of the mechanism through which Pcdhs establish single-cell identity, the nature of Pcdh recognition units and the mechanism of their interactions remains uncertain.

α, β, γ and C-type Pcdhs mediate specific homophilic recognition

We showed that Pcdh isoforms from all three gene clusters (α, β and γ) can mediate highly specific homophilic interactions (Figures 1D, 3B, S3D and S3F). Striking examples of this *trans* homophilic specificity are provided by the observation that Pcdh isoform pairs with as great as 91-97% identity in their EC2-EC3 recognition domains (α7-α8, β6-β8 and γA8-γA9) do not engage in heterophilic interactions (Figures 2C, 2D, and 3E). While PcdhαC1 does not interact homophilically in the aggregation assay (Figures 1D and S3F), a chimeric construct containing the PcdhαC1 EC1-EC3 domains can mediate homophilic interactions (Figure S4E (i)). Thus, it seems likely that the function of PcdhαC1 involves self-recognition, although the biological context is not yet understood. We note that, unlike the other Pcdhs, the calcium-binding motif DRE is not present in the EC3 domain of PcdhαC1 (Figure S1A). Rather, this motif is replaced by the sequence GPP, which is conserved in the PcdhαC1s in other species. We therefore speculate that the unique behavior of PcdhαC1 in the cell aggregation assay may result from differences in protein structure due to the absence of the calcium-binding motif, as, for example, in DN-cadherin (Jin et al., 2012).

Evidence for Pcdh *cis* interactions

Definitive evidence for *cis* interactions between distinct Pcdh isoforms is lacking. However, a number of experimental observations provide strong support for this possibility. First, we observe an altered recognition specificity when multiple Pcdh isoforms are expressed, a property thus far unique to Pcdhs. It is difficult to imagine how this could occur without *cis* interactions. Second, Pcdh β, γ and certain C-type

isoforms deliver Pcdhα proteins to the cell surface in a process that requires membrane proximal domains (EC5 and EC6) of the carrier proteins, which are likely to be involved in *cis* interactions (Figure 4A). Third, distinct Pcdh-α, β and γ isoforms can be co-immunoprecipitated (Han et al., 2010; Murata et al., 2004; Schalm et al., 2010; Schreiner and Weiner, 2010) (Figures S3B and S6F). Fourth, multiple Pcdh isoforms are found in high molecular weight detergent-solubilized Pcdh complexes from the brain (Han et al., 2010).

Analysis of domain deletion and substitution experiments revealed a critical role of Pcdhα EC6 domains in the inhibition of cell surface delivery. Differential cell surface localization functions of EC6 domains may be reflected in amino acid sequence differences between them. The EC6 domains are the most highly conserved within the Pcdh -β and -γ subfamilies (Figures S2A and 4D), but differ from the EC6 domains of the Pcdhα isoforms (Figures 4D and 4E). The correspondence between membrane-delivery phenotypes and distinct EC6 sequence signatures suggests that the carrier function is a conserved property of clustered Pcdhs. The question of whether Pcdh *cis* complexes are stable on the cell surface, or can exchange *cis* partners in the plasma membrane remains to be determined. Reassortment of multimeric complexes on the cell surface would have obvious implications for Pcdh cell surface diversity and combinatorial specificity.

Combinatorial homophilic interactions between Pcdh α, β & γ isoforms

The key findings of cell aggregation assays can be interpreted in terms of the differential adhesion hypothesis (Foty and Steinberg, 2005) and the relationship

between molecular binding affinities and the strength of cell-cell adhesion (Katsamba et al., 2009). Specifically, the aggregates we observe are likely the consequence of maximizing the number of favorable protein-protein interactions between cells. For example, cells expressing five Pcdh isoforms will prefer to form homophilic aggregates with cells expressing identical isoforms rather than to intermix with cells expressing only four of the five isoforms (Figure 6C). The cells expressing four Pcdh isoforms would similarly be expected to form homophilic aggregates with each other, in order to maximize the number of protein-protein interactions. However, one would also expect the two types of homophilic aggregates to adhere to one another, again to maximize favorable protein-protein contacts, as was observed in the experiments with N-cad and Pcdh(s) (Figures 5B, S6D, S6E and S6G). Remarkably, contact between aggregates expressing distinct Pcdh isoforms does not occur, suggesting a mechanism in which mismatched isoforms interfere with intercellular interactions. Indeed in all cases tested here (Figures 5 and 6), even a single Pcdh mismatch is sufficient to prevent the two types of homophilic aggregates from adhering to each another.

What is the maximum fraction of expressed isoforms that two cells can have in common before incorrectly recognizing each other as self? Our results suggest that at least in the cases examined, up to 80% (4/5) of the Pcdh common-isoforms can be shared between two cell populations without triggering co-aggregation (Figure 6C). In contrast, Schreiner and Weiner (2010) reported a graded recognition in which expression of 50% (1/2) and 75% (3/4) common-isoforms resulted in corresponding percentage of binding (~30-50% and ~70% respectively). These differences are likely to be due, at least in part, to different experimental approaches. Specifically, we used

direct visualization to assess the specificity of cell-cell interactions and to determine which types of aggregates are formed (Figure S5G and S6G). In contrast, the previous report utilized an indirect colorimetric assay in which different types of aggregates could not be distinguished.

Theoretical analysis of Pcdh-mediated neuronal diversity

The prevailing model for generating neuronal diversity by Pcdhs involves the existence of discrete tetrameric recognition units formed by random combinations of Pcdh proteins that interact in *cis* (Schreiner and Weiner, 2010; Yagi, 2012). To consider this model in detail and to evaluate the implications of the high level of common-isoform tolerance identified in our study, we carried out an independent analysis of the factors that may contribute to Pcdh-mediated neuronal identity. Our analysis is based in part on earlier studies on Dscam1 by (Hattori et al., 2009) and by (Forbes et al., 2011), but focuses on the issue of isoform tolerance and introduces a factor not addressed previously; specifically how do neurites of the same neuron recognize that they are "the same"? We believe that the *cis*-tetramer model fails to answer this question. We begin with an analysis of isoform tolerance, which is key to understanding neuronal non-self discrimination.

For both Pcdhs and invertebrate Dscam1, the probability of errors in non-self discrimination depends on three parameters: the total number of potential isoforms, the number of distinct isoforms expressed per cell, and the tolerance for common isoforms between cells in contact (Hattori et al., 2009). Common-isoform tolerance is defined as the maximum percentage of common isoforms that can be present in two cells in

contact without incorrect recognition as self. Based on this model (Hattori et al., 2009), if two cells have a higher fraction of common isoforms than the tolerance, they will inappropriately recognize each other as self. Hattori et al., (2009) assumed low tolerance for Dscam1 (10-20%), which is intuitively reasonable, since two cells expressing larger fractions of common isoforms would be expected to bind to one another. However to our knowledge, no experimental measure of tolerance has been reported for Dscam1. The results of the work presented here reveal much higher common-isoform tolerance levels for Pcdhs than assumed for Dscam1. This difference is likely the consequence of homophilic interactions between Pcdh *cis* multimers, in contrast to the Dscam1 isoforms which appear to interact as monomers. In the following section, we present an analysis of the inter-related effects of isoform diversity and isoform tolerance on non-self recognition. This in turn makes it possible to discuss Pcdhs and Dscam1 within a common framework.

Figure 7A shows the probabilities that two cells stochastically expressing different numbers of Pcdh isoforms will improperly recognize each other as self. Given the total number of possible isoforms, the number of isoforms expressed per cell (the x-axis in the figure), and a common-isoform tolerance, analytical expressions (Forbes, 2011) or Monte-Carlo simulations (Hattori et al., 2009) can be used to calculate these probabilities (See Supplemental Information). Results for Dscam1 were reported for a 5000-member isoform pool with a 15% tolerance (Hattori et al., 2009). In the case of Pcdhs, we made the conservative assumption of 67% tolerance (2/3 as observed in Figure 6A), and a 58-member isoform pool. Remarkably, even with a 67% common-isoform tolerance for clustered Pcdhs, the probabilities of incorrect recognition are as

low as those for Dscam1 isoforms over much of the region that includes the expected number of isoforms (estimated at about 15 for Pcdhs and 10–50 for Dscam1) (Hattori et al., 2009; Yagi, 2012). These results suggest that a mechanism for achieving extremely high common-isoform tolerance is a key factor explaining how only 58 Pcdhs may be sufficient to mediate non-self discrimination in vertebrates.

Combinatorial specificity of Pcdh interactions based on the assembly of multimeric *cis* Pcdh recognition units containing isoforms from all three gene clusters provides a possible mechanism to achieve the observed high level of tolerance. To illustrate this, we consider a model similar to that proposed for *cis*-tetramers (Schreiner and Weiner, 2010; Weiner et al., 2013; Yagi, 2012). We noted that the *cis*-tetramer model was based on a molecular weight estimate from size-exclusion chromatography (Schreiner and Weiner, 2010). However, the molecular weight of elongated proteins such as Pcdhs cannot be rigorously determined by this method, nor can it distinguish between *cis* and *trans* multimers. We therefore did not assume a specific multimeric state in our analysis.

A specific case where two cells each express one common and one different isoform and engage in cell-cell interactions through monomer, dimer, trimer or tetramer recognition units is illustrated in Figure 7B. As the multimer size increases, the fraction of common-recognition units decreases. This behavior is generalized in Figure 7C, which shows that at the same common-isoform tolerance, larger multimers will have a lower common-recognition unit tolerance, thus increasing cell surface diversity. For example, assuming tetrameric recognition units with 67% common isoforms (2/3) between two cells, only 20% of the recognition units will be shared, well within the range

assumed for Dscam1 monomers (Hattori et al., 2009). This result highlights the essential feature of the tetramer model (Yagi, 2012). A tetrameric recognition unit implies that different neurons will have only a small fraction of recognition units in common even if they have a high fraction of common isoforms. In this way, mismatched isoforms will interfere with cell-cell recognition by diluting the number of common recognition units between two contacting cells (Yagi, 2013). However, this analysis did not consider the effect of dilution on self-recognition.

Randomly assembled tetrameric recognition units in which all Pcdh isoforms form multimers with equal probability cannot explain how two neurites from the same cell body are able to recognize each other as self. The point can be easily seen by calculating the average number of copies of each multimeric recognition unit per cell as a function of the number of copies of each Pcdh isoform in a cell. Figure 7D reports these numbers for the case of 15 different isoforms expressed per cell. A striking conclusion is that, for tetramers, there would be an unacceptably small number of copies of each recognition unit per neuron. For example, assuming that there are 5,000 copies each of 15 distinct Pcdh isoforms in an individual cell (75,000 Pcdhs total - in the range estimated for cells overexpressing classical cadherins (Duguay et al., 2003)), 12,720 unique tetramers could form (Yagi, 2012) and there would thus be fewer than two copies (approximately 75,000/(12,720*4)=1.4) of each unique recognition unit per neuron. This number is clearly insufficient for self-recognition by neurons with many neurites. This self-recognition problem is reduced but not eliminated for trimeric and dimeric recognition units (Figure 7D).

These considerations bring into question the validity of the tetramer model in which all isoforms have equal probability of participation. This would be less of a problem if only certain combinations of Pcdh isoforms could assemble into multimers. For example, our data indicates that Pcdhα isoforms may form obligate complexes with Pcdh-β or –γ isoforms, or with constitutively expressed C-type isoforms to function on the cell surface. The obligate assembly could also determine the nature of the multimeric complexes. Another possibility is that like classical cadherins (Harrison et al., 2011), Pcdhs could form junction-like structures involving *cis* and *trans* interactions, which require a minimal percentage of matched isoforms to mediate stable adhesion. With such a mechanism, an excess of mismatched isoforms in contacting cells would reduce the number of favorable interactions so as to prevent junction formation.

We conclude that specific models of Pcdh combinatorial homophilic interactions cannot be rigorously supported at the present time. Neither the physical properties of the proposed multimeric complexes nor the mechanism of their interactions can be discerned on the basis of the currently available data. Nevertheless, it seems highly likely that the clustered Pcdhs play a fundamental role in intercellular recognition in the vertebrate nervous system based on the extraordinary diversity of their single cell expression, and the highly specific homophilic interactions between individual as well as combinations of all three families of clustered Pcdh isoforms as shown here. Most remarkable in this regard is interference in the interactions between cells each expressing five distinct Pcdh isoforms only one of which differs in the two cell populations. Based on these observations, and the demonstrated role of the Pcdhs in

dendritic self-avoidance, these cell surface proteins clearly play a role in the establishment and maintenance of complex neural circuits in the brain.

Experimental procedures

Plasmid construction. Coding sequence of each clustered Pcdh isoform was PCR amplified from C57BL/6 genomic DNA or brain cDNA, cloned into modified Gateway vectors to generate C-terminal mCherry- or mVenus-tagged Pcdh proteins. Domain deletions, substitutions, or insertion of an extracellular c-Myc tag were created by overlapping PCR. See Extended Experimental Procedures for details.

Cell aggregation assay. Expression constructs were transfected into K562 cells (human leukemia cell line, ATCC CCL243) by electroporation using Amaxa 4D-Nucleofactor (Lonza). After 24 hours in culture, the transfected cells were allowed to aggregate for one to three hours on a rocker kept inside the incubator. The cells were then fixed in 4% PFA for 10 minutes, washed in PBS, and cleared with 50% glycerol for imaging. Quantification of the sizes of cell aggregates was described in Extended Experimental Procedures.

Immunostaining. K562 cells were transfected as described above. After 24 hours, FITC-conjugated anti-c-Myc antibodies were added to the cells and then incubated with shaking for one hour. Cells were then fixed with 4% PFA and washed in PBS. Fixed single cells or aggregates were collected on glass coverslips by using a cell concentrator (StatSpin) at 1000 rpm for 10 minutes. Images were collected with an Olympus Fluoview FV1000 confocal microscope.

Binding specificity assay for cells expressing single or multiple Pcdh isoform(s). Differentially tagged Pcdh isoforms were transfected into K562 cells as described above. Transfected cell populations expressing mCherry- or mVenus-tagged Pcdh(s) were mixed after 24 hours by shaking for one to three hours. Images of cell aggregates were imported into ImageJ (http://rsbweb.nih.gov/ij), and the number of aggregates containing red cells only (R), green cells only (G), and both red and green cells (RG) were counted for analysis of binding specificity. See Extended Experimental Procedure for details.

Supplemental information

Supplemental information includes Extended Experimental Procedures, seven supplementary figures and spreadsheet for raw data.

Author contributions

C.A.T. designed and performed experiments. W.V.C. co-initiated the study, and contributed to the design and establishment of experimental approaches. R.R. led the computational and statistical analysis. K.F predicted the EC domain alignment. M.C. provided technical support. J.C.T. contributed to image analysis. C.A.T., W.V.C., R.R., H.W.,L.S.,B.H.,T.M analyzed the data and wrote the paper.

Acknowledgments

We thank Drs. Richard Axel, Charles Zuker, Wesley Grueber and Hemali Phatnani for critical reading of the manuscript, and valuable comments. We also thank Lin Jin and Angelica Struve for technical assistance and members of the Maniatis, Shapiro, and Honig labs for discussion and comments. We thank Drs. Joshua Sanes, Julie Lefebvre, Stefanie S. Schalm and Steven Vogel for providing reagents and plasmids. This work was supported by the National Institutes of Health grant to T.M. (2R56NS043915-33A1), joint NIH grant to T.M. and L.S. (1R01GM107571-01), National Science Foundation grant to B.H (MCB-0918535), and NIH Training Programs to H.W. (T32GM008281) and K.F. (T32GM082797).

References

Bonn, S., Seeburg, P.H., and Schwarz, M.K. (2007). Combinatorial expression of alpha- and gamma-protocadherins alters their presenilin-dependent processing. Mol Cell Biol *27*, 4121-4132.

Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. Development *140*, 3297-3302.

Duguay, D., Foty, R.A., and Steinberg, M.S. (2003). Cadherin-mediated cell adhesion and tissue segregation: qualitative and quantitative determinants. Dev Biol *253*, 309-323.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S., and Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. Nat Genet *37*, 171-176.

Forbes, E.M., Hunt, J.J., and Goodhill, G.J. (2011). The combinatorics of neurite self-avoidance. Neural computation *23*, 2746-2769.

Foty, R.A., and Steinberg, M.S. (2005). The differential adhesion hypothesis: a direct evaluation. Dev Biol *278*, 255-263.

Han, M.H., Lin, C., Meng, S., and Wang, X. (2010). Proteomics analysis reveals overlapping functions of clustered protocadherins. Mol Cell Proteomics *9*, 71-83.

Harrison, O.J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M.*, et al.* (2011). The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. Structure *19*, 244-256.

Hattori, D., Chen, Y., Matthews, B.J., Salwinski, L., Sabatti, C., Grueber, W.B., and Zipursky, S.L. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature *461*, 644-648.
Hattori, D., Millard, S.S., Wojtowicz, W.M., and Zipursky, S.L. (2008). Dscam-mediated cell recognition regulates neural circuit formation. Annu Rev Cell Dev Biol *24*, 597-620.
Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron diversity generated by Protocadherin-beta cluster in mouse central and peripheral nervous systems. Front Mol Neurosci *5*, 90.
Jin, X.S., Walker, M.A., Felsovalyi, K., Vendome, J., Bahna, F., Mannepalli, S., Cosmanescu, F., Ahlsen, G., Honig, B., and Shapiro, L. (2012). Crystal structures of Drosophila N-cadherin ectodomain regions reveal a widely used class of Ca2+-free interdomain linkers. P Natl Acad Sci USA *109*, E127-E134.
Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., and Yagi, T. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. J Biol Chem *281*, 30551-30560.
Katsamba, P., Carroll, K., Ahlsen, G., Bahna, F., Vendome, J., Posy, S., Rajebhosale, M., Price, S., Jessell, T.M., Ben-Shaul, A.*, et al.* (2009). Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. Proc Natl Acad Sci U S A *106*, 11594-11599.
Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature *488*, 517-521.
Miura, S.K., Martins, A., Zhang, K.X., Graveley, B.R., and Zipursky, S.L. (2013). Probabilistic splicing of dscam1 establishes identity at the level of single neurons. Cell *155*, 1166-1177.
Murata, Y., Hamada, S., Morishita, H., Mutoh, T., and Yagi, T. (2004). Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. J Biol Chem *279*, 49508-49516.
Neves, G., Zucker, J., Daly, M., and Chess, A. (2004). Stochastic yet biased expression of multiple Dscam splice variants by individual cells. Nat Genet *36*, 240-246.
Reiss, K., Maretzky, T., Haas, I.G., Schulte, M., Ludwig, A., Frank, M., and Saftig, P. (2006). Regulated ADAM10-dependent ectodomain shedding of gamma-protocadherin C3 modulates cell-cell adhesion. J Biol Chem *281*, 21735-21744.
Ribich, S., Tasic, B., and Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. Proc Natl Acad Sci U S A *103*, 19719-19724.
Schalm, S.S., Ballif, B.A., Buchanan, S.M., Phillips, G.R., and Maniatis, T. (2010). Phosphorylation of protocadherin proteins by the receptor tyrosine kinase Ret. Proc Natl Acad Sci U S A *107*, 13894-13899.
Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes Dev *23*, 147-156.
Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671-684.

Schreiner, D., and Weiner, J.A. (2010). Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. Proc Natl Acad Sci U S A *107*, 14893-14898.
Sun, W., You, X., Gogol-Doring, A., He, H., Kise, Y., Sohn, M., Chen, T., Klebes, A., Schmucker, D., and Chen, W. (2013). Ultra-deep profiling of alternatively spliced Drosophila Dscam isoforms by circularization-assisted multi-segment sequencing. EMBO J *32*, 2029-2038.
Tasic, B., Nabholz, C.E., Baldwin, K.K., Kim, Y., Rueckert, E.H., Ribich, S.A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. Mol Cell *10*, 21-33.
Wang, X., Su, H., and Bradley, A. (2002). Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. Genes Dev *16*, 1890-1905.
Weiner, J.A., Jontes, J.D., and Burgess, R.W. (2013). Introduction to mechanisms of neural circuit formation. Front Mol Neurosci *6*, 12.
Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell *118*, 619-633.
Wojtowicz, W.M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S.L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell *130*, 1134-1145.
Wu, Q. (2005). Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. Genetics *169*, 2179-2188.
Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell *97*, 779-790.
Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M.*, et al.* (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. Genome Res *11*, 389-404.
Wu, Y., Vendome, J., Shapiro, L., Ben-Shaul, A., and Honig, B. (2011). Transforming binding affinities from three dimensions to two with application to cadherin clustering. Nature *475*, 510-513.
Wu, Y.H., Jin, X.S., Harrison, O., Shapiro, L., Honig, B.H., and Ben-Shaul, A. (2010). Cooperativity between trans and cis interactions in cadherin-mediated junction formation. P Natl Acad Sci USA *107*, 17592-17597.
Yagi, T. (2012). Molecular codes for neuronal individuality and cell assembly in the brain. Front Mol Neurosci *5*, 45.
Yagi, T. (2013). Genetic basis of neuronal individuality in the mammalian brain. Journal of neurogenetics *27*, 97-105.
Zhan, X.L., Clemens, J.C., Neves, G., Hattori, D., Flanagan, J.J., Hummel, T., Vasconcelos, M.L., Chess, A., and Zipursky, S.L. (2004). Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies. Neuron *43*, 673-686.
Zipursky, S.L., and Grueber, W.B. (2013). The molecular basis of self-avoidance. Annual review of neuroscience *36*, 547-568.
Zipursky, S.L., and Sanes, J.R. (2010). Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. Cell *143*, 343-353.

Figure Legends

Figure 1. The Pcdh gene cluster encodes a large repertoire of cell surface recognition proteins.

(A) Schematic representation of the mouse *Pcdh-a, -b* and *-g* gene clusters. Variable exons of each subtype are differentially color-coded. *Pcdha* and *Pcdhg* variable exons are joined via *cis*-splicing to three constant exons. An example is shown for *Pcdha9*. Each variable exon encodes six EC domains, a TM, and a short cytoplasmic extension. The constant exons encode the common ICD domain.

(B) Schematic diagrams representing the four major subtypes of Pcdhs are shown.

(C) Schematic diagram of the cell aggregation assay. mCherry-tagged Pcdh proteins are expressed in K562 cells to assay for their ability to induce cell aggregation. As

shown in the examples, cells expressing mCherry alone do not aggregate, while robust cell aggregation is observed with cells expressing PcdhγC3-mCherry.

(D) Survey of homophilic binding properties of all 58 Pcdh isoforms in the cell aggregation assay. Scale bar, 50 μm. (See also Figure S1B).

Figure 2. Pcdh-β, –γ, and C-type isoforms engage in specific homophilic interactions.

(A) Schematic diagram of the binding specificity assay. Cells expressing differentially tagged Pcdh isoforms are then mixed and assayed for homophilic or heterophilic interactions. A strict homophilic interaction is indicated by mixed red-and-green co-aggregates between cells expressing only the identical isoforms and segregation of separate red and green aggregates between cells expressing different isoforms.

(B) Heat map of pair-wise protein sequence identities of the EC2-EC3 domains of Pcdh isoforms and their evolutionary relationship is presented. Subsets of the isoforms within the boxed region were assayed. (See also Figure S2B).

(C-E) Pairwise combinations within each subtype Pcdhβ (C), Pcdhγ (D), and C-type (E) isoforms were assayed for their binding specificity. Scale bar, 50 μm. (See also Figure S3C).

Figure 3. Pcdhα isoforms engage in specific homophilic interactions when delivered to the cell surface by co-expressed Pcdh-β or -γ isoforms.

(A) Surface expression of mCherry-tagged Pcdh constructs bearing an extracellular c-Myc tag were shown. White arrows indicate the c-Myc staining at cell-cell contacts. Scale bar, 10 μm.

(B) Cells transfected with single Pcdhα isoforms (upper panels) and cells co-transfected with Pcdhα isoforms and PcdhγB6ΔEC1 (lower panels) were assayed for aggregation. Scale bar, 50 μm.

(C) Cells expressing ΔEC1-Pcdhs alone (upper panels), Pcdhα4/ΔEC1-Pcdhs (middle panels) were assayed for aggregation. Cells co-expressing Pcdhα4 and a carrier ΔEC1-Pcdhs do not interact with cells expressing only the wild-type carrier Pcdhs (lower panels). Scale bar, 50 μm. (See also Figure S3C).

(D) Heat map of pairwise sequence identities of the EC2-EC3 domains of Pcdhα isoforms. The boxed region shows Pcdh-α4, -α7 and -α8, which share a high level of sequence conservation.

(E-F) Cells co-expressing pairs of differentially tagged Pcdhαs and ΔEC1-Pcdhs were assayed for co-aggregation.

Figure 4. The role of EC6 domains in membrane delivery.

(A) Mapping the minimum binding region of carrier Pcdhs. Cells expressing PcdhγB6 mutants alone (Upper panels) and with Pcdhα4 (Lower panels) were assayed for aggregation. Cell surface expression of Myc-tagged Pcdhα were shown (xv-xvi). (See also Figure S4G)

(B) Schematic representation of chimeric proteins and the results of homophilic binding assays are presented. All of the chimeras bearing the EC6 domain from PcdhγC3

(yellow) mediate cell aggregation. All of the chimeras bearing the EC6 domain of Pcdhα4 (red) fail to mediate cell aggregation. Scale bar, 50 μm. (See also Figure S4B and S4H).

(C) Cells expressing Pcdhα4 EC6/ICD domain deletion mutants are tested for aggregation (i-ii). Surface expression of Pcdhα constructs were shown on the right (iii-iv). (See also Figure S4I)

(D) Heat map of pairwise sequence identities for EC6 domains. The EC6 domain is highly conserved in alternate Pcdhβ and Pcdhγ isoforms, but the EC6 domains of alternate Pcdhα isoforms are less conserved.

(E) Multiple sequence alignment of EC6 domains of membrane-delivered Pcdhs (light gray) and non-membrane delivered Pcdhs (dark grey). Residues conserved within only one group are highlighted in blue and invariant residues in red.

White arrows indicate the c-Myc staining at cell-cell contacts in A and C. Scale bar, 50 μm.

Figure 5. Co-expression of two distinct Pcdh isoforms generates a unique cell surface identity.

(A) Cells co-expressing two distinct mCherry-tagged Pcdh isoforms were assayed for interaction with cells expressing an mVenus-tagged Pcdh isoform or identical pairs. Pcdhα4^{+} is efficiently membrane-delivered and it possess the EC6 domain from PcdhγC3. (See also Figures S5A, S5C, S5G)

(B) Cells expressing mCherry-tagged N-cad were assayed for interaction with cells expressing single Pcdh isoform (Upper panels). Cells co-expressing pair of mCherry-tagged N-cad and Pcdh isoform were assayed for interaction with cells expressing an

mVenus-tagged N-cad or Pcdh isoform (Middle and last panels). (See also Figures S5D, S5G, S6D, S6E and S6G)

(C-G) Cells co-expressing different combinations of differentially tagged Pcdh/Pcdh pairs were mixed and assayed for their interaction. (See also Figure S6B)

(H) Illustration of the outcome of cell-cell interaction dictated by combinatorial homophilic specificity of two distinct Pcdh isoforms (e.g a-c). Illustration of the outcome of cell-cell interaction dictated by cells co-expression N-cad and single Pcdh isoform (e.g d-g). This schematic diagram presented here does not reflect the *cis*-dimer and asterisk represents the non-matching Pcdhγ.

Figure 6. Combinatorial co-expression of multiple Pcdh isoforms generates unique cell surface identities.

(A-C) Cells co-expressing an identical or a distinct set Pcdh-α, $-\beta$, and $-\gamma$ isoforms (A) and with C-type isoforms (B-C) were assayed for co-aggregation. Pcdh$\alpha 4^{+}$ is efficiently membrane-delivered and it possess the EC6 domain from PcdhγC3. The non-matching isoforms between two cell populations were underlined. (See also Figure S6A-C and S6G)

(D) Illustration of the different behaviors of cell-cell interaction generated by combinatorial homophilic specificity of distinct sets of multiple Pcdh isoforms. (See also Figure S5G)

Figure 7. Probabilistic analysis of Pcdh and Dscam1 recognition

(A) Probabilities of incorrect non-self recognition between two cells as a function of the number of isoforms expressed per cell. Lines in the plot appear jagged due to the integer number of tolerated isoforms.

(B) Schematic representation of recognition units for different *cis*-multimeric states. Two cells share one common Pcdh isoform (blue) and one distinct isoform (red and yellow). Unique multimers are shown and the number of permutations for each multimer (e.g. ×2) are given.

(C) The relationships between common-isoforms and common-recognition units for different multimeric states. Vertical dotted lines mark the cases of 67% common Pcdh isoforms and show the corresponding percentage of common recognition units for monomers, dimers, trimers and tetramers. For the same percentage of common isoforms, larger multimers have a smaller percentage of common recognition unit.

(D) Monte-Carlo simulations were used to estimate the average number of copies of each multimer in a single cell. For the case of 15 Pcdh isoforms expressed per cell, the average number of copies of each multimeric recognition unit generated by the stochastic assembly of Pcdh isoforms into multimers is shown as a function of the number of copies of each isoform.

Figure 1

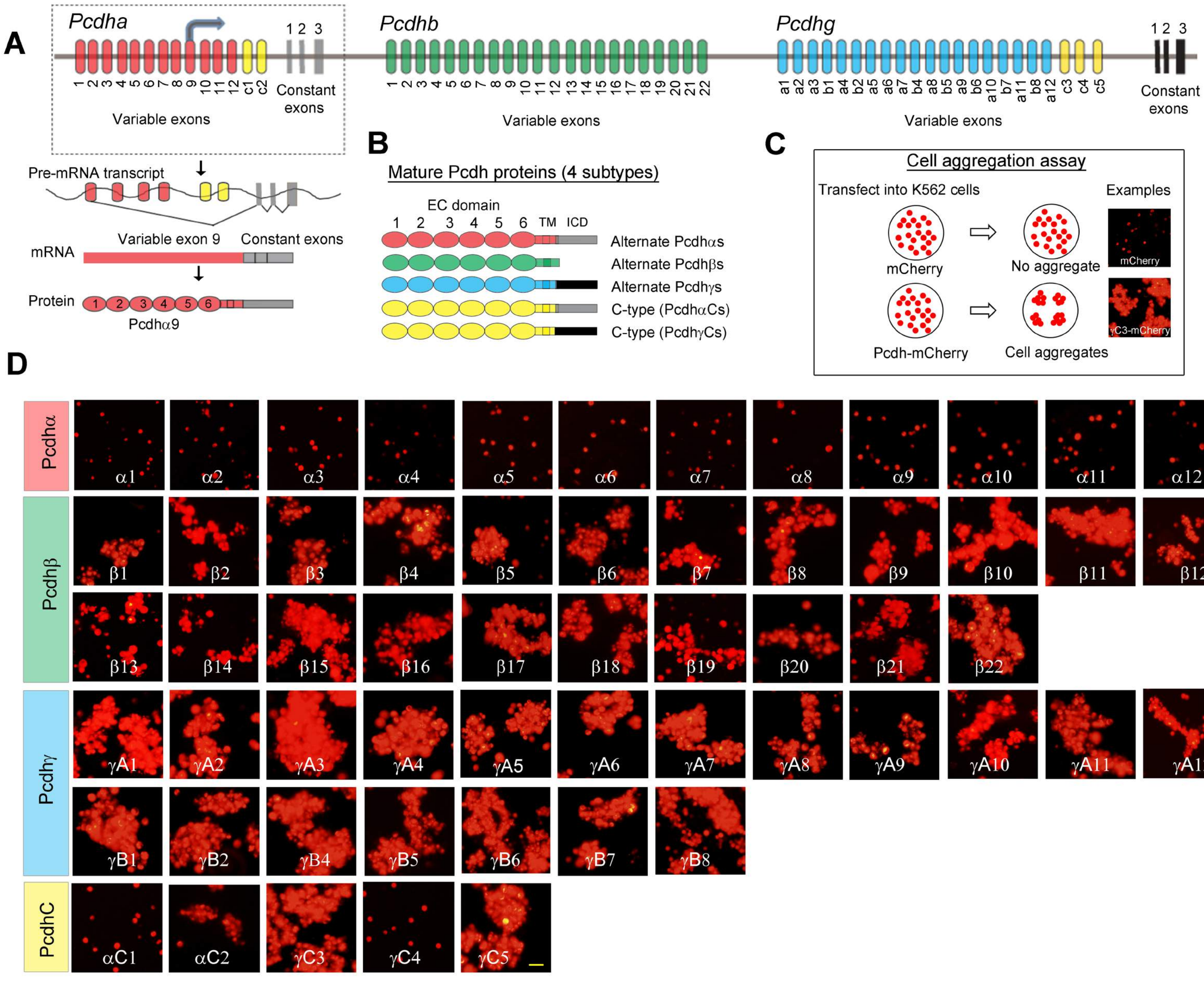

A
Pcdha
1 2 3
1 2 3 4 5 6 7 8 9 10 11 12 c1 c2
Variable exons
Constant exons
Pcdhb
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Variable exons
Pcdhg
a1 a2 a3 b1 a4 b2 a5 a6 a7 b4 a8 b5 a9 b6 a10 b7 a11 b8 a12 c3 c4 c5
Variable exons
1 2 3
Constant exons
Pre-mRNA transcript
Variable exon 9
Constant exons
mRNA
Protein
1 2 3 4 5 6
Pcdhα9
B
Mature Pcdh proteins (4 subtypes)
EC domain
1 2 3 4 5 6 TM ICD
Alternate Pcdhαs
Alternate Pcdhβs
Alternate Pcdhγs
C-type (PcdhαCs)
C-type (PcdhγCs)
C
Cell aggregation assay
Transfect into K562 cells
Examples
mCherry
No aggregate
mCherry
Pcdh-mCherry
Cell aggregates
γC3-mCherry
D
Pcdhα
α1 α2 α3 α4 α5 α6 α7 α8 α9 α10 α11 α12
Pcdhβ
β1 β2 β3 β4 β5 β6 β7 β8 β9 β10 β11 β12
β13 β14 β15 β16 β17 β18 β19 β20 β21 β22
Pcdhγ
γA1 γA2 γA3 γA4 γA5 γA6 γA7 γA8 γA9 γA10 γA11 γA12
γB1 γB2 γB4 γB5 γB6 γB7 γB8
PcdhC
αC1 αC2 γC3 γC4 γC5

Figure 2

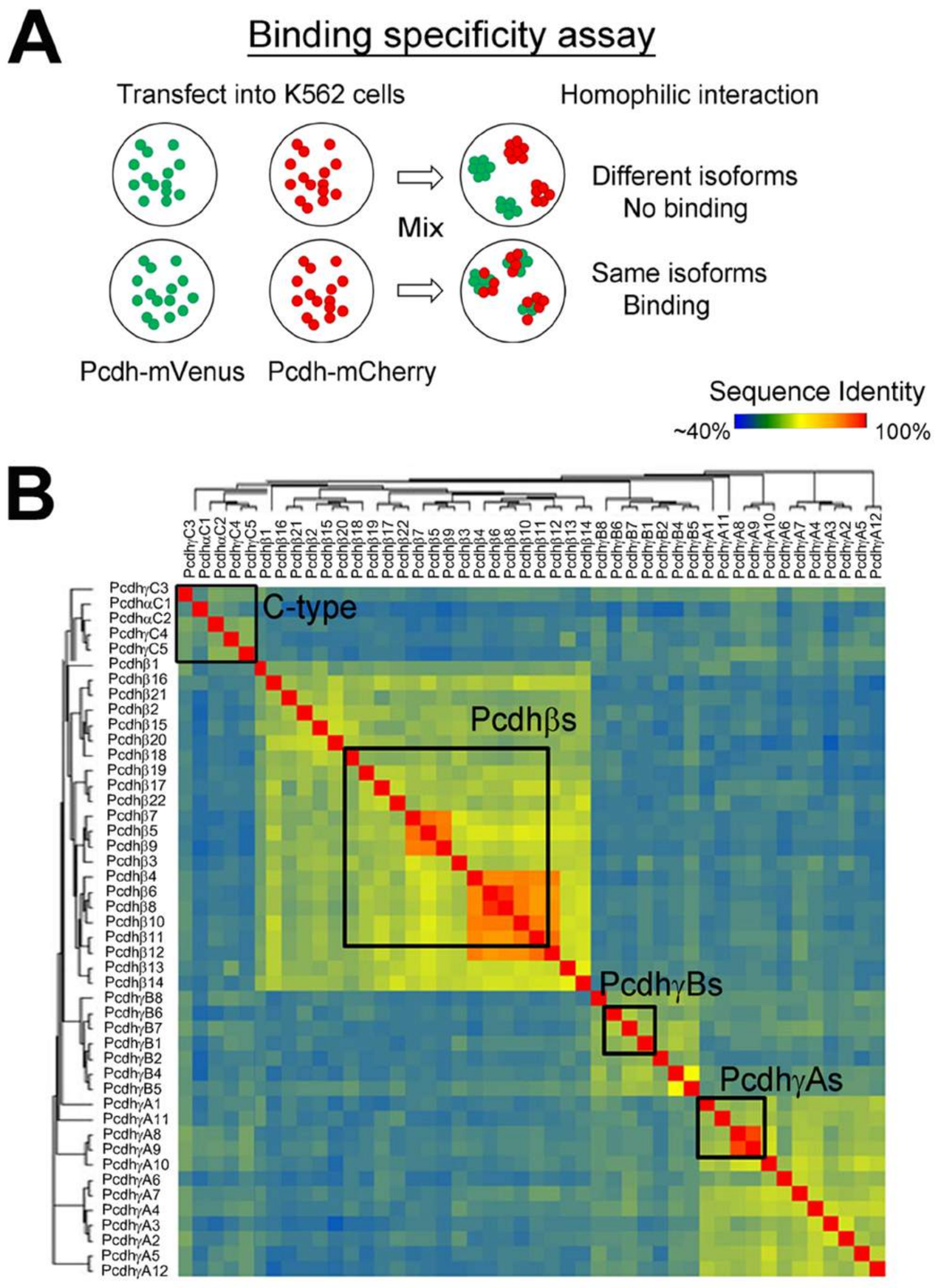

A
Binding specificity assay
Transfect into K562 cells
Homophilic interaction
Mix
Different isoforms
No binding
Same isoforms
Binding
Pcdh-mVenus
Pcdh-mCherry
Sequence Identity
~40%
100%
B
C-type
Pcdhβs
PcdhγBs
PcdhγAs
D
Pcdh-mCherry
γA1
γA8
γA9
γA11
γB1
γB6
γB7
Pcdh-mVenus
C
Pcdh-mCherry
β3
β4
β5
β6
β8
β9
β11
β17
β18
β22
Pcdh-mVenus
E
Pcdh-mCherry
αC2
γC3
γC5
Pcdh-mVenus

Figure 3

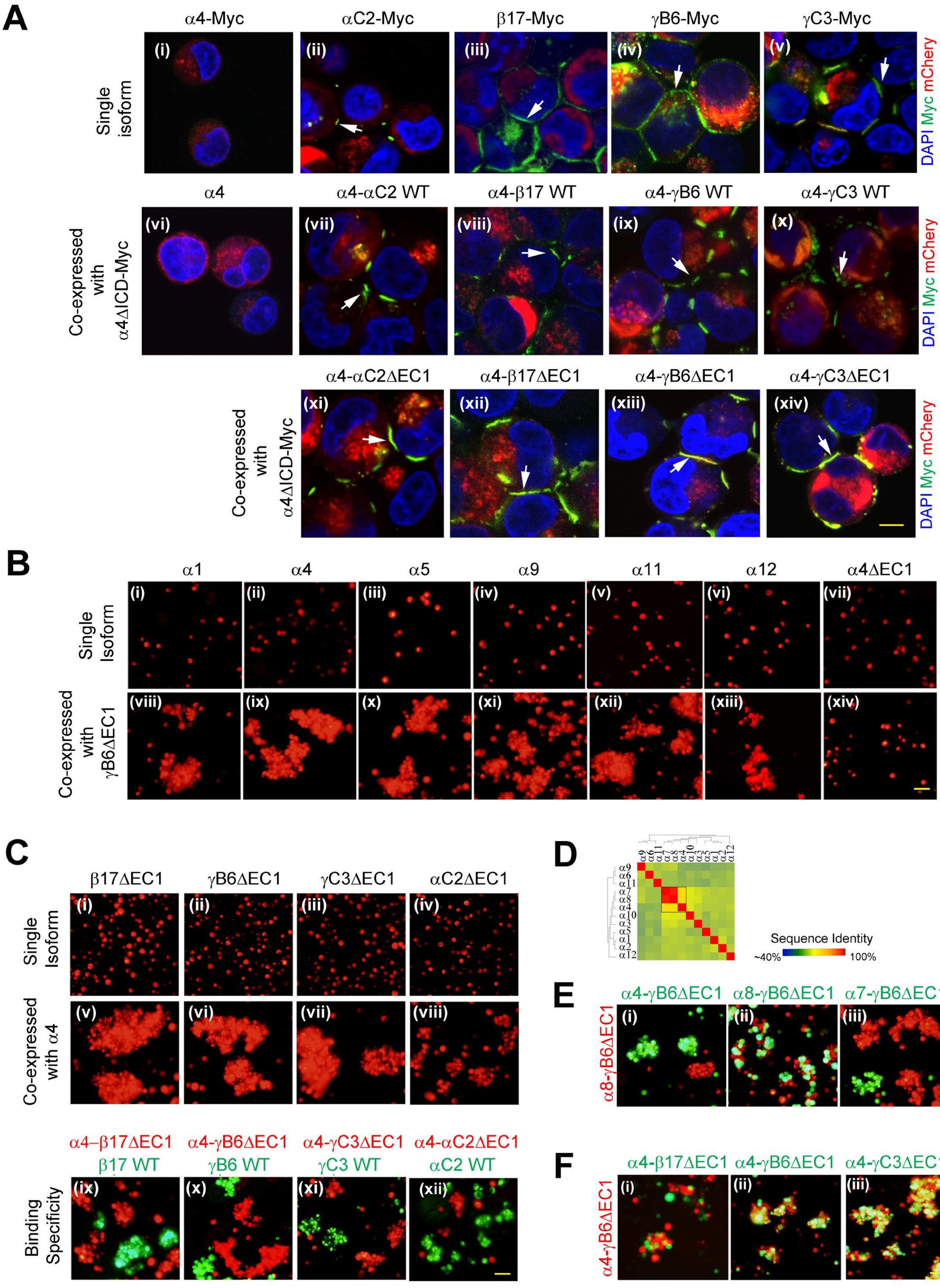

A
α4-Myc
αC2-Myc
β17-Myc
γB6-Myc
γC3-Myc
Single isoform
(i)
(ii)
(iii)
(iv)
(v)
DAPI Myc mChery
α4
α4-αC2 WT
α4-β17 WT
α4-γB6 WT
α4-γC3 WT
Co-expressed with α4ΔICD-Myc
(vi)
(vii)
(viii)
(ix)
(x)
DAPI Myc mChery
α4-αC2ΔEC1
α4-β17ΔEC1
α4-γB6ΔEC1
α4-γC3ΔEC1
Co-expressed with α4ΔICD-Myc
(xi)
(xii)
(xiii)
(xiv)
DAPI Myc mChery
B
α1
α4
α5
α9
α11
α12
α4ΔEC1
Single Isoform
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
Co-expressed with γB6ΔEC1
(viii)
(ix)
(x)
(xi)
(xii)
(xiii)
(xiv)
C
β17ΔEC1
γB6ΔEC1
γC3ΔEC1
αC2ΔEC1
Single Isoform
(i)
(ii)
(iii)
(iv)
Co-expressed with α4
(v)
(vi)
(vii)
(viii)
α4–β17ΔEC1 β17 WT
α4-γB6ΔEC1 γB6 WT
α4-γC3ΔEC1 γC3 WT
α4-αC2ΔEC1 αC2 WT
Binding Specificity
(ix)
(x)
(xi)
(xii)
D
α9
α6
α11
α7
α8
α4
α10
α3
α5
α1
α2
α12
Sequence Identity
~40%
100%
E
α4-γB6ΔEC1
α8-γB6ΔEC1
α7-γB6ΔEC1
α8-γB6ΔEC1
(i)
(ii)
(iii)
F
α4-β17ΔEC1
α4-γB6ΔEC1
α4-γC3ΔEC1
α4-γB6ΔEC1
(i)
(ii)
(iii)

Figure 4

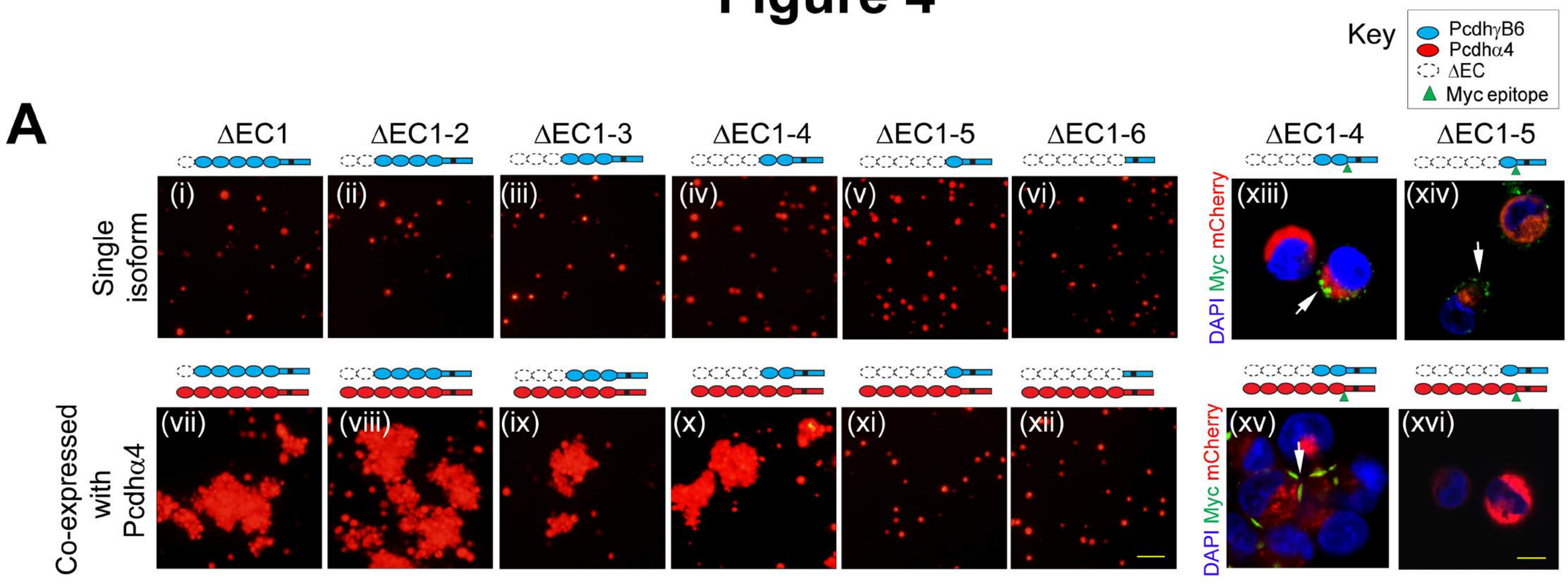

A
Key
PcdhγB6
Pcdhα4
ΔEC
Myc epitope
ΔEC1
ΔEC1-2
ΔEC1-3
ΔEC1-4
ΔEC1-5
ΔEC1-6
ΔEC1-4
ΔEC1-5
Single isoform
Co-expressed with Pcdhα4
DAPI Myc mCherry
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
(ix)
(x)
(xi)
(xii)
(xiii)
(xiv)
(xv)
(xvi)
B
Pcdhα-Pcdhγ chimeras
Pcdhγ-Pcdhα chimeras
Homophilic interaction
EC domain
1 2 3 4 5 6 TM ICD
(i) α4(EC1)-γC3 Yes
(ii) α4(EC1-2)-γC3 Yes
(iii) α4(EC1-3)-γC3 Yes
(iv) α4(EC1-4)-γC3 Yes
(v) α4(EC1-5)-γC3 Yes
(vi) α4-γC3(ICD) No
(vii) α4-γC3(EC6)-α4 Yes
(viii) α4 No
(ix) γC3(EC1)-α4 No
(x) γC3(EC1-2)-α4 No
(xi) γC3(EC1-3)-α4 No
(xii) γC3(EC1-4)-α4 No
(xiii) γC3(EC1-5)-α4 No
(xiv) γC3-α4(ICD) Yes
(xv) γC3-α4(EC6)-γC3 No
(xvi) γC3 Yes
Key
Pcdhα4
PcdhγC3
Homophilic interaction
Pcdh chimeras
α4-γC3
γC3-α4
α4 WT
γC3 WT
C
ΔEC6
ΔICD
ΔEC6(Myc)
ΔICD(Myc)
Pcdhα4
DAPI Myc mCherry
D
α C β γA & γB
EC6 domain sequence identity
~40%
100%
E
A B C D E F G
Pcdhβ1 RPMILYP-LQNGTLPCNDLVPRSAEAGYLVTKVVAVDGDSGQNSWLSYHLL--KATDLGLFSVQRQNGEIRTLRQISERDPMMQKLVILVQDHGQPALSTTASLNILL
Pcdhβ13 APFVLYP-MQNASAPCTELLPRAAEPGYLVTKVVAVDRDSGQNAWLSFQLF--KATEPGLFSVWAHNGEVRTTRLLSERDVPKHRLLLLVKDNGEPPRSASVTLHVLL
PcdhγA6 VPEILYPVLPTDGSTGVELAPRSAEPGYLVTKVVAVDKDSGQNAWLSYRLI--KASEPGLFSVGLHTGEIRTARVLLDRDALKQSLVVAVQDHGQPPLSATVTLTVAV
PcdhγB8 APRVLYPTLEPDGSALFDMVPRAAEPGYLVTKVVAVDADSGHNAWLSYHVL--QASDPGLFSLGLRTGEVRTARALGDRDSARQRLLVAVRDGGQPPLSATATLHLIF
PcdhαC2 APHILYP-TSTNTSAAIEMVPRTAPAGYLVTKVIAMDSDSGQNAWLFYHLV--PTSDSDLFKVELHTGEIRITRKIGDESGTTFNLTVVVRDNGEPPLSATVAITVAV
PcdhγC3 APQVLYP---RPGQSSVEMLPRGTAAGHVVSRVVGWDADAGHNAWLSYSLL--GAPNQSLFAVGLHTGQISTARPIQDTDSPRQILTVLISDSGEPLLSTTATLTVSV
PcdhγC5 APAVLHP-RPGREFSAPQRLPRSAPPGSLVTKVTAVDADAGHNAWLSYSLLP-QSTAPGLFLVSAHTGEVRTARALLEDDSDTQQVVVLVRDNGDPSLSSTATVLLVL
Pcdhα2 APLLLPQGAVGVGGSVSELVPRSVGSGHVVTKVRAVDADSGYNAWLSYELQSTVG-MRSLFRVGLYTGEISTTRALDEVDAPRQRLLVLVKDHGEPSLTATATVLVSL
Pcdhα4 APTLLEPEAGVSGGIVSRLVSRSVGAGHVVAKVRAVDADSGYNAWLSYELQSSEGNSRSLFRVGLYTGEISTTRILDEADSPRQRLLVLVKDHGDPAMIVTATVLVSL
Pcdhα7 APTLLGPLANGVGGTMNEVVSRALVPGQVVAKVRAVDEDSGYNAWLSFELQPVAGGVRSPFRVGLYTGEISTTRALEETDALRQCLLVLVKDHGEPSLTATATVLLSL
Pcdhα11 APTLLPHGAVGAGGAVSELVSRSVGSGHVVTKVRAVDADSGYNAWLSYEVQLGTGGVRSPFRVGLYTGEISTTRALDEVDSPRQTLLVLVKDHGEPALTATATVLVSL
PcdhγC4 APAVLRP-RARPGSLCPQALPPSVGAGHLVTKVTAVDLDSGYNAWVSYQLL--EAPDPSLFAVSRYAGEVRTA-VPIPADLPPQKLVIVVKDSGSPPLSTSVTLLVSL

Figure 5

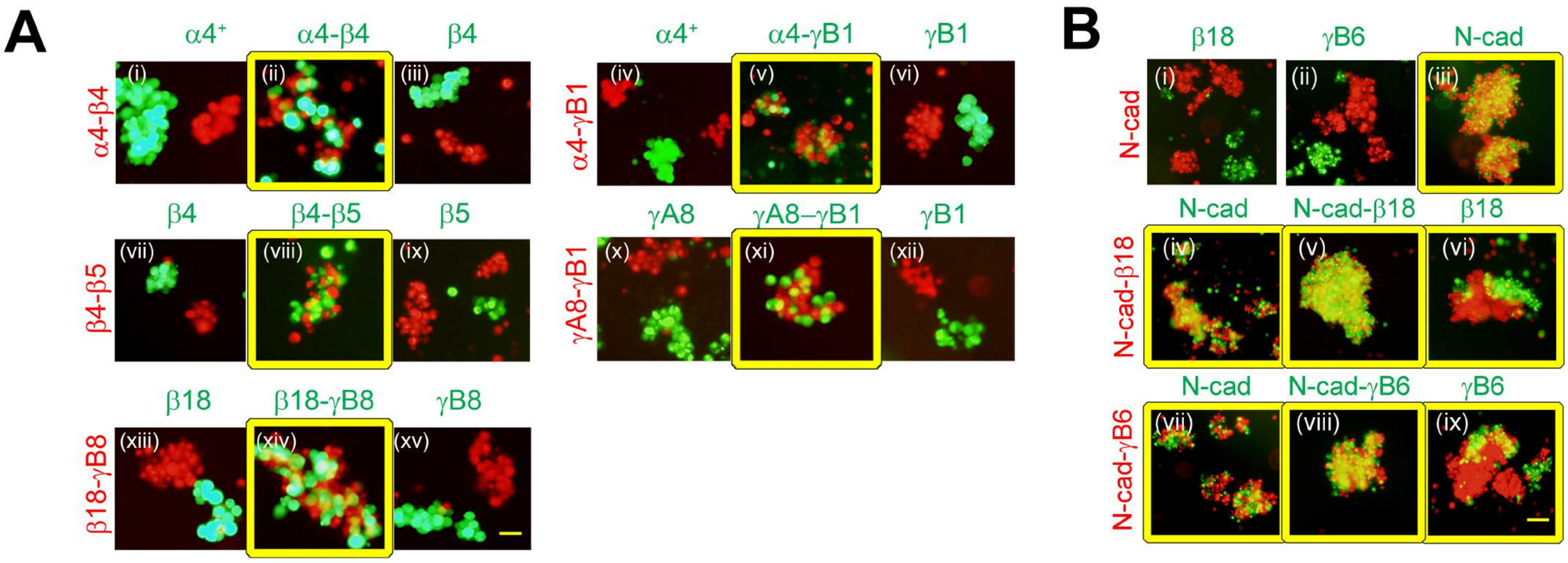

A
α4+
α4-β4
β4
α4-β4
β4
β4-β5
β5
β4-β5
β18
β18-γB8
γB8
β18-γB8
α4+
α4-γB1
γB1
α4-γB1
γA8
γA8–γB1
γB1
γA8-γB1
B
β18
γB6
N-cad
N-cad
N-cad-β18
β18
N-cad-β18
N-cad
N-cad-γB6
γB6
N-cad-γB6
C
Pcdh- αβ Pairs
Pcdh-mChery
α4–β6
α4–β4
α9–β6
α9–β4
Pcdh-mVenus
D
Pcdh- αγ Pairs
Pcdh-mCherry
α4–γA8
α4–γB1
α9–γA8
α9–γB1
Pcdh-mVenus
E
Pcdh- βγ Pairs
Pcdh-mCherry
β18–γB6
β18–γB8
β17–γB6
β17–γB8
Pcdh-mVenus
F
Pcdh- ββ Pairs
Pcdh-mCherry
β5–β6
β5–β12
β6–β12
Pcdh-mVenus
G
Pcdh- γγ Pairs
Pcdh-mCherry
γB1–γB6
γB6–γA8
γA8–γB1
Pcdh-mVenus
H
(a)
(b)
(c)
α
γ
γ'
*
Interaction
No
Yes
No
(d)
(e)
(f)
(g)
N-cad
Cell membrane
Cell membrane
Yes
Yes
Yes
Yes
Homophilic binding through Pcdh-Pcdh pairs
Homophilic binding through N-cad/Pcdhs

Figure 6

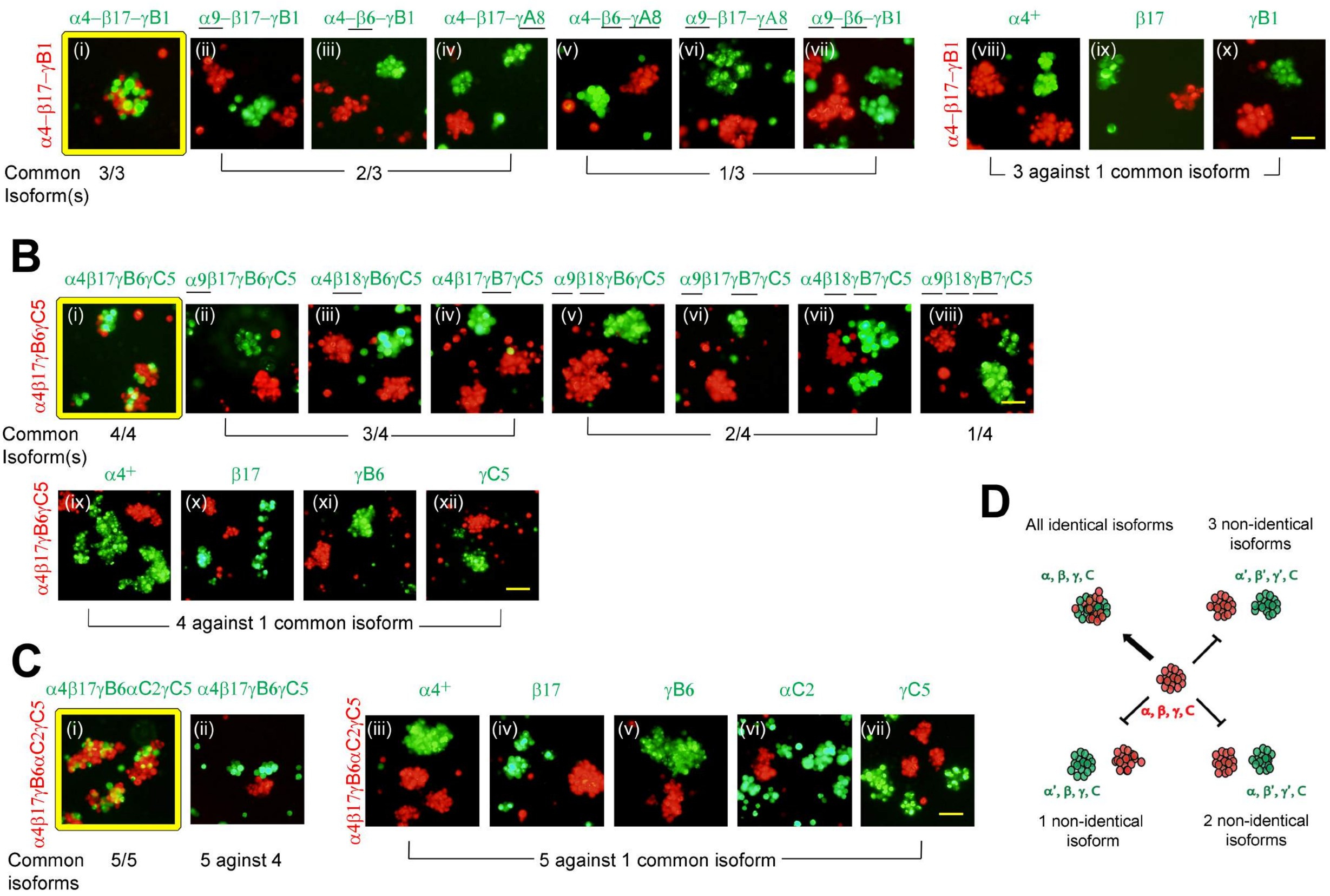
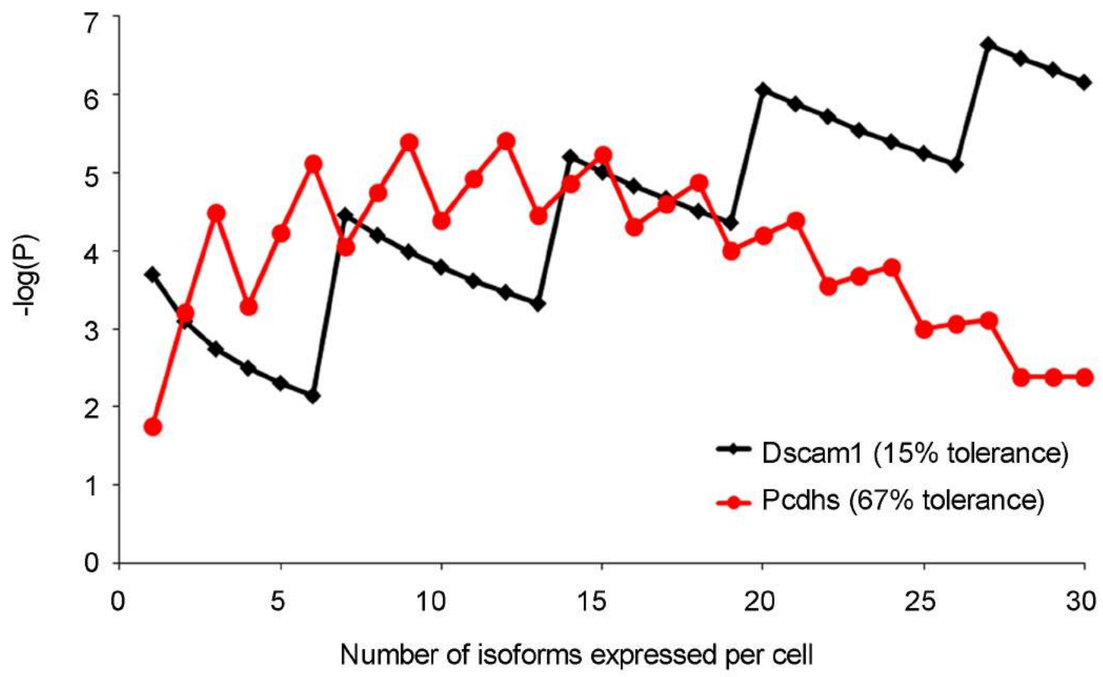
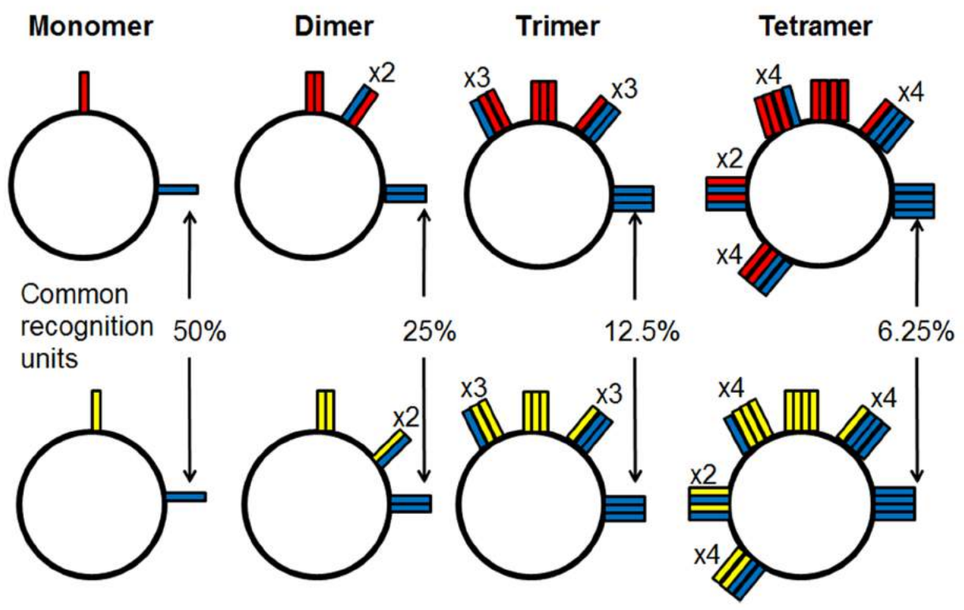
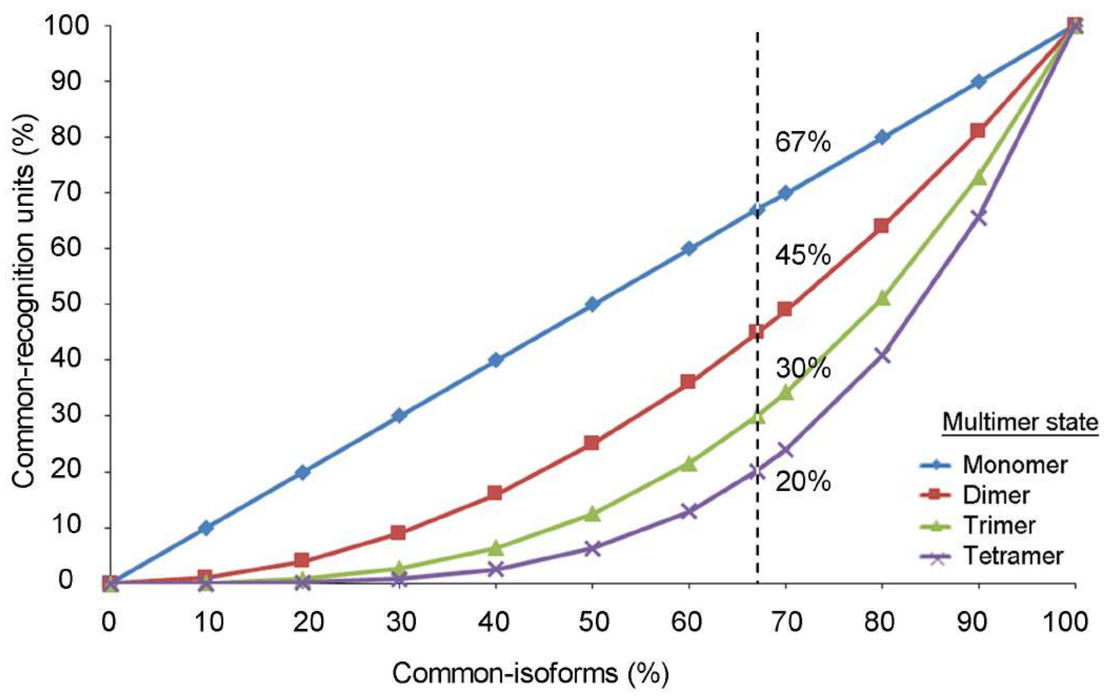
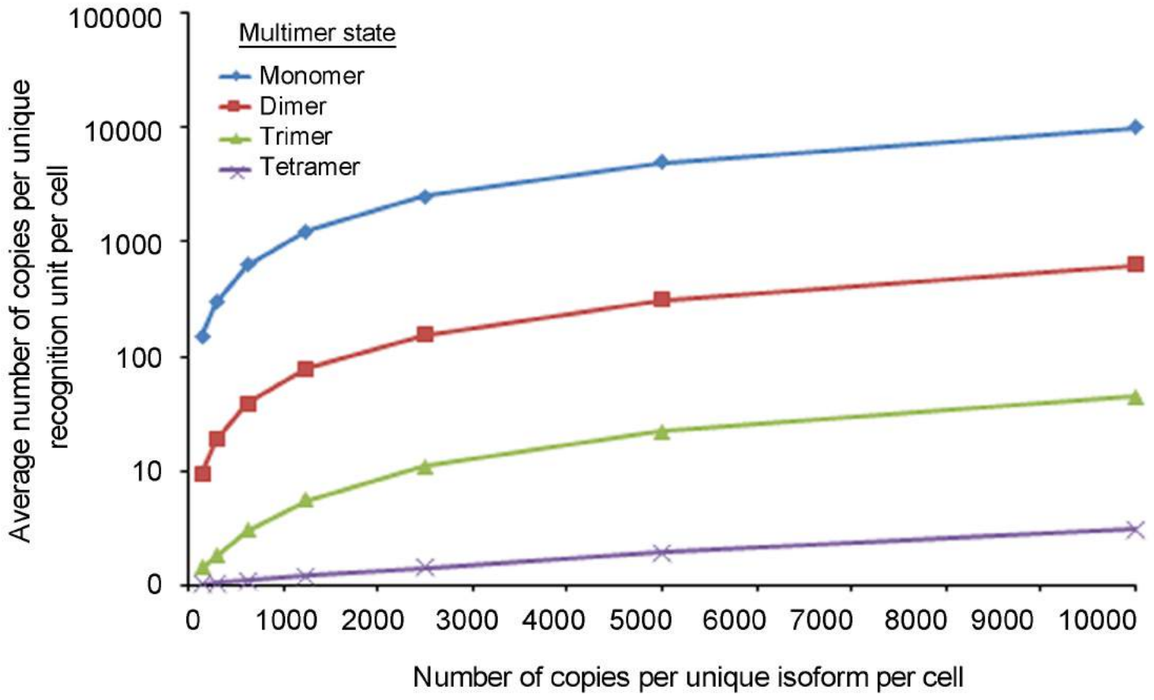

A
α4–β17–γB1
α9–β17–γB1
α4–β6–γB1
α4–β17–γA8
α4–β6–γA8
α9–β17–γA8
α9–β6–γB1
α4+
β17
γB1
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
(ix)
(x)
Common Isoform(s) 3/3
2/3
1/3
3 against 1 common isoform
B
α4β17γB6γC5
α9β17γB6γC5
α4β18γB6γC5
α4β17γB7γC5
α9β18γB6γC5
α9β17γB7γC5
α4β18γB7γC5
α9β18γB7γC5
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
(viii)
Common Isoform(s) 4/4
3/4
2/4
1/4
α4+
β17
γB6
γC5
(ix)
(x)
(xi)
(xii)
4 against 1 common isoform
C
α4β17γB6αC2γC5
α4β17γB6γC5
αC2
(i)
(ii)
(iii)
(iv)
(v)
(vi)
(vii)
Common isoforms 5/5
5 aginst 4
5 against 1 common isoform
D
All identical isoforms
3 non-identical isoforms
α, β, γ, C
α', β', γ', C
α', β, γ, C
α, β', γ', C
1 non-identical isoform
2 non-identical isoforms

Figure 7


A
-log(P)
Number of isoforms expressed per cell
Dscam1 (15% tolerance)
Pcdhs (67% tolerance)


B
Monomer
Dimer
Trimer
Tetramer
Common recognition units
50%
25%
12.5%
6.25%


C
Common-recognition units (%)
Common-isoforms (%)
67%
45%
30%
20%
Multimer state
Monomer
Dimer
Trimer
Tetramer


D
Average number of copies per unique recognition unit per cell
Number of copies per unique isoform per cell
Multimer state
Monomer
Dimer
Trimer
Tetramer

Supplemental information

Legends for supplementary Figures

Figure S1. Protein sequence alignments of the clustered Pcdhs (Related to Figure 1)

(A) Amino acid sequence alignment of representative clustered Pcdh isoforms. Sequences were aligned with Muscle (Edgar, 2004) and divided into individual EC domains based on the DXNDN-XPXF motif (Posy et al., 2008). The locations of β-strands are indicated above the alignment, with the conserved residues that form the hydrophobic core of the EC domain highlighted in blue. Red boxes indicate the canonical Ca^{2+} binding motifs and are labeled according to which inter-domain Ca^{2+} binding site they belong to. The N-terminus based on signal peptide prediction is shown with a grey arrow. The XPXF motifs at the start of the ECs are highlighted in green. The end of EC6 domain is predicted using a cadherin database created in the lab of Dr.Barry Honig (http://cadherindb.c2b2.columbia.edu/). The "DRE" motif of EC3 domain of PcdhαC1 is replaced with "GPP", which is highlighted in red box.

(B) Quantification of the sizes of cell aggregates – A graph showing the size distribution of cell aggregates generated by cells expressing a single Pcdh isoform. All 12 alternate Pcdhαs fail to mediate cell aggregation. Among 22 members of Pcdhβs, Pcdhβ7 and Pcdhβ19 expressing cells generate very small aggregates. All 19 members of PcdhγAs or PcdhγBs expressing cells generate relatively large aggregaes. Among 5 C type isoforms, PcdhαC1 and PcdhγC4 expressing cells fail to mediate cell aggregation, whereas PcdhaC2, PcdhgC3 and PcdhgC5 do. The cell aggregates were classified according to the number of cells they contain - small size (<10 cells), medium size (10-30 cells), and large size (>30 cells). The representative

bar graph shown here quantifies the types of cell aggregates from one representative experiment. The cell aggregation experiments were performed at least three times for each of the Pcdh isoforms.

Figure S2. Sequence identity of EC domains and Pcdh binding specificities. (Related to Figure 2)

(A) Average pair-wise sequence identity of individual EC domains of Pcdh subtypes. The left panel shows the average pairwise sequence identity of individual EC domains of all four Pcdh subtypes. Among the six EC domains, EC2 and EC3 are the most divergent, and EC6 is highly conserved (except in the case of the C-type).

(B) Heat map of the pair-wise protein sequence identity of the EC2 and EC3 domains. A subset of Pcdh -α, -β or -γ isoforms (shaded in red) shares a high level of sequence identity in their EC2 and EC3 domains.

(C) PcdhγA8 and PcdhγA9 do not engage in heterophilic interactions. Low magnification views for differentially-tagged PcdhγA8 and PcdhγA9 tested for binding specificity are shown. Cells expressing PcdhγA8 fail to co-aggregate with cells expressing PcdhγA9 (top panel). Cells expressing PcdhγA9 fused to mCherry co-aggregate with cells expressing PcdhγA9 fused to mVenus (lower panel). These results demonstrate the high level of homophilic specificity of interactions of closely related Pcdh isoforms. Scale bar, 50 μm.

(D) Binding specificity of cells expressing different Pcdh subtypes. Representative members of each Pcdh subtype - Pcdhβ17, -γA8, -γB6, and C-type γC3 were tested for binding specificity. Co-aggregation was only observed for cells expressing the same Pcdh isoform (i and xiii), but not for cells expressing different Pcdh isoforms (ii-vii). Scale bar, 50 μm.

Figure S3. Alternate Pcdhαs and a subset of C-type isoforms engage in specific homophilic interactions when delivered to the cell surface by a carrier Pcdh. (Related to Figure 3)

(A) The EC1 domain is required for homophilic interaction, but not for membrane delivery. PcdhγB6 constructs bearing an extracellular c-Myc tag inserted after EC6 domain were used. c-Myc-tagged PcdhγB6ΔEC1 failed to mediate cell aggregation (i) but was delivered to the cell surface (ii). White arrows indicate c-Myc staining at cell-cell contacts. Scale bar, 50 μm (i), 10 μm (ii).

(B) *Cis*-interaction between Pcdhα4 and wild-type or ΔEC1 carrier Pcdhs. Wild-type or ΔEC1 carrier Pcdh-αC2, -β17, -γB6 or -γC3 (fused to mCherry) are co-expressed with Pcdhα4-TAP (one HA epitope tag, followed by two TEV protease cleavage sites and two FLAG epitope tags) in HEK293 cells. The cells were lysed, and Pcdhα4 was immunoprecipitated using anti-FLAG beads. Western blotting with anti-mCherry antibodies shows the interaction between Pcdhα4 and wild type or ΔEC1 carrier Pcdh isoforms. Western blotting with anti-Flag antibodies shows that Pcdhα4 is present in all co-IP experiments. IP,immunoprecipitation; IB, immunoblotting.

(C) Quantification of sizes of cell aggregates. Data presented in graph show the size distribution of the number of cell aggregates generated by singly transfected cells expressing carrier Pcdhs (ΔEC1-Pcdh-αC2, -β17, -γB6 or -γC3), or cells co-expressing ΔEC1-Pcdhs with Pcdhα4 from Figure 3C (i-viii). According to the number of cells they contain, cell aggregates are classified into small size (<10 cells), medium size (10-30 cells), and large size (>30 cells). Scale bar, 50 μm.

(D) Alternate Pcdhα isoforms mediate homophilic interactions when co-transfected with different carrier Pcdhs. Cells transfected with single Pcdhα isoforms are shown

(first row). Aggregation is observed for cells co-transfected with Pcdhβ17ΔEC1 (second row), PcdhγC3ΔEC1 (third row), or PcdhαC2ΔEC1 (fourth row) and an alternate Pcdhα isoform. Cells expressing PcdhαC2ΔEC1 and an alternate Pcdhα isoform generate smaller aggregates.

(E) Quantification of cell aggregate size. Data presented in graph show the size distribution of the quantified number of cell aggregates generated as shown in the third row of Figure S3D. According to the number of cells they contain, cell aggregates are classified into small size (<10 cells), medium size (10-30 cells), and large size (>30 cells).

(F) PcdhγC4 mediates homophilic interactions when co-transfected with carrier Pcdhs, whereas PcdhαC1 does not. Cells transfected with single ΔEC1-carrier Pcdhs failed to mediate cell aggregation (first row). PcdhγC4 and PcdhαC1 failed to mediate homophilic interactions when transfected alone (v and x). Cells co-expressing PcdhγC4 and ΔEC1-carrier Pcdhs mediate cell aggregation (second row). The sizes of aggregates are relatively small and it may be due to the low surface expression of PcdhγC4 isoforms and/or the proteins display a weak trans-binding affinity. Cells co-expressing PcdhαC1 and ΔEC1-carrier Pcdhs failed to mediate cell aggregation (third row). Scale bar, 50 μm.

Figure S4. The universal role of EC6 domains in cell surface delivery for all Pcdh subtypes. (Related to Figure 4)

(A) EC5-EC6 domains of Pcdhβ are minimum region required for Pcdhα cell surface delivery. Schematic representations of Pcdh constructs are shown above the images. Immunostaining of extracellular c-Myc-tagged Pcdhs was used to visualize cell surface localization. While Pcdhα4 does not localize to cell surface (iii) and fails to engage in

homophilic interaction (i), co-expression of Pcdhβ17(ΔEC1-4) facilitates the membrane delivery of Pcdhα4 (iv) and subsequent cell aggregation (ii). White arrows indicate c-Myc staining at cell-cell contacts.

(B) EC6 domains of Pcdhα or PcdhγC3 isoforms inhibit or promote membrane delivery respectively. Schematic representations of Pcdh constructs are shown above the images. Cell surface localization of c-Myc-tagged Pcdhs was used to visualize cell surface localization. Chimeras of Pcdhα4 bearing the EC6 domain from PcdhγC3 localize to cell membrane (iv) and mediate robust cell aggregation (ii). Chimeras of PcdhγC3 bearing the EC6 domain of Pcdhα4 fail to localize to the cell membrane (iii) and fail to mediate cell aggregation (i). White arrows indicate the c-Myc staining at cell-cell contacts.

(C) ΔEC1-carrier Pcdhs promote the aggregation activity of Pcdhγ chimeras bearing the EC6 domain from Pcdhαs. Schematic representations of Pcdh constructs are shown above the images. Chimeras of Pcdhγs bearing the EC6 domain of Pcdhα4 fail to mediate cell aggregation (upper panels, i-vi) but they mediate robust cell aggregation with co-transfected carrier PcdhγB6ΔEC1 (lower panels, vii-xii).

(D) The EC6 domains of Pcdhβ17, PcdhγB6, or PcdhγC3 can rescue the inability of Pcdhα isoforms to localize to the membrane. Schematic representations of Pcdh constructs are shown above the images. Chimeras of Pcdhα isoforms bearing the EC6 domains of Pcdhβ17, PcdhγB6, or PcdhγC3 mediate robust cell aggregation (upper panels, i-iii). Chimeras of Pcdh isoforms bearing the EC6 domain of Pcdh-α4 or -α8 fail to mediate cell aggregation (lower panels, iv-vi).

(E) Domain swapping experiments with C-type Pcdh isoforms. Schematic representations of chimeras are shown above the cell aggregation results. Chimeric

PcdhαC1(EC1-3)-γC3 mediates cell aggregation (i). However, chimeric PcdhαC1(EC1-5)-γC3 fail to mediate cell aggregation (ii). Similarly, cell aggregation was observed for chimeric PcdhγC4(EC1-3)- γC3 and PcdhγC4(EC1-5)- γC3 (iii-iv) but not for the PcdhγC4 chimera bearing the ICD domain of PcdhγC3 (v). Conversely, all the chimeras of PcdhγC3 proteins bearing the EC6 domain of PcdhαC1 or PcdhγC4 fail to mediate cell aggregation (vi-ix). PcdhγC3 chimera bearing the ICD domain of PcdhγC4 mediates homophilic interaction (x).

(F) The EC6 domain of PcdhγC4 inhibits membrane delivery, whereas the EC6 domain of PcdhγC3 promotes cell surface localization. Schematic representations of Pcdh constructs are shown above the images. Chimeras of PcdhγC4 bearing the EC6 domain from PcdhγC3 localize to the plasma membrane at a level comparable to that of wild-type PcdhγC3 (i and iv). Chimeras of PcdhγC3 bearing the EC6 domain of PcdhγC4 failed to localize to the cell surface, as observed for wild-type PcdhγC4 (ii and iii). White arrows indicate c-Myc staining at cell-cell contacts.

(G-I) Quantification of cell aggregate size. Data presented in graph show the size distribution of the quantified number of cell aggregates generated by experiments shown in Figure 4A (G), Figure 4B (H), and Figure 4C (I). According to the number of cells they contain, cell aggregates are classified into small size (<10 cells), medium size (10-30 cells), and large size (>30 cells).

Figure S5. Combinatorial binding specificity displayed by co-expression of two distinct Pcdh isoforms. (Related to Figures 5 and 6)

(A) Quantification of co-aggregation of two differentially labeled cell populations from Figure 5A. The percentages of red and green co-aggregates are quantified and illustrated as bar graph. (See supplemental spreadsheet for raw data)

(B) Quantification of co-aggregation of two differentially labeled cell populations. The percentages of red and green co-aggregates from Figures 5C-5G are quantified and illustrated as a heat map. (See also supplemental table 1 for raw data)

(C-D) Low magnification views for cell aggregates of differentially-tagged Pcdh isoforms (C) and N-cad-Pcdh isoform (D) were shown. Cells expressing Pcdh/Pcdh pairs only interact homophilically with the identical pairs (Third row). However, in all cases, cells expressing N-cad/Pcdh pairs interact with cells expressing any combination of N-cad or Pcdh isoform. Scale bar, 50 μm.

(E) K562 cells were used to co-express N-cad and Pcdh isoforms or two different Pcdh isoforms. The cells were lysed, and mVenus-tagged Pcdh proteins were immunoprecipitated using anti-GFP beads. Western blotting was carried out using anti-mCherry and anti-GFP antibodies, which shows that Pcdhβ17 was co-immunoprecipitated by PcdhγB1, whereas N-cad was not co-immunoprecipitated by PcdhγC3. IP,immunoprecipitation; IB, immunoblotting.

(F) Cells expressing differentially tagged N-cad and Pcdh isoforms were mixed with cells expressing combinations of N-cad with the same or different Pcdh isoform (i-iii). Even in the presence of the mismatched Pcdh as shown in the case of (ii), red and green cells co-aggregate. However, singly transfected cells expressing different Pcdh isoform does not interact with each other (iv).

(G) Schematic diagram represents the outcome of co-aggregation of the two cell populations expressing mCherry-tagged proteins and mVenus-tagged proteins. Three types of cell aggregate behaviors are presented and the results of interaction between

red and green cells from the cell aggregation experiments (Figures 5A, 5B, 5E, 5H and 5I) were summarized below.

Figure S6. Combinatorial binding specificity displayed by co-expression of up to five distinct Pcdh isoforms. (Related to Figure 6)

(A-C) Quantification of co-aggregation of two differentially labeled cell populations. The percentages of red and green co-aggregates from Figures 6A-6C are quantified. (See supplemental spreadsheet for raw data)

(D) *Cis*-interaction between Pcdh α, β and γ isoforms. Pcdhα4-TAP, Pcdhβ17-mCherry and PcdhγB1 mVenus were co-expressed in K562 cells. The cells were lysed, and Pcdhα4 was immunoprecipitated using anti-FLAG beads and PcdhγB1 was immunoprecipitated using anti-GFP beads. Western blotting with anti-mCherry, anti-FLAG and anti-GFP antibodies shows the interaction between all three isoforms. Western blotting with anti-FLAG antibodies shows that Pcdhα4 is present in all co-IP experiments. TAP tag contains one HA epitope tag, followed by two TEV protease cleavage sites and two FLAG epitope tags. IP,immunoprecipitation; IB, immunoblotting.

(E) Cells co-expressing mVenus-tagged N-cad and two different Pcdh isoforms were mixed with cells expressing identical compositions of all three proteins, single N-cad/Pcdh or combinations of two Pcdhs. Three different behaviors of cell aggregates were observed between two cell populations. Transfected with 1μg plasmid input of N-cad and 3μg plasmid input of Pcdhs (Upper panels). Transfected with 1μg plasmid input of N-cad, 3μg plasmid input of Pcdhs (Lower panels).

(F) Higher magnification images from experiment transfected with 0.25μg plasmid input of N-cad and 3μg plasmid input of Pcdhs. Three different behaviors of cell aggregates were observed between red and green cells.

(G) Schematic diagram showing the outcome of co-aggregation of the two cell populations expressing mCherry-tagged proteins and mVenus-tagged proteins. Three types of cell aggregate behaviors are presented and the results of interaction between red and green cells were summarized below. N-cad(H)-Pcdhβ-Pcdhγ where "H" indicates the cells transfected with high amount of plasmid encoding N-cad (S6E Upper panels). N-cad (L)-Pcdhβ-Pcdhγ indicates the cells transfected with lower amount of plasmid encoding N-cad (S6E Lower panels).

Figure S7. **Experimental procedure for testing combinatorial specificity (Related to Figure 6 and methods)**

(A) Schematic diagram of testing combinatorial specificity of cells co-expressing multiple isoforms. To ensure the equal functional contribution of each Pcdh isoform in co-transfection experiments, three steps of binding assay were performed. First, the expression level of an individual isoform in one population of cells was tested by mixing with cells expressing individual common isoform. Only mCherry positive cells do not mix with cells expressing a single isoform but mix substantially with the identical sets of isoforms. These mCherry positive cells are then used for the second step of binding assay, in which mCherry positive cells were mixed with cells expressing distinct sets of Pcdh isoforms including one, two or three non-matching mVenus tagged isoforms. These same population of mVenus positive cells were then mixed with cells expressing mCherry tagged Pcdh isoform(s).

(B) Illustration of the control cell aggregation experiments for Figure 6B. mVenus positive cells expressing non-matching Pcdh isoforms used in Figure 6B were assayed for their interaction with mCherry positive cells expressing the common sets of Pcdh isoforms.

Extended Experimental Procedures

Plasmids. DNA fragments encoding full length Pcdh isoforms or variable exons were amplified by PCR using genomic DNA or brain cDNA isolated from C57BL/6J mice (The Jackson Laboratory). PCR products were inserted into the entry pDONR™/Zeo vector by using BP clonase enzyme as described (Invitrogen). To construct pmax-mCherry gateway expression vector, the Ubc promoter from pUbc-RfC1-mCherry (a gift from Dr. Joshua Sane, Harvard University) was replaced with immediate early promoter of cytomegalovirus with intron element (PCMV-IE) from pmaxCloning™ vector (Lonza) by using Xba1 and Sac1 restriction sites. A 798 base pair long promoter sequence (PCMV-IE) was amplified by PCR using forward primer containing an XbaI site (5'- CCCTCTAGATCAATATTGGCCATTAGCCA-3') and reverse primer containing SacI site (5'-TTTGAGCTCCCCTGTGGAGAGAAAGGCAA-3'). To construct pmax-mVenus gateway expression vector, a DNA fragment encoding the mCherry gene is replaced with the DNA fragment encoding the mVenus gene by using Ale1 restriction sites in the pmax-mCherry gateway expression vector. The template of DNA encoding mVenus gene was amplified by PCR using pEYFP-C1 (Addgene Plasmid 27794) (Koushik et al., 2006). To generate extracellular c-Myc-tagged Pcdhs, the DNA sequence encoding peptide of c-Myc (EQKLISEEDL) was inserted after EC6 domains of Pcdhs by using overlapping PCR. The domain deletion and substitution between different Pcdh isoforms were also made by using overlapping PCR. The PCR products were then sub-cloned into gateway entry vectors and expression vectors. EC domains were predicted by conserved protein sequence encoding DXNDN/XPXF motif (Posy et al., 2008) and cadherin database created by Barry Honig lab (Examples are shown in Figure S1A) (http://cadherindb.c2b2.columbia.edu/). Transmembrane domaisn (TM) were predicted by using TMHMM2.0

(http://www.cbs.dtu.dk/services/TMHMM-2.0/). Signal peptides (SP) were predicted by Signal P prediction tools from Signal IP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/). Primer sequences used for PCR amplifications will be provided upon request.

Antibodies. The following commercial antibodies were used: anti-FLAG (Sigma), anti-mCherry (Clontech), anti-GFP (MBL) and anti-c-Myc-FITC (Miltenyi Biotec).

Binding assay to test combinatorial specificity of multiple Pcdh isoforms or N-cad and Pcdh isoforms. The binding assay was performed as described in experimental procedure section. We noted that surface expression of different subtype of Pcdh isoforms varies when they are overexpressed in K562 cells. To overcome the variability in the level of cell surface expression of individual Pcdh isoforms among individual cells, we selected Pcdh isoforms that display similar sized cell aggregates for testing the combinatorial specificity (Refer to Figures 1D and S1B). Among four subtypes of Pcdh isoforms, surface expression of alternate Pcdhαs requires the co-expression of other isoforms (Pcdhβ/γ/C-type). We observed that Pcdhα's surface expression is lower than those of Pcdhβ or Pcdhγ isoforms in the co-expression experiments (Staining data in Figure 3A and FACS experiment was used to quantify the surface expression of different Pcdh isoforms by using the fluorescent labeling of extracellular epitope of the proteins (Data not shown)) Thus, in the co-expression experiments, plasmids encoding Pcdhα:Pcdhβ:Pcdhγ isoforms were used in 2:1:1 ratio to normalize the equal surface delivery of Pcdhα isoforms to other Pcdh subtypes. In these multiple isoforms transfected cells, equal contribution of individual Pcdh isoform in combinatorial homophilic interaction is ensured by performing the stepwise

cell co-aggregation assays as illustrated in Figure S7A. If cells co-expressing multiple isoforms interact preferentially with cells expressing a single common isoform, it indicates dominant expression of a particular Pcdh isoforms and those co-transfected cells were discarded from binding analysis.

Cells expressing N-cad induce very robust cell aggregations and a dominant binding effect of N-cad was observed in binding assay shown Figure S6D and S6E. Surface expression of N-cad is also higher compared to those of Pcdh isoforms when over-express in K562 cells (Data not shown). To achieve a similar surface expression level of N-cad and Pcdhs, amount of plasmids encoding N-cad was titrated in the co-transfection experiments. For double transfections as shown in Figure 5B, plasmids encoding N-cad and Pcdh were used in 1:2 or 1:3 ratio. For triple isoforms transection, plasmids encoding N-cad and two Pcdhs were used in 1:3:3 or 1:12:12 ratio.

Immunoprecipitation and western blotting. HEK293T cells were transfected by using Lipofectamine 2000 Reagent as described (Invitrogen). Transfected cells were homogenized in IP lysis buffer (Pierce) composed of 25 mM Tris-HCl (pH7.4), 150 mM NaCl, 1 mM EDTA, 1% NP-40, and 5% glycerol, supplemented with the Complete Protease Inhibitor Cocktail and the PhosSTOP Phosphatase Inhibitor Cocktail (Roche). The cells were incubated in lysis buffer for 30 minutes at 4˚C. The supernatant was collected by centrifugation at 10,000 rpm for 20 minutes. The proteins of interest were immunoprecipitated by incubating the supernatant with anti-FLAG M2 Affinity Gel (Sigma) or anti-GFP magnetic beads (MBL) for 2 hours using end-to-end rotator at 4˚C. SDS-PAGE and Western blotting with respective antibodies were performed according to standard procedures.

Quantification of the size of cell aggregates using Matlab. The images of cells from two independent aggregation experiments were used to quantify the relative size of cell aggregates generated by each Pcdh isoform. The images were first converted to black and white formats with 1530x2040 pixels. Objects with 3500 or fewer pixels were categorized as small (< 10 cells), objects between 3500 pixels and 12000 pixels were categorized as medium (10 to 30 cells), and objects larger than 12000 pixels were categorized as large (> 30 cells). The number of aggregates of each size category was then counted for analysis.

Statistical Analysis performed for Figure 7

The probability that two neurons incorrectly recognize each other as self (Figure 7A).

The probability that a pair of neurons has more than *t* isoforms in common (ie. the number of shared isoforms that can be tolerated), given different number of isoforms (*m*) stochastically expressed from total isoform pool available (i) is calculated using the following formula:

$$P_{i,m,t} = 1 - \sum_{k=0}^{t} \binom{m}{k} \frac{\binom{i-m}{m-k}}{\binom{i}{m}} \text{ (1).}$$

For Figure 7A we used 58 as the size of the isoform pool (i), and the tolerance (*t*) was assumed to be 67% based on data presented in Figure 6A. Then for a given number of isoforms expressed per cell (*m*), *t* is calculated as the integer part of 0.67 × m. Using these values in formula (1) the probability of two neurons incorrectly recognizing each other is calculated as follows: $P_m = 1 - \sum_{k=0}^{\left\lfloor \frac{2m}{3} \right\rfloor} \binom{m}{k} \frac{\binom{58-m}{m-k}}{\binom{58}{m}}$ (2).

Using a similar statistical framework designed to estimate DSCAM1 diversity, we show here that the probabilities estimated for DSCAM1 are comparable to the probabilities estimated for Pcdhs. A similar statement was made previously (Yagi,

2012), however the probabilities they were referring to are different than the probabilities we refer to here. In the previous study, the author did not calculate the probabilities of incorrect recognition as self between two contacting cells, but rather calculated the fractions of common tetramers between two such cells for specific cases of common isoforms. For example in the case of two cells each expressing 15 isoforms 5 of which are in common 1.3% common tetramer will be generated. However, since the probability of these two cells to have 5 common isoforms was not taken into account, 1.3% is not the probability of incorrect recognition between two cells and cannot be compared to probabilities estimated for DSCAM1, or the probabilities calculated here for Pcdhs.

The relationships between common-isoforms and common-recognition units for different multimeric states (Figure 7C). If a cell has *i* isoforms, *j* of which are in common with a contacting cell, then the percentage of common recognition units between the contacting cells can be calculated as $(\frac{j}{i})^n$ where *n* is 1, 2, 3, or 4 for monomer, dimer, trimer, or tetramer, respectively.

Calculation of the average number of copies of each type of multimer expressed in a single cell (Figure 7D). Monte-Carlo simulations were used to estimate the average number of copies of each multimeric recognition unit generated by the random selection and assembly of Pcdh isoforms into multimers (dimers, trimers, and tetramers) per cell. For each case, it was assumed that there are an equal number of copies of each of 15 Pcdhs expressed per cell. A striking result was observed for the tetramer recognition unit. If we assume that there are 5,000 copies each of 15 distinct

Pcdh isoforms expressed in an individual cell (75,000 Pcdhs total), there will be fewer than two copies of each unique recognition tetramer per neuron (Figure 7D). Although, this low number of copies may seem surprising, one needs to consider that 15 distinct Pcdhs can, in principle, generate 12,720 distinct tetramers (Yagi, 2012). In the example described above, the total number of tetramers will be 18,750 (75,000 total Pcdh copies/4 Pcdhs per tetramer). If we approximate that these 18,750 tetramers are randomly distributed among the 12,720 possible distinct tetramers, this will result in an average of only 1.5 (18,750/12,720) copies of each specific tetramer per cell.

Supplemental References

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

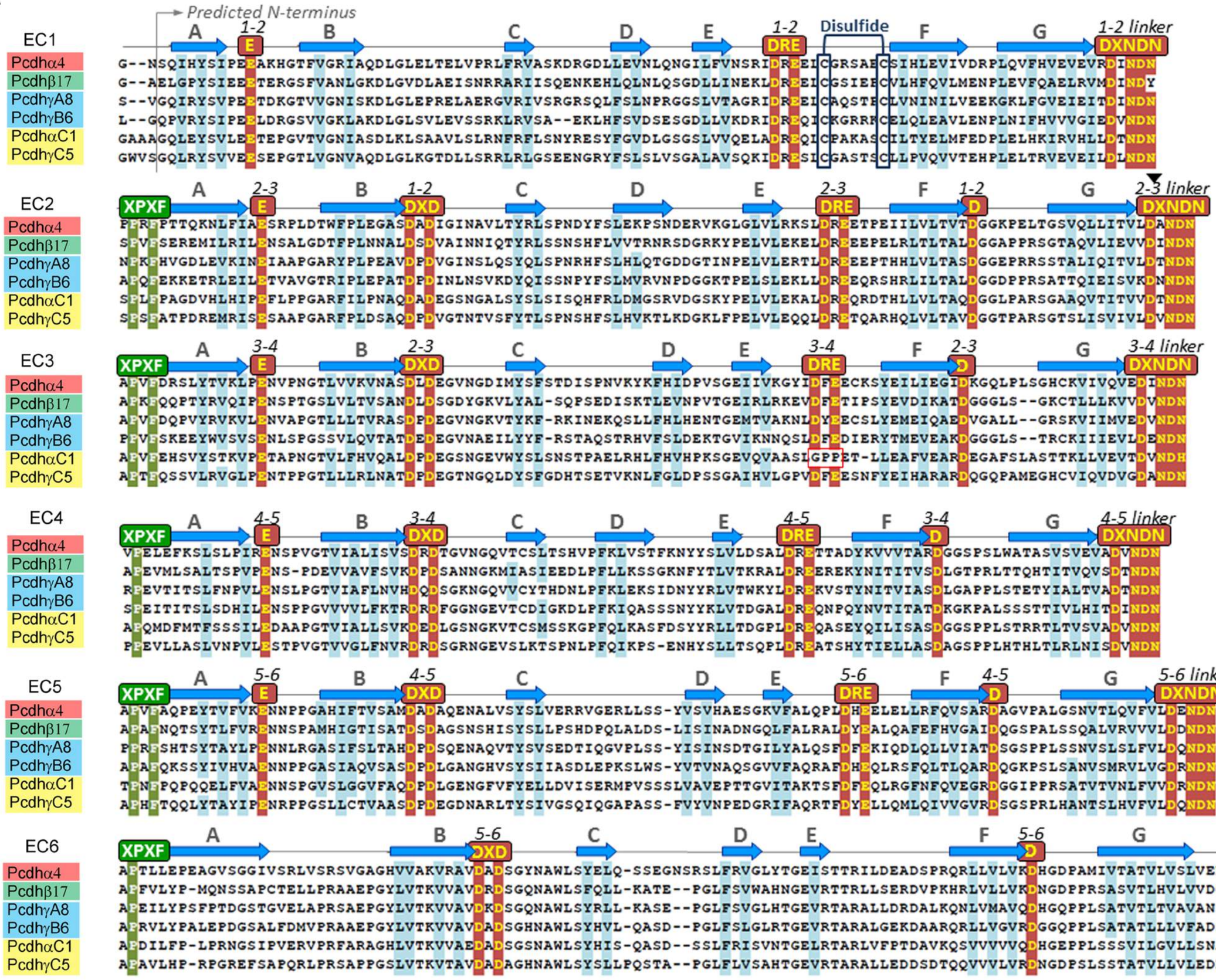
Koushik, S.V., Chen, H., Thaler, C., Puhl, H.L., 3rd, and Vogel, S.S. (2006). Cerulean, Venus, and VenusY67C FRET reference standards. Biophysical journal *91*, L99-L101.

Posy, S., Shapiro, L., and Honig, B. (2008). Sequence and structural determinants of strand swapping in cadherin domains: do all cadherins bind through the same adhesive interface? J Mol Biol *378*, 954-968.

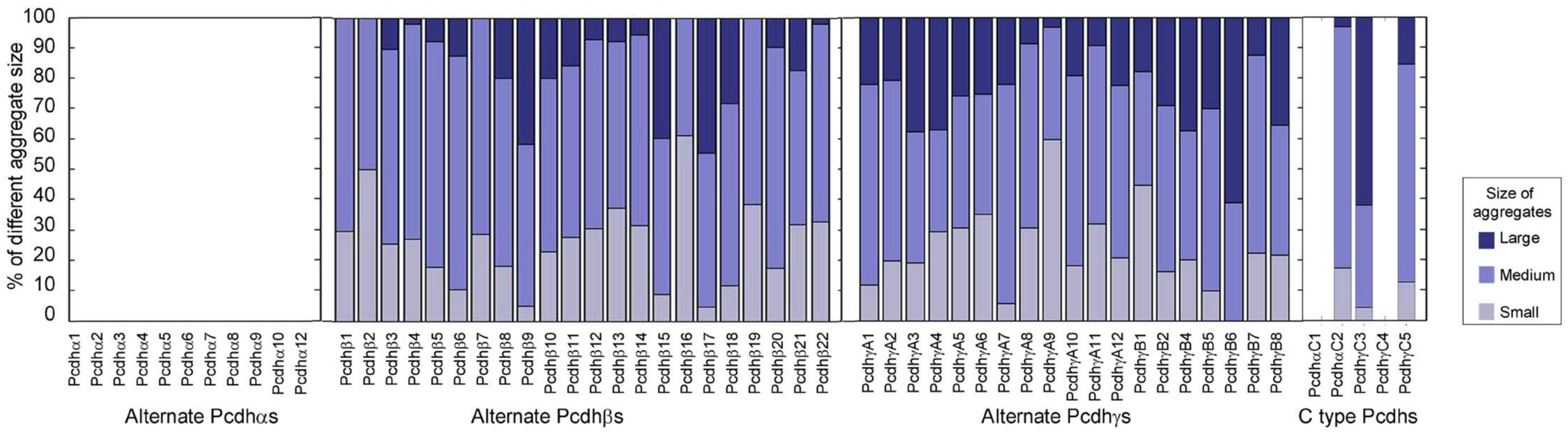
Yagi, T. (2012). Molecular codes for neuronal individuality and cell assembly in the brain. Front Mol Neurosci *5*, 45.

Supplementary Figure S1

A



B

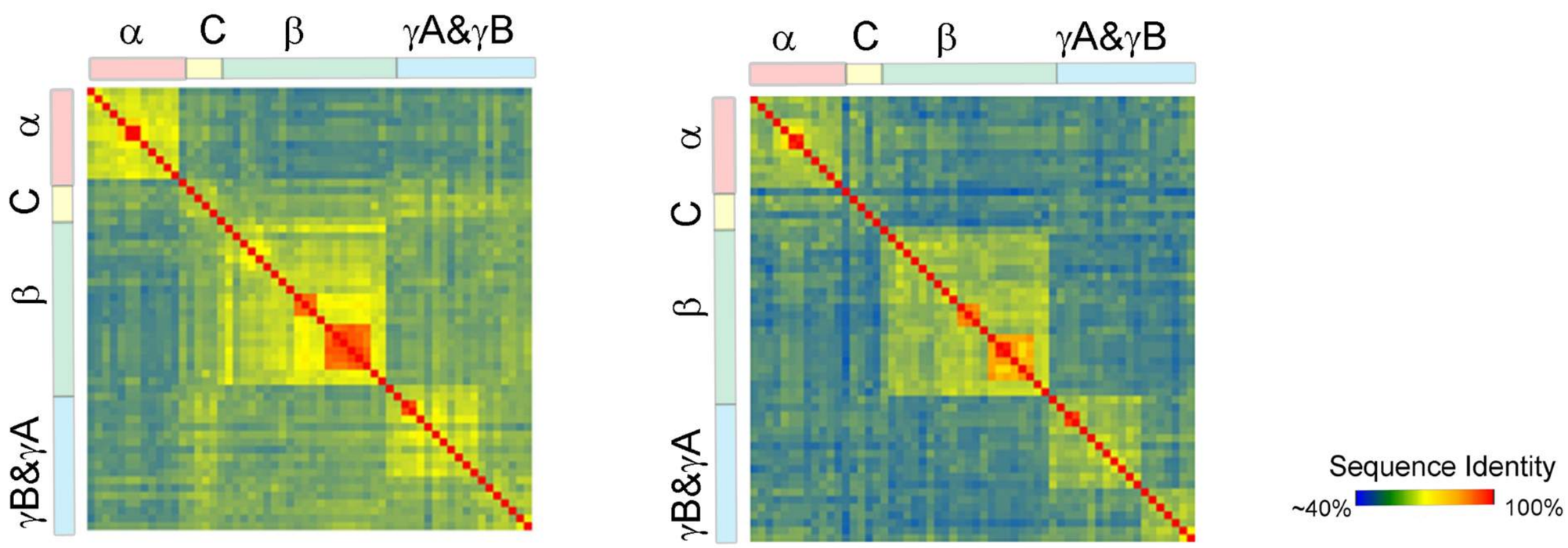


Supplementary Figure S2

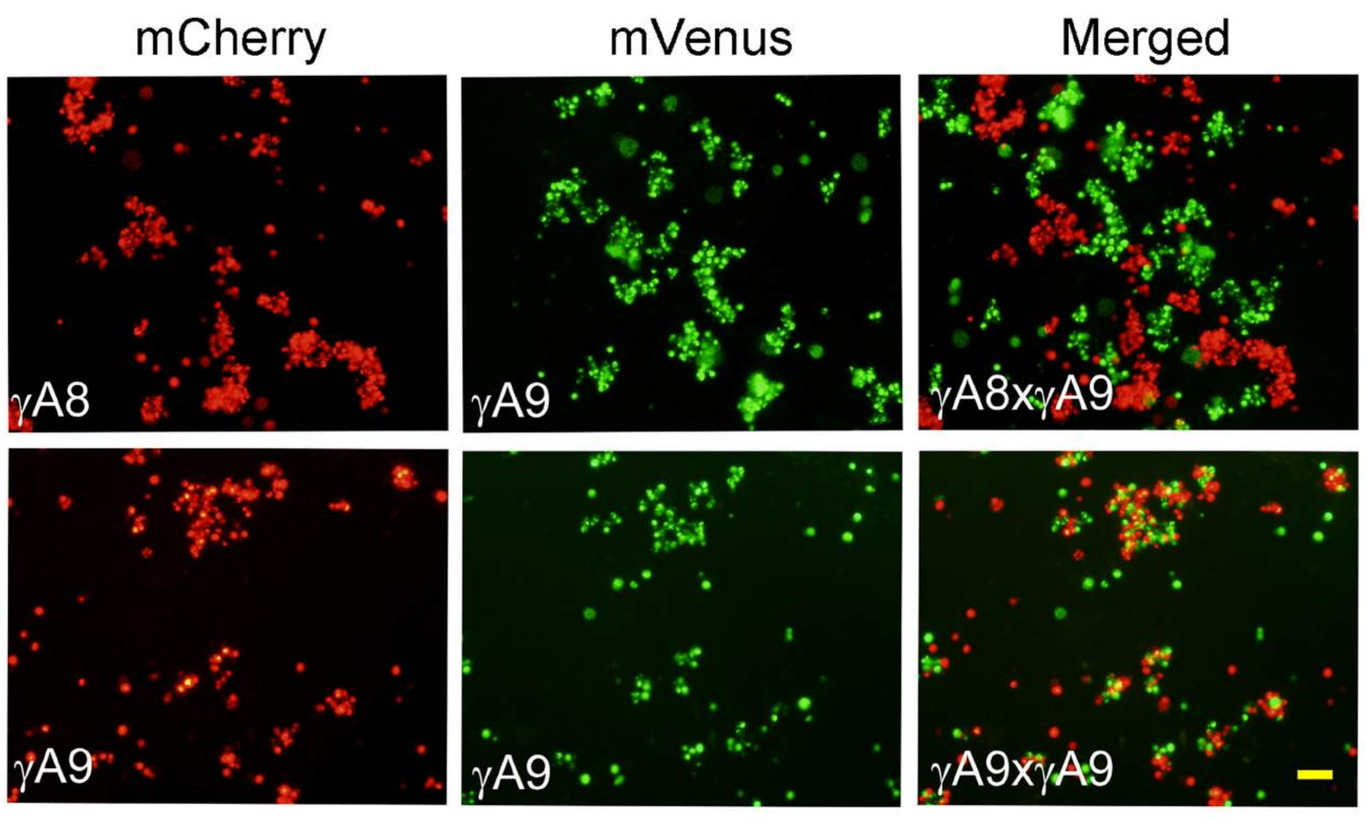
A

Average Pairwise Sequence Identity						
Subtypes	**EC1**	**EC2**	**EC3**	**EC4**	**EC5**	**EC6**
Alternate Pcdhαs	0.94	0.64	0.54	0.81	0.97	0.78
Alternate Pcdhβs	0.57	0.62	0.56	0.68	0.84	0.9
Alternate Pcdhγs	0.69	0.59	0.54	0.62	0.78	0.92
C-type Pcdhs	0.43	0.64	0.44	0.48	0.4	0.45

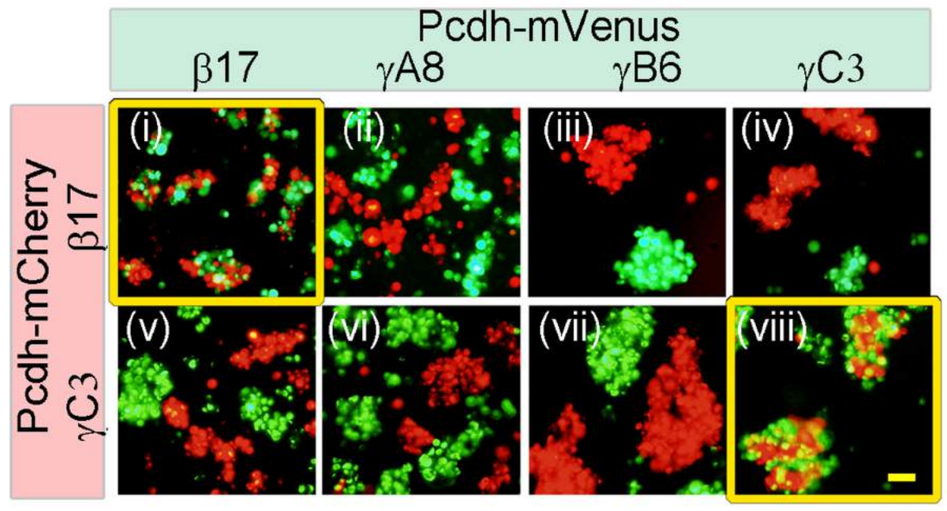
B



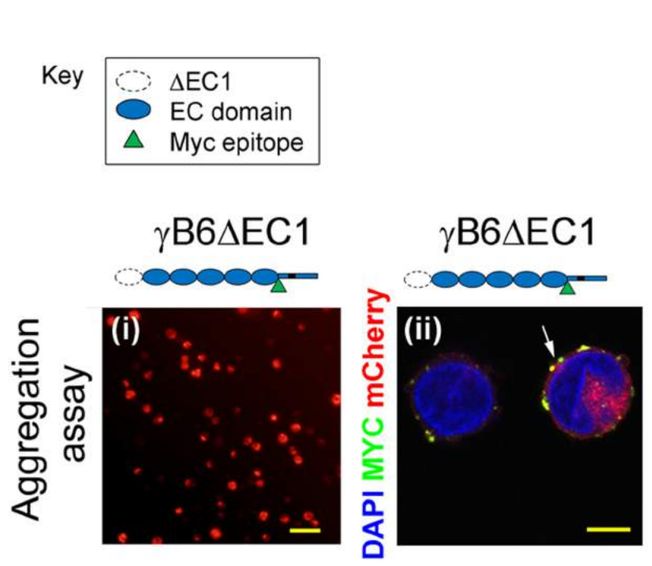
C

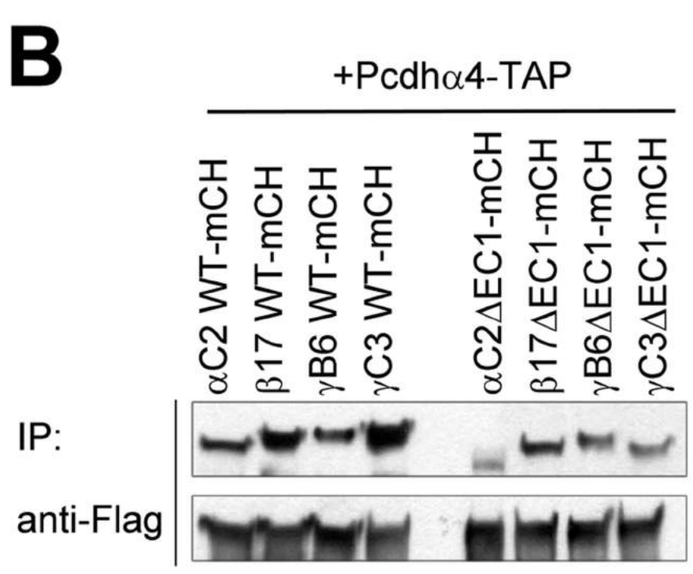


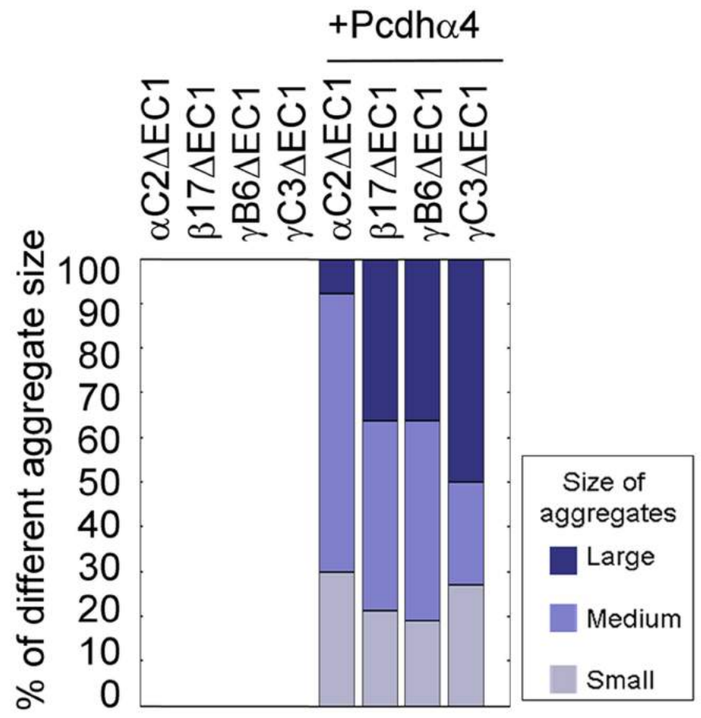
D

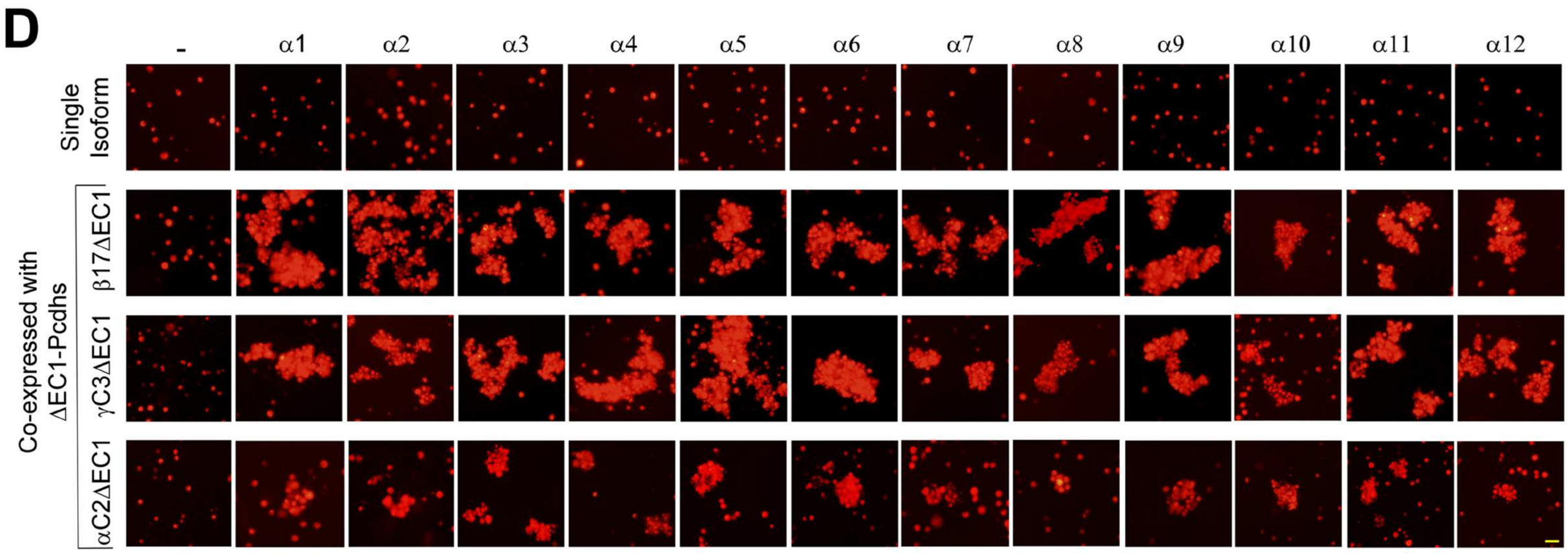


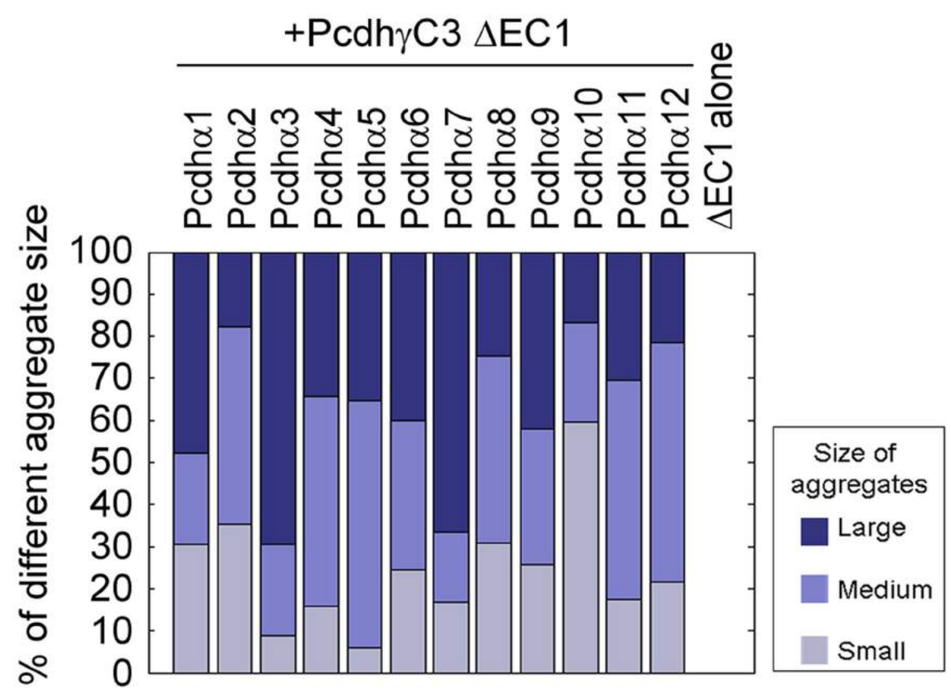
Supplementary Figure S3

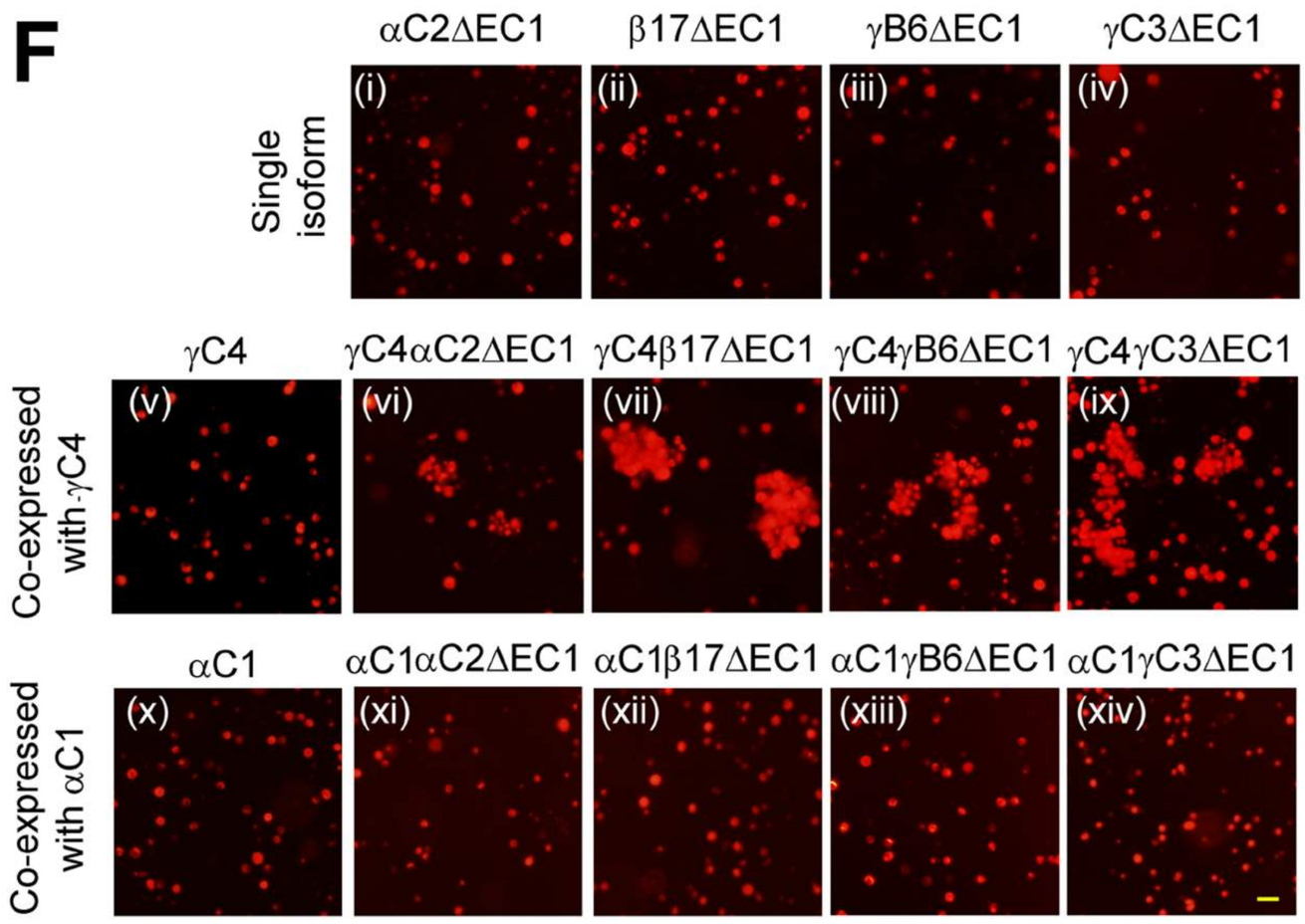




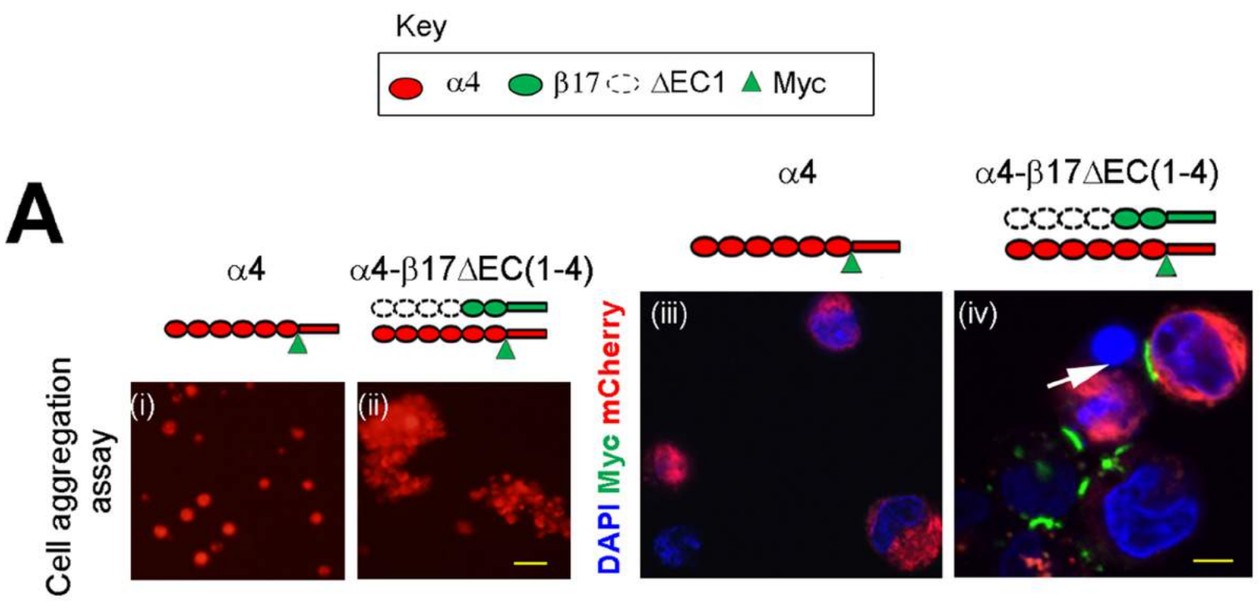


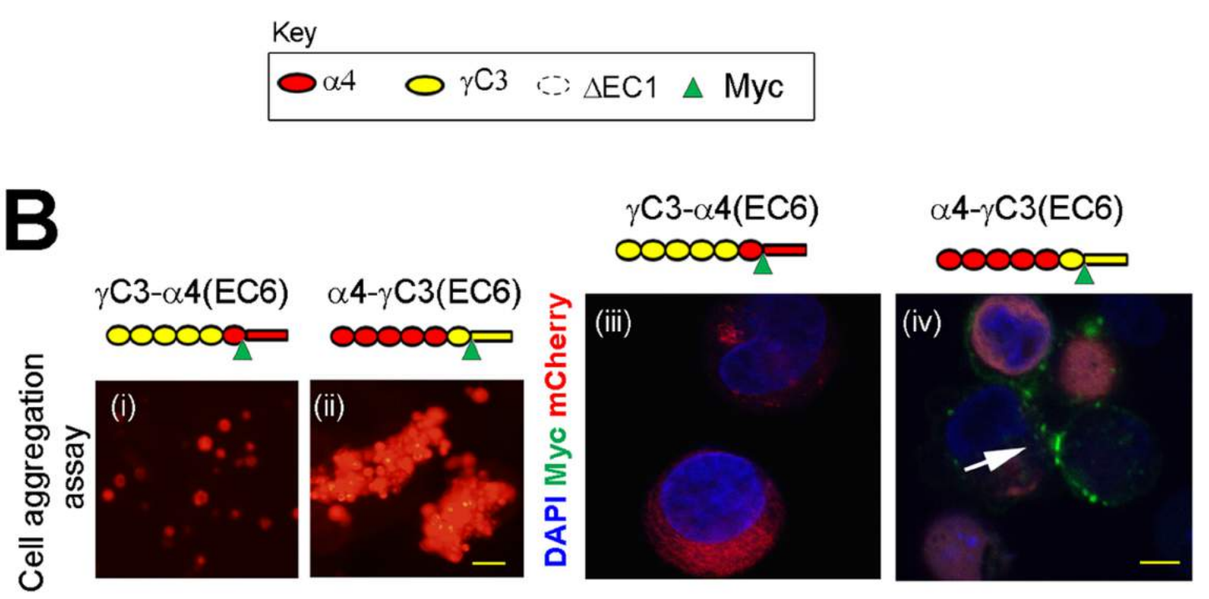


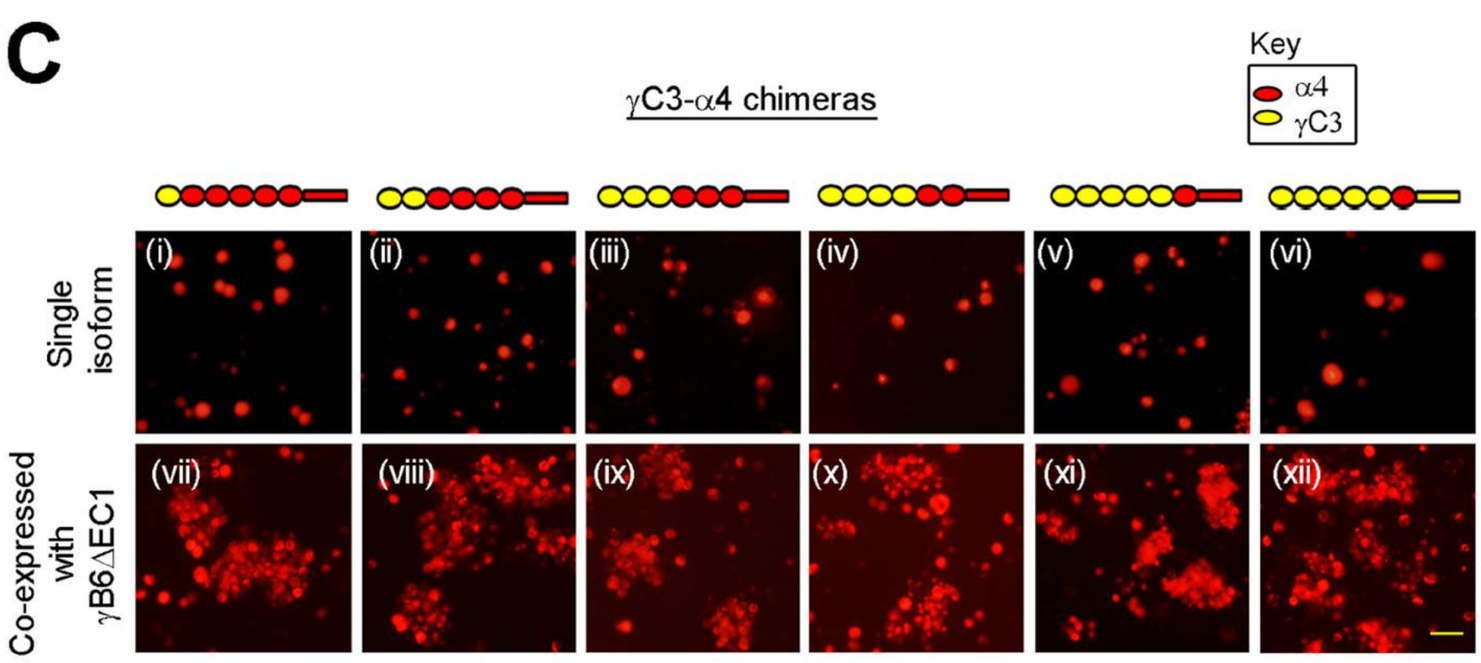


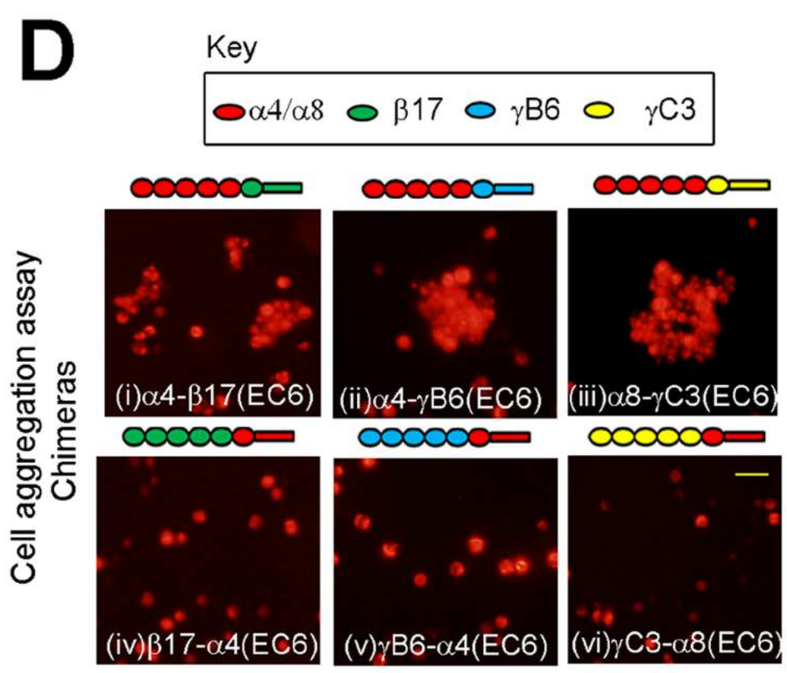


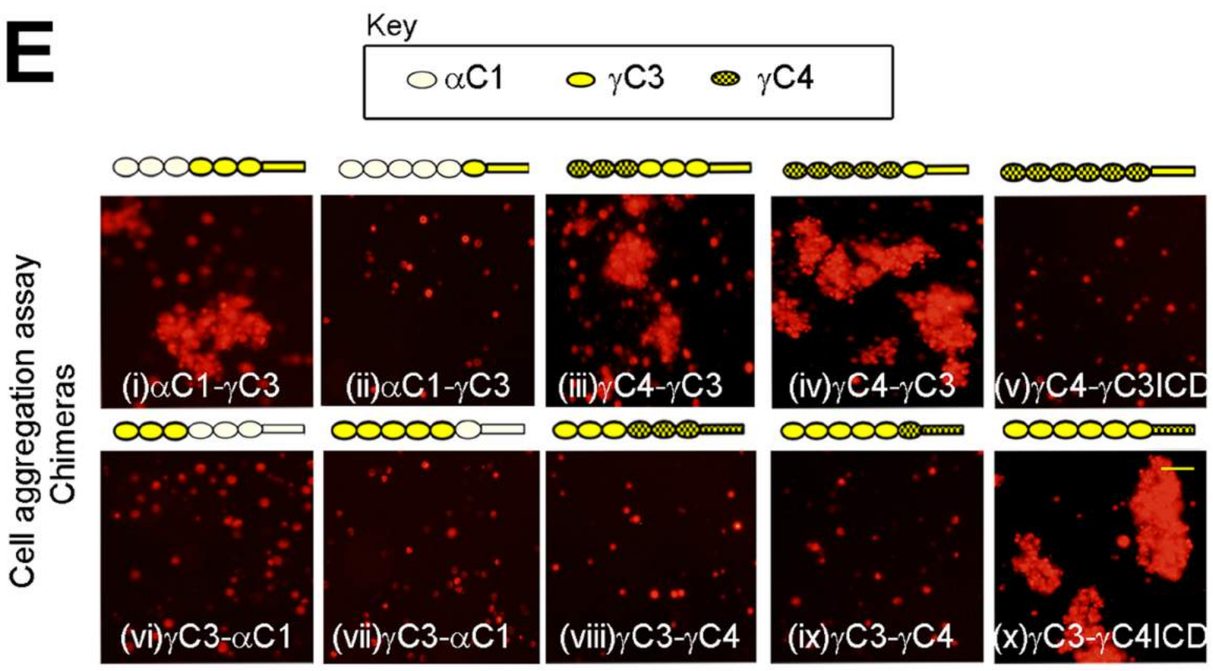
Supplementary Figure S4

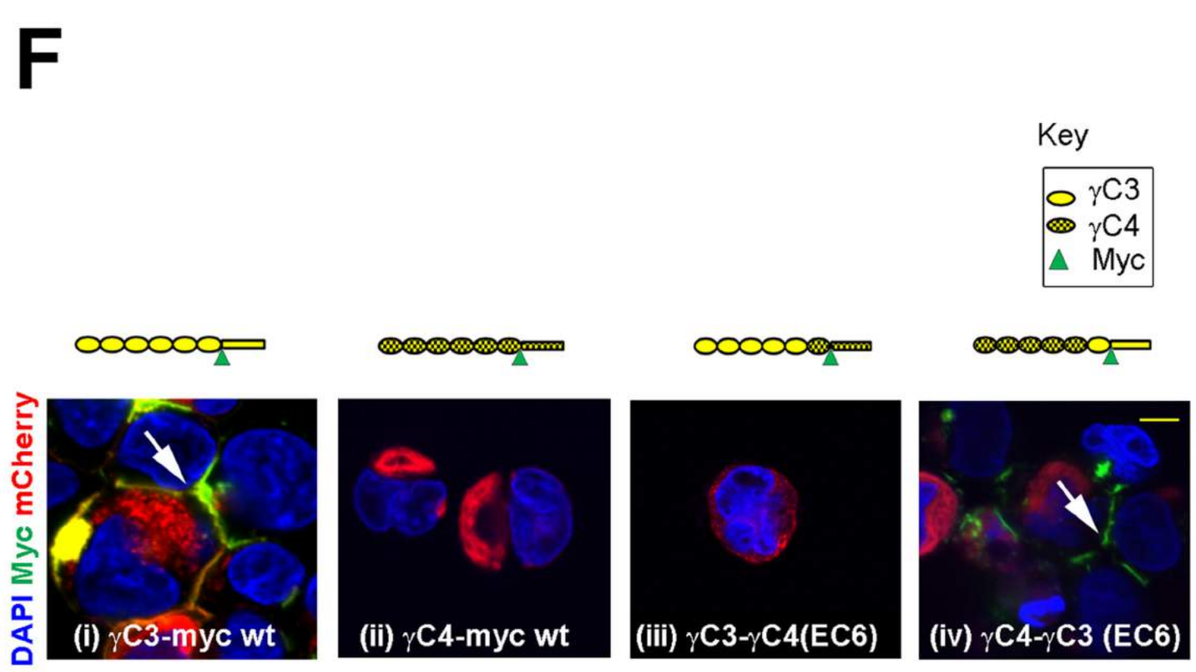


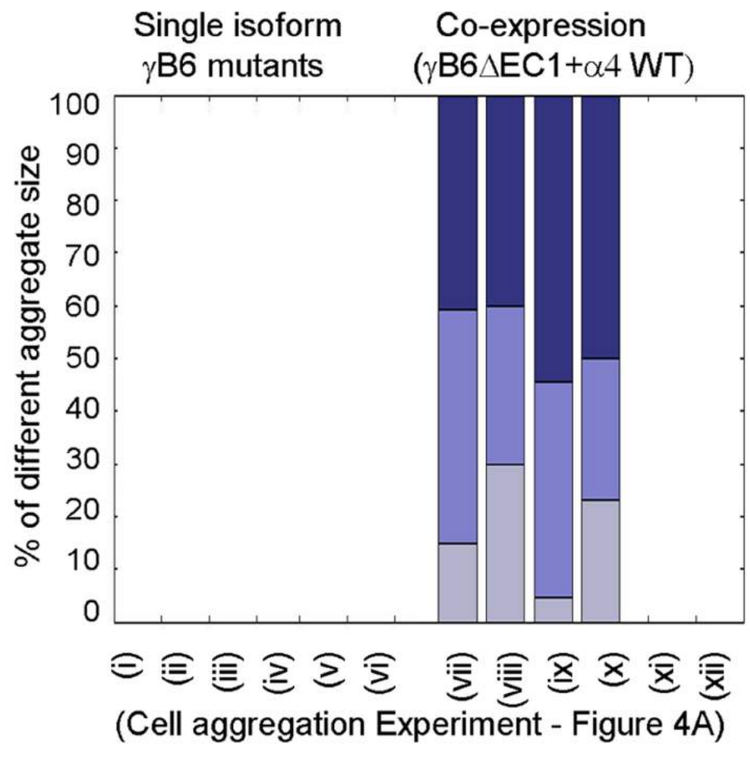


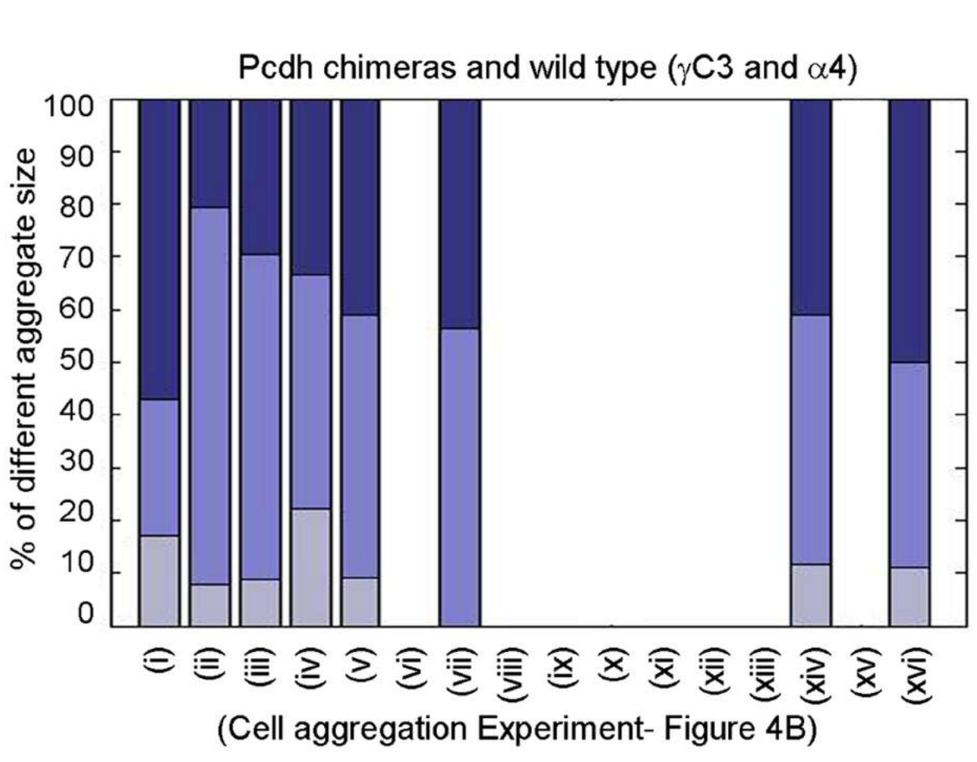


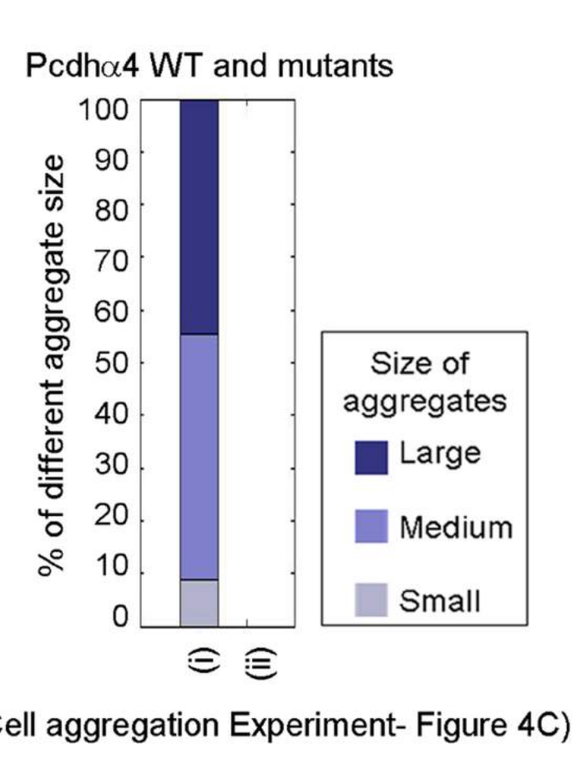




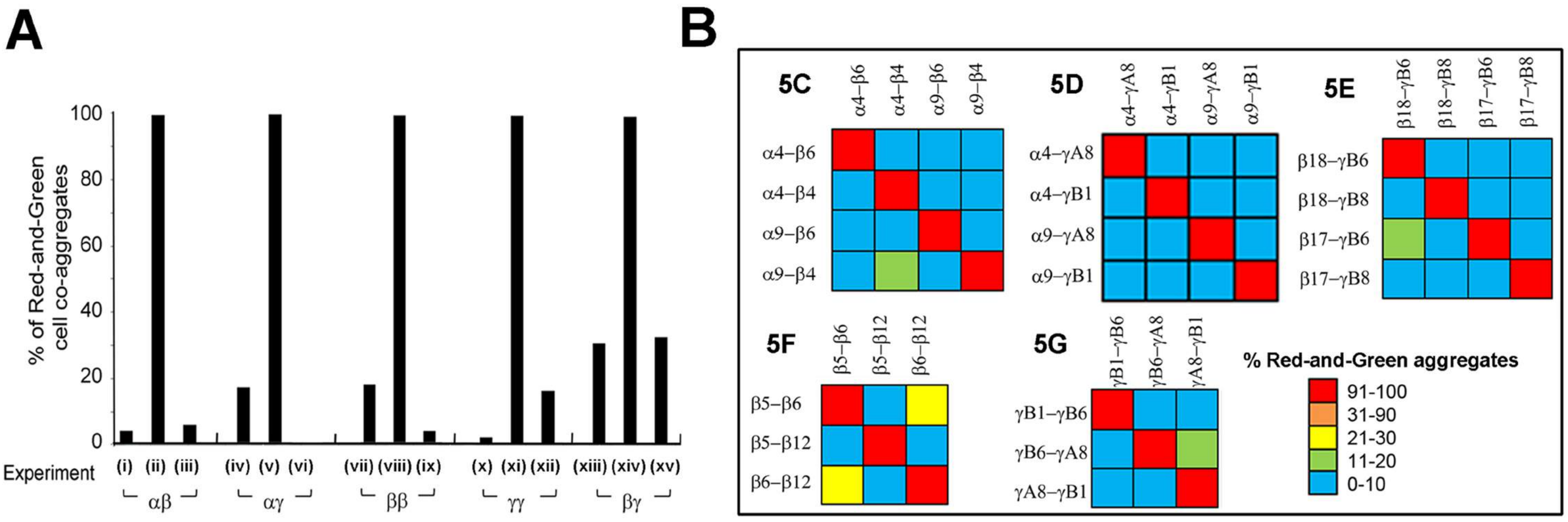


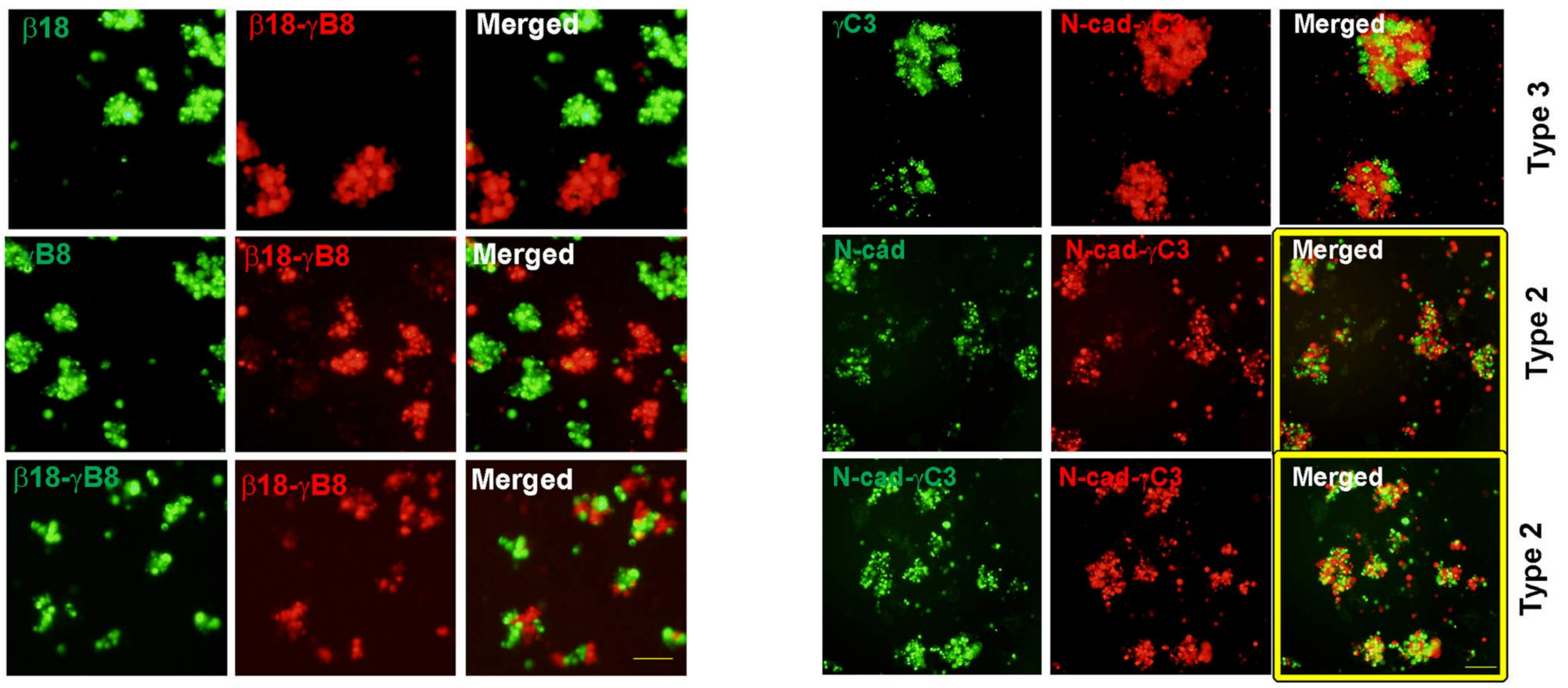


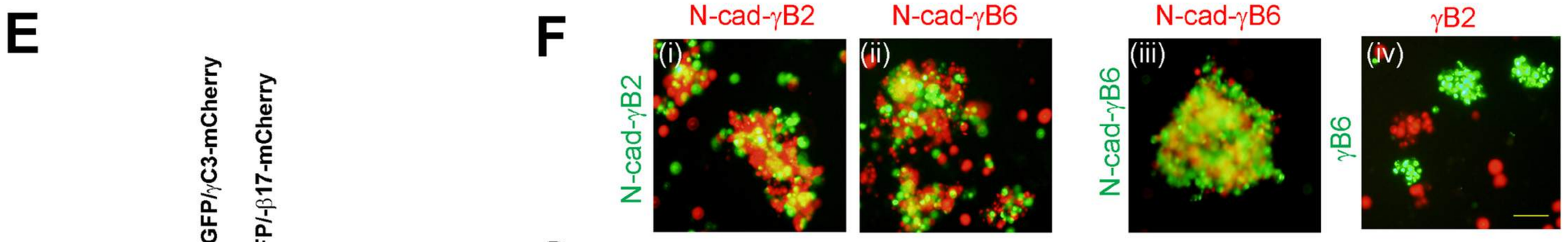


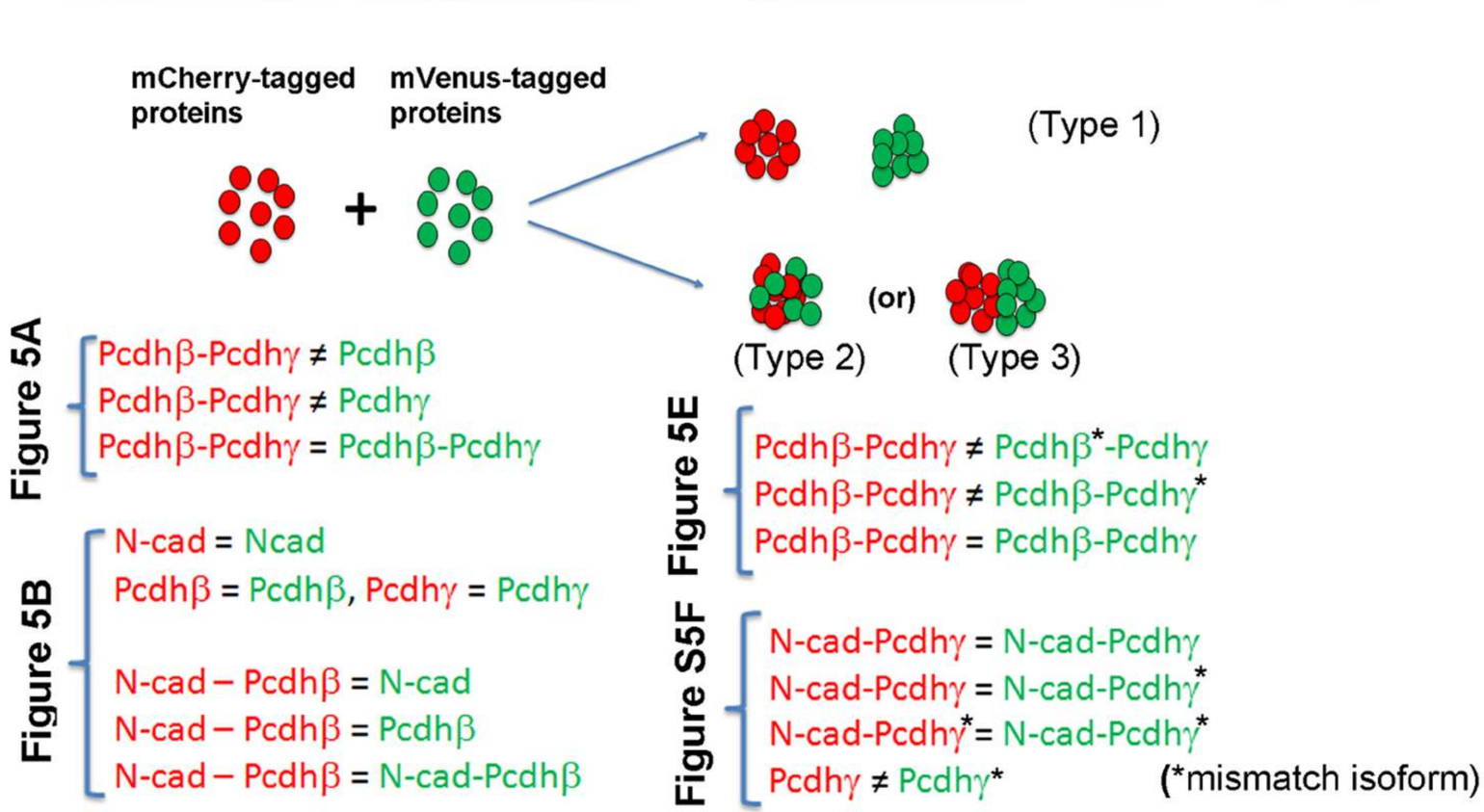


Supplementary Figure S5

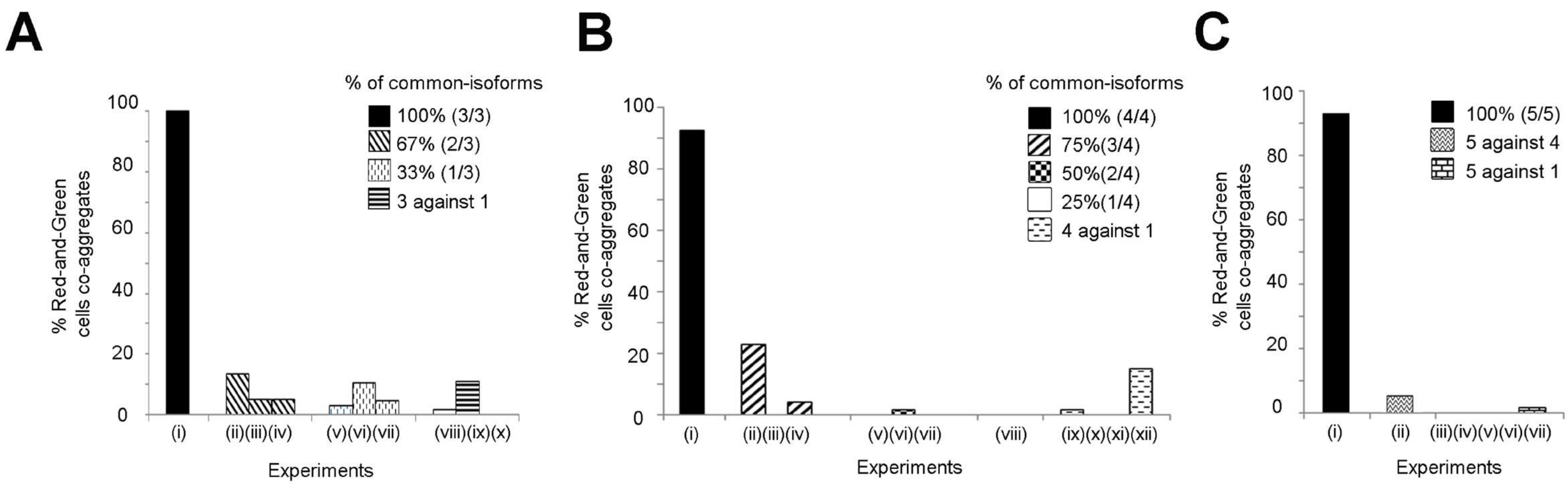


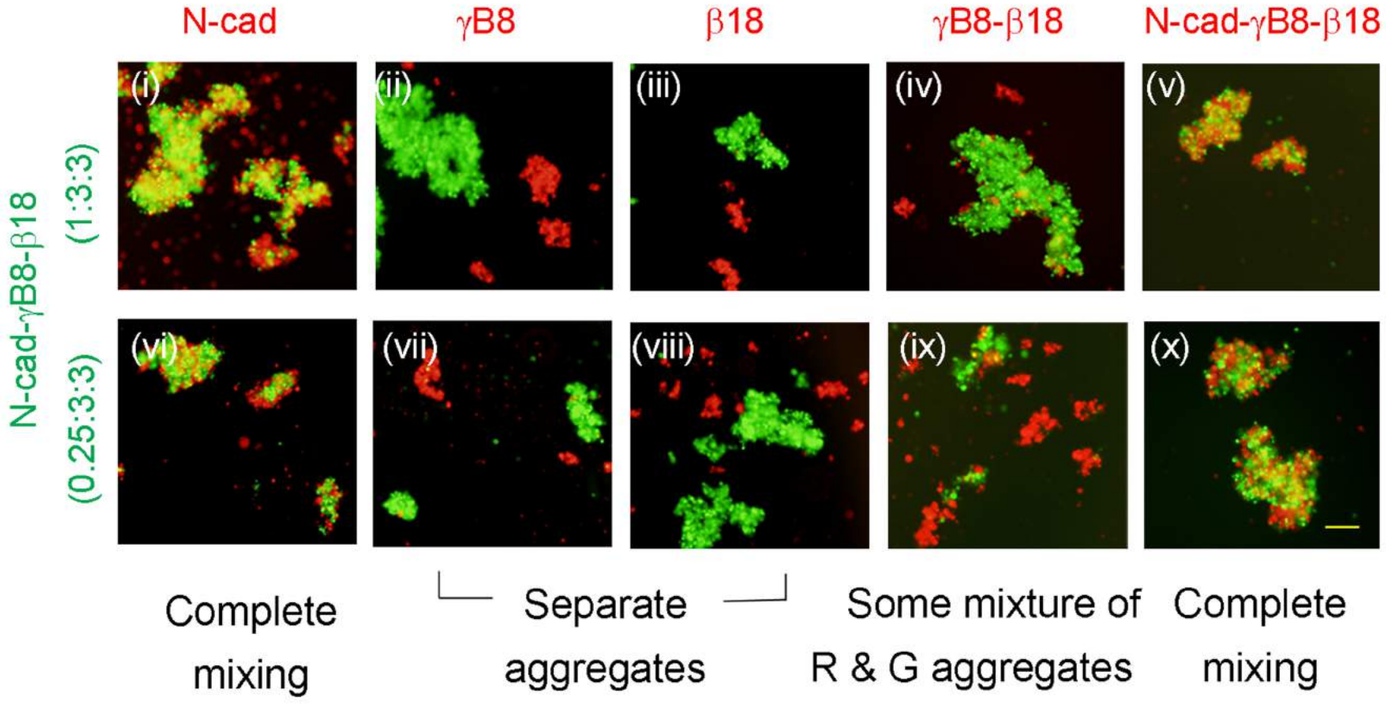






Supplementary Figure S6





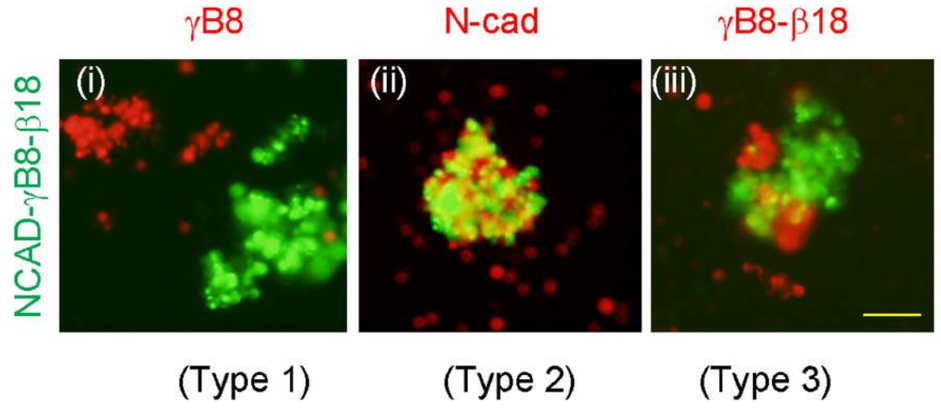
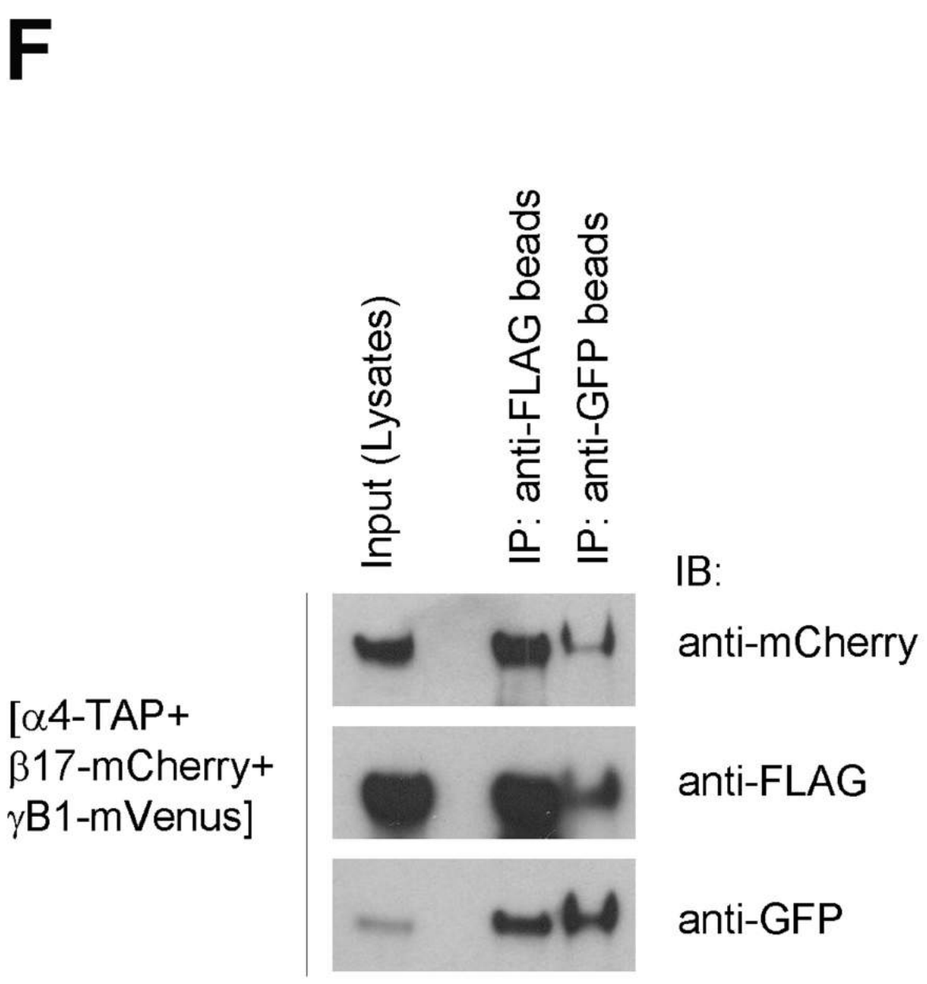


Figure S6D High magnification



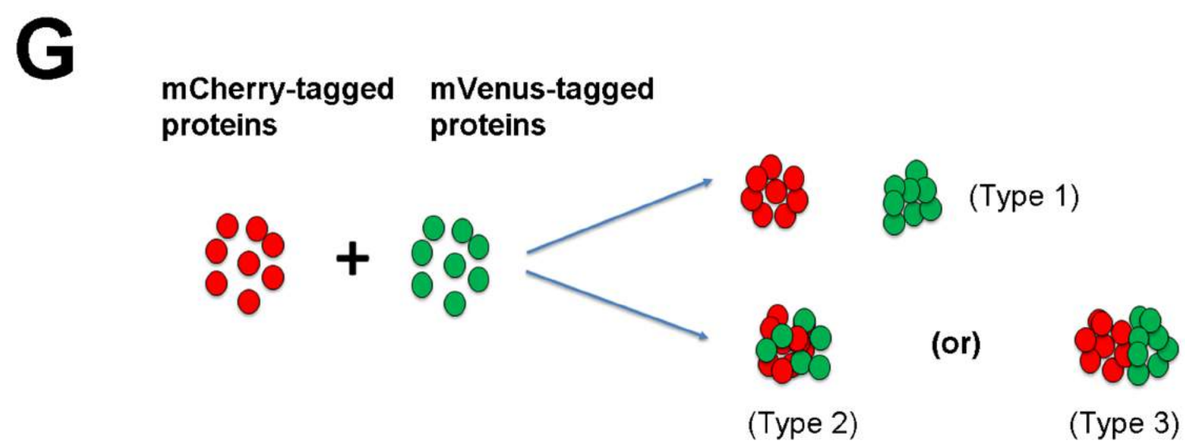


Figure 6A

Pcdhα-Pcdhβ-Pcdhγ ≠ Pcdhα or Pcdhβ or Pcdhγ
Pcdhα-Pcdhβ-Pcdhγ ≠ Pcdhα*-Pcdhβ - Pcdhγ
Pcdhα-Pcdhβ-Pcdhγ ≠ Pcdhα- Pcdhβ*-Pcdhγ
Pcdhα-Pcdhβ-Pcdhγ ≠ Pcdhα-Pcdhβ-Pcdhγ*
Pcdhα-Pcdhβ-Pcdhγ = Pcdhα-Pcdhβ-Pcdhγ

Figure S6D

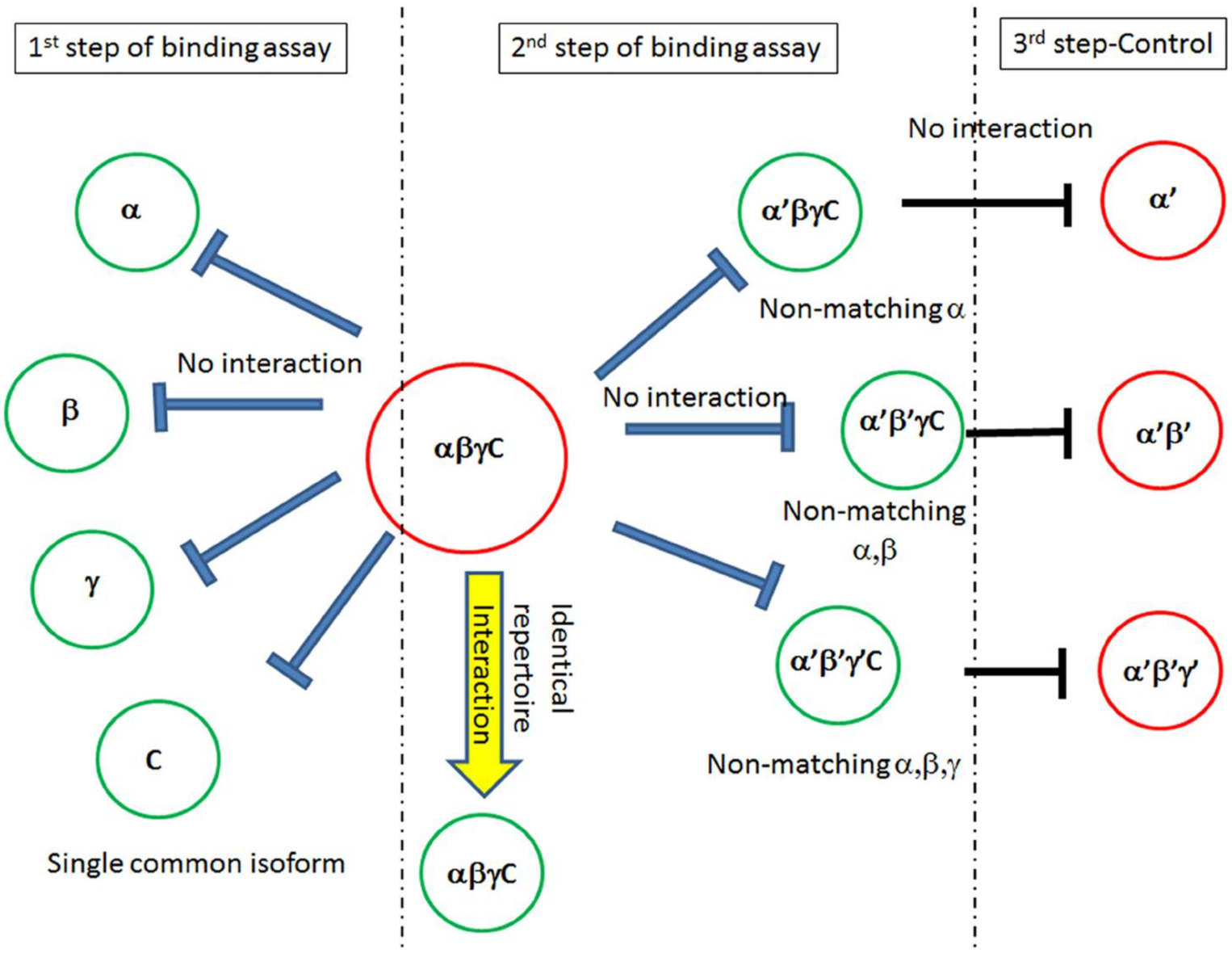
N-cad (H) – Pcdhβ – Pcdhγ = N-cad
N-cad (H) – Pcdhβ – Pcdhγ ≠ Pcdhβ
N-cad (H) – Pcdhβ – Pcdhγ ≠ Pcdhγ
N-cad (H) – Pcdhβ – Pcdhγ = Pcdhβ-Pcdhγ (Less binding)
N-cad (H) – Pcdhβ – Pcdhγ = N-cad (H) – Pcdhβ – Pcdhγ

N-cad (L) – Pcdhβ – Pcdhγ = Ncad
N-cad (L) – Pcdhβ – Pcdhγ ≠ Pcdhβ
N-cad (L) – Pcdhβ – Pcdhγ ≠ Pcdhγ
N-cad (L) – Pcdhβ – Pcdhγ = Pcdhβ-Pcdhγ (More binding)
N-cad (L) – Pcdhβ – Pcdhγ = N-cad (L) – Pcdhβ – Pcdhγ

* Mismatch isoform
(H: High, L:Low)

Supplementary Figure S7

A



B