Co-cultures of cerebellar slices from mice with different \textit{reelin} genetic backgrounds as a model to study cortical lamination - Supplementary Material 5

\textbf{Spatial statistics tools}

Computer-based GIS technologies are used to store, view, analyze, and interpret geographic data. The topographical location of features (i.e. the objects of interest in a space represented by a map) is determined by geographic data, often known as spatial or geospatial data. The toolbox for spatial statistics includes statistical methods for examining geographical patterns, distributions, linkages, and processes related to the map features. Although the ideas and goals of spatial and nonspatial (conventional) statistics may be similar, spatial statistics are distinctive in that they were created especially for use with geographic data. They directly incorporate space information (proximity, area, connectedness, and/or other spatial interactions) into their mathematics, in contrast to conventional nonspatial statistical methods. As microscopic images can be represented in an X-Y Cartesian coordinate system, GIS can be used to study biological features in these images once they are converted into a "geographic" map.

We briefly report a few fundamental notions on the applications of some ArcMap tools we used in the main paper (Figure S5-1). For its purpose, the feature under study is the spatial distribution of the PNs.

\textit{Measuring Geographic Distributions tools}

These tools are used to calculate a value that represents certain characteristics of the distribution of the objects under study, such as the center, compactness, or orientation. The \textit{Central Feature tool} identifies the most centrally located feature (in our case a PN), the \textit{Mean Center tool} identifies the geographic center (or the center of concentration) for the set of PNs, and the \textit{Median Center tool} identifies the location that minimizes the overall Euclidean distance of the PNs. The \textit{Standard Distance tool} measures the degree to which features are concentrated or dispersed around the geometric mean center. The \textit{Directional Distribution tool} creates standard deviational ellipses or ellipsoids to summarize the spatial trend in the distribution of the PNs. The input and output of these tools are shown graphically in the first row of Figure S5-1.

\textit{Analyzing Patterns tools}

This set of inferential statistics tools analyzes the pattern of distribution of the feature(s) under study calculating a statistic for its quantification. These tools are based on the concept of \textit{spatial autocorrelation} which is a measure of the correlation of an observation with other observations across space. The values of the observations in a given sample are assumed to be independent of one another in most conventional statistical analyses. This assumption is violated in the presence of spatial autocorrelation since, in this case, samples gathered from close sites are related to one another and hence statistically not independent of one another. Some Analyzing Patterns tools in ArcMap are designed to assess the presence of spatial autocorrelation between the observations and to analyze the variations of this relationship over the space (spatial heterogeneity). Spatial autocorrelation among observations can be positive or negative. Positive spatial autocorrelation arises when observations with identical values are close to one another (i.e.
clustered), whereas negative spatial autocorrelation occurs when observations with differing values are close to one another (i.e. random) 1.

The **Average Nearest Neighbor** tool returns five values: Observed Mean Distance (OMD), Expected Mean Distance (EMD), Nearest Neighbor Index, z-score, and p-value (see Table 1 in *Supplementary Material 6*). The values are written as messages at the bottom of the **Geoprocessing pane** during tool execution and issued as derived output values. The nearest neighbor (NN) ratio (R) is the ratio between the OMD and EMD among the feature(s) of interest, in our study the PNs. EMD is the mean distance between these neurons calculated on a hypothetical random distribution of the same number of cells covering the same total area. The value of R = 1 indicates randomness; R = 0 indicates maximum aggregation; and R = 2.149 indicates maximum possible dispersion. For the theory behind NN analysis see 2.

The **High/Low Clustering (Getis-Ord General G)** tool returns four values: Observed General G, Expected General G, z-score, and p-value (see Tables 2 and 3 in *Supplementary Material 6*). The values are written as messages at the bottom of the **Geoprocessing pane** during tool execution and issued as derived output values. This spatial statistics tool gives a measure of spatial autocorrelation related to the concentration of high or low values for a given feature in the entire area of study. In our case, the tool measures the number of PNs in each hexagon of the tessellation (see Figure 6 in the main text) and calculates whether hexagons with high or low numbers of PNs are prevalent in certain areas of the section or are rather randomly distributed. The null hypothesis for the High/Low Clustering statistic states that there is no spatial clustering of feature values. When the p-value is small and statistically significant, the null hypothesis can be rejected and the sign of the z-score becomes important. If the z-score value is positive, the observed General G index is larger than the expected General G index, indicating that high values for the attribute (number of PNs/hexagon) are clustered in the study area. If the z-score value is negative, the observed General G index is smaller than the expected index, indicating that low values are clustered in the study area. The General G index ranges from 0 to 1 3.

The **Spatial Autocorrelation (Global Moran's I)** tool returns five values: the Moran's I Index, Expected Index, Variance, z-score, and p-value (see Tables 4 and 5 in *Supplementary Material 6*). The values are written as messages at the bottom of the **Geoprocessing pane** during tool execution and issued as derived output values. With the Generate Report box ticked the tool creates an HTML report file with a graphical summary of results (see Figure 6I in the main text). Moran's I index is a measure of spatial autocorrelation that can vary from -1 to 1, with 0 indicating perfect randomness. When the p-value or z-score indicates statistical significance, a positive Moran's I index value indicates a tendency toward clustering, while a negative Moran's I index value indicates a tendency toward dispersion 1.

The **Multi-distance Spatial Cluster Analysis (Ripley's K Function)** tool output is a table with fields: ExpectedK and ObservedK containing the expected and observed K values, respectively. Because the L(d) transformation is applied, the ExpectedK values will always match the Distance value. A field named DiffK contains the Observed K values minus the Expected K values. As a confidence interval option is specified, the Output Table will include two additional fields named LwConfEnv and HiConfEnv. When the observed K value is larger than the expected K value for a particular distance, the distribution is more clustered than a random distribution at that distance. When the observed K value is smaller than the expected K value, the distribution is more dispersed than a random distribution at that distance. When the observed K value is larger than the HiConfEnv value, spatial clustering for that distance is statistically significant. When the observed K value is smaller than the LwConfEnv value, spatial dispersion for that distance is statistically
significant. Data can be represented graphically as reported in Figures S6-1 to 3 in Supplementary Material.

Mapping Clusters tools

The Cluster and Outlier Analysis (Anselin Local Moran’s I) tool (Figure S5.1 second row) and the Optimized Outer Outlier Analysis tool (Figure S5.1 third row) calculate a local Moran’s I value (LM index the figure), a z-score, a pseudo-p-value, and a COType (Cluster/Outlier Type) field in the Output Feature Class (map layer) that identifies the cluster with a color code (last panel on the right) for each statistically significant feature. The z-scores and pseudo-p-values represent the statistical significance of the computed index values. A high positive z-score for a feature indicates that the surrounding features have similar values (either high values or low values). The COType field in the Output Feature Class will be HH for a statistically significant cluster of high values and LL for a statistically significant cluster of low values. A low negative z-score for a feature indicates a statistically significant spatial data outlier. The COType field (last panel on the right) also identifies statistically significant high and low outliers (HL and LH), indicating if the feature has a high value and is surrounded by features with low values (HL) or if the feature has a low value and is surrounded by features with high values (LH). The difference between these two tools is that the Optimized Outlier Analysis is corrected for multiple testing and spatial dependence using the False Discovery Rate (FDR) correction method. By the Optimized Outlier Analysis, the COType field will always indicate statistically significant clusters and outliers based on an FDR-corrected 95% confidence level.

The Hot Spot Analysis (Getis-Ord G*) tool (Figure S5 fourth row) and the Optimized Hot Spot Analysis tool (Figure S5 fifth row) create a new Output Feature Class (map layer) with confidence level bin (G* in the figure) a z-score, p-value, and a graphic representation of Hot and Cold Spots (last panel of the row) for each feature in the Input Feature Class (the Joint Count panel in the figure). The resultant z-scores and p-values show where either high or low values cluster spatially. These tools operate by examining each characteristic in light of its surrounding features. Even while a feature with a high value may interesting, it could not be a statistically significant hot spot. A feature must have a high value and be surrounded by additional features that have high values to be a statistically significant hot spot. A statistically significant z-score is produced when the local sum for a feature and its neighbors deviates significantly from the expected local sum and deviates by an amount that is too great to be the product of random chance. The FDR adjustment in the Optimized Hot Spot Analysis tool adjusts statistical significance to take multiple testing and spatial dependency into account.

Figure S5: Example of the output of some ArcGIS tools using as an input an image of 100 randomly generated points.