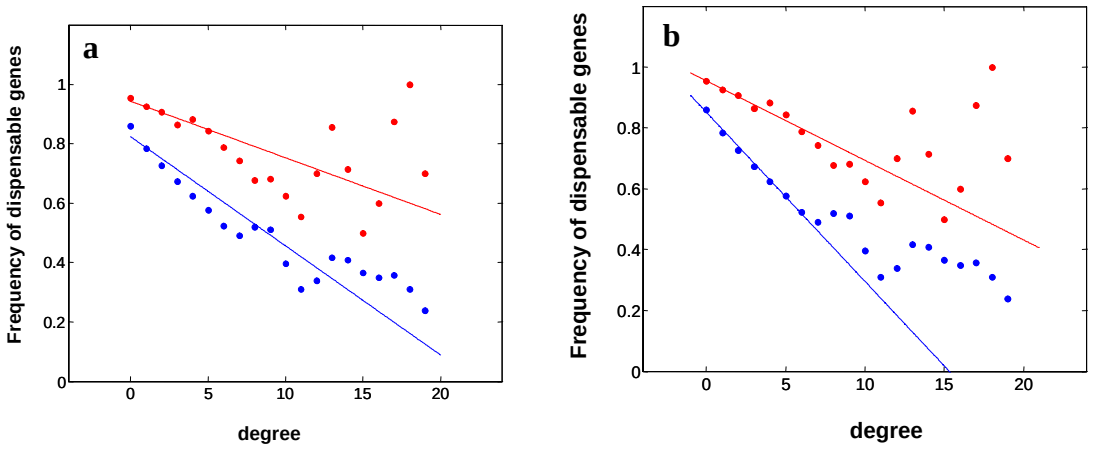
SI Appendix 4

The GRID protein interaction dataset is noisy and contaminated with many false positive interactions (1). One way to circumvent this problem has been performed by (1) by filtering out interactions that have not been validated by at least two independent sources. This procedure yielded a multi-validated dataset termed the “high-confidence” (HC) protein interaction dataset (1).

To further test the conclusions of our study we examined whether our results were also consistent with the HC dataset instead of using the unfiltered GRID. As expected, using a significantly smaller set of interactions increased the noise and reduced the statistical strength of the results. However, all general trends we report have remained qualitatively unchanged (see Figs. S4 a and c, below) by this confinement of the dataset thereby strengthening our overall conclusions of preferential backup of highly connected proteins.



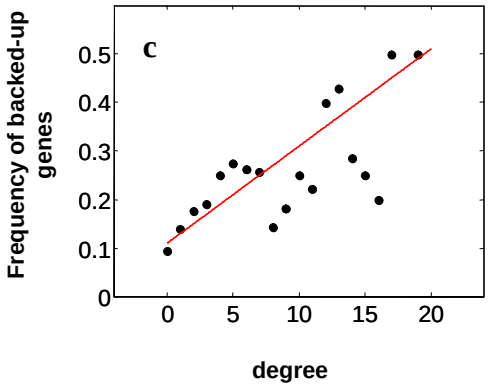


Fig. S4. Degree vs. dispensability using the HC dataset of high-quality protein-protein interactions (1). Fits were computed by using weighted least squares on all data points (a) or only on the first 10 datapoints (b) and p-values (see below) were calculated based on a logistic regression analysis. (c) frequency of backed up duplicates (calculated as detailed in *Materials and Methods*) as a function f the degree based on the HC dataset.

Calculated for degree < 20	Slope (linear regression)	r^2 (linear regression)	P value (**logistic regression**)
Singletons	-0.03 ± 0.007	0.78	1e-51
Duplicates	-0.009 ± 0.01	0.11	1e-21

1. Batada NN, ●●●●●●●, *et al.* (2006) *PLoS Biol* 4:●●-●●.[1]

[1] Please supply the complete reference (first 10 author names if there are more than 11), including the full page range.