Directional alignment of different cell types organizes planar cell polarity

Masaki Arata[1,*], Hiroshi Koyama[1,2], and Toshihiko Fujimori[1,2,*].

[1] Division of Embryology, National Institute for Basic Biology, Okazaki, Aichi, Japan.

[2] SOKENDAI (The Graduate University for Advanced Studies), Okazaki, Aichi, Japan.

Corresponding author: Masaki Arata and Toshihiko Fujimori

Email: fujimori@nibb.ac.jp (T.F.), arata@nibb.ac.jp (M.A.)

This PDF file includes:

Main Text
Figures 1 to 6

Abstract

Planar cell polarity (PCP) refers to unidirectional coordination of cell polarity in the plane of epithelial tissues, whose proper formation is essential for development in various tissues. While imbalances in the amount of core PCP proteins (core proteins) between adjacent cells disrupt PCP in some tissues, other tissues maintain PCP despite containing multiple cell types with different levels of core proteins. How such tissues tolerate these imbalances remains unclear. Here, we theoretically analyzed the contribution of spatial distribution of different cell types to PCP maintenance. We adopted a previously established PCP model on two-dimensional cell sheets, where local protein-protein interactions were assumed. Our systematic simulations revealed that the patterns of cell-type distribution significantly affect the accuracy of PCP maintenance under the condition of imbalanced PCP proteins. To identify the critical patterns, we applied both deep learning techniques and statistical modeling to the simulation data. Consequently, these analyses revealed that orientation of cell-type alignment is a key aspect of cell-type distribution that affects PCP. Such directional cell-type alignment was observed in the mouse oviduct. Our findings highlight the overlooked contribution of spatial distribution of cell types to PCP maintenance in tissues with multiple cell-types.

Main Text

Introduction

The proper function of various organs depends on planar cell polarity (PCP), the tissue-level coordination of cell polarity within the plane of epithelial tissue (1–8). For example, the transport of oocytes is driven by the ovary-to-uterus-directed beating of cilia in the oviduct epithelium (9–12), and misoriented hair cells in the cochlea are associated with hearing deficit (13).

Previous studies have identified an evolutionally conserved group of proteins, known as core PCP proteins (hereafter referred to as core proteins), that play key roles in PCP establishment (1–7). Core proteins form two types of complexes at cell edges (Fig. 1A): one comprised of CELSR and FZD (FRIZZLED), and another comprised of CELSR and VANGL (Van Gogh-like; hereafter referred to as the FZD complex and the VANGL complex, respectively). In various developing organs, the FZD and VANGL complexes show asymmetric distributions at cell edges along body axes (1, 4, 8). For example, in the mouse oviduct, the FZD complex is localized to the uterus side, while the VANGL complex is localized to the ovary side of the cell (10). This asymmetry is established before ciliogenesis, and the loss of CELSR1 results in misoriented ciliary beating (9), suggesting that the asymmetry is a prerequisite for PCP establishment. Although details remain unclear, experimental and theoretical studies suggest that the establishment of the asymmetry

requires intracellular and intercellular interactions between the FZD and VANGL complexes (14–17).

Epithelial tissues are typically composed of multiple cell types. In such cases, how is PCP formed or maintained? The mouse oviduct epithelium consists of two cell types: multiciliated cells (MCCs) and secretory cells (SCCs, Fig. 1A). Notably, these two cell types differ in their levels of core proteins; CELSR1 is enriched at the cell edges of MCCs but is present at lower levels in SCCs (Fig 1B-B", (9)). This imbalance in core protein levels between adjacent cells (hereafter referred to as core protein imbalance) has the potential to disrupt PCP. Indeed, when cells lacking Frizzled (Fz, a fly homolog of FZD), Van Gogh (Vang, a fly homolog of VANGLs), or Flamingo (Fmi, also known as Starry night, a fly homolog of CELSR) are experimentally produced in the *Drosophila* wing, polarities in surrounding wild-type cells are reorganized (6, 7, 18). Nevertheless, in the oviduct, both the ciliary beating orientation and core protein distribution remain biased along the ovary-uterus axis (Fig. S1B-B"). Such core protein imbalances have also been observed in various organs, including the mouse inner ear (19), airway (20), and *Drosophila* eye disc (21). However, the mechanisms by which tissues overcome core protein imbalance to maintain PCP remain unclear.

Mathematical models have been widely used to dissect the mechanisms of PCP establishment and maintenance. Several types of PCP models exist. Mechanistic models explicitly simulate the dynamics of molecules relevant to individual core proteins or their complexes (14, 15, 22–25). On the other hand, phenomenological models abstract these molecular details and represent cell polarity as a vector that evolves according to effective parameters, such as neighbor interactions, global cues, or cell geometry (26–28). These models have been used to investigate core protein functions by performing *in silico* perturbation experiments in which clusters of mutant cells lacking core proteins are embedded within a wild-type cell sheet. These studies primarily focused on elucidating the molecular functions of core proteins, rather than studying PCP establishment and maintenance in tissues of multiple cell-types with various spatial distributions. Therefore, these studies considered just idealized configurations of mutant cell clusters: a linear cell row, a rectangular cluster (14, 15, 22). Consequently, it remains largely unclear how various spatial patterns of cell types with differing core protein levels—beyond a simple mutant cluster—affect PCP in normal tissues.

In this study, to consider the core protein imbalances, we adopted a mechanistic mathematical model in which concentrations of the FZD and VANG complexes are explicitly represented. We systematically varied the spatial patterns of core protein levels in a two-dimensional cell sheet and analyzed the resulting PCP outcomes. To identify key spatial features that promote PCP maintenance, we combined simulations with deep learning, and statistical modeling. Our analysis

revealed that the directional alignment of cell types, specifically, the alignment of cell types along the tissue axis, contributes to preservation of PCP in the presence of core protein imbalance. In consistent with this theoretical finding, we found that SCCs are predominantly aligned along the ovary–uterus axis in the mouse oviduct, and PCP defects emerge in regions where this alignment is disrupted. Our study offers crucial insights into how the spatial patterning of multiple cell types influences PCP establishment and maintenance.

Results

Modeling PCP in a 2D cell sheet

To examine how distributions of cell types with different core protein levels affect PCP, we constructed a mathematical model (see also Supporting Information), which builds upon previous frameworks (15, 23). In our model, we explicitly implemented concentrations of the FZD and VANGL complexes at cell edges and within the cytoplasmic region (Fig. 1B). This allows us to directly address the effects of intercellular imbalances in core protein levels—an aspect not captured by phenomenological vector-based models (26–28). The epithelial sheet is represented as an array of regular hexagonal cells (Fig. 1B-C"), where the FZD complex and the VANGL complex shuttle between cell edges and the cytoplasmic region (Fig. 1B, blue arrows)

$$F_{\mathrm{C}_k} \underset{R_{\mathrm{F}_{ki}}}{\overset{K_{\mathrm{F}_{ki}}}{\rightleftarrows}} F_{\mathrm{M}_{ki}} \tag{1}$$

$$V_{\mathrm{C}_k} \underset{R_{\mathrm{V}_{ki}}}{\overset{K_{\mathrm{V}_{ki}}}{\rightleftarrows}} V_{\mathrm{M}_{ki}} \tag{2}$$

Here, F_{C_k} and V_{C_k} denote the concentration of the FZD and VANGL complexes in the cytoplasmic region of *k*th cell, respectively. The concentration of the FZD and VANGL complexes at *i* th cell edge in the *k*th cell are represented as $F_{\mathrm{M}_{ki}}$ and $V_{\mathrm{M}_{ki}}$, respectively; $i = 1{\sim}6$. In addition, concentrations of the FZD and VANGL complexes were assumed to be uniform within each cell edge and cytoplasmic region. The FZD complex shuttles between cell edge i and a cytoplasmic region at rates $K_{\mathrm{F}_{ki}}$ and $R_{\mathrm{F}_{ki}}$, while the Vangl complex undergoes similar transport at rates $K_{\mathrm{V}_{ki}}$ and $R_{V_{ki}}$. Therefore, the dynamics of $F_{\mathrm{M}_{ki}}$ is described as $dF_{\mathrm{M}_{ki}}/dt = -R_{\mathrm{F}_{ki}}F_{\mathrm{M}_{ki}} + K_{\mathrm{F}_{ki}}F_{\mathrm{C}_{ki}}$; the differential equations for other components are shown in Supporting Information.

According to previous studies, to establish the planar polarized distributions of the two complexes on cell edges, the formation of intercellular bridges composed of the two complexes (the FZD-VANGL bridges) was required, where the FZD-VANGL bridges have activities of both autoactivation and mutual inhibition (Fig. 1A'; (14–17)). We basically adopted the theoretical formulation in these studies to model the above shuttling rates as follows:

$$K_{\mathrm{F}_{ki}} = \rho_{\mathrm{F}} + \omega_{\mathrm{F}}\left(F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}} + d_{\mathrm{A_F}}\sum_{j=i-1,i+1} F_{\mathrm{M}_{kj}}V'_{\mathrm{M}_{kj}}\right) \tag{3}$$

$$R_{\mathrm{F}_{ki}} = \mu_{\mathrm{F}} + \varphi_{\mathrm{F}}\left(V_{\mathrm{M}_{ki}}F'_{\mathrm{M}_{ki}} + d_{\mathrm{I_V}}\sum_{j=i-1,i+1} V_{\mathrm{M}_{kj}}F'_{\mathrm{M}_{kj}}\right) \tag{4}$$

$$K_{\mathrm{V}_{ki}} = \rho_{\mathrm{V}} + \omega_{\mathrm{V}}\left(V_{\mathrm{M}_{ki}}F'_{\mathrm{M}_{ki}} + d_{\mathrm{A_V}}\sum_{j=i-1,i+1} V_{\mathrm{M}_{kj}}F'_{\mathrm{M}_{kj}}\right) \tag{5}$$

$$R_{V_{ki}} = \mu_{\mathrm{V}} + \varphi_{\mathrm{V}}\left(F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}} + d_{\mathrm{I_F}}\sum_{j=i-1,i+1} F_{\mathrm{M}_{kj}}V'_{\mathrm{M}_{kj}}\right) \tag{6}$$

where concentrations of the FZD and VANGL complexes on a neighboring cell's edge juxtaposing edge i are denoted as $F'_{\mathrm{M}_{ki}}$ and $V'_{\mathrm{M}_{ki}}$,respectively (Fig. 1B). The intercellular bridges are formed between the FZD and VANGL complexes: $F_{\mathrm{M}_{ki}} - V'_{\mathrm{M}_{ki}}$ and $V_{\mathrm{M}_{ki}} - F'_{\mathrm{M}_{ki}}$. For simplicity, we assumed that the concentrations of the FZD-VANGL bridges were proportional to the terms of $F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}}$ and $F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}}$ in the above equations (15). The first term in the parenthesis of eq.3 and 5 correspond to the autoactivation activity: $K_{\mathrm{F}_{ki}}$and $K_{\mathrm{V}_{ki}}$ increase when $F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}}$ and $V_{\mathrm{M}_{ki}}F'_{\mathrm{M}_{ki}}$ are elevated, respectively. The first term in the parenthesis of eq.4 and 6 corresponds to the mutual inhibition activity: $R_{\mathrm{F}_{ki}}$and $R_{\mathrm{V}_{ki}}$ increase when $V_{\mathrm{M}_{ki}}F'_{\mathrm{M}_{ki}}$ and $F_{\mathrm{M}_{ki}}V'_{\mathrm{M}_{ki}}$ are elevated, respectively. In addition, this model assumed a non-local effect, where the shuttling rates on cell edge i was affected by core proteins on neighboring edges $i-1$ and $i+1$ (Fig. S2A and A'; the second term in the parenthesis of eq.3-6; see Supporting Information for details). A constant d determines the strength of the non-local effects (Fig. S2A and A'). When d_{A} > 0, the complexes on edge $i-1$ and $i+1$ elevate $K_{\mathrm{F}_{ki}}$and $K_{\mathrm{V}_{ki}}$. Similarly, when d_{I} > 0, the complexes on edge $i-1$ and $i+1$ elevate $R_{\mathrm{F}_{ki}}$and $R_{\mathrm{V}_{ki}}$. Although the non-local effect has not been experimentally validated, theoretical studies suggest it is necessary for robust PCP establishment (15, 29).

The parameter ρ represents the basal internalization rate of the FZD and VANGL complexes from the cell edge to the cytoplasm, while μ denotes the basal trafficking rate of the FZD and VANGL

complexes from the cytoplasm to the cell edge; both parameters were set to be constant. The values of parameters used in our model are listed in Table S1. We selected these values to reproduce the non-autonomous effects of mutant cells lacking core proteins, as observed *in vivo* (Fig. S3; see details in the Supporting Information).

Setting of simulation and definition of polarity index

Before showing systematic analyses of simulations, we explain the setting of simulations. First, we prepared a cell sheet in which PCP was already established (Fig. 1C, left panel). Second, core protein levels were reduced in selected cells (low cells; gray cells in Fig. 1C; remaining cells are referred to as high cells in the following). The magnitude of the core protein imbalance was defined as a constant $balance_{\mathrm{LH}}$, which ranges from 0 to 1. When $balance_{\mathrm{LH}} = 1$, low and high cells contain equal amounts of core proteins, whereas, when $balance_{\mathrm{LH}} = 0$, low cells have no core proteins. Then, we performed simulations to analyze how the distribution of core proteins was reorganized. Fig. 1C is an example in the case that low cells ($balance_{\mathrm{LH}} = 0.5$) are randomly distributed. At the initial state, distributions of core proteins at cell edges were unidirectionally biased, with the FZD and VANGL complexes enriched at right and left ends of each cell, respectively (Fig. 1C and S2B). In our data, we represent each cell's polarity vector, which indicates an orientation of FZD complex-enriched cell edges, with an orange-headed bar (Fig. 1C", see details in Supporting Information). The bar's length corresponds to the polarity magnitude, and its angle corresponds to the polarity direction. $\overline{\Delta angle_{\mathrm{high}}}$ is the mean absolute change in polarity vector angles of high cells in a cell sheet. Since the value of $\overline{\Delta angle_{\mathrm{high}}}$ is 0 at the initial state of the simulation, the value of $\overline{\Delta angle_{\mathrm{high}}}$ at the final state represents the extent of polarity reorganization. In Fig. 1C, $\overline{\Delta angle_{\mathrm{high}}}$ is 36.0 degrees at the final state, indicating that the core protein imbalance disrupts PCP.

Systematic simulations under various cell type distribution

To gain insights into which aspects of cell-type distributions influence PCP, we generated cell-type distributions by randomly placing low cells, expecting these distributions to encompass diverse features of cell-type distributions. Through our simulations, we evaluated their effects on PCP by calculating $\overline{\Delta angle_{\mathrm{high}}}$ and searched for features of cell-type distributions that significantly impact PCP.

We began by generating random distributions with low-cell frequencies ($freq_{\mathrm{low}}$) ranging from 0.05 to 0.95 in increments of 0.05. At $freq_{\mathrm{low}} = 0$, all cells are high cells, whereas at $freq_{\mathrm{low}} = 1$, all

167 cells are low cells. For each value of $freq_{\mathrm{low}}$, we generated 10 random distributions and performed
168 simulations (Fig. 2A-B). Fig. 2A-A" shows examples of our simulations at $balance_{\mathrm{LH}} = 0.5$. $freq_{\mathrm{low}}$
169 significantly influenced the values of $\overline{\Delta angle_{\mathrm{high}}}$ (Fig. 2B; see the boxed column for these values
170 at $balance_{\mathrm{LH}} = 0.5$), indicating that $freq_{\mathrm{low}}$ is an essential parameter influencing PCP. Additionally,
171 changes in $balance_{\mathrm{LH}}$ significantly altered the value of $\overline{\Delta angle_{\mathrm{high}}}$ (Fig. 2B and S4A-B'; e.g.,
172 compare Fig S4A with S4A'). We further noted that the influence of low cells on $\overline{\Delta angle_{\mathrm{high}}}$ was
173 particularly evident when $balance_{\mathrm{LH}}$ was below 0.6 (Fig. 2B). This indicates that PCP is affected by
174 low cells even when core proteins are only partially lost in these cells.

175 Although $freq_{\mathrm{low}}$ and $balance_{\mathrm{LH}}$ appeared to be essential parameters influencing $\overline{\Delta angle_{\mathrm{high}}}$,
176 they might not be the only determinants of $\overline{\Delta angle_{\mathrm{high}}}$. For example, among the 10 simulations
177 under $freq_{\mathrm{low}} = 0.4$ and $balance_{\mathrm{LH}}$ = 0.5 with different cell-type distributions, the maximum and the
178 minimum values of $\overline{\Delta angle_{\mathrm{high}}}$ were 36.4 and 16.3 degrees, respectively (variance: 40.5 degrees;
179 Fig. S4B). Variances of $\overline{\Delta angle_{\mathrm{high}}}$ at different values of $freq_{\mathrm{low}}$ and $balance_{\mathrm{LH}}$ are summarized in
180 Fig. S4C. We expected that such a variance can be explained by differences in features of cell-
181 type distribution between cell sheets. One candidate is the orientation of the alignment of low and
182 high cells. As highlighted by boxes in Fig. 2A-A', polarity vectors of high cells were strongly
183 reoriented where these cells were aligned along the Y-axis of the cell sheets. In addition, we noticed
184 that when high cells were dominant in a cell sheet, polarity vectors of high cells were reoriented
185 only in the vicinity of low cells (see a boxed region in Fig. 2A"), suggesting that differences in the
186 frequency of edges between high and low cells underlie the variations in $\overline{\Delta angle_{\mathrm{high}}}$ between cell
187 sheets. This is consistent with a previous theoretical study, which report that that core proteins
188 were eliminated from cell edges between Fmi (Flamingo, a fly homolog of CELSRs)-deficient cells
189 and wild type cells (22). We analyzed effects of these parameters on the values of $\overline{\Delta angle_{\mathrm{high}}}$ in
190 later sections.

191

192 **Cell type distribution is effective on PCP maintenance**

193 To examine whether features of cell-type distribution can explain differences in values of $\overline{\Delta angle_{\mathrm{high}}}$
194 between cell sheets, we set out to train a neural network to predict the extent of PCP
195 disorganization in a cell sheet using only information on cell-type distributions (Fig. 3A). If the
196 network was successfully trained, it indicates that the cell-type distribution is a crucial factor in
197 explaining the value of $\overline{\Delta angle_{\mathrm{high}}}$ in a cell sheet. Otherwise, it implies that the cell-type distribution
198 has little impact.

To this end, we first generated a large number of simulation data with $freq_{\mathrm{low}} = 0.4$ and $balance_{\mathrm{LH}} = 0.5$, as we anticipated that this parameter set would produce a wide range of $\overline{\Delta angle_{\mathrm{high}}}$ values. A total of 6,000 random cell-type distributions were generated, and those in the top 12.5% and bottom 12.5% of $\overline{\Delta angle_{\mathrm{high}}}$ were labeled as “top” and “bottom” distributions, respectively (Fig. 3A and B). In addition to $\overline{\Delta angle_{\mathrm{high}}}$, we calculated the averaged angle of polarity vectors over low cells ($\overline{\Delta angle_{\mathrm{low}}}$) and all cells ($\overline{\Delta angle_{\mathrm{all}}}$). However, using $\overline{\Delta angle_{\mathrm{low}}}$ or $\overline{\Delta angle_{\mathrm{all}}}$ for labeling did not significantly alter the classification of each distribution as “top” or “bottom”, since these values were closely correlated with $\overline{\Delta angle_{\mathrm{high}}}$ (Fig. S4D and E).

Then, a ResNet18-based neural network (30) was trained to categorize distributions as either "top" or "bottom" (Fig. 3A). Among the 750 "top" distributions and 750 "bottom" distributions, 100 from each group were used for validation, while the remaining distributions were used for training. After training, the model achieved a classification accuracy of 0.925 (Fig. 3C), indicating that $\overline{\Delta angle_{\mathrm{high}}}$ is determined by cell-type distributions.

Deep learning-based exploration of features of cell type distributions that affect PCP

To find what kind of spatial cell-type distributions is critical for PCP maintenance, we used an explainable artificial intelligence technique, Grad-CAM (31), by which we can visualize regions of a cell sheet where the network focused its attention to make predictions (Fig. 3A). We expected that this technique would allow us to identify key features of cell-type distributions that influence $\overline{\Delta angle_{\mathrm{high}}}$. Notably, in distributions predicted as “top,” Grad-CAM highlighted regions where low cells were aligned along the Y-axis of the cell sheet (Fig. 3D). In contrast, in distributions predicted as “bottom,” Grad-CAM highlighted regions where low cells were aligned along the X-axis of the cell sheet (Fig. 3D’). To quantify these patterns, we measured the total lengths of horizontal and vertical alignment lengths of low cells adjacent to high cells (Fig. 3E). For each pair of adjacent low cells whose shared vertex was also shared by a high cell, we measured the alignment angle θ between the two low cells (Fig. 3E). Here, θ is defined as the angle between the line connecting the centroids of the two low cells and the X-axis of the cell sheet (Fig. 3E). We then calculated the total low-cell alignment as follows:

$$L_{\mathrm{h}} = \sum_{\text{cell sheet}} |\cos\theta|$$

$$L_{\mathrm{v}} = \sum_{\text{cell sheet}} |\sin\theta|$$

where L_{h} and L_{v} are total horizontal and vertical alignment lengths of low cells adjacent to high cells, respectively. Hereafter, we refer to alignment of low cells adjacent to high cells simply as "low-cell alignment."

The values of vertical alignment (L_{V}) of "top" distributions were significantly higher than those of "bottom" distributions (p<0.01; Student's *t*-test; Fig. 3F). In contrast, the values of horizontal alignment (L_{H}) of "top" distributions were significantly lower than those of "bottom" distributions (p<0.01; Student's *t*-test; Fig. 3F). As discussed earlier, we expected that the frequency of edges between high cells and low cells ($freq_{\mathrm{edge}}$) would influence $\overline{\Delta angle_{\mathrm{high}}}$. Although the difference was subtle, we found that the value of $freq_{\mathrm{edge}}$ was higher in "top" distributions than in "bottom" distributions (p<0.05; Student's *t*-test; Fig. 3F'). As discussed later, the contribution of this parameter to $\overline{\Delta angle_{\mathrm{high}}}$ is not negligible.

GLM-based evaluation of factors affecting PCP

Finally, to dissect effects of orientations of low-cell alignment and other aspects of cell-type distributions on PCP, we constructed a generalized linear model (GLM), a framework for modeling the relationship between an object variable and explanatory variables. Here, the object variable was $\overline{\Delta angle_{\mathrm{high}}}$ and candidates for explanatory variables included L_{h}, L_{v}, $balance_{\mathrm{LH}}$, $freq_{\mathrm{low}}$, and $freq_{\mathrm{edge}}$. In short, by using GLM, we can expect to evaluate to what extent each explanatory variable contributes to the $\overline{\Delta angle_{\mathrm{high}}}$. Based on the value of AIC (Akaike Information Criterion), we selected the model that best predicted $\overline{\Delta angle_{\mathrm{high}}}$ in our simulation data (Fig. 4). Data with $balance_{\mathrm{LH}}$ >0.6 were excluded from the GLM analysis, as $\overline{\Delta angle_{\mathrm{high}}}$ was weakly affected by cell distributions when the $balance_{\mathrm{LH}}$ was high. In addition, since L_{h} and L_{v} were strongly correlated (i.e., positive correlation) in the dataset, their direct inclusion in the model was problematic. To address this, we converted them into a single variable, $L_{\mathrm{h/v}} = L_{\mathrm{h}}/L_{\mathrm{v}}$, which represents the anisotropy of low-cell alignment. The optimal model selected based on the value of AIC was as follows:

$$\log \overline{\Delta angle_{\mathrm{high}}} = -0.834 + 0.615 \cdot freq'_{\mathrm{low}} - 0.0390 \cdot L'_{\mathrm{h/v}} + 0.327 \cdot freq'_{\mathrm{edge}} - 0.355 \cdot balance'_{\mathrm{LH}}$$

where values of explanatory variables were normalized so that their mean and SD were 0 and 1, respectively. Normalized variables are marked with an apostrophe (e.g., $freq'_{\mathrm{low}}$). The values of $\overline{\Delta angle_{\mathrm{high}}}$ predicted by this model were closely correlated with actual values of $\overline{\Delta angle_{\mathrm{high}}}$ (Fig. 4A; Pearson correlation coefficient: 0.859). However, the model's accuracy decreased when the

frequency of low cells was high, particularly at $freq_{\mathrm{low}}$= 0.95 (yellow dots in Fig. 4A'). Here, the absolute value of each coefficients indicates the strength of the influence of explanatory variables on $\overline{\Delta angle_{\mathrm{high}}}$. The coefficients indicate that low cell frequencies ($freq_{\mathrm{low}}$) and magnitude of core protein imbalances ($balance_{\mathrm{LH}}$) have a significant impact on $\overline{\Delta angle_{\mathrm{high}}}$, as expected. In addition, the model suggests that the frequency of edges between low and high cells ($freq_{\mathrm{edge}}$) is a key parameter. Although the absolute value of the coefficient for the anisotropy of low-cell alignment ($L'_{\mathrm{h/v}}$) was smaller than those of the other explanatory variables, a model excluding $L_{\mathrm{h/v}}$ was less optimal according to AIC. Since the coefficient of $L_{\mathrm{h/v}}$ is negative, the model suggests that an increase in the ratio of total lengths of horizontal to vertical alignment of low cells decreases $\overline{\Delta angle_{\mathrm{high}}}$. These results indicate that the frequency of low cells ($freq_{\mathrm{low}}$), the frequency of edges between high and low cells ($freq_{edge}$), and the anisotropy of low-cell alignment ($L_{\mathrm{h/v}}$) are key aspects of cell-type distributions that influence PCP.

The direction of low-cell alignment is essential for mitigating the effect of low cells on PCP

To further address whether the direction of low-cell alignment affects PCP, we placed rectangular low-cell clusters within a cell sheet and analyzed the degree of PCP reorganization after simulations. First, low-cell clusters were aligned in a lattice at intervals of 3 cells and $balance_{\mathrm{LH}}$ was set to 0 (i.e., no core proteins in the low cells). We then examined the effects of low-cell clusters with varying values of l_{h} and l_{v} (Fig. 5A-A'). Here, l_{h} and l_{v} represent the lengths of the cluster in the horizontal and vertical directions, respectively, relative to the X-axis, which corresponds to the initial direction of PCP. Our analysis revealed that when low-cell clusters were elongated perpendicular to the X-axis, the value of $\overline{\Delta angle_{\mathrm{high}}}$ increased (Fig. 5A, compare $l_{\mathrm{h}} = 3$, $l_{\mathrm{v}} = 5$ with $l_{\mathrm{h}} = 3$, $l_{\mathrm{v}} = 3$). In contrast, when low-cell clusters were elongated parallel to the X-axis, the values of $\overline{\Delta angle_{\mathrm{high}}}$ did not increase (Fig. 5A, compare $l_{\mathrm{h}} = 5$, $l_{\mathrm{v}} = 3$ with $l_{\mathrm{h}} = 3$, $l_{\mathrm{v}} = 3$).

Next, we introduced stochastic shifts from the lattice points (Fig. 5B-B'). The parameter $\sigma_{\max}$ determines the magnitude of this shift. When $\sigma_{\max} = 1$, low cell clusters could be displaced by up to one cell along both the x- and y-axis. Fig. 3B shows that elongation of low-cell clusters along the Y-axis increased the mean $\overline{\Delta angle_{\mathrm{high}}}$, while elongation of low-cell clusters along the X-axis did not ($l_{\mathrm{h}} = 3$, $l_{\mathrm{v}} = 5$ vs. $l_{\mathrm{h}} = 5$, $l_{\mathrm{v}} = 3$). We obtained similar results under the conditions of various parameter values, $\sigma_{\max}$ (1~3), l_{h}, l_{v}, and $balance_{\mathrm{LH}}$ (Fig. 5C). When low cell clusters were regularly aligned, misorientation of polarity vectors was observed only in the immediate vicinity of low cell clusters (Fig. 5A). In contrast, when clusters were shifted, misorientation of polarity vectors was not

limited to the vicinity of low cell clusters (Fig. 5B, middle panel), suggesting that shifting positions of clusters disrupts PCP.

These findings support our hypothesis that the direction of low-cell alignment plays a key role in mitigating the disruptive effects of low cells on PCP. However, in our analysis, modifying the lengths of low-cell clusters inevitably altered other aspects of cell-type distributions. For instance, changes in l_{h} and l_{v} also affected the frequency of low cells. To dissect effects of low-cell alignment and other aspects of cell-type distributions on PCP, we constructed a GLM. Here, the object variable was $\overline{\Delta angle_{\mathrm{high}}}$ and candidates for explanatory variables included $balance_{\mathrm{LH}}$, the frequency of low cells ($freq_{\mathrm{low}}$), σ_{max}, l_{h}, and l_{v}. Based on the value of AIC, we selected the model that best predicted $\overline{\Delta angle_{\mathrm{high}}}$ in our simulation data as follows.

$$\log \overline{\Delta angle_{\mathrm{high}}} =$$

$$-0.951 + 0.529 \cdot freq'_{\mathrm{low}} - 0.237 \cdot l'_{\mathrm{h}} + 0.289 \cdot l'_{\mathrm{v}} - 0.118 \cdot balance'_{\mathrm{LH}} + 0.0418 \cdot \sigma'_{\mathrm{max}}$$

where all explanatory variables are normalized. The Pearson correlation coefficient between the actual and predicted values of $\overline{\Delta angle_{\mathrm{high}}}$ was 0.822 (Fig. 5D), indicating that the model accurately described the relationship between the object variable and explanatory variables. Among all coefficients, the absolute value of the coefficient of $freq'_{\mathrm{low}}$ is the largest, indicating that frequency of low cells has the most significant effect on $\overline{\Delta angle_{\mathrm{high}}}$. The model shows that l_{h} and l_{v} are also essential parameters that affect PCP, as their coefficients have the second and third highest absolute values, respectively. As expected, the model shows that an increase in horizontal length (l_{h}) reduces $\overline{\Delta angle_{\mathrm{high}}}$, an increase in vertical length (l_{v}) elevates $\overline{\Delta angle_{\mathrm{high}}}$. In comparison, the absolute value of the coefficient for σ'_{max} is much smaller than those of l'_{h} and l'_{v}. Therefore, the effect of cluster distribution (σ_{max}) on $\overline{\Delta angle_{\mathrm{high}}}$ is less significant than that of cluster lengths (l_{h} and l_{v}). These results were conserved in a different type of cell array (Fig. S2B' and S5). Taken together, we concluded that the directional low-cell alignment is a key aspect of cell-type distributions that mitigates the effects of low cells on PCP.

Orientation of PCP is correlated with the direction of low-cell alignment *in vivo*

To evaluate the relevance of our theoretical findings to living systems, we analyzed the mouse oviduct epithelium, which consists of only two cell types: secretory cells (SCCs) and multiciliated cells (MCCs). Because the frequency and distributions of SCCs are varied across different regions

322 in the oviduct, we simultaneously analyzed both the direction of PCP and the distribution patterns
323 of SCCs. We focused on two regions: infundibulum and ampulla.

324 In the infundibulum of the adult mouse oviduct, previous studies reported that SCCs comprise
325 24% ($freq_{\mathrm{low}}$ = 24%) of the epithelial cells (11 weeks old; 1, 3), where distributions of core proteins
326 are biased along the ovary-uterus axis, and that CELSR1 levels are lower in SCCs than in MCCs
327 (Fig. S1B-B", 1). First, we quantitatively evaluated the imbalance in CELSR1 levels between SCCs
328 and MCCs. To this end, we measured the intensities of CELSR1 immunofluorescent signals at cell
329 edges. We found that CELSR1 signals at edges between MCCs were approximately twice as high
330 as those at edges between SCCs (Fig. S6A-C', $balance_{\mathrm{LH}}$ = 0.5). We also validated this conclusion
331 by using transgenic mice expressing EGFP-fused VANGL2 ubiquitously (Fig. S6D-E'). These
332 observations suggest that SCCs have fewer levels of core proteins than MCCs and do not establish
333 the polarized distribution of core proteins at cell edges.

334 Next, we analyzed the distribution of SCCs in the infundibulum. For this purpose, we used
335 previously-published images in which cell edges and MCCs were fluorescently labeled (Fig. 6A and
336 A'; images were adopted from (11)). In the oviduct epithelium, SCCs form small clusters and these
337 clusters scatter around the epithelium (Fig. 6A and A'). We measured the length of each SCC
338 cluster by placing a bounding rectangle whose horizontal edges were aligned parallel to the ovary-
339 uterus axis (Fig. 6A'). The mean l_{h} was 2.45 cells (Standard deviation, SD: 1.16), which was
340 significantly larger than the mean l_{v}, 1.85 cells (SD: 0.862; $p<0.01$, Wilcoxon rank sum test, Fig.
341 6B and B'). This result indicates that SCCs preferentially align along the direction of PCP in the
342 infundibulum.

343 In the ampulla, another region of the oviduct, MCCs are sparsely distributed (Fig. 6C-C"). Although
344 the distribution of SCCs was varied, we sometimes found regions where SCCs were
345 circumferentially aligned. In these regions, we analyzed the distribution of VANGL2:EGFP in MCCs
346 adjacent to these SCCs, and found that VANGL2:EGFP was enriched at cell edges running along
347 the ovary-uterus axis in adjacent MCCs (white arrowheads in Fig. 6C"). Note that, in regions where
348 SCCs were aligned along the ovary-uterus axis, VANGL2:EGFP in MCCs was enriched at cell
349 edges oriented circumferentially (yellow arrowheads in Fig. 6C"). Quantitative analysis confirmed
350 that VANGL2:EGFP localization in the ampulla was misaligned relative to that in the infundibulum
351 (Fig. 6D-D"). Together with the results in the infundibulum, the direction of PCP in the oviduct was
352 consistent with our theoretical outcomes.

353

354 **Discussion**

355 In the present study, we theoretically investigated how PCP is maintained in tissues comprising
356 multiple cell types, one of which expresses lower level of core PCP proteins, called low cell. Our
357 systematic simulations suggested that spatial distribution of the low cells, as well as their core
358 protein expression levels and their population, affects proper PCP maintenance. Moreover, deep
359 learning–based analyses using Grad-CAM revealed the impact of spatial distribution: PCP is less
360 affected when low cells are aligned parallel to the tissue polarity. Finally, through generalized linear
361 modeling (GLM), we successfully quantified the contribution of spatial distribution to PCP
362 maintenance. Our results showed that the alignment of low cells has significant impact on PCP
363 maintenance, similar to that of expression levels and population size. From these results, we
364 propose that the spatial distribution of the low cells can be an essential factor for maintaining
365 polarity of tissues composed of multiple cell types. Consistent with our theoretical findings, PCP
366 orientations in distinct regions of the mouse oviduct varied in accordance with the alignment of low
367 cells.

368 The establishment of PCP requires interplays of core proteins between adjacent cells. Previous
369 studies showed that experimentally induced low cells disrupt polarity of surrounding wild-type cells
370 (18). Nevertheless, several tissues naturally exhibit imbalances in core protein levels between cells,
371 while still maintaining PCP. Prior theoretical work has examined how mutant cells lacking core
372 proteins influence neighboring wild-type cells, focusing primarily on the molecular mechanisms of
373 PCP establishment. These investigations typically employed simplified spatial configurations, a
374 linear cell row, or a rectangular cluster (14, 15, 22), without explicitly analyzing the effect of cell-
375 type distribution. Consequently, the effect of complex, physiologically relevant spatial
376 arrangements on PCP maintenance has remained largely unexplored. Our theoretical analysis
377 revealed that the impact of low cells on PCP is strongly influenced by their spatial distribution. This
378 suggests that spatial configuration acts as a regulatory factor in PCP maintenance, functioning
379 cooperatively with molecular interactions between core proteins. In other words, cell-type
380 distribution itself can be considered an integral component of the PCP formation system. In *Celsr1*
381 mutant epithelium, anisotropy of SCC alignment was lost (Fig. S7), suggesting that CELSR1 is
382 required for proper low-cell alignment. If CELSR1 directory regulates SCC alignment, there might
383 be a feedback regulation between core proteins and cell-type distributions. However, the
384 mechanisms governing SCC alignment along the tissue axis remain unclear.

385

386 To examine how the degree of core protein imbalance influences PCP, we introduced a parameter,
387 $balance_{\mathrm{LH}}$, which quantifies the relative level of core proteins in low versus high cells. This allowed
388 us to move beyond binary mutant/wild-type modeling and analyze graded imbalances in protein

389 levels. While our study focused on a biologically relevant case in which both FZD and VANGL
390 complex levels are equally reduced in low cells, our model framework is flexible and can
391 accommodate other combinations of imbalance, for example, unequal reductions in the two
392 complexes or mixed populations of different low-cell types that might be found *in vivo*. A full
393 exploration of these parameter spaces would require extensive simulations and is beyond the
394 scope of the current study. However, our approach can be readily extended to such scenarios in
395 future work.

396 Because our mathematical model is deterministic, the outcome of each simulation is determined
397 by the initial configuration of cell types. This feature allowed us to apply deep learning techniques
398 to systematically identify spatial patterns that influence PCP maintenance. Specifically, we used a
399 neural network classifier to screen diverse configurations, and subsequently employed generalized
400 linear modeling (GLM) to evaluate the contributions of interpretable spatial features, such as
401 alignment of low cells, to polarity outcomes. Although the main focus of this study is the relationship
402 between cell-type distribution and PCP maintenance, this hybrid computational framework may be
403 broadly applicable to other problems in tissue patterning. For example, it could be adapted to
404 investigate the effects of spatial heterogeneity in signaling molecule levels beyond core proteins,
405 offering a general strategy for bridging complex spatial input patterns with emergent tissue-level
406 patterning.

407 While our study focused on cell-type distribution, we do not exclude the possibility that additional
408 mechanisms help mitigate the effects of core protein imbalance on PCP. For example, in the mouse
409 airway, the frequency of MCCs (0.37 on average; (32)) is much lower than in the oviduct (0.76 on
410 average; (9)). If non-ciliated cells in the airway are low cells, our model suggests that PCP
411 maintenance would not be feasible in the airway epithelium. Several studies have reported that
412 extrinsic mechanical cues, such as fluid flow can align cilia beating orientation in the ventricle (33),
413 the airway (34), and the *Xenopus* skin (35). These mechanisms might also play a role in
414 compensating for core protein imbalance (36).

415

416 **Materials and Methods**

417

418 **Simulation**

419 We used Octave (36) for numerical simulations and visualized the data by using Fiji (37). The
420 epithelial tissue was modeled as an array of regular hexagonal cells, with a width and height of 24
421 cells each. The boundary conditions were periodic in all directions. See Supporting Information for
422 details on our mathematical modeling and simulation.

Exploring features of cell-type distribution that affect PCP by deep learning

1. Data generation and labeling

For generating random distribution of low cells in Fig. 2, random numbers that range from 0 to 1 were assigned to each cell, and cells with top 100*$freq_{\mathrm{low}}$ % values were defined as low cells. Threshold values of $\overline{\Delta angle_{\mathrm{high}}}$ for labeling "top" or "bottom" were determined by the accuracy of prediction made by the trained network. More specifically, at first, 3,000 distributions were generated at $freq_{\mathrm{low}} = 0.4$ and $balance_{\mathrm{LH}} = 0.5$ and data that took top 25% and bottom 25% values of $\overline{\Delta angle_{\mathrm{high}}}$ were labeled as "top" and "bottom", respectively. We used these data for training the network, but the accuracy of the prediction made by the network did not exceed 0.85. We additionally generated 3,000 distributions and selected distributions that took top 12.5% and bottom 12.5% values of $\overline{\Delta angle_{\mathrm{high}}}$ for deep learning.

Data augmentation is a technique to increase the number and diversity of training data. Since our simulation adopted the periodic edge condition in all directions, shifting a cell-type distribution along the X and Y axis did not affect the value of $\overline{\Delta angle_{\mathrm{high}}}$. Taking advantage of this, we generated three cell-type distributions from one distribution: one shifted 12 cells along the X-axis, one shifted 12 cells along the Y-axis, and one shifted 12 cells along both axes.

2. Training a neural network

We used the ResNet18 architecture (30) that was pre-trained with the ImageNet-1K dataset for training. We re-trained the network (i.e., transfer learning) by using Google colaboratory. Hyperparameters used for training were as follows: the optimizer was SGD with a learning rate of 0.01 and a momentum of 0.9. batch size, 50; epoch number, 10. Since ResNet18 accepts images as input, each cell-type distribution was converted to an RGB image with 224 pixels in height and 224 pixels in width each. Since each cell was drawn as a hexagon, the converted image had margins at edges. The red channel was assigned to high cells, the green to low cells, and the blue to the margin. Then intensities of images were normalized so that the mean and SDs of intensities of each channel matched with these of the ImageNet-1K dataset. The accuracy of the trained network was evaluated by counting the number of validation data whose labels were correctly predicted by the network.

3. Grad-CAM visualization of the neural network attention

Grad-CAM is an explainable AI technique that can visualize which regions of images the network focuses on when making a classification of an input image (31). We used pytorch-gradcam package

455 in Python to visualize the network attention. Layer2[-1] of PyTorch ResNet18 were selected as a
456 target layer to compute the attention map.

457

458 **Generalized linear model**

459 All following procedures were performed in R. Generalized linear models (GLM) were fit to data by
460 using *glm* function of *stats* package (version 4.4.1). Explanatory variables were standardized before
461 fitting GLM by using *scale* function of *base* package (version 4.4.1), which enabled comparisons of
462 coefficients of explanatory variables. Gamma distribution was selected as a probability distribution,
463 because $\overline{\Delta angle_{\mathrm{high}}}$ takes positive continuous values. The object variable was related to
464 explanatory variables via a log link function. To avoid the multicollinearity, we calculated variance
465 inflation factors (VIF) by using *vif* function of *car* package (version 3.1-2). We selected explanatory
466 variables so that the value of VIF does not to exceed 10. Then, models were evaluated based on
467 values of AIC (Akaike information criterion) by using *dredge* function of *MuMIN* package (version
468 1.48.4). The model that took the lowest value of AIC was selected and used in the manuscript. We
469 used *r2* function of package *performance* version 0.12.4 for calculating Pearson correlation
470 coefficient.

471

472 **Quantification of low-cell distributions *in vivo***

473 In Fig. 6A-B', data adopted from (11) were reanalyzed. In Fig. S7, data adopted from (9) were
474 reanalyzed. Cluster lengths were measured as follows: First, cells were segmented using
475 Cellpose2 (38), and the segmented data were saved as ImageJ ROIs. ROIs marking SCCs were
476 selected based on signals of MCC markers (γ-TUBULIN or acetylated-TUBULIN). These ROIs
477 were then dilated and filled with white, resulting in fusions of areas of adjacent SCCs. The images
478 were then rotated so that the ovary-uterus axis aligned with the X-axis of the image. In the rotated
479 images, white areas were converted to ROIs using the analyze particle function in Fiji. Using these
480 ROIs, the O-U and circumferential lengths of each cluster were measured by the measurement
481 function in Fiji. Clusters that crossed the image edges or epithelial edges were excluded from our
482 analysis. Since oviduct epithelial cells elongate along the ovary-uterus axis (9), we normalized the
483 O-U and circumferential lengths of each SCC cluster. To achieve this, we divided cluster lengths
484 by the average length of SCCs that comprise each cluster. These values were plotted in Fig. 6B
485 and B''' and Fig. S7C-D.

486 In Fig. S7, frequencies of MCCs were measured by calculating the ratio of MCCs to all cells in
487 each image. To measure the frequency of edges between SCCs and MCCs, we counted the
488 number of these edges and divided the value by the number of all cell edges in each image.

489

490 **Imaging fluorescent signals**

491 Oviducts were dissected and opened longitudinally along the ovary uterus axis. For imaging
492 VANGL2:EGFP signals, oviducts were fixed in 4% PFA in PBS (-) at 4℃ overnight. After washing,
493 samples were incubated with Blocking One (Nacalai Tesque, 03953-95) at room temperature for
494 30 min. The samples were then incubated with anti-acetylated-TUBULIN antibody (1/500, Sigma-
495 Aldrich, T7451) in Blocking One (Nacalai Tesque, 03953-95) at 4℃ overnight. After washing,
496 samples were incubated with phalloidin (1/500, Thermo Fisher Scientific, A12381) and anti-Mouse
497 IgG antibody (1/500, Thermo Fisher Scientific, A21235) in Blocking One (Nacalai Tesque, 03953-
498 95) at 4℃ overnight. Finally, after washing, samples were mounted in Fluoromount G
499 (SouthernBiotech, 0100-01). All washing steps were performed as follows: 3 x rinse and 3 x 10 min
500 washes with 0.1% Triton X-100 in PBS (-) at room temperature. Images were acquired using Nikon
501 ECLIPSE Ti2 microscope equipped with CSUW1 spinning disk confocal unit (Yokogawa).

502

503 **Quantification of VANGL2:EGFP distributions**

504 In Figure 6, the distribution of VANGL2:EGFP at cell edges of individual cells was quantitatively
505 evaluated using QuantifyPolarity v2.1 (39). The principal component analysis (PCA) method
506 provided by QuantifyPolarity was used to determine both the magnitude and angle of polarity in the
507 VANGL2:EGFP distribution. The angle represents the orientation of the signal-enriched region. The
508 magnitude reflects the degree of directional bias in the signal distribution; higher values (depicted
509 as longer bars in Fig. 6) indicate stronger enrichment in a specific direction along the cell periphery.

510

511 **Mouse**

512 Female mice (*R26-VANGL2-EGFP* transgenic mice, (10)) were used to study VANGL2:EGFP
513 distributions in wild-type oviducts. Details of *R26-VANGL2-EGFP* transgenic mice (10) were
514 described in these papers. Animal care and experiments were conducted following the Guidelines
515 of Animal Experimentation of National Institutes for Natural Sciences. All animal experiments were
516 approved by either the Animal Research Committee of National Institutes for Natural Sciences.

Animals were maintained in a light- and temperature-controlled room using a 12 h:12 h light:dark cycle at 23±2°C.

Acknowledgments

We thank Dr. Goshi Ogita for discussion; members of the T.F laboratory for discussion and technical assistance. This work was supported by JSPS KAKENHI Grant Numbers, 21H02494, 24K02039, 22H05168 (to T.F.), 22K15130 and the Takeda Science Foundation (to M.A.).

Declaration of interests

The authors declare no competing interests.

Author Contributions

M.A., H.K. and T. F. designed research; M.A. performed research; M.A. analyzed data; and M.A., H.K. and T. F. wrote the paper.

Use of generative AI and AI-assisted technologies

Gemini (Google) was used for writing R, Fiji, and Python scripts. ChatGPT (Open AI) was used for improving language and readability of the manuscript. After using these tools, we reviewed and edited the content of the scripts and the manuscript.

Reference

1. M. T. Butler, J. B. Wallingford, Planar cell polarity in development and disease. *Nat Rev Mol Cell Biol* 18, 375–388 (2017).
2. J. D. Axelrod, Planar cell polarity signaling in the development of left–right asymmetry. *Curr Opin Cell Biol* 62, 61–69 (2020).
3. P. N. Adler, The frizzled/stan Pathway and Planar Cell Polarity in the Drosophila Wing. *Curr Top Dev Biol* 101, 1–31 (2012).
4. Y. Koca, G. M. Collu, M. Mlodzik, Wnt-frizzled planar cell polarity signaling in the regulation of cell motility. *Curr Top Dev Biol* 150, 255–297 (2022).
5. W. Y. Aw, D. Devenport, Planar cell polarity: global inputs establishing cellular asymmetry. *Curr Opin Cell Biol* 44, 110–116 (2017).
6. H. Strutt, D. Strutt, How do the Fat-Dachsous and core planar polarity pathways act together and independently to coordinate polarized cell behaviours? *Open Biol* 11, 200356

550 (2021).
551 7. P. A. Lawrence, J. Casal, Planar cell polarity: Two genetic systems use one mechanism to
552 read gradients. *Development (Cambridge)* 145, dev168229 (2018).
553 8. H. Koyama, D. Shi, T. Fujimori, Biophysics in oviduct: Planar cell polarity, cilia, epithelial
554 fold and tube morphogenesis, egg dynamics. *Biophys Physicobiol 16, 89–107 (2019).*
555 9. D. Shi, *et al.*, Celsr1 is required for the generation of polarity at multiple levels of the mouse
556 oviduct. *Development (Cambridge)* 141, 4558–4568 (2014).
557 10. D. Shi, *et al.*, Mechanisms of Development Dynamics of planar cell polarity protein Vangl2
558 in the mouse oviduct epithelium. *MOD* 141, 78–89 (2016).
559 11. F. M. Usami, *et al.*, Intercellular and intracellular cilia orientation is coordinated by CELSR1
560 and CAMSAP3 in oviduct multi-ciliated cells. *J Cell Sci* 134, jcs257006 (2021).
561 12. M. Arata, F. M. Usami, T. Fujimori, Coordination of Cilia Movements in Multi-Ciliated Cells.
562 *J Dev Biol* 10, 47 (2022).
563 13. C. O. Copley, J. S. Duncan, C. Liu, H. Cheng, M. R. Deans, Postnatal refinement of auditory
564 hair cell planar polarity deficits occurs in the absence of Vangl2. *Journal of Neuroscience*
565 33, 14001–14016 (2013).
566 14. K. Amonlirdviman, *et al.*, Mathematical modeling of planar cell polarity to understand
567 domineering nonautonomy. *Science (1979)* 307, 423–426 (2005).
568 15. K. Abley, *et al.*, An intracellular partitioning-based framework for tissue cell polarity in plants
569 and animals. *Development (Cambridge)* 140, 2061–2074 (2013).
570 16. B. Cho, G. Pierre-louis, A. Sagner, S. Eaton, Clustering and Negative Feedback by
571 Endocytosis in Planar Cell Polarity Signaling Is Modulated by Ubiquitinylation of Prickle.
572 *PLoS Genet.* 11(5), e1005259-29 (2015).
573 17. S. J. Warrington, H. Strutt, K. H. Fisher, D. Strutt, A Dual Function for Prickle in Regulating
574 Frizzled Stability during Feedback-Dependent Amplification of Planar Polarity. *Current*
575 *Biology* 27(18), 2784-2797.e3. (2017).
576 18. W. Chen, *et al.*, Asymmetric Homotypic Interactions of the Atypical Cadherin Flamingo
577 Mediate Intercellular Polarity Signaling. *Cell* 133, 1093–1105 (2008).
578 19. J. S. Duncan, *et al.*, Celsr1 coordinates the planar polarity of vestibular hair cells during
579 inner ear development. *Dev Biol* 423, 126–137 (2017).
580 20. E. K. Vladar, R. D. Bayly, A. M. Sangoram, M. P. Scott, J. D. Axelrod, Microtubules enable
581 the planar cell polarity of airway cilia. *Current Biology* 22, 2203–2212 (2012).
582 21. G. Das, J. Reynolds-Kenneally, M. Mlodzik, The Atypical Cadherin Flamingo Links Frizzled
583 and Notch Signaling in Planar Polarity Establishment in the Drosophila Eye. *Dev Cell* 2,
584 655–666 (2002).
585 22. K. H. Fisher, D. Strutt, A. G. Fletcher, Experimental and Theoretical Evidence for

586 Bidirectional Signaling via Core Planar Polarity Protein Complexes in Drosophila. *iScience*
587 17, 49–66 (2019).
588 23. B. Aigouy, *et al.*, Cell Flow Reorients the Axis of Planar Polarity in the Wing Epithelium of
589 Drosophila. *Cell* 142, 773–786 (2010).
590 24. J. F. Le Garrec, P. Lopez, M. Kerszberg, Establishment and maintenance of planar
591 epithelial cell polarity by asymmetric cadherin bridges: A computer model. *Developmental*
592 *Dynamics* 235, 235–246 (2006).
593 25. S. Schamberg, P. Houston, N. A. M. Monk, M. R. Owen, Modelling and analysis of planar
594 cell polarity. *Bull Math Biol* 72, 645–680 (2010).
595 26. L. D. Hazelwood, J. M. Hancock, Functional modelling of planar cell polarity : an approach
596 for identifying molecular function. *BMC Dev Biol.* 14:13:20 (2013).
597 27. H. Zhu, Is anisotropic propagation of polarized molecular distribution the common
598 mechanism of swirling patterns of planar cell polarization? *J Theor Biol* 256, 315–325
599 (2009).
600 28. Y. Wang, T. Badea, J. Nathans, Order from disorder : Self-organization in mammalian hair
601 patterning. *Proc Natl Acad Sci U S A.* 103(52), 19800–5 (2006).
602 29. S. Shadkhoo, M. Mani, The role of intracellular interactions in the collective polarization of
603 tissues and its interplay with cellular geometry. *PLoS Comput Biol* 15, e1007454 (2019).
604 30. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition. *2016*
605 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV,*
606 *USA* 770–778 (2015).
607 31. R. R. Selvaraju, *et al.*, Grad-CAM: Visual Explanations from Deep Networks via Gradient-
608 Based Localization. *Int J Comput Vis* 128 (2020).
609 32. G. R. Ramirez-San Juan, *et al.*, Multi-scale spatial heterogeneity enhances particle
610 clearance in airway ciliary arrays. *Nat Phys* 16, 958–964 (2020).
611 33. B. Guirao, *et al.*, Coupling between hydrodynamic forces and planar cell polarity orients
612 mammalian motile cilia. *Nat Cell Biol* 12, 341–350 (2010).
613 34. N. Sone, *et al.*, Multicellular modeling of ciliopathy by combining ips cells and microfluidic
614 airway-on-a-chip technology. *Sci Transl Med* 13, eabb1298 (2021).
615 35. B. Mitchell, R. Jacobs, J. Li, S. Chien, C. Kintner, A positive feedback mechanism governs
616 the polarity and motion of motile cilia. *Nature* 447, 97–101 (2007).
617 36. J. W. Eaton, D. Bateman, S. Hauberg, R. Wehbring, GNU Octave version 8.2.0 manual: a
618 high-level interactive language for numerical computations.
619 https://www.gnu.org/software/octave/doc/v8.2.0/ (2023).
620 37. J. Schindelin, *et al.*, Fiji: An open-source platform for biological-image analysis. *Nat*
621 *Methods* 9, 676-82 (2012).

38. M. Pachitariu, C. Stringer, Cellpose 2.0: how to train your own model. *Nat Methods* 19, 1634–1641 (2022).

39. S. E. Tan, W. Tan, K. H. Fisher, D. Strutt, QuantifyPolarity, a new tool-kit for measuring planar polarized protein distributions and cell properties in developing tissues. *Development (Cambridge)* 148, 1–47 (2021).

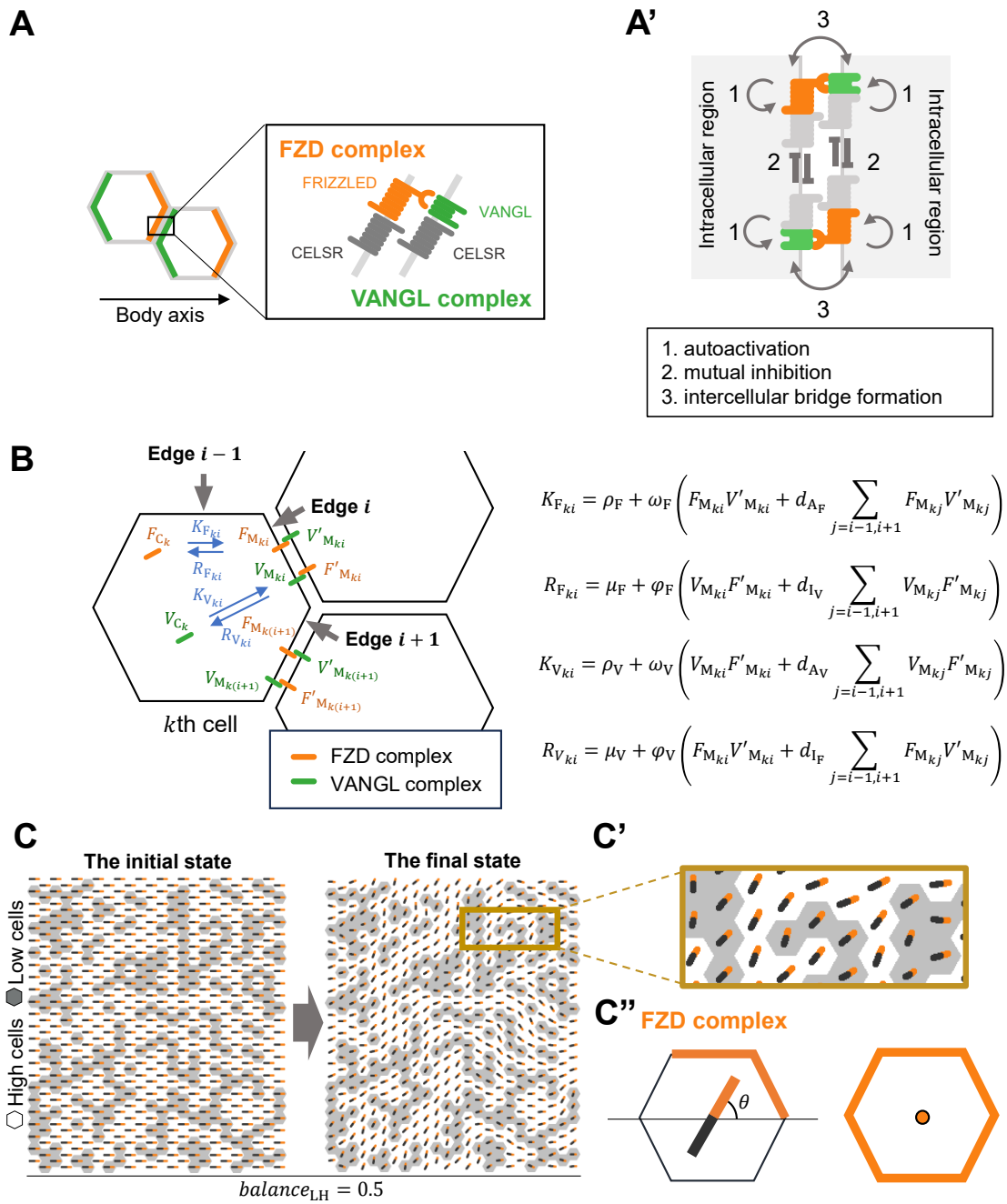

Figure 1. Mathematical modeling for evaluating effects of spatial organization of core protein imbalance on PCP

(**A**) A diagram illustrating the asymmetric distribution of the FZD and VANGL complexes on cell edges. (**A'**) Three key interactions between core-protein complexes implemented in our mathematical model: 1. Autoactivation: the FZD and VANG complexes stabilize complexes of the same type. 2. Mutual inhibition: the FZD and VANGL complexes mutually destabilizes complexes

of the opposite type. 3. Intercellular bridge formation: the FZD and VANG complexes form bridges, intercellularly. (**B**) Diagram illustrating an overview of the mathematical model. See the main text and Supporting Information for details. (**C-C”**) Example analysis of how cell-type distribution affects PCP. White hexagons: high cells; gray hexagons: low cells. (**C**) The unidirectionally biased distribution of the FZD and VANGL complexes was reorganized by core protein imbalance ($balance_{\mathrm{LH}} = 0.5$, compare the initial and final states). A boxed region in (**C**) is enlarged and shown in (**C’**). (**C”**) A bar centered in each cell represents polarity of cell-edge distribution of the FZD complexes. its length represents the magnitude of polarity, and its angle represents the orientation (see Supplementary Information for details).

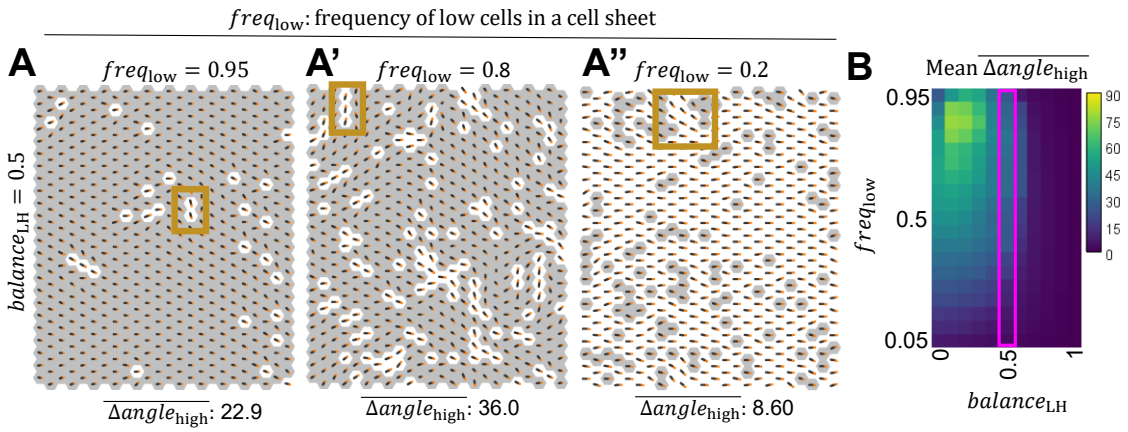

Figure 2. Systematic evaluation of effects of random low-cell distributions and magnitude of core protein imbalance on PCP

(**A-A"**) Examples showing the effects of random low-cell distributions on PCP. $balance_{\mathrm{LH}} = 0.5$. $freq_{\mathrm{low}} = 0.95$ (**A**), 0.8 (**A'**), and 0.95 (**A"**). The resulting values of $\overline{\Delta angle_{\mathrm{high}}}$ are shown at the bottom of each panel. Boxes in **A** and **A'** highlight alignment of high cells along the Y-axis of the cell sheet. A box in **A"** highlight the misoriented polarity vectors of high cells in the vicinity of low cell clusters. (**B**) For each $freq_{\mathrm{low}}$ value, 10 random distributions of low cells were generated and their effects on PCP were evaluated by simulation. Heatmap colors represent the average $\overline{\Delta angle_{\mathrm{high}}}$ values over 10 random distributions for each combination of $balance_{\mathrm{LH}}$ and $freq_{\mathrm{low}}$ (unit: degrees). The boxed column represents data at $balance_{\mathrm{LH}} = 0.5$.

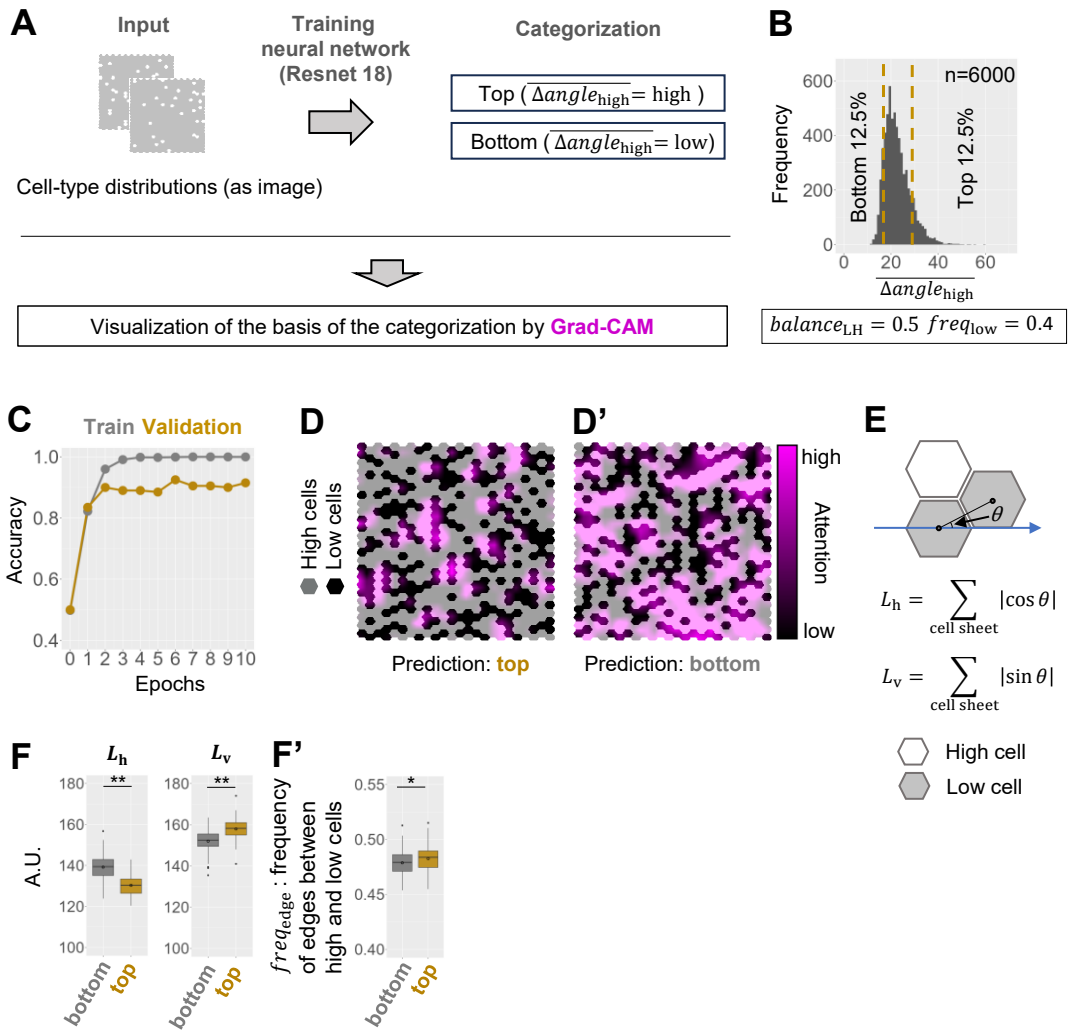

Figure 3. Exploring features of cell-type distribution affecting PCP by deep learning techniques (**A**) Diagram illustrating the strategy for deep learning analysis. Cell-type distributions were provided as images for the input into the neural network. No information about polarity vectors was included in the input images. The neural network (ResNet18) was trained to predict whether a given cell-type distribution corresponds to a "top" ($\overline{\Delta angle_{\mathrm{high}}}$ = high) or "bottom" ($\overline{\Delta angle_{\mathrm{high}}}$ = low) classification. Finally, Grad-CAM was applied to visualize the regions in the input images that contributed to the prediction. (**B**) A total of 6,000 random distributions were generated at $freq_{\mathrm{low}} = 0.4$, and their effects on PCP were simulated at $balance_{\mathrm{LH}} = 0.5$. The variation in the resulting $\overline{\Delta angle_{\mathrm{high}}}$ values is shown as a histogram. Distributions within the top 12.5% and bottom 12.5% $\overline{\Delta angle_{\mathrm{high}}}$ values were labeled "top" and "bottom", respectively. (**C**) Prediction accuracy of the neural network during training. The gray and gold lines show changes in prediction accuracy for the training and validation data, respectively. (**D-D'**) Grad-CAM attention maps highlighting regions

of input distributions where the network emphasized during predictions. Black and gray hexagons represent low and high cells, respectively. Magenta brightness indicates attention magnitude. (**D**) A distribution predicted as “top”. Attention is concentrated at vertical edges between low cells and high cells. (**D’**) A distribution predicted as “bottom”. Attention is concentrated around low cells adjacent to high cells. (**E**) Diagram illustrating the definition of L_{h} and L_{V}. Gray and white cells represent low and high cells, respectively. Blue arrow indicates the X-axis of the cell sheet. (**F**) Boxplots showing L_{h} and L_{V} values for top and bottom validation data (**: $p<0.01$; Wilcoxon rank-sum test). (**F’**) Boxplots showing frequencies of edges between low cells and high cells for the top and bottom validation data (*: $p<0.05$; Wilcoxon rank-sum test).

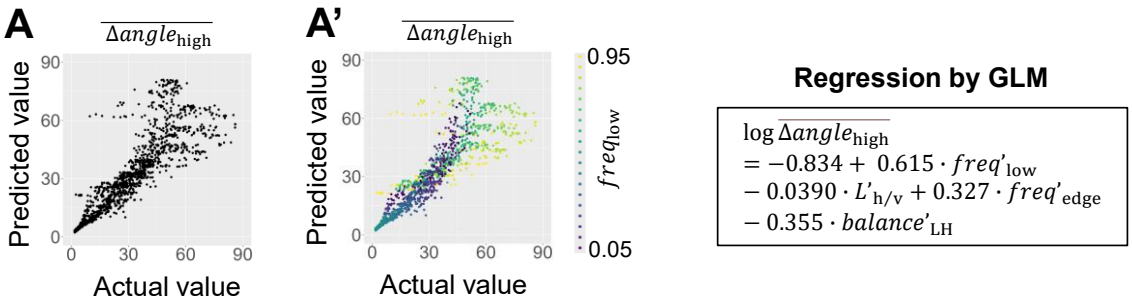

Figure 4. Regression analysis of features of cell-type distributions by Generalized Linear Model (GLM).

(**A-A'**) Scatter plots showing the correlation between actual $\overline{\Delta angle_{\mathrm{high}}}$ values and those predicted by the GLM. Each dot represents a data point shown in Figure 2**B**. $freq_{\mathrm{low}}$ values are color-coded in (**A'**). The detail of the model is shown in the right panel. Pearson correlation coefficient: 0.859.

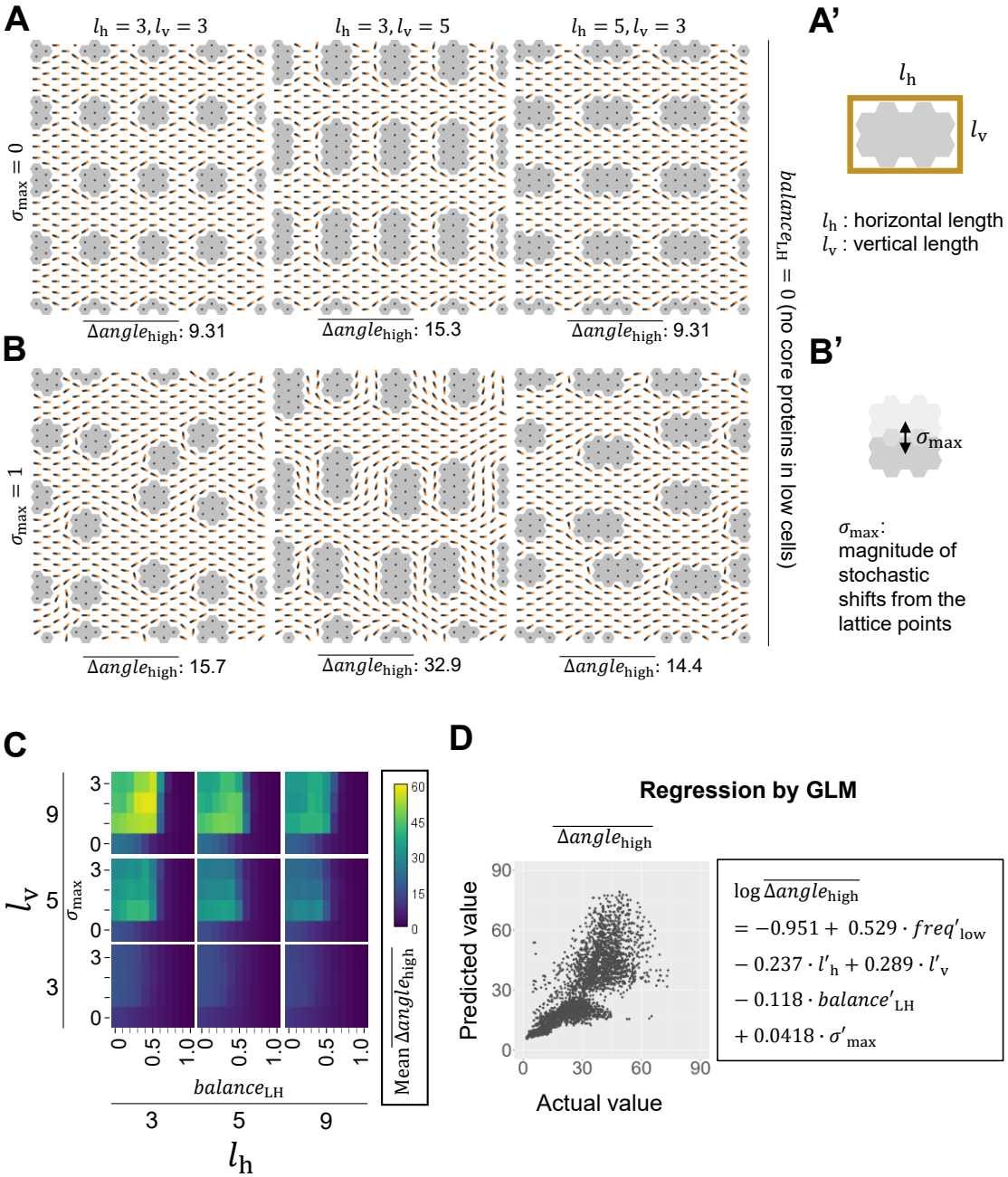

Figure 5. Evaluating the effects of directions of low-cell alignment on PCP

(**A-B'**) Examples of simulations that evaluate the effects of low-cell cluster lengths (**A'**) on PCP. The resulting $\overline{\Delta angle_{high}}$ values are shown at the bottom of each panel. Clusters were arranged in a lattice (**A**) and were randomly shifted up to σ_{max} cell(s) along the X and Y axis of each cell sheet (**B** and **B'**). (**C**) Heatmaps showing mean values of $\overline{\Delta angle_{high}}$, averaged over 10 distributions for each combination of $balance_{LH}$, σ_{max}, l_h, and l_v. Mean values of $\overline{\Delta angle_{high}}$ are color coded as shown in the right lookup table (unit: degrees). (**D**) Scatter plot showing a correlation

between actual values of $\overline{\Delta angle_{\mathrm{high}}}$ and those predicted by the GLM (right panel). Pearson correlation coefficient: 0.822. Each dot represents a single simulation.

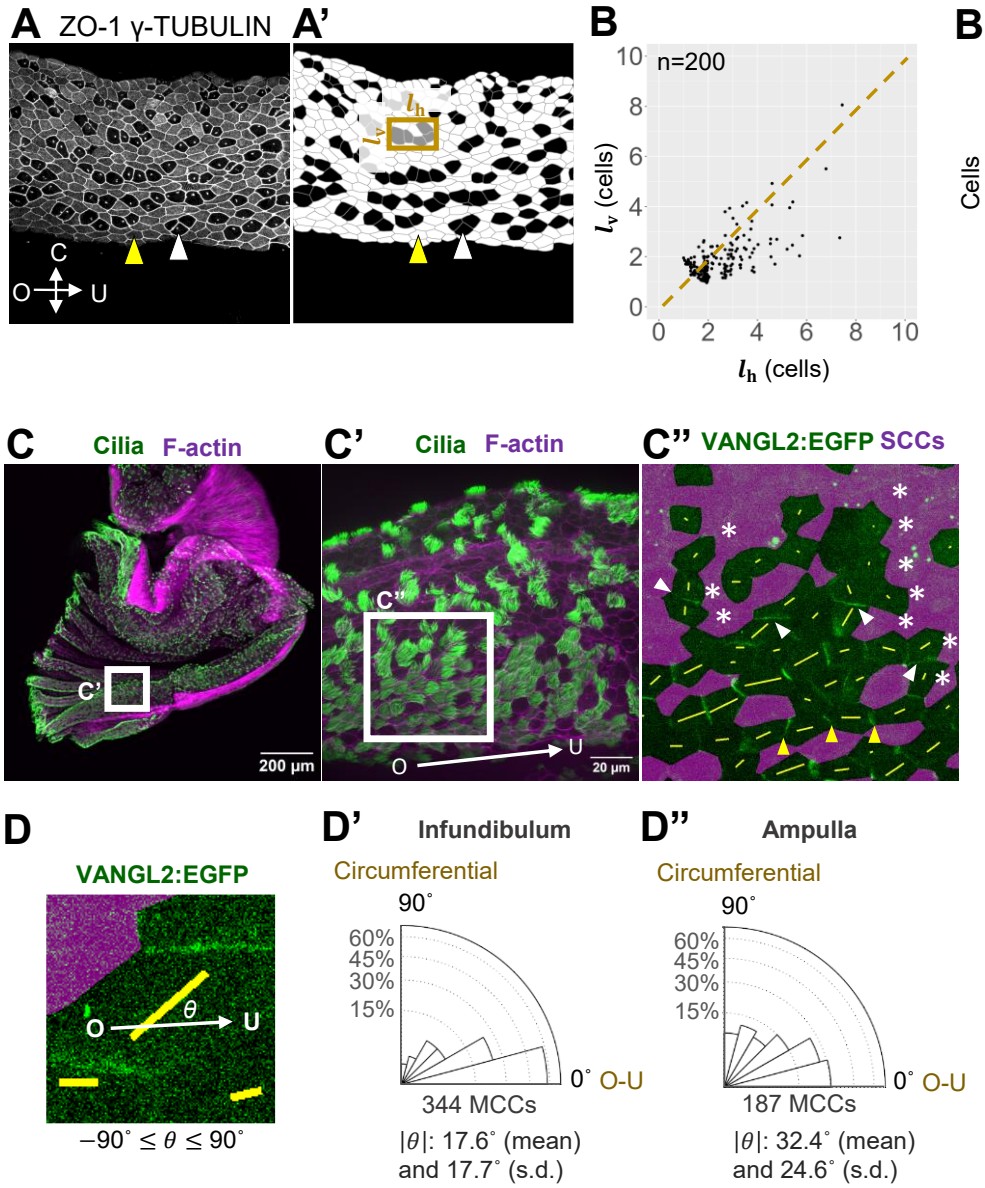

Figure 6. Direction of alignment of secretory cells was correlated with the orientation of core protein distribution

(**A-B'**) Analysis of SCC distribution in the infundibulum region of the oviduct. (**A**) Oviduct stained for ZO1 (a marker for cell edges) and γ-TUBULIN (a marker for MCCs). White and yellow arrowheads point to a SCC and MCC, respectively. (**A'**) Segmented image: gray lines denote cell edges; white cells are MCCs; black cells are SCCs. The ovary-uterus (O-U) and circumferential (C) lengths of each SCC cluster (l_h and l_v, respectively) were measured using a bounding rectangle (gold box). (**B-B'**) l_h and l_v values were analyzed using a scatter plot (**B**) and a boxplot (**B'**). The dotted line in (**B**) represents $l_h = l_v$. Open circles in **B'** indicate average values. **: p<0.01

715 (Wilcoxon signed-rank test). Isolated SCCs were excluded from the analysis. N: 200 clusters; 3
716 mice. (**C-C”**) The ampulla region of the oviduct was opened longitudinally and stained for
717 acetylated-TUBULIN (green, a marker for cilia) and F-actin (magenta, a marker for cell edges). The
718 boxed regions in **C** and **C’** are shown in **C’** and **C”** respectively. The direction of the ovary-uterus
719 axis is indicated by an arrow in **C’**. (**C”**) SCCs are filled with magenta. VANGL2:EGFP signals are
720 shown in green. Asterisks mark examples of circumferential SCC alignment. White arrowheads
721 mark VANGL2:EGFP signals enriched at cell edges running along the ovary-uterus axis. These
722 enrichments were observed in MCCs adjacent to SCCs aligned circumferentially. In the same view
723 field, enrichments of VANGL2:EGFP signals at cell edges running circumferentially were also
724 observed (yellow arrowheads). These enrichments were found near low cells aligned along the
725 ovary-uterus axis (e.g., compare yellow arrowheads with white arrowheads in **C”**). Yellow bars in
726 C” represent polarity of VANGL2:EGFP distributions at cell edges. The length and angle of each
727 bar indicate the magnitude and direction of VANGL2:EGFP polarity, respectively (see
728 Supplementary Information for details). Compare the orientations of bars near the white
729 arrowheads with those near the yellow arrowheads (other examples are shown in Figure S7). (**D-**
730 **D”**) Quantification of VANGL2:EGFP polarity in the infundibulum (**D’**) and the ampulla (**D”**). Bar
731 angles relative to the O-U axis (**D**) were calculated and plotted as rose diagrams (circular
732 histograms).
733