# SUPPLEMENTARY MATERIALS

## Review of node degree bias in protein-protein interaction networks

As mentioned in the main text, it is important to be aware of underlying biases in the interaction data when working with protein-protein interaction (PPI) networks, as false conclusions may otherwise be reached when conducting the downstream analysis.

An important factor to consider, is that certain proteins have the problem of being associated with an inflated number of interaction partners. As a result, they may connect otherwise unrelated parts of the global network, and purely due to their "over-connectedness" appear to be interacting with key genes/proteins in a particular biological system under investigation (in this case ALS). Two major factors should be taken into consideration: 1) Experimental artifacts and 2) Study bias.

Important experimental artifacts to consider in this type of data are randomly distributed noise and unspecific binding of proteins. As with all other experimental data, the determination of protein-protein interactions is affected by experimental noise (false positives and false negatives). The degree of the noise depends on the nature of the experimental system (e.g. yeast two-hybrid (Y2H) and affinity-purification mass spectrometry (AP-MS)) as well as the nature of the study [77]. Small-scale studies with few interaction partners are often considered less noisy compared to large scale studies, where less attention to detail may be given to individual interactions. inBio Map [13] handles the uncertainty in the underlying experimental data, by assigning a confidence score to each interaction, which includes a measurement of the number of independent studies supporting the interaction as well as the nature of the study (small-scale vs. large-scale) - see Li et al [13], supplement for all details. This allows us to rank all interactions based on experimental support, and only use the high-confidence subset of interactions [13]. This precaution addresses the part of the problem, that is due to randomly distributed noise in the data, which can indeed be substantial when integrating 40,000+ experimental studies.

However, in AP-MS experiments certain proteins and protein families tend to be co-purified with other proteins under investigation due to unspecific binding. This can even be unspecific binding to the affinity-purification column itself. The CRAPome project [15] has investigated this in great detail, and across 300+ negative controls (AP-MS experiments run without the "bait" protein present) more than 100 proteins were seen in more than 50% of the experiments. Worst offenders are heat-shock proteins (HSPs) with more than 90% presence, tubulins/actins with more that 80% presence, and generally highly expressed proteins such as histones and elongation factors with more that 50% presence. To complicate the matters further, certain post-translational modifications (PTM) often show up as false positives in mass-spectrometry based studies, for example when a (small) protein or peptide is covalently attached to another protein. This will appear as if the attached protein/peptide is interacting with the other protein. Common examples of this includes UBB, UBC, UBA52, RPS27A, SUMO1, SUMO2, SUMO3 and SUMO4, which are often covalently attached to other proteins in an unspecific manner. Unspecific should in this context be understood in the following way: Protein-protein interactions are most often between proteins that work in the same biology, whereas peptides flagging other proteins for degradation can be covalently linked to a broad range of proteins not likely to be otherwise related. It should be noted that UBB and UBC are also detected in more than 50% of the negative controls in the CRAPome study. Ideally, AP-MS experiments should correct for unspecific binding before reporting the results, but this is not always the case [15], and it may be difficult to completely eliminate the false positive signal.

Finally, a fundamentally different issue that needs to be considered is study bias. While some protein-protein interaction experiments lend themselves well to genome-wide studies (e.g. Y2H screens), the AP-MS class of experiments are fundamentally built to identify interaction partners for sets of proteins of particular interest. In essence, the study bias in the proteome (with much of the proteome being understudied) is carried over into study of the interactome.

Good examples of the issues of overrepresentation of highly connected proteins in the PPI resource is the heat-shock protein HS90A and many of the ubiquitin-like modifiers which are some of the most studied proteins in literature (top 10 proteins with most connections in our PPI network are shown in Supplementary Table S 5. This is important, since network analysis based studies that are built around the concept of the "*centrality-lethality rule*" [16] where the node degree (number of interaction partners) is important, are likely to pick up exactly this class of over-connected proteins, unless a node-degree normalization scheme is used.

### Supplementary Table S 1

SupplementaryTableS1\_ALS-gene\_list-nw\_support.xlsx

### Supplementary Table S 2

SupplementaryTableS4\_ALS\_protein\_list\_new2020.xlsx

### Supplementary Table S 3

SupplementaryTableS3\_GO\_CLASS+SUPERCLASS.xlsx

### Supplementary Table S 4

SupplementaryTableS2\_TM\_top50\_table.xlsx

### Supplementary Table S 5

SupplementaryTableS5\_Most\_Connected\_Proteins.xlsx

### Supplementary File S 1

SupplementaryFileS1\_ALS\_core\_modules.cys

### Supplementary File S 2

SupplementaryFileS2\_ALS\_disease\_modules.cys

# Description of ALS network collection files

The ALS core modules (Supplementary File S 1) as well as the entire set of 282 ALS seeded networks (Supplementary File S 2), are available for download as Cytoscape session files.

In order to aid the interpretation of the networks, the following meta-data has been added/visualized:

* Known ALS-associated genes/proteins, including sources
* Biological information: Receptors, secreted proteins and nuclear proteins
* Proteins associated with a collection of 50 other disease often co-mentioned with ALS in the literature.

Supplementary Figure S 1: Size distribution of the ALS network collection

