Surface delivery quantification reveals distinct trafficking efficiencies among clustered protocadherin isoforms

Elizabeth J. May[1] and Rachelle Gaudet*

Department of Molecular and Cellular Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA

*Correspondence: Rachelle Gaudet, Harvard University, 52 Oxford St, Cambridge, MA 02138
gaudet@mcb.harvard.edu

[1] Present Address: California Institute for Quantitative Biosciences, University of California, Berkeley, Stanley Hall, Berkeley, CA

Author contributions statement:

E.J.M and R.G. designed the research, wrote the paper, and prepared the figures. E.J.M performed experiments, developed the analysis and wrote the code, and analyzed the data.

Classification:
Major: Biological Sciences
Minor: Cell Biology

Keywords: clustered protocadherins, cell surface trafficking, plasma membrane proteins

This PDF file includes:
Main text
Figures 1 to 5

Abstract

Proteins that transmit molecules and signals across the plasma membrane are crucial in cell biology because they enable cells to sense and respond to their surroundings. A major challenge for studying cell-surface proteins is that often they do not fold or traffic properly to the plasma membrane when produced in heterologous cells. We developed a strategy for quantifying surface localization from fluorescence microscopy images of surface-stained cells. Using clustered protocadherins, a protein family important for cell-cell recognition during neuronal development, we found that surface delivery levels vary among clustered protocadherin isoforms and between wildtype and engineered variants. Quantifying these differences provides evidence that *cis* dimerization is not tightly coupled to surface delivery for clustered protocadherins. This work establishes a generalizable framework for screening proteins and variants of interest for proper cell surface localization.

Significance

Surface proteins allow cells to interact with their environments, and their activities are often regulated by their delivery to and removal from the plasma membrane. We developed a strategy to quantitatively compare the surface delivery of proteins based on established epitope tag-based surface staining methods. Using natural and engineered variants of clustered protocadherins, cell-surface proteins essential for neuron development, we show that quantitative comparisons of surface trafficking levels facilitate the interpretation of mutational effects and can shed light on key regulatory mechanisms. We find that surface trafficking levels differ between variants and that, contrary to what was previously thought, a domain that inhibits surface delivery in some clustered protocadherins does so without directly relying on its protein-protein interface.

Introduction

Clustered protocadherins (cPCDHs) are a large family of single-pass transmembrane proteins that function as adhesion and signaling molecules in brain development. They are primarily expressed in the nervous system (1-4) and their genetic deletion in mice is lethal at birth due to extensive neuronal cell death (5-8). At the cellular scale, cPCDH perturbations lead to a range of defects in neuronal cell morphology and connectivity, and their roles in dendrite complexity and self-avoidance are the most studied and best understood (9-14). These essential neuronal activities rely on homophilic cPCDH interactions at cell-cell contacts, so proper localization to the plasma membrane is a prerequisite for cPCDH function.

In mammalian genomes, including mice and humans, more than fifty genes encode a diverse set of paralogous cPCDH isoforms (4, 15, 16). Each isoform contains six extracellular cadherin (EC) domains, a transmembrane helix, and an intracellular region **(Figure 1A**). The cPCDH genomic locus includes clustered arrays of exons that encode the N-terminal extracellular domains and the first ~100 amino acids of the intracellular region for each isoform. From there, the sequences differ only by subfamily: α-PCDHs and γ-PCDHs have α- and γ-specific C-terminal sequences encoded by shared exons, while β-PCDHs do not have shared exons and thus have shorter intracellular regions (**Figure S1A**) (3, 4, 16-18). At the cell surface, cPCDHs form strictly homophilic *trans* dimers between molecules on juxtaposed membranes and preferentially heterophilic *cis* dimers between isoforms on the same membrane (19-31). Defining these cPCDH interactions has led to a proposed mechanism wherein cPCDH oligomerization via both *cis* and *trans* dimers serves as a self-contact signal that ultimately leads to self-avoidance through an as-yet-unknown signaling pathway (**Figure S1B**) (20, 22, 23, 26-30, 32). Cell surface localization and *cis* and *trans* dimerization are all required for this mechanism.

72 In the brain and in cultured neurons, cPCDHs are not uniformly distributed on the cell surface;
73 instead, they are largely intracellular, have a punctate distribution, and are often enriched at cell-
74 cell contact sites (31, 33-39). When transfected into heterologous cells, α-PCDHs do not reach the
75 cell surface, while β- and γ-PCDHs do (21). This differential trafficking has been attributed to *cis*
76 dimerization (21, 26, 28, 40). The cPCDH *cis* interaction involves EC6 from one subunit and both
77 EC5 and EC6 (EC5-6) from the other subunit, forming an asymmetric dimer (28). Previous work
78 identified several conditions that induced α-PCDHs to go to the cell surface: (i) deletion of the
79 EC6 domain, (ii) replacement of the α-PCDH EC6 domain with a β- or γ-PCDH EC6 domain, (iii)
80 co-transfection with a full-length β- or γ-PCDH, or (iv) co-transfection with an EC5-6 fragment
81 (but not an EC6-only fragment) of a β- or γ-PCDH (**Figure S1C**) (21). Disruptive mutations at *cis*
82 interface positions in β- and γ-PCDHs also interfered with their surface delivery (26, 28). These
83 observations led to the conclusion that *cis* dimerization is required for cPCDH surface delivery
84 (21, 26, 28, 40), although the mechanism of this regulation remains unclear. In addition to the role
85 of the extracellular domains described above, ubiquitination or phosphorylation of the γ-PCDH
86 variable intracellular region may also play a regulatory role in surface trafficking (31, 38, 39, 41-
87 44).

88 Several previous studies have included observations of cPCDH subcellular localization in
89 heterologous cells (1, 19, 21, 33-35, 41, 45-50), but few have made quantitative surface delivery
90 measurements (31, 49). To identify cPCDH features that influence their surface trafficking, we
91 quantified cell surface localization using antibody detection of an extracellular epitope tag and
92 developed generalizable metrics that summarize multiple aspects of diverse surface delivery
93 phenotypes. We tested mutation-containing and domain-swapped α-PCDH variants to identify
94 molecular features that enable α-PCDH cell surface trafficking. We found that while replacing the

95 EC6 domain of an α-PCDH with EC6 of a β- or γ-PCDH increased surface localization compared
96 to the wildtype α-PCDH, introducing γ-PCDH *cis* interface residues into an α-PCDH did not. EC6-
97 swapped α-PCDH constructs also had lower surface delivery levels than wildtype β- and γ-PCDHs,
98 and *cis* dimerization affinity did not correlate with wildtype β- and γ-PCDH surface trafficking
99 levels. Our results indicate that features of EC6 beyond the *cis* dimer interface contribute to the
100 surface delivery of α-PCDHs, and they suggest that *cis* dimerization-independent mechanisms also
101 regulate cPCDH surface trafficking.

102

103 **Results**

104 **Defining metrics for surface delivery quantification**

105 To investigate the surface trafficking of cPCDHs, we pursued a quantitative approach that would
106 facilitate comparisons of wildtype cPCDH isoforms and engineered variants. Surface staining by
107 antibody detection of extracellular epitope tags is a common approach to assess a protein's surface
108 localization that has previously been applied to cPCDHs in heterologous cell lines and cultured
109 neurons (21, 34, 48-51). We used a derivative of the 293T cell line that lacks the adhesion protein
110 N-cadherin (293T-ΔNC); these cells are adherent but detach easily and can be used for cell
111 aggregation assays (52). Using cPCDH constructs containing an extracellular Myc tag and an
112 intracellular fluorescent protein (FP; **Figure S1A**), we added fluorescent anti-Myc antibodies (Ab)
113 to transfected cells and collected fluorescence microscopy images. We first compared two wildtype
114 cPCDH isoforms known to have different surface trafficking behaviors: α4, which does not traffic
115 to the plasma membrane, and γC3, which does (**Figure 1B**) (21). The FP served as a transfection
116 marker, and transfection efficiencies were similar across samples (**Figure S1D**). Notably, the
117 cellular distribution of the FP signal was indistinguishable between α4 and γC3, although the

118 overall FP intensity was different. Ab signal was undetectable in the α4-transfected sample (**Figure**
119 **1C**), confirming the non-surface localization of α4. As expected, the Ab stained the cell periphery
120 and cell-cell junctions at contacts between transfected cells in the Myc-tagged γC3 sample (**Figure**
121 **1C**).

122 Although the overall surface delivery behaviors of α4 and γC3 were apparent in our images, we
123 noticed that some γC3-expressing cells had very little or no measurable surface staining (**Figure**
124 **1C, arrowheads**). We therefore devised metrics to account for the phenotypes of individual cells
125 and evaluate overall surface delivery based on the distribution of cellular phenotypes. For each
126 sample, we collected both fluorescence images in the FP and Ab channels and brightfield images
127 (**Figure 2A**). We used automated segmentation in brightfield to identify cells and calculated the
128 mean FP and Ab signal per cell (FP_{cell} and Ab_{cell}) from the fluorescence channels (**Figure 2A**). We
129 identified transfected cells using a threshold FP_{cell} value determined by an untransfected control
130 (**Figure 2B**). We then generated surface stain histograms of the Ab_{cell} values of transfected cells to
131 visualize the spread of surface delivery phenotypes within samples (**Figure 2C**).

132 Using a γC3 control construct lacking the extracellular Myc tag (**Figure 2C**, gray histogram), we
133 determined a threshold Ab_{cell} value (**Figure 2C**, cyan vertical line) and calculated the fraction of
134 transfected cells with measurable surface staining, f^+ (**Figure 2C**, cyan shaded region). For γC3,
135 f^+ was 0.62 ± 0.07 (mean ± SD), while for α4 f^+ was close to zero (0.01 ± 0.01; **Figure 2D**). We
136 also calculated the mean Ab_{cell} value, $\overline{Ab}$, which is proportional to the average cell-surface cPCDH
137 concentration for a given construct (**Figure 2C**, black vertical line). As expected, f^+ and $\overline{Ab}$ for
138 γC3 were significantly higher than for α4 ($p = 2\times10^{-12}$ for f^+ and $p = 5\times10^{-7}$ for $\overline{Ab}$; **Figure 2E**).

139 We next investigated how the amount of protein on the cell surface depends on expression level.
140 Taking advantage of the variation in expression generated by transient transfection and reported

141 by our transfection marker, we constructed dose-response curves, which exhibited approximately
142 power law scaling when plotted linearly (**Figure S1E**). To group cells by expression level, we
143 binned by FP_{cell} and calculated the mean Ab_{cell} for each bin (**Figure 2F**; gray dots). On a log scale,
144 the slope of a linear fit to these data represents a surface delivery "efficiency" ($E_{surface}$) that is
145 related to the fraction of protein produced that localized to the cell surface. By accounting for
146 protein expression level, $E_{surface}$ facilitates meaningful comparisons of surface trafficking between
147 constructs even when expression levels vary. $E_{surface}$ was positive for γC3, indicating that cells
148 making more γC3 protein overall had more γC3 on the surface than cells with lower γC3 expression
149 levels, whereas $E_{surface}$ for α4 was close to zero, showing that there was little to no surface staining
150 of α4-transfected cells at any expression level (**Figure 2G**).

151

152 **Quantification can distinguish different surface delivery levels**

153 To validate our quantification strategy and investigate the relationship between cPCDH *cis*
154 dimerization and surface delivery, we focused on the *cis* interface domain EC6. We tested EC6-
155 swapped α4 constructs (α4-γB6$_{EC6}$, α4-β17$_{EC6}$, and α4-γC3$_{EC6}$), and their wildtype counterparts
156 (γB6, β17, and γC3) for surface delivery in 293T-ΔNC cells (**Figure 3A**). We chose these variants
157 because they were all previously shown to mediate cell aggregation in K562 cells (21, 22, 26).
158 Cells transfected with EC6-swapped α4 variants had dim, punctate Ab signal (**Figure S2A**), often
159 at contact points between transfected cells (**Figure S2B**). Wildtype β- and γ-PCDHs had more
160 uniform surface staining with bright signal at cell-cell contacts, and β17 and γC3 were brighter
161 than γB6 (**Figure S2A**). These samples thus presented a range of surface delivery phenotypes.

162 All the Ab_{cell} histograms at least partially overlapped with the untagged γC3 distribution, showing
163 that every sample contained a subpopulation of transfected cells without any surface stain signal

164 (**Figure S3**). The EC6-swapped α4 variants had f^{+} between 0.2 and 0.4, while f^{+} for the wildtype
165 β and γ isoforms was generally higher and ranged from 0.4 to 0.7 (**Figure 3B**). $\overline{Ab}$ values for the
166 wildtype β17 and γC3 isoforms were significantly higher than $\overline{Ab}$ for wildtype α4 ($p \leq 4\times10^{-10}$;
167 **Figure 3C**). The $\overline{Ab}$ metric emphasized the ultra-bright surface staining of some cells in the β17
168 and γC3 samples (**Figure S3**), indicating that β17 and γC3 traffic to the cell membrane particularly
169 well for cPCDHs. Compared to α4, $E_{surface}$ was positive and significantly higher for all other
170 constructs ($p \leq 3\times10^{-9}$), and $E_{surface}$ was highest for the wildtype β and γ isoforms (**Figure 3D**).

171 Overall, our quantification indicates that surface levels of the EC6-swapped variants were higher
172 than wildtype α4 in 293T-ΔNC cells, in agreement with previous findings that the same variants
173 mediated cell aggregation in K562 cells (21, 22, 26). Furthermore, our quantification revealed that
174 the wildtype isoforms (γB6, β17, and γC3) generally trafficked better than the engineered α4
175 variants (α4-γB6$_{EC6}$, α4-β17$_{EC6}$, and α4-γC3$_{EC6}$).

176 Finally, to validate our use of the 293T-ΔNC cell line for surface trafficking measurements, we
177 also measured surface delivery in K562 cells. K562 cells grow in suspension, so we measured
178 FP_{cell} and Ab_{cell} using flow cytometry. Using α4, α4-γC3$_{EC6}$, and γC3, we confirmed that the trends
179 we observed were consistent across cell types and measurement modalities (**Figure S4**). As in
180 293T-ΔNC cells, both α4-γC3$_{EC6}$ and γC3 were significantly higher than α4 in K562 cells (p =
181 5×10^{-11} and 0 for $E_{surface}$, respectively; **Figure S4B**), and γC3 was significantly higher than α4-
182 γC3$_{EC6}$ ($p = 5\times10^{-7}$ for $E_{surface}$; **Supplementary Dataset S1**).

183

184 **The γ-PCDH *cis* interface is insufficient to enable α-PCDH surface delivery**

185 We next explored the link between cPCDH *cis* dimerization and surface delivery to investigate
186 why α-PCDHs do not localize to the cell surface. Previous studies used a loss-of-function
187 approach—introducing disruptive mutations into the *cis* interface of representative β- and γ-
188 PCDHs—to show that some, but not all, *cis* interface mutations prevented cell aggregation in K562
189 cells (26, 28). This finding supported the conclusion that *cis* dimerization is necessary for cPCDH
190 surface delivery (21, 26, 28, 40). To clarify the role of *cis* dimerization in α-PCDH surface delivery,
191 we took a complementary gain-of-function approach—introducing β- and γ-PCDH amino acids
192 into an α-PCDH at differentially conserved *cis* interface positions—to test whether *cis* interface
193 residues are sufficient to increase α-PCDH surface delivery.

194 To identify differentially conserved *cis* interface positions, we aligned all wildtype mouse cPCDH
195 isoforms using their extracellular sequences (EC1-6) and divided the sequences into two
196 alignments according to their cell aggregation behavior as reported by Thu et al. (21): cell surface
197 trafficking isoforms (β- and γ-PCDHs, αC2, γC3, and γC5) and non-trafficking isoforms (α-
198 PCDHs, αC1, and γC4). We calculated the amino acid frequencies at each position in EC6
199 separately for the two alignments and compared the resulting distributions using the Kullback-
200 Leibler divergence (KL_{div}; **Figure 4A**). KL_{div} scores are high for positions that are conserved within
201 each alignment but differ between the two alignments, and low for positions that are either variable
202 or conserved across both alignments.

203 We mapped the KL_{div} scores of EC6-side *cis* interface residues onto the γB7 *cis* dimer structure
204 (28) and the predicted α4 model from the AlphaFold Protein Structure Database (53) (**Figure 4B**).
205 In general, the residues in the top and center of the *cis* interface are strictly conserved among all
206 cPCDHs, while differentially conserved positions are at the sides and bottom of the interacting
207 surface, as previously noted (28). Several of the most differentially conserved positions in the *cis*

208 interface have different sidechain properties (e.g., R595 in α4 versus V560 in γB7, or Y627 in α4
209 versus R590 in γB7; **Figure 4B**). To test whether these differences underlie subfamily-specific
210 surface delivery, we introduced conserved β- and γ-PCDH *cis* interface residues into the
211 corresponding positions of an α-PCDH.

212 We continued using α4 as a representative non-trafficking α-PCDH isoform. To make α4 more
213 likely to *cis* homodimerize, we introduced residues from a γB-subfamily isoform because surface
214 plasmon resonance measurements of cPCDH *cis* heterodimers suggested that γB-subfamily
215 cPCDHs are the most amenable to participating as either the EC6-only or EC5-6 sides of the *cis*
216 interface (29). We chose γB6 because it is highly similar to γB7, the only isoform for which a *cis*
217 dimer structure is available (28), with 97% sequence identity in EC6. Of the 28 EC6 interface
218 residues, fourteen differ between α4 and γB6. We mutated the top five (α4-5x$_{\gamma B6}$), top ten (α4-
219 10x$_{\gamma B6}$), and top twelve (α4-12x$_{\gamma B6}$) most differentially conserved amino acids in α4 to the
220 corresponding γB6 amino acid, as well as making all possible EC6 interface mutations (α4-14x$_{\gamma B6}$;
221 **Table S1; Figure 4C**). Contrary to our hypothesis that these would be gain-of-function mutations,
222 none of the mutation-containing chimeras showed improved surface delivery compared to
223 wildtype α4, even though the EC6 domain-swapped construct α4-γB6$_{EC6}$ did (**Figure 4D-F**). To
224 verify that the mutations did not destabilize the protein, we co-expressed α4 and the chimeric
225 constructs with untagged γA11 (**Figure S5A**). Wildtype α4 and all the mutation-containing
226 variants except α4-14x$_{\gamma B6}$ were detected on the surface to similar extents in the co-transfection
227 experiment ($p \leq 0.03$ comparing $E_{surface}$ of each co-transfection experiment to α4 expressed on its
228 own; **Figure S5B-D**). Although α4-14x$_{\gamma B6}$ was expressed at similar levels to surface-trafficking
229 constructs such as α4-γB6$_{EC6}$ (**Figure S5E**), it is possible that α4-14x$_{\gamma B6}$ does not fold properly,
230 and that at least some α4 EC6 *cis* interface positions are important for the protein's stability.

231 Overall, we found that changing differentially conserved EC6 *cis* interface positions of α4 to match
232 γB6 was insufficient to increase surface delivery compared to wildtype α4. This was surprising
233 because α-PCDHs were previously thought to not reach the cell surface because they were unable
234 to form *cis* homodimers (28), which implied that the *cis* dimer interface in the γB6 EC6 domain
235 was responsible for the surface delivery of α4-γB6$_{EC6}$. In contrast, our results suggest that EC6
236 features away from the *cis* dimer interface explain the domain-swap-rescue phenotype for α-
237 PCDHs.

238

239 **Surface levels of EC6-swapped α-PCDHs are lower than wildtype β- and γ-PCDHs**

240 To further explore the EC6 features that explain why a domain-swap can rescue surface delivery
241 of α-PCDHs, we generated an expanded set of EC6-swapped α4 constructs using isoforms from
242 the different cPCDH subfamilies (**Table S1; Figure 5A**). We also tested the corresponding
243 wildtype β or γ isoforms, and an α4 variant lacking EC6 altogether (α4ΔEC6; **Figure S6A**). The
244 α4ΔEC6 construct showed a significantly higher $E_{surface}$ than wildtype α4 ($p = 2\times10^{-5}$; **Figure 5B**),
245 and most EC6-swapped α4 constructs trafficked like α4ΔEC6 ($p \geq 0.2$ for $E_{surface}$ for all EC6-
246 swapped α4 constructs compared to α4ΔEC6), except for α4-γC3$_{EC6}$, which had higher $E_{surface}$ than
247 α4ΔEC6 ($p = 0.0002$; **Figures 5B and S6B-C**). This suggests that surface delivery is enabled by
248 the absence of the α4 EC6 domain, not the presence of a β- or γ-PCDH EC6 domain, except for
249 the γC3 EC6 domain possibly further enhancing surface trafficking.

250 Our results agree with previous observations in K562 cells suggesting that the α-PCDH EC6
251 domain inhibits surface trafficking, including that α4ΔEC6 trafficked to the cell surface and
252 mediated cell aggregation, and that replacing γC3's EC6 with α4's EC6 interfered with surface
253 trafficking and cell aggregation (21). Because co-expression of α-PCDHs with β- or γ-PCDHs led

254 to α-PCDH surface trafficking (21, 22, 26, 28), a finding which we reproduced here in 293T-ΔNC
255 cells (**Figure S5**), it was previously inferred that the α-PCDH *cis* dimer interface contained a
256 negative regulator that was responsible for the inhibitory activity of EC6, and that *cis*
257 heterodimerization with a β- or γ-PCDH could mask this inhibitory retention signal. Together with
258 our observation that extensive mutagenesis of the α4 EC6 *cis* dimerization interface did not
259 increase surface trafficking, our quantitative surface-trafficking results instead suggest that the α-
260 PCDH EC6 inhibits surface trafficking through an unknown mechanism that does not require the
261 *cis* dimer interface.

262 We next compared the surface delivery of the EC6-swapped α4 variants to that of the originating
263 wildtype isoforms. For all constructs except αC2, the EC6-swapped variant trafficked less than the
264 corresponding wildtype ($p \leq 1\times10^{-6}$ for $E_{surface}$) (**Figure 5C, Figure S6D-E**). αC2 is part of the α-
265 PCDH genomic cluster but is phylogenetically distinct from the α-PCDHs, like the other C-type
266 isoforms (16). Wildtype αC2 trafficked poorly, with f^{+}, $\overline{Ab}$, and $E_{surface}$ similar to α4 ($p = 0.2$ for
267 $E_{surface}$; **Figure S4**), while α4-αC2$_{EC6}$ trafficked better than α4 ($p = 1\times10^{-14}$ for $E_{surface}$; **Figure 5B**),
268 similar to the other EC6-swapped variants (**Figure 5B-C, Figure S6B-E**). This suggests that the
269 αC2 EC6 does not share α-PCDH EC6's inhibitory feature. We note that in previous work, the
270 same αC2 construct was detected on the cell surface and mediated cell aggregation in K562 cells
271 (21, 22), albeit at low levels. In our hands, full-length and cytoplasmic region-deleted αC2
272 trafficked as poorly as α4 in both K562 and 293T-ΔNC cells (**Figure S4**). However, the overall
273 trend that α4ΔEC6 and EC6-swapped α4 variants did not traffic as well as wildtype β- and γ-
274 PCDHs suggests that additional features beyond EC6 also contribute to determining α-PCDH
275 surface delivery levels.

276 As cPCDH *cis* dimerization is asymmetric, with EC6 of one subunit binding EC5 and EC6 of the
277 other subunit, we wondered whether α4 constructs with both EC5 and EC6 swapped would traffic
278 better than EC6-swapped α4 variants. We generated EC5-6-swapped α4 variants that represented
279 each cPCDH subfamily: α4-γA11$_{EC5-6}$, α4-γB6$_{EC5-6}$, α4-β17$_{EC5-6}$, and α4-γC3$_{EC5-6}$ (**Table S1;**
280 **Figure 5A, Figure S6A**). The EC5-6-swapped α4 constructs showed no improvement over their
281 EC6-swapped counterparts; rather, they were also similar to α4ΔEC6 ($p \geq 0.07$ for $E_{surface}$; **Figure**
282 **5B**). Therefore, features of EC6, but not EC5, contribute to determining α-PCDH surface delivery
283 levels.

284

285 **Surface delivery levels are independent of *cis* dimerization affinity**

286 Previous studies using *cis*-dimer-disrupting mutations in β- and γ-PCDHs revealed that the *cis*
287 dimer interface is necessary for β- and γ-PCDH surface delivery (26, 28). We investigated the
288 relationship between *cis* binding affinity and surface delivery levels in wildtype β- and γ-PCDH
289 isoforms. We looked at whether our quantitative surface trafficking measurements of wildtype β-
290 and γ-PCDHs correlate with published *cis* homodimerization affinity values (or that of the closest
291 paralog for which an affinity has been reported (22, 26-29); **Figure S6F)**. If *cis* dimerization is the
292 primary determinant of surface trafficking levels, we expect isoforms that dimerize with higher
293 affinity to traffic better than lower affinity isoforms. We found no convincing correlation between
294 affinity and any of our surface delivery metrics (**Figure 5D, Figure S6G-H)**. Overall, *cis*
295 dimerization affinity was a poor predictor of surface delivery levels for β- and γ-PCDH isoforms,
296 suggesting that it is not the sole determinant of β- and γ-PCDH export.

297

298 **Discussion**

299 In heterologous expression systems, surface delivery is a useful metric to validate proper surface
300 localization for proteins of interest, to account for differences in surface protein levels, and to
301 identify protein modifications that disrupt trafficking. We developed a robust method to quantify
302 the surface delivery of plasma membrane proteins using epitope-tag-based surface staining. Our
303 metrics describe three axes of variation common to any surface staining experiment: the fraction
304 of cells with detectable levels of surface protein, relative surface protein abundance, and relative
305 surface trafficking efficiency. We show that our quantification strategy is compatible with data
306 acquired either by fluorescence microscopy or flow cytometry, and our scripts are available for
307 researchers to modify or apply directly to their own datasets. To validate our approach, we applied
308 it to cPCDHs and found that the quantification recapitulated known subfamily-specific trafficking
309 trends observed qualitatively via K562 cell aggregation assays: β- and γ-PCDHs reach the plasma
310 membrane, while α-PCDHs do not (21, 22, 26, 34, 45). We also found that while removing EC6
311 from PCDHα4 enabled its surface localization, replacing α4's EC6 or EC5-6 with β- or γ-PCDH
312 domains imparted no additional increase in surface trafficking. Our measurements further showed
313 varying trafficking levels among wildtype and engineered cPCDH isoforms, enabling analyses that
314 tested the relationship between cPCDH *cis* dimerization and surface delivery.

315 *Cis* dimerization is essential for cPCDH function in self- versus non-self recognition (54), and
316 previous studies investigating why α-PCDHs do not traffic to the cell membrane in heterologous
317 cells concluded that *cis* dimerization is required for cPCDH surface localization (21, 26, 28). Two
318 key observations supported this conclusion: (1) co-expressing an α-PCDH with a β- or γ-PCDH
319 enabled surface localization of the α-PCDH (21, 26, 28, 54), and (2) dimer-disrupting mutations
320 in γ-PCDHs, including mutations introducing α-specific residues, prevented their surface

321 trafficking (26, 28). We were therefore surprised to find that incorporating γ-PCDH *cis* interface
322 residues into an α-PCDH was insufficient to enable surface delivery. We also found that the *cis*
323 dimerization properties of our constructs—EC6 identity for domain-swapped α-PCDH variants
324 and binding affinity for wildtype isoforms—were not correlated with surface delivery levels. As a
325 whole, while previous work indicated that *cis* dimerization is necessary for cPCDH surface
326 delivery (21, 26, 28, 54), our data suggest that at least for α-PCDHs, *cis* dimerization is not
327 sufficient for surface trafficking, and that *cis* dimerization is not the sole determinant of cPCDH
328 export levels. Interestingly, this shifted interpretation re-opens the question of whether α-PCDHs
329 can form *cis* homodimers in cells, as the conclusion that they cannot has been based on their surface
330 trafficking behavior in non-neuronal cells.

331 Previous research with α4ΔEC6 in K562 cells led to the conclusion that the α-PCDH EC6 domain
332 inhibits surface delivery (21). In replicating these findings in 2963T-ΔNC cells, we showed that
333 inhibition by the α-PCDH EC6 domain was consistent across cell types. Our finding that replacing
334 EC6 in α4 with EC6 from a β-, γ-, or C-type cPCDH did not improve surface delivery compared
335 to α4ΔEC6 also substantiates that idea, and it emphasizes that the lack of the α-PCDH EC6 matters
336 more for the surface delivery of these variants than particular features of β- and γ-PCDH EC6
337 domains. Furthermore, our α4-γB6 *cis* interface chimeras showed that the *cis* interface is not the
338 α-PCDH EC6 region that inhibits surface trafficking, and our affinity-versus-surface-delivery
339 analysis suggests that *cis* dimerization is not the sole determinant for β- and γ-PCDH export.

340 Investigations into the cellular mechanisms of cPCDH trafficking suggest potential alternative
341 models for how surface delivery might be controlled independently of *cis* dimerization. The
342 cPCDH EC domains are O-mannosylated, an unusual type of glycosylation in metazoans catalyzed
343 by four glycosyltransferases dedicated to cadherin superfamily proteins, including cPCDHs (55).

344 As protein glycosylation and folding are tightly linked quality control mechanisms in the secretory
345 pathway, differences in glycosylation site motifs or glycosyltransferase interactions could
346 determine the trafficking efficiency of a particular cPCDH isoform. Although we focused on
347 extracellular interaction interfaces here, a motif in the variable intracellular region of γ-PCDHs
348 has been shown to also regulate their trafficking in a ubiquitinylation- and phosphorylation-
349 dependent manner in 293T cells (38, 39, 43, 44). Finally, correlative light and electron microscopy
350 studies have identified specialized intracellular compartments that appear when γ-PCDHs, but not
351 α-PCDHs, are overexpressed in 293T cells, raising the possibility that α- and γ-PCDHs may be
352 trafficked by distinct intracellular pathways (39, 41). Quantification of surface delivery with the
353 tools presented here will facilitate future investigations of the complex cellular regulation of
354 cPCDH surface delivery in both heterologous cells and neurons.

355 In the cPCDH field, cell aggregation assays have played a crucial role in deciphering the molecular
356 rules of cPCDH-mediated cell recognition (20-22, 26, 28, 54). Perhaps most importantly, sorting
357 occurs in cell aggregation assays, with cells preferring to bind others with completely matching
358 sets of cPCDH isoforms, the same scenario that leads to self-recognition and avoidance in neurons
359 (21, 22, 54). Notably, Wiseglass and colleagues recently demonstrated that *cis* dimerization is
360 essential for cPCDH-mediated sorting in cell aggregation assays (54). In these experiments, the
361 extent of cell aggregation and sorting depended on relative isoform expression levels (20, 21, 54).
362 Direct measurements of surface trafficking could provide insight into the relative surface
363 concentrations of different isoforms in co-transfection experiments and distinguish how each
364 isoform contributes to the energetics of competitive binding.

365 A limitation of studying membrane trafficking regulation in heterologous cells is that regulation
366 may differ in the native context. Overexpression may overwhelm aspects of the secretory pathway,

367 which could obscure which regulatory features are crucial for export. Of our three surface
368 trafficking metrics, the mean Ab_{cell} signal is highly dependent on expression level, while the
369 threshold-based stained cell fraction metric, f^{+}, is less so. Both of these metrics should be evaluated
370 cautiously in contexts where expression levels vary, such as in transient transfections or when
371 drawing comparisons between different proteins. In contrast, the $E_{surface}$ metric directly accounts
372 for expression levels by measuring how strongly the amount of surface protein depends on the
373 amount of protein expressed, making it independent of expression level.

374 Beyond cPCDHs, our method provides a general-purpose, quantitative way to screen plasma
375 membrane proteins and variants of interest for surface delivery or secretion via either microscopy
376 or flow cytometry. For attached cells, tissues, and adhesion protein studies, microscopy is
377 preferable to avoid disrupting surface proteins when detaching cells from the culture vessel and
378 each other. For suspended cells, flow cytometry has the advantage of being extremely high-
379 throughput and therefore can detect rare events. Our surface delivery summary metrics are
380 compatible with either approach. Cell surface proteins including ion channels, G-protein coupled
381 receptors, and transporters represent highly important drug targets, and their binding pockets,
382 gating and regulatory mechanisms are often mapped through structure determination and
383 mutational scans (56-63). Membrane protein instability and mis-localization are common in
384 heterologous expression systems (61, 64), and few screening studies validate loss-of-function hits
385 by evaluating surface delivery (57, 59, 60, 63). Implementing quantitative surface delivery analysis
386 in combination with activity screens will facilitate clearer interpretation of these experiments and
387 improve investigations of surface protein functions.

388

389 **Materials and Methods**

390 **Differential conservation sequence analysis**

391 The sequences of all mouse cPCDHs were manually gathered from UniProt (release 2023_05)
392 (65), aligned using MUSCLE v.5.1 (66), and truncated to EC6 (Geneious Prime 2023.1.2
393 [http://www.geneious.com]). The subsequent alignment manipulations were performed in Python
394 3 using bioviper v.0.20 (67). The alignment was separated into two sets of sequences: α-PCDHs,
395 αC1, and γC4 in one set, and β-PCDHs, γ-PCDHs, αC2, γC3, and γC5 in the other set. For each
396 sub-alignment, the amino acid frequencies were calculated and the frequency distributions at each
397 position were compared between the two alignments using the entropy function in scipy v.1.11.4
398 (68) to calculate the Kullback-Leibler divergence, which we refer to as the KL_{div} score.

399 **Molecular cloning**

400 The wildtype-like constructs used in this study are illustrated in **Figure S1A**. The myc-tagged α4,
401 α4ΔEC6, β17, αC2, and γC3 plasmids, and the untagged γC3 plasmid, were provided by Tom
402 Maniatis's lab (21). The α4 and α4ΔEC6 proteins, and all α4 variants used in this study, lacked the
403 C-terminal intracellular region. The β17 construct lacked its C-terminal 25 intracellular residues,
404 and αC2, γC3, and untagged γC3 were full length. The γA3, γA11, γB2, and γB6 constructs lacked
405 the C-terminal intracellular region. All derived plasmids were cloned into the pmax-mCherry
406 backbone of the α4 plasmid using Gibson assembly (Gibson Assembly Master Mix, NEB E2611S).
407 For derived constructs including the myc tag, the tag was positioned in the stalk region between
408 EC6 and the transmembrane helix (**Figure S1A**), equivalently to previously validated constructs
409 (21). Inserts were generated by DNA synthesis (IDT gBlocks or eBlocks) or PCR (Q5 polymerase,
410 NEB M0491S). The complete plasmid sequence for α4 and the full coding region sequences for
411 all other plasmids are provided in .fasta format on Zenodo
412 (https://doi.org/10.5281/zenodo.13345292). For all plasmids generated, DNA from single clones

413 was purified (E.Z.N.A. Plasmid DNA Mini Kit I, Omega Bio-Tek D6942) and sequenced (Whole
414 Plasmid Sequencing, Plasmidsaurus) to verify the integrity of the constructs.

415 **293T-ΔNC cell culture and transfection**

416 The 293T N-cadherin knock-out cell line (293T-ΔNC) (52) was provided by Joshua Sanes's lab.
417 Cells were cultured in the adherent state in a 1:1 mixture of Dulbecco's Modified Eagle's Medium
418 and Ham's F-12 medium containing L-glutamine and without phenol red (DMEM/F-12; Corning,
419 16-405-CV) supplemented with 10% fetal bovine serum (FBS; Corning, 35-011-CV) and 1X non-
420 essential amino acids (NEAA; Lonza, 13-114E). Cells were maintained in exponential growth by
421 passaging every 3-4 days. A day before transfection, 293T-ΔNC cells were seeded at 100,000
422 cells/well in a total volume of 400 µL medium in glass-bottom culture chambers (Cellvis, C8-1.5-
423 N). The cells adhered to the culture chambers overnight in an incubator maintained at 37°C with
424 5% CO_2.

425 Stocks of DNA for 293T-ΔNC transfections were prepared by diluting plasmids to 50 ng/µL with
426 DNA elution buffer (10 mM Tris-HCl pH 8.5; Omega Bio-Tek, D6942-02) so that each transfection
427 reaction mix contained the same amount of DNA and salts. Transfection was carried out using
428 LipoFectMax 3000 Transfection Reagent (ABP Biosciences, FP318). For each sample, one well
429 of cells was transfected with 0.3 µg DNA in 25 µL total volume of transfection mixture containing
430 a 1:3:2 mass ratio of DNA, Component A, and Component B in Opti-MEM I Reduced Serum
431 Medium (Opti-MEM; Gibco, 31985-70) per transfection reaction. The transfection reaction
432 mixtures were incubated at room temperature for 10-15 minutes, then added directly to the medium
433 of the cells in the culture chambers. Transfection proceeded overnight in a 37°C, 5% CO_2
434 incubator. The seeded 293T-ΔNC cells remained attached to the glass substrate throughout the
435 transfection process.

436 **Sample preparation and staining**

437 Live cells were stained 20-24 hours post-transfection, and the 293T-ΔNC cells remained adhered
438 to the glass substrate throughout staining and imaging. A FITC-conjugated antibody recognizing
439 the Myc epitope tag (Miltenyi Biotec, 130-116-485) was diluted 1:50 in fresh medium. Three
440 hundred and fifty microliters of culture medium were removed from each well and replaced with
441 50 µL diluted antibody solution (final antibody working concentration: 1:100 dilution). Unstained
442 samples were treated identically, except that the spent culture medium was replaced with 50 µL
443 fresh medium instead of antibody solution. Samples were incubated at 37°C, 5% CO_2 for one hour,
444 then 300 µL fresh medium were added to each well to dilute the antibody. Two sets of three washes
445 were subsequently performed, with a 30-minute incubation at 37°C, 5% CO_2 between the two sets
446 of washes. For each wash, 200 µL medium were removed from each well and replaced with 200
447 µL fresh medium (antibody diluted 2X per wash; final antibody dilution 256X after all washes).
448 Cells were imaged live immediately following the final wash.

449 **Microscopy**

450 Samples were imaged using a Zeiss LSM 880 laser scanning confocal microscope using a Plan-
451 Apochromat 10x/0.45 NA air objective. The sample was brought into focus and the pinhole was
452 set to correspond to a 16 µm optical section, or approximately the height of a cell. Three fields of
453 view were manually defined using only the brightfield channel for each sample well, and the fields
454 of view were acquired sequentially using automated stage positioning with hardware autofocus to
455 maintain the same focal plane at each position. Images were acquired in bidirectional scanning
456 mode with a 1 µsec pixel dwell time and 4x line-by-line averaging. The two fluorescence channels
457 were acquired sequentially on a PMT detector, and the brightfield channel was acquired in
458 transmission mode. For the transfection marker mCherry (FP) channel, we used a 561 nm

459 excitation laser and collected emitted light of wavelengths 578-696 nm. For the FITC-conjugated
460 Myc antibody (Ab) channel, we used a 488 nm excitation laser and collected emitted light of
461 wavelengths 493-592 nm.

462 **Image processing and quantification**

463 Image processing was performed in Python 3 using czifile v.2019.7.2 to load the images and
464 metadata. Cells were segmented in the brightfield channel with cellpose v.2.2.3 (69) using a GPU
465 on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing
466 Group at Harvard University. For both fluorescence channels, the background was estimated for
467 each individual image as the median intensity of all non-cell pixels and subtracted. FP_{cell} and Ab_{cell}
468 were calculated with the single_cell.average function in microutil v.0.4.0 (70). Cells containing
469 any saturated pixels were excluded from downstream analysis. The transfection threshold was set
470 as the 0.995^{th} quantile of the pooled FP_{cell} values from the mock-transfected samples in all
471 biological replicate experiments. Thus, the probability that a cell that we designated as transfected
472 (having an FP_{cell} value greater than the transfection threshold) was actually not transfected was
473 less than 0.005. The surface stain threshold was calculated similarly, as the 0.995^{th} quantile of
474 pooled Ab_{cell} values from samples transfected with untagged γC3 or untagged γA11. Histograms
475 were calculated and displayed after applying a logicle transform (71) using FlowKit v.1.0.1 (72)
476 and seaborn v.0.13.0 (73). For each sample, f^+ was calculated as the fraction of transfected cells
477 with Ab_{cell} values higher than the surface stain threshold, and $\overline{Ab}$ was calculated as the average of
478 all transfected cell Ab_{cell} values. To calculate $E_{surface}$, log-spaced bins were defined using the
479 minimum and maximum FP_{cell} values for all transfected cells (across biological replicate
480 experiments and samples). Cells were assigned to bins based on their FP_{cell} values, and FP_{bin} and
481 Ab_{bin} were then calculated individually for each sample as the mean FP_{cell} and Ab_{cell} values per

482 bin, respectively. $E_{surface}$ was calculated as the slope of the least-squares linear regression of the
483 log of FP_{bin} and Ab_{bin} using the linregress function in scipy (68), excluding bins with negative Ab_{bin}
484 values (this only occurred for a few bins in untagged isoform samples). Of note, the data provide
485 single-cell measurements for FP_{cell} and Ab_{cell}, so the linear regression and slope can be computed
486 either with or without binning. We chose to use bins as they reduce the contribution of the many
487 transfected cells with very low expression levels. P-values were calculated using the statistical
488 tests indicated in the text or figure legends.

489 **K562 cell culture, electroporation, and staining**

490 K562 cells were provided by Daniel A. Fletcher's lab and the UC Berkeley Cell Culture Facility
491 and were maintained in suspension between 1×10^5 and 1×10^6 cells/mL in Roswell Park Memorial
492 Institute 1640 medium (RPMI; Corning 10-040-CV) supplemented with 10% FBS and 1 mM
493 sodium pyruvate (Gibco 11360070). Twenty µL of cells at 1×10^7 cells/mL were electroporated
494 with two µL of DNA at 500–1000 ng/µL concentration with two 10-ms width, 1450 V pulses using
495 a Neon electroporation system and a 10 µL Neon tip (Invitrogen MPK1025). After electroporation,
496 cells recovered for 18-24 hours in complete medium at 37°C with 5% CO_2. For staining, cells were
497 centrifuged at 500 × g for 5 minutes and resuspended in complete medium containing the anti-
498 Myc antibody diluted 1:100. After incubating for one hour at 37°C, 5% CO_2, the cells were
499 centrifuged and resuspended in complete medium without antibody three times to wash. Cells were
500 analyzed on an Attune CytPix Flow Cytometer, and surface delivery quantification was performed
501 as described for imaging data.

502
503 **Acknowledgements**

The authors thank Barry Honig, Tom Maniatis, and Erin Flaherty for supplying plasmids, Josh Sanes for supplying 293T-ΔNC cells, Daniel A. Fletcher for supplying K562 cells, and John Russell, José Velilla, and Sam Berry for helpful feedback on the manuscript. We also thank Brenda Chiang and Nhu Dang for early work on the project and the Harvard Center for Biological Imaging (RRID:SCR_018673) for infrastructure and support. R.G. acknowledges funding from a Harvard Brain Science Initiative Bipolar Disorder Seed Grant and NIH grant R01GM120996. E.J.M. thanks the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard, award number #1764269 and the Harvard Quantitative Biology Initiative, Harvard Physics of Living Systems, the Aramont Fund for Emerging Science Research, and the Simmons family for generous funding.

Data availability:

Microscopy datasets, plasmid sequences, alignment files, and raw data files containing values for the figures in the manuscript are available on Zenodo (https://doi.org/10.5281/zenodo.13345292). Python scripts are available on GitHub (https://github.com/emay2022/surface-trafficking). Plasmids are available from the authors upon reasonable request.

References

1. K. Sano *et al.*, Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J* **12**, 2249-2256 (1993).
2. N. Kohmura *et al.*, Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* **20**, 1137-1151 (1998).
3. S. Obata *et al.*, A common protocadherin tail: multiple protocadherins share the same sequence in their cytoplasmic domains and are expressed in different regions of brain. *Cell Adhes Commun* **6**, 323-333 (1998).
4. Q. Wu, T. Maniatis, A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**, 779-790 (1999).
5. X. Wang *et al.*, Gamma protocadherins are required for survival of spinal interneurons. *Neuron* **36**, 843-854 (2002).

533 6. W. R. Mancia Leon *et al.*, Clustered gamma-protocadherins regulate cortical interneuron
534 programmed cell death. *Elife* **9**, e55374 (2020).
535 7. J. A. Weiner, X. Wang, J. C. Tapia, J. R. Sanes, Gamma protocadherins are required for
536 synaptic development in the spinal cord. *Proc Natl Acad Sci U S A* **102**, 8-14 (2005).
537 8. J. L. Lefebvre, Y. Zhang, M. Meister, X. Wang, J. R. Sanes, gamma-Protocadherins regulate
538 neuronal survival but are dispensable for circuit formation in retina. *Development* **135**,
539 4141-4151 (2008).
540 9. J. L. Lefebvre, D. Kostadinov, W. V. Chen, T. Maniatis, J. R. Sanes, Protocadherins mediate
541 dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517-521 (2012).
542 10. M. J. Molumby, A. B. Keeler, J. A. Weiner, Homophilic Protocadherin Cell-Cell
543 Interactions Promote Dendrite Complexity. *Cell Rep* **15**, 1037-1050 (2016).
544 11. G. Mountoufaris *et al.*, Multicluster Pcdh diversity is required for mouse olfactory neural
545 circuit assembly. *Science* **356**, 411-414 (2017).
546 12. D. M. Steffen *et al.*, A Unique Role for Protocadherin gammaC3 in Promoting Dendrite
547 Arborization through an Axin1-Dependent Mechanism. *J Neurosci* **43**, 918-935 (2023).
548 13. A. M. Garrett, D. Schreiner, M. A. Lobas, J. A. Weiner, gamma-protocadherins control
549 cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling
550 pathway. *Neuron* **74**, 269-276 (2012).
551 14. A. B. Keeler, D. Schreiner, J. A. Weiner, Protein Kinase C Phosphorylation of a gamma-
552 Protocadherin C-terminal Lipid Binding Domain Regulates Focal Adhesion Kinase
553 Inhibition and Dendrite Arborization. *J Biol Chem* **290**, 20674-20686 (2015).
554 15. H. Sugino *et al.*, Genomic organization of the family of CNR cadherin genes in mice and
555 humans. *Genomics* **63**, 75-87 (2000).
556 16. Q. Wu *et al.*, Comparative DNA sequence analysis of mouse and human protocadherin
557 gene clusters. *Genome Res* **11**, 389-404 (2001).
558 17. Q. Wu, T. Maniatis, Large exons encoding multiple ectodomains are a characteristic feature
559 of protocadherin genes. *Proc Natl Acad Sci U S A* **97**, 3124-3129 (2000).
560 18. B. Tasic *et al.*, Promoter choice determines splice site selection in protocadherin alpha and
561 gamma pre-mRNA splicing. *Mol Cell* **10**, 21-33 (2002).
562 19. S. Obata *et al.*, Protocadherin Pcdh2 shows properties similar to, but distinct from, those
563 of classical cadherins. *J Cell Sci* **108 (Pt 12)**, 3765-3773 (1995).
564 20. D. Schreiner, J. A. Weiner, Combinatorial homophilic interaction between gamma-
565 protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc*
566 *Natl Acad Sci U S A* **107**, 14893-14898 (2010).
567 21. C. A. Thu *et al.*, Single-cell identity generated by combinatorial homophilic interactions
568 between alpha, beta, and gamma protocadherins. *Cell* **158**, 1045-1059 (2014).
569 22. R. Rubinstein *et al.*, Molecular logic of neuronal self-recognition through protocadherin
570 domain interactions. *Cell* **163**, 629-642 (2015).
571 23. J. M. Nicoludis *et al.*, Structure and Sequence Analyses of Clustered Protocadherins Reveal
572 Antiparallel Interactions that Mediate Homophilic Specificity. *Structure* **23**, 2087-2098
573 (2015).
574 24. J. M. Nicoludis *et al.*, Antiparallel protocadherin homodimers use distinct affinity- and
575 specificity-mediating regions in cadherin repeats 1-4. *Elife* **5** (2016).
576 25. J. M. Nicoludis *et al.*, Interaction specificity of clustered protocadherins inferred from
577 sequence covariation and structural analysis. *Proc Natl Acad Sci U S A* **116**, 17825-17830
578 (2019).

26. K. M. Goodman *et al.*, gamma-Protocadherin structural diversity and functional implications. *Elife* **5** (2016).
27. K. M. Goodman *et al.*, Structural Basis of Diverse Homophilic Recognition by Clustered alpha- and beta-Protocadherins. *Neuron* **90**, 709-723 (2016).
28. K. M. Goodman *et al.*, Protocadherin cis-dimer architecture and recognition unit diversity. *Proc Natl Acad Sci U S A* **114**, E9829-E9837 (2017).
29. K. M. Goodman *et al.*, How clustered protocadherin binding specificity is tuned for neuronal self-/nonself-recognition. *Elife* **11**, e72416 (2022).
30. J. Brasch *et al.*, Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **569**, 280-283 (2019).
31. M. Fernandez-Monreal, S. Kang, G. R. Phillips, Gamma-protocadherin homophilic interaction and intracellular trafficking is controlled by the cytoplasmic domain in neurons. *Mol Cell Neurosci* **40**, 344-353 (2009).
32. N. Boni, L. Shapiro, B. Honig, Y. Wu, R. Rubinstein, On the formation of ordered protein assemblies in cell-cell interfaces. *Proc Natl Acad Sci U S A* **119**, e2206175119 (2022).
33. G. R. Phillips *et al.*, Gamma-protocadherins are targeted to subsets of synapses and intracellular organelles in neurons. *J Neurosci* **23**, 5096-5104 (2003).
34. Y. Murata, S. Hamada, H. Morishita, T. Mutoh, T. Yagi, Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. *J Biol Chem* **279**, 49508-49516 (2004).
35. M. Frank *et al.*, Differential expression of individual gamma-protocadherins during mouse brain development. *Mol Cell Neurosci* **29**, 603-616 (2005).
36. S. Kallenbach *et al.*, Changes in subcellular distribution of protocadherin gamma proteins accompany maturation of spinal neurons. *J Neurosci Res* **72**, 549-556 (2003).
37. M. Fernandez-Monreal *et al.*, gamma-protocadherins are enriched and transported in specialized vesicles associated with the secretory pathway in neurons. *Eur J Neurosci* **32**, 921-931 (2010).
38. R. O'Leary *et al.*, A variable cytoplasmic domain segment is necessary for gamma-protocadherin trafficking and tubulation in the endosome/lysosome pathway. *Mol Biol Cell* **22**, 4362-4372 (2011).
39. G. R. Phillips, N. LaMassa, Y. M. Nie, Clustered protocadherin trafficking. *Semin Cell Dev Biol* **69**, 131-139 (2017).
40. R. Rubinstein, K. M. Goodman, T. Maniatis, L. Shapiro, B. Honig, Structural origins of clustered protocadherin-mediated neuronal barcoding. *Semin Cell Dev Biol* **69**, 140-150 (2017).
41. H. H. Hanson *et al.*, LC3-dependent intracellular membrane tubules induced by gamma-protocadherins A3 and B2: a role for intraluminal interactions. *J Biol Chem* **285**, 20982-20992 (2010).
42. S. S. Schalm, B. A. Ballif, S. M. Buchanan, G. R. Phillips, T. Maniatis, Phosphorylation of protocadherin proteins by the receptor tyrosine kinase Ret. *Proc Natl Acad Sci U S A* **107**, 13894-13899 (2010).
43. A. Shonubi, C. Roman, G. R. Phillips, The clustered protocadherin endolysosomal trafficking motif mediates cytoplasmic association. *BMC Cell Biol* **16**, 28 (2015).
44. A. Ptashnik *et al.*, Ubiquitination of the protocadherin-gammaA3 variable cytoplasmic domain modulates cell-cell interaction. *Front Cell Dev Biol* **11**, 1261048 (2023).

624 45. T. Mutoh, S. Hamada, K. Senzaki, Y. Murata, T. Yagi, Cadherin-related neuronal receptor
625 1 (CNR1) has cell adhesion activity with beta1 integrin mediated through the RGD site of
626 CNR1. *Exp Cell Res* **294**, 494-508 (2004).
627 46. H. Sago *et al.*, Cloning, expression, and chromosomal localization of a novel cadherin-
628 related protein, protocadherin-3. *Genomics* **29**, 631-640 (1995).
629 47. M. Blank, G. B. Triana-Baltzer, C. S. Richards, D. K. Berg, Alpha-protocadherins are
630 presynaptic and axonal in nicotinic pathways. *Mol Cell Neurosci* **26**, 530-543 (2004).
631 48. G. B. Triana-Baltzer, M. Blank, Cytoplasmic domain of protocadherin-alpha enhances
632 homophilic interactions and recognizes cytoskeletal elements. *J Neurobiol* **66**, 393-407
633 (2006).
634 49. S. Bonn, P. H. Seeburg, M. K. Schwarz, Combinatorial expression of alpha- and gamma-
635 protocadherins alters their presenilin-dependent processing. *Mol Cell Biol* **27**, 4121-4132
636 (2007).
637 50. I. G. Haas, M. Frank, N. Veron, R. Kemler, Presenilin-dependent processing and nuclear
638 function of gamma-protocadherins. *J Biol Chem* **280**, 9313-9319 (2005).
639 51. B. Hambsch, V. Grinevich, P. H. Seeburg, M. K. Schwarz, gamma-Protocadherins,
640 presenilin-mediated release of C-terminal fragment promotes locus expression. *J Biol*
641 *Chem* **280**, 15888-15897 (2005).
642 52. K. M. Goodman *et al.*, Molecular basis of sidekick-mediated cell-cell adhesion and
643 specificity. *Elife* **5** (2016).
644 53. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
645 583-589 (2021).
646 54. G. Wiseglass, N. Boni, K. Smorodinsky-Atias, R. Rubinstein, Clustered protocadherin cis-
647 interactions are required for combinatorial cell-cell recognition underlying neuronal self-
648 avoidance. *Proc Natl Acad Sci U S A* **121**, e2319829121 (2024).
649 55. I. S. B. Larsen *et al.*, Discovery of an O-mannosylation pathway selectively serving
650 cadherins and protocadherins. *Proc Natl Acad Sci U S A* **114**, 11163-11168 (2017).
651 56. C. B. Billesbolle *et al.*, Structural basis of odorant recognition by a human odorant receptor.
652 *Nature* **615**, 742-749 (2023).
653 57. H. J. Ellis *et al.*, Deep Mutagenesis of a Transporter for Uptake of a Non-Native Substrate
654 Identifies Conformationally Dynamic Regions [Preprint]. *bioRxiv*
655 10.1101/2021.04.19.440442, 2021.2004.2019.440442 (2024).
656 58. A. M. Glazer *et al.*, Deep Mutational Scan of an SCN5A Voltage Sensor. *Circ Genom*
657 *Precis Med* **13**, e002786 (2020).
658 59. J. D. Heredia *et al.*, Mapping Interaction Sites on Human Chemokine Receptors by Deep
659 Mutational Scanning. *J Immunol* **200**, 3825-3839 (2018).
660 60. W. Coyote-Maestas, D. Nedrud, Y. He, D. Schmidt, Determinants of trafficking,
661 conduction, and disease within a K(+) channel revealed through multiparametric deep
662 mutational scanning. *eLlife* **11** (2022).
663 61. S. W. Yee *et al.*, The full spectrum of SLC22 OCT1 mutations illuminates the bridge
664 between drug transporter biophysics and pharmacogenomics. *Mol Cell* **84**, 1932-1947
665 e1910 (2024).
666 62. E. M. Jones *et al.*, Structural and functional characterization of G protein-coupled receptors
667 with deep mutational scanning. *Elife* **9** (2020).
668 63. M. K. Howard *et al.*, Molecular basis of proton-sensing by G protein-coupled receptors
669 [Preprint]. *bioRxiv* 10.1101/2024.04.17.590000 (2024).

64. K. Ikegami *et al.*, Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proc Natl Acad Sci U S A* **117**, 2957-2967 (2020).
65. C. UniProt, UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023).
66. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
67. S. Berry, S. Ganesh (2023) bioviper. (Zenodo, https://zenodo.org/doi/10.5281/zenodo.12549380).
68. P. Virtanen *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020).
69. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100-106 (2021).
70. J. Russell, I. Hunt-Isaak, D. Hekstra (2021) microutil. (GitHub, https://github.com/Hekstra-Lab/microutil).
71. W. A. Moore, D. R. Parks, Update for the logicle data scale including operational code implementations. *Cytometry A* **81**, 273-277 (2012).
72. S. White *et al.*, FlowKit: A Python Toolkit for Integrated Manual and Automated Cytometry Analysis Workflows. *Front Immunol* **12**, 768541 (2021).
73. M. L. Waskom, seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021).

Main Text Figures and Legends

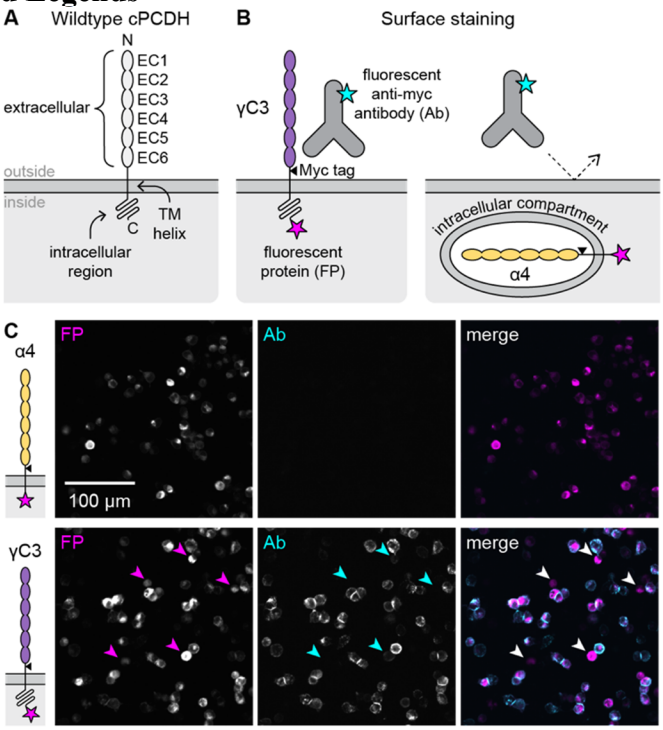

Figure 1. Surface staining by extracellular epitope detection of cPCDHs in 293T-ΔNC cells is the basis of surface delivery quantification. (A) Schematic showing the domain arrangement and topology of a generic cPCDH isoform. From N to C terminus: six extracellular cadherin (EC) domains are tethered to the membrane by a transmembrane helix (TM) followed by an unstructured intracellular region. (B) Diagram of experimental approach. cPCDH isoforms modified to contain an extracellular Myc epitope tag (triangle) and a C-terminal fluorescent protein (magenta star) are expressed in 293T-ΔNC cells. A fluorescent (cyan star) anti-Myc antibody recognizes the Myc tag to detect proteins that reach the cell surface, like γC3. The antibody does not bind proteins that are retained intracellularly, like α4, because it does not cross the plasma membrane. (C) Microscopy images of surface-stained control samples. Top row: An α-PCDH isoform, α4, is produced in transfected 293-ΔNC cells (signal in FP channel) but does not reach the cell surface (no signal in Ab channel). Bottom row: A C-type isoform, γC3, is produced in transfected 293T-ΔNC cells (signal in FP channel), and the anti-Myc antibody stains the periphery of cells with plasma membrane-localized protein (Ab channel). Arrowheads in γC3 images indicate examples where transfected cells have little to no Ab staining, highlighting the heterogeneity of surface trafficking.

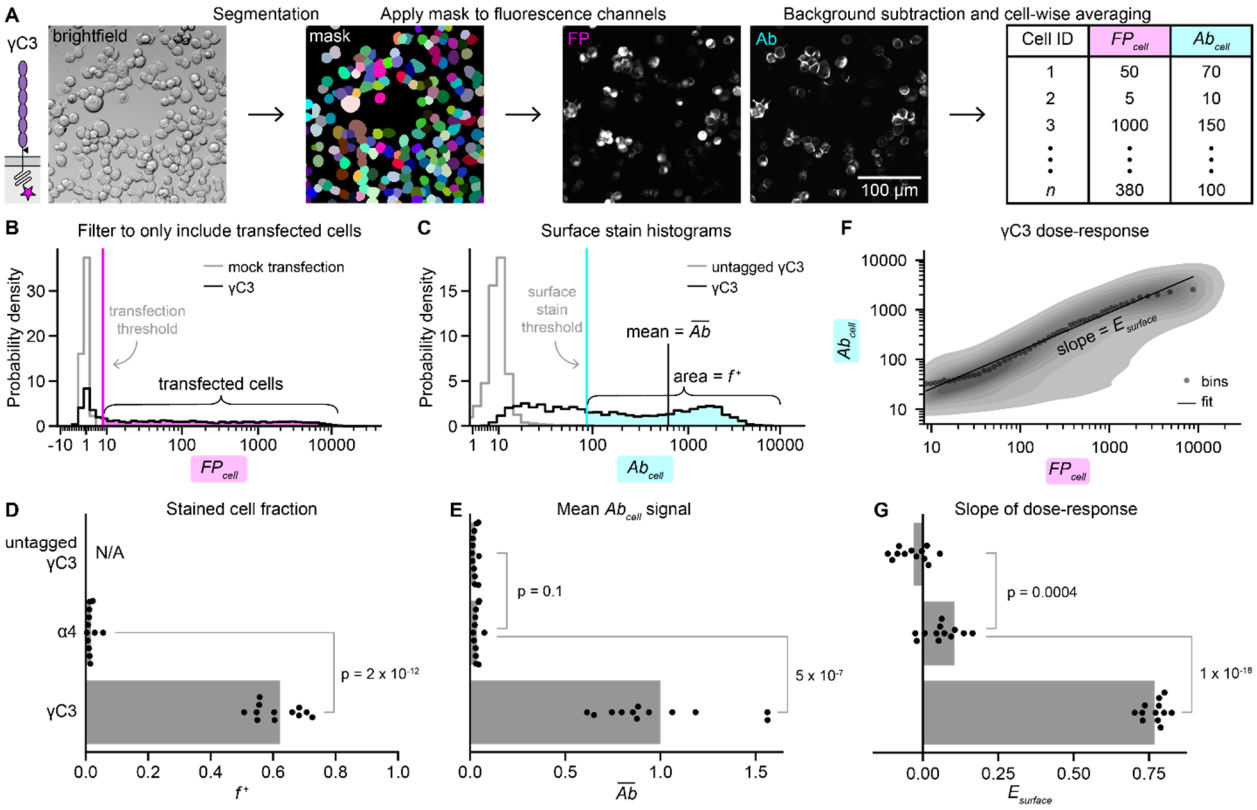

Figure 2. Quantification reveals the heterogeneity of surface trafficking across cells. (A) Initial image processing steps for an example sample, γC3. Automated cell segmentation using the brightfield channel from a particular field of view produces a mask that assigns each pixel in the field of view to a cell (or background). To perform background subtraction, we applied the mask to the FP and Ab images from the same field of view to assign background pixels, and we subtracted the average background pixel values across the image. Cell-wise averaging of background-subtracted pixel intensities produces FP_{cell} and Ab_{cell} values for each of n cells in the field of view. (B) A histogram of FP_{cell} values, where FP_{cell} is a measure of cellular protein levels of the transfected cPCDH. The data are filtered to only include transfected cells (magenta highlight) using a transfection threshold (magenta vertical line) determined using a sample of mock-transfected cells. (C) A histogram of the filtered Ab_{cell} values, where Ab_{cell} a measure of surface abundance of the transfected cPCDH. The mean of the entire Ab_{cell} distribution, $\overline{Ab}$ (black vertical line), corresponds to the average surface stain signal for all transfected cells in the sample. The fraction of Ab-stained cells, f^{+} (cyan highlight), is calculated using a threshold Ab_{cell} value (cyan vertical line) determined using cells transfected with an untagged version of γC3. (D-E) Graphs of f^{+} (D) and $\overline{Ab}$ (E) for untagged γC3, α4, and γC3. In panel E, $\overline{Ab}$ values are normalized to the mean $\overline{Ab}$ of Myc-tagged γC3. (F) Dose-response plot of the γC3 data. The contour plot displays the density of FP_{cell} and Ab_{cell} values for each cell. The data were binned by FP_{cell} and the means of each bin are shown (gray points). The bins were scaled to maintain the same number of cells in each bin, which resulted in approximately log-scaled bins. We define $E_{surface}$ as the slope of a linear fit to the data (black line). (G) Graph of $E_{surface}$ for untagged γC3, α4, and γC3. In panels D, E, and G, points are values for biological replicates (n = 12), and bars show the mean values. P-values were calculated using a one-tailed Welch's t-test.

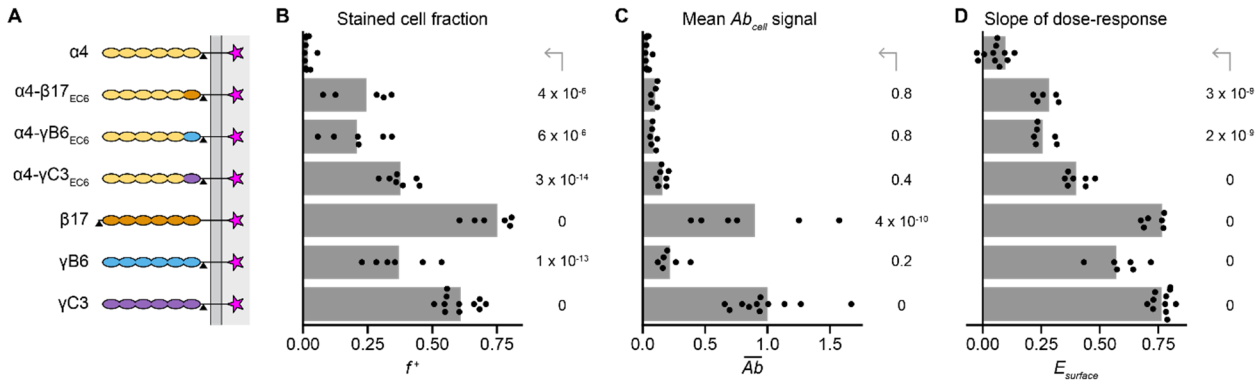

Figure 3. Quantification can distinguish samples with intermediate surface delivery phenotypes. (A) Schematics of tested constructs; domains are color-coded by originating isoform. (B-D) Quantification of f^{+} (B), $\overline{Ab}$ (C), and $E_{surface}$ (D) for the constructs shown in A. In B, $\overline{Ab}$ values are normalized to the mean $\overline{Ab}$ of the positive control construct, Myc-tagged γC3. Points are biological replicates (n = 11 for α4 and γC3; 5 for α4-β17$_{EC6}$; 6 for α4-γB6$_{EC6}$, β17, and γB6; and 7 for α4-γC3$_{EC6}$), and bars are mean values across replicates. P-values for each sample compared to α4 (top row) were calculated using a one-tailed Dunnett's test.

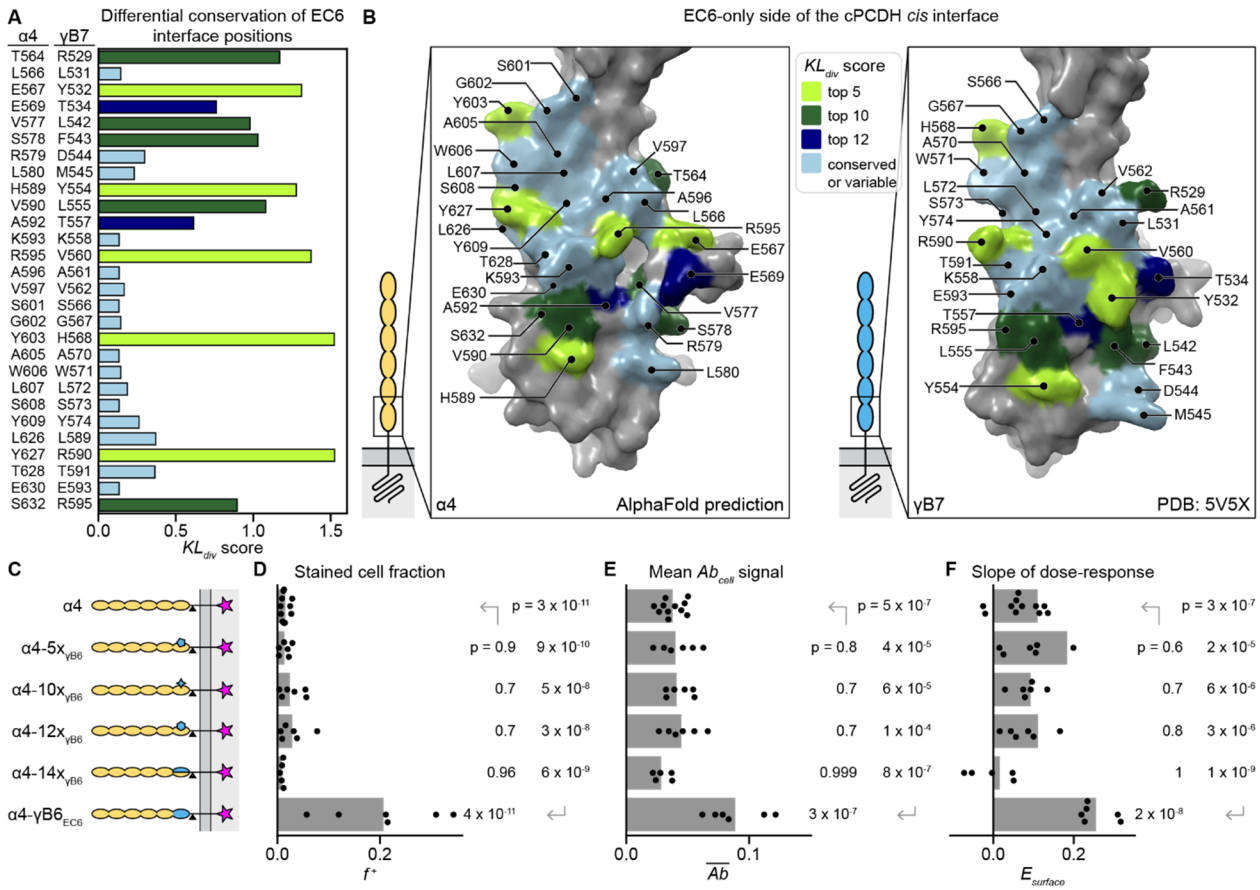

Figure 4. The EC6 *cis* interface of γB6 is insufficient for surface delivery of α4. (A) Differential conservation for all EC6 *cis* interface positions quantified using Kullback-Leibler divergence (KL_{div}) score, which compares the conservation of a position among isoforms that traffic to the cell surface to that position's conservation among isoforms that do not reach the surface. The color of the bar indicates how positions were grouped by KL_{div} score for chimeric constructs (see legend in panel B). (B) Surface representations of the EC6 *cis* interface for α4 (left; structure prediction from AlphaFold Protein Structure Database [accession code: AF-O88689-F1-v4] (53)) and γB7 (right; experimentally-determined structure [PDB: 5V5X] (28)). *Cis* interface residues are colored according to their KL_{div} score categories (legend top middle). (C) Schematics of the chimeric α4 constructs quantified in panels D-F. (D-F) Quantification of f^{+} (D), $\overline{Ab}$ (E), and $E_{surface}$ (F) for the constructs shown in panel C. In E, $\overline{Ab}$ values are normalized to the mean $\overline{Ab}$ of the positive control construct γC3 (not shown in the figure). Points are biological replicates (n = 11 for α4; 5 for α4-14x$_{\gamma B6}$; and 6 for α4-5x$_{\gamma B6}$, α4-10x$_{\gamma B6}$, α4-12x$_{\gamma B6}$, and α4-γB6$_{EC6}$), and bars are replicate means. P-values for each sample compared to α4 were calculated using a one-tailed Dunnett's test. P-values for each sample compared to α4-γB6$_{EC6}$ were calculated using a two-tailed Dunnett's test.

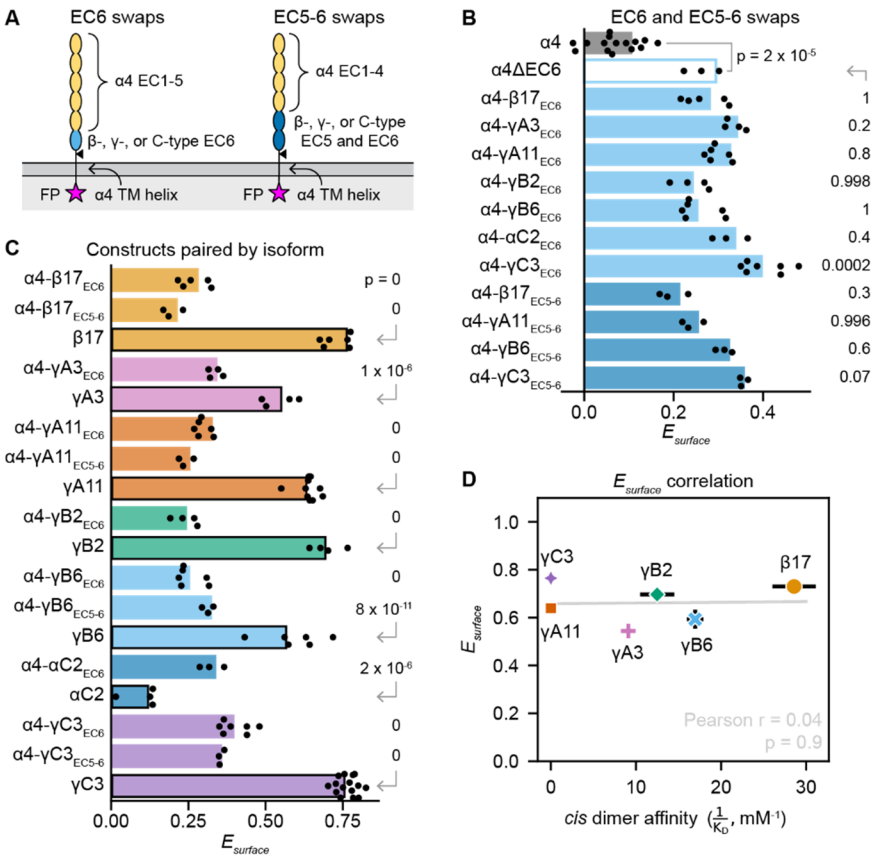

Figure 5. Removing or replacing EC6 of α4 increases its surface delivery, but not to wildtype β- and γ-PCDH levels. (A) Schematics of the types of constructs compared in B-D. EC6-swapped variants have EC1-5 and the transmembrane helix of α4, and EC6 of a different isoform. EC5-6 swaps have EC1-4 and the transmembrane helix of α4, and EC6 of a different isoform. (B) Graph comparing $E_{surface}$ for wildtype α4 (gray bar) to α4 variants with EC6 deleted (unfilled light blue bar), EC6 swapped (lighter blue bars), or EC5-6 swapped (darker blue bars). Points are biological replicates (n = 15 for α4 and γC3; 3 for α4ΔEC6, α4-αC2$_{EC6}$, α4-β17$_{EC5-6}$, α4-γA11$_{EC5-6}$, α4-γB6$_{EC5-6}$, and γC3$_{EC5-6}$; 4 for α4-γA3$_{EC6}$ and α4-γB2$_{EC6}$; 5 for α4-β17$_{EC6}$; 6 for α4-γA11$_{EC6}$ and α4-γB6$_{EC6}$; and 7 for α4-γC3$_{EC6}$), and bars are replicate means. The p-value for α4ΔEC6 compared to α4 was calculated using a one-tailed Welch's t-test; p-values for domain-swapped samples compared to α4ΔEC6 were calculated using a two-tailed Dunnett's test. (C) Graph of $E_{surface}$ comparing constructs grouped by isoform, with bars color-coded by isoform. Bars without borders are α4 variants (EC6 or EC5-6 swaps), while bars with black borders are wildtype cPCDHs. Points are biological replicates (n = 15 for α4 and γC3; 3 for α4-αC2$_{EC6}$, α4-β17$_{EC5-6}$, α4-γA11$_{EC5-6}$, γB6, α4-γB6$_{EC5-6}$, and γC3$_{EC5-6}$; 4 for αC2, γA3, α4-γA3$_{EC6}$, γB2, and α4-γB2$_{EC6}$; 5 for α4-β17$_{EC6}$; 6 for β17, α4-γA11$_{EC6}$, γB6, and α4-γB6$_{EC6}$; 7 for α4-γC3$_{EC6}$; and 10 for γA11), and bars are replicate means. P-values for each α4 variant compared to the corresponding wildtype isoform were calculated using Tukey's test. (D) Scatter plot of wildtype cPCDH *cis* dimerization affinity versus surface delivery, as measured by $E_{surface}$. See Figure S6F for affinity values used in this plot. Points are shape- and color-coded by isoform. The Pearson r statistic and its associated p-value reflect the degree of correlation.