

Gene essentiality, gene duplicability and protein connectivity in human and mouse

Han Liang and Wen-Hsiung Li

Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA
Corresponding author: Li, W.H. (Email: whli@uchicago.edu).

Introduction

To test the robustness of our observations, we extended our analyses in two directions. First, in view of considerable noise in the Y2H protein interaction data, we performed the analysis using a literature-curated dataset and a multi-validated dataset, separately. As annotated in Rual et al. (Ref.17), the literature-curated dataset includes 3,610 interactions (self-binding interactions were excluded). The multi-validated dataset includes 528 interactions (self-binding interactions were excluded), where each interaction was supported by at least two literature sources or identified by the Y2H assay and a literature source. Second, to examine the effect of our definition of “essential” genes, we performed the analysis considering the genes whose single-gene deletions have a lethal effect and the genes whose single-gene deletions cause infertility separately. In all the analyses, we obtained the same results as we presented in the main text (see below).

Gene Essentiality vs. Protein Connectivity

In humans, using the literature-curated dataset, we found that the protein connectivity of an essential gene is significantly greater than that of a non-essential gene (Wilcoxon rank test, $P = 4 \times 10^{-5}$, and Fig. S1). Moreover, we obtained the same result using a multi-validated dataset (Wilcoxon rank test, $P = 1 \times 10^{-3}$).

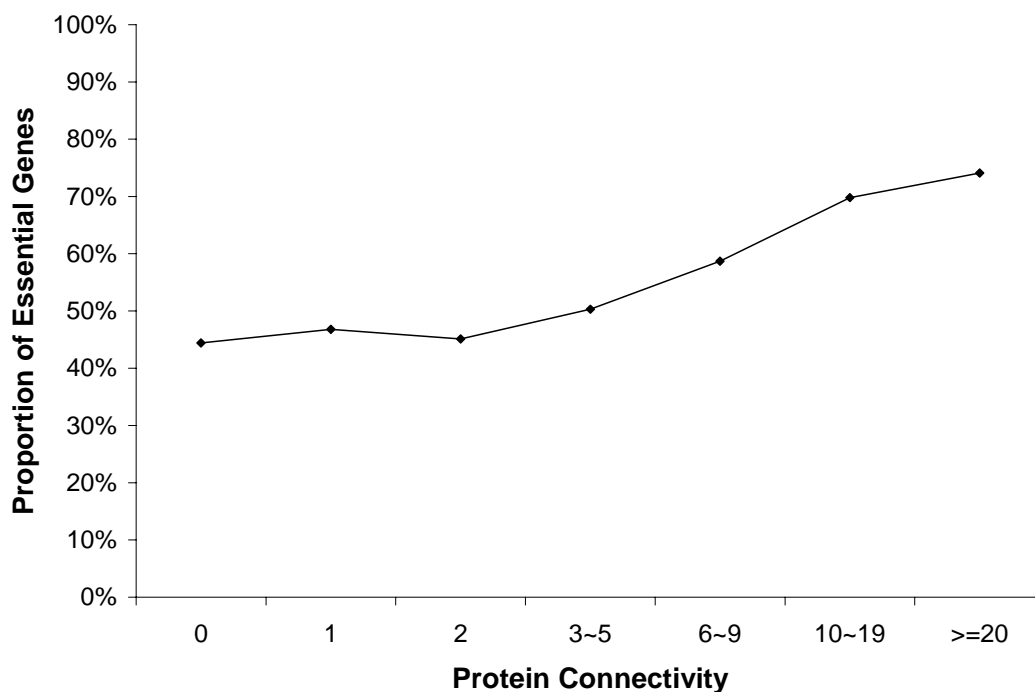


Fig. S1

We also found that compared with non-essential genes, the genes whose deletions have a lethal effect tend to have a higher connectivity (Wilcoxon rank test, $P = 1 \times 10^{-7}$; and Fig. S2). Similarly, compared with non-essential genes, the genes whose deletions causes infertility tend to have a higher connectivity (Wilcoxon rank test, $P = 2 \times 10^{-2}$).

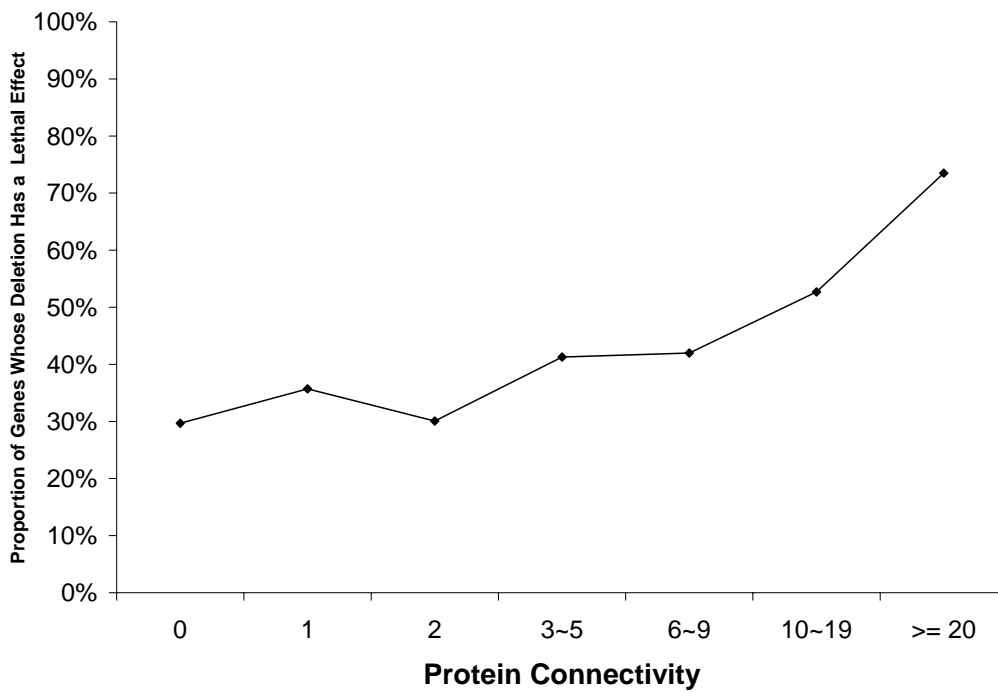


Fig. S2

Protein Connectivity vs. Gene Duplicability

In humans, using the literature-curated dataset, we found that gene duplicability is positively correlated with protein connectivity (Fig. S3) and the number of paralogs per gene is positively correlated with protein connectivity (Spearman rank test $R = 0.35$, $N = 5,530$, $P < 7 \times 10^{-151}$). Moreover, we obtained the same result using a multi-validated dataset (Spearman rank test $R = 0.48$, $N = 5,530$, $P < 5 \times 10^{-280}$).

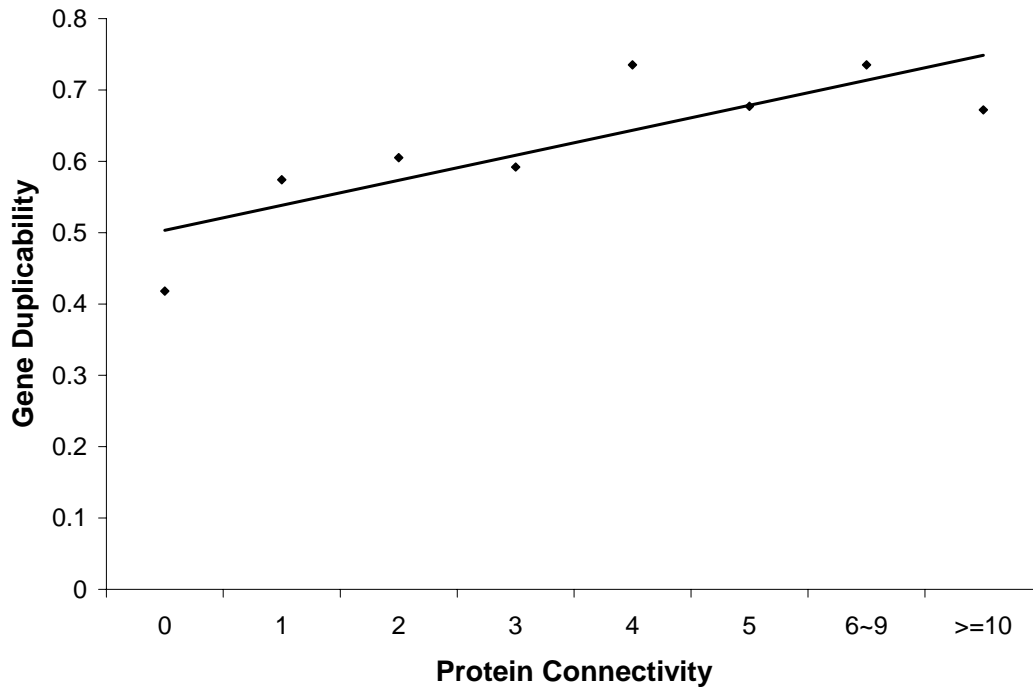


Fig. S3

Gene Fitness Effect vs. Gene Duplication

In mice, we found that the proportion of essential genes does not differ between singletons and duplicates (48.6% vs. 46.2%, $\chi^2 = 1.3$, $P = 0.3$, Table S1).

Table S1

	Singletons	Duplicates	Total
Essential Genes	355	1001	1356
Non-Essential Genes	376	1167	1543
Total	731	2168	2899

Moreover, we found that the proportion of singletons does not differ between the genes whose deletion has a lethal effect and non-essential genes (25.2% vs. 24.4%, $\chi^2 = 0.2$, $P = 0.6$, Table S2). Also, there is no difference between the genes whose deletions have a lethal effect and non-essential genes in terms of family size distribution (Wilcoxon rank test, $P = 0.1$).

Table S2

	Singletons	Duplicates	Total
genes whose deletion has a lethal effect	268	796	1064
Non-Essential Genes	376	1167	1543
Total	644	1963	2607

Similarly, we also found that the proportion of singletons does not differ between the genes whose deletions cause infertility and non-essential genes (26.4% vs. 24.4%, $\chi^2 = 0.8$, $P = 0.4$, Table S3). Moreover, there is no difference between the genes whose deletion causes infertility and non-essential genes in terms of family size distribution (Wilcoxon rank test, $P = 0.2$).

Table S3

	Singletons	Duplicates	Total
genes whose deletion causes infertility	128	357	485
Non-Essential Genes	376	1167	1543
Total	504	1524	2028

Analysis of Gene Duplicability of Mouse Singleton Genes in the Human Genome

Using a similar approach as in He and Zhang et al. (Ref. 25), we focused on mouse singletons by examining the duplicability of their orthologs in the human genome. We found that the essential-gene proportion of mouse singletons that were duplicated in human is higher than that of mouse singletons that were not duplicated (58.7% vs. 47.9%, $P < 0.09$, one-tailed, Table S4), suggesting that in mammals important genes tend to be duplicated more often than unimportant genes.

Table S4

	Duplicated in Human	Not Duplicated in Human	Total
Mouse Essential Singleton Genes	27	328	355
Mouse Non-Essential Singleton Genes	19	357	376
Total	46	685	731

Potential Biases and Caveats in Our Analyses

Our current analyses are limited by several factors. First, the human protein-protein interaction data is far from being complete, which might bias our results to some extent. Second, it is known that there is considerable noise (i.e. a high false positive rate) in high throughput protein interaction data. However, the consistent results across various datasets suggest that this factor has no serious effects. Third, phenotypic data might be limited by current annotation methods, and biased by discovery-oriented studies. Finally, to study the correlation between protein connectivity and gene essentiality, we inferred the fitness effect of human genes based on the human-mouse orthology, which might bring additional noise. Some gene deletions in mouse have no apparent phenotype, but have phenotypic effects in human. In the future, when better gene and protein interaction data become available, these analyses should be repeated to check the validity of our conclusions.