Supplemental Methods

Pair-Location Conditional Probability Matrix (PLCP)

We used the protein network along with the known protein locations to calculate a **PLCP** matrix. This matrix captures the conditional probability of a protein having a location l_j given that its interaction partner has location l_i:

$$p(l_j \mid l_i) = \frac{I_{ij}}{\sum_k I_{ik}},$$

where I_{ij} is the normalized number of interactions between protein pairs spanning (l_i and l_j). I_{ij} is defined as:

$$I_{ij} = \frac{\sum_{a \in l_i, b \in l_j (a \neq b)} \frac{\phi(a,b)}{N_a \times N_b}}{I(l_i) \times I(l_j)}$$

where $I(l_i)$ is the total number of interactions of proteins with location l_i, N_a is the number of locations assigned to protein a, and $\phi(a,b)$ is "1" if there is an interaction between proteins a and b; otherwise, zero.

Generation of protein features

Briefly mentioning, three kinds of amino acid features was used, including amino acid composition frequencies, pair-coupled amino acid frequencies, and pair-coupled amino acid frequencies with a gap (length = 1). Additionally, four kinds of chemical amino acid compositions using normalized hydrophobicity, hydrophilicity, side-chain mass, or those combined features pseudo-amino acid composition. Finally, motif and GO term features was also used for describing each protein. Note that we only used GO terms with computational evidence codes. For conditional network features, we generated two major kinds of feature

vector $\boldsymbol{N^D}$ and $\boldsymbol{L^D}$ for interaction neighbors up to network distance D=2 from the protein. $\boldsymbol{N^D}$ of protein P_i is defined as the weighted average of the $\boldsymbol{S}$ feature vectors of proteins up to distance D from P_i in the network, including P_i itself (called the D^{th} neighborhood of P_i and represented by the variable $\boldsymbol{C_i^D}$):

$$N_i^D = \frac{1}{\sum_{P_k \in C_i^D} \Phi(P_k, P_i)} \sum_{P_k \in C_i^D} \Phi(P_k, P_i) S_k \, .$$

where $\Phi(P_k, P_i)$ is the expression coherence score between P_k and P_i. The second network feature vector, $\boldsymbol{L_i^D}$, represents the probability that P_i has each of the K=13 distinct locations, considering the probabilities of interaction between proteins in distinct location pairs:

$$\boldsymbol{L}_i^D = \left[L_i^D(1), \ldots L_i^D(y), \ldots L_i^D(K) \right]$$

$$L_i^D(y) = \frac{1}{\sum_{P_k \in C_i^D} \Phi(P_k, P_i)} \sum_{P_k \in C_i^D} \left[\Phi(P_k, P_i) \max_{f_x \in \boldsymbol{\Gamma}_k} p\left(l_y \mid l_x\right) \right]$$

where l_x is one element of the location set $\boldsymbol{\Gamma}_k$ of P_k, and $p(l_y \mid l_x)$ is the conditional probability of location label l_y given the label l_x (from the **PLCP** matrix).

Generating a confidence score using a DC-kNN classifier

The DC-kNN classifier has been described by Lee *et al*. (Lee et al. 2008). Briefly, it has three main steps: dividing, choosing, and synthesizing. In the dividing step, the full feature vector is divided into m (here, 29) feature sub-vectors (here, 9 ***S***s, the 18 ***N***s, and the 2 ***L***s). In the choosing step, k (here, 5) nearest neighbors are chosen for a protein using each sub-vector. Finally, the synthesizing step averages the m sets of k neighbors with a weight for each set, and generates a confidence c_l for each label l by means of a normalization process involving m and k:

$$c_l = \left[\frac{1}{k} \sum_m n_{ml} \times \phi_m \right]^{\frac{1}{\sqrt{m}}},$$

where n_{ml} is the number of *k* nearest neighbors that have the label *l* according to the sub-vector *m*. ϕ_m ($\sum_m \phi_m = 1$) is the weight of the m^{th} sub-vector. Note that, instead of using all sub-vectors, DC-kNN finds a feasible combination of feature sub-vectors for each label based on a forward approach involving cross-validation with an AUC performance measure. At each iteration, DC-kNN chooses the most predictive feature sub-vector among those remaining. In the first iteration, feature sub-vectors are used individually to find the most predictive one. For the weights, ϕ_m, DC-kNN uses the AUC obtained using each feature sub-vector alone. In each training round, a LTOCV approach is used, with one being used for feature selection and the other for performance reporting. DC-kNN produces a confidence degree (0 ~ 1).

Brain tissue preparation and primary cell culture

Three micron sections were prepared on a cryostat (Leica). Primary cells were prepared directly from human normal brain and glioma tissues. The brain and glioma tissue collection and usage were approved by the Ethics Committee of the Seoul National University College of Medicine, Seoul, Korea. Small pieces of human brain cortexes were incubated with phosphate-buffered saline (PBS) containing 0.25% trypsin and 40 mg/ml DNase I for 30 min at 37 °C. Dissociated cells were suspended in DMEM, supplemented with 5% FBS, 5% horse serum, 20 mg/ml gentamicin, and 2.5 mg/ml amphotericine B (feeding medium), plated in culture dishes, and maintained at 37 °C in an incubator under 5% CO2/95% air atmosphere. After 2–3 weeks of in vitro culture, cells were prepared by harvesting the cells in culture dishes and replating the primary cells on Lab-Tek II Chamber Slide System (2 x 104

cells/wells, Nunc) for immunocytochemistry.

Proximity ligation assay (PLA)

Two primary antibodies raised in different species recognize the target antigens. Species-specific secondary antibodies, called PLA probes, bind to the primary antibodies. When the PLA probes are in close proximity, the DNA strands can interact two other circle-forming DNA oligonucleotides. After adding of enzymatic ligation, they are amplified via rolling circle amplification. After the amplification reaction, DNA circles are labeled by red dye labeled oligomer. The result in each single-molecule amplification product is easily visible as a red spot when viewed with a fluorescence microscope.