Protocol: Comparison between consensus localization patterns using a Distance Matrix

Summary: The computation of consensus localizations maps facilitates a virtual colocalization measurement: the average dynamic localization of any two proteins can be compared. (It is important to note that since the cell-to-cell average localization is compared, proteins that do colocalize but either do not appear localized in the cell or are not localized to a consistent physical location in the cell will not be detected. For instance, two proteins that form a focus that is not positioned in the cell would not be detected.) To compare localization patterns, we define a distance metric on the localization patterns and then compute of distance metric which represents the distance or difference between any two localization patterns. For each protein, we generate a list of the ten proteins with the most similar localization patterns and place these proteins in the Most Similar Localization Pattern table.

1. Distance Matrix.
   a. Compute the Consensus localization Image Towers (Protocol) to give image matrix for each fusion/condition
      \( \tilde{I}_{ij} \)
      where the index J denotes strain and condition and the indices i and j refer to the rows and columns of the image matrix.
   b. We define the metric tensor:
      \[ g_{ijkl} = \delta_{ik}\delta_{jl} \]
      where \( \delta \) is the kronecker delta. We define the inner product between two images:
      \[ \tilde{I}^J \cdot \tilde{I}^K = \sum_{i,j,k,l} g_{ijkl} \tilde{I}^J_{ij} \tilde{I}^K_{kl} = \sum_{i,j} \tilde{I}^J_{ij} \tilde{I}^K_{ij} \]
   c. We define the normalized images matrices:
      \[ I^J = \tilde{I}^J / (\tilde{I}^J \cdot \tilde{I}^J)^{1/2} \]
   d. The distance matrix is defined:
      \[ M^{IJ} = [(I^I - I^J) \cdot (I^I - I^J)]^{1/2} \]
2. Most Similar Localization Pattern
   a. To generate the Most Similar Localization Pattern Matrix, we find the ten smallest entries in each row, not counting the protein itself.
   b. If there are multiple entries for a protein, as is usually the case—representing different induction conditions—we choose the ten smallest distances overall.