## Supplementary data

## Distinct Properties of Human Disease Genes in Protein Interaction Networks

James J. Cai<sup>1,2,\*</sup>, Elhanan Borenstein<sup>3,4</sup> and Dmitri A. Petrov<sup>1</sup>

<sup>1</sup>Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA <sup>2</sup>Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA <sup>3</sup>Department of Genome Sciences, University of Washington, 1705 NE Pacific St, Seattle, WA 98195, USA

<sup>4</sup>Santa Fe Institute

\*Corresponding author: James J. Cai (jcai@tamu.edu)

Figure S1. Integrated protein interaction network. Interaction data was obtained from the study of (Bossi and Lehner 2009). (**A**) The component consists of 10,042 proteins and 80,543 connections. The layout was done using Pajek (http://pajek.imfm.si/). Mendelian disease genes in hOMIM data set (Blekhman et al. 2008) are depicted as red dots. (**B**) Log-log plot for frequency of degree P(k) and degree k. Estimated  $\gamma = 2.1$ .





Figure S2. . (*A*) Phylostratigraphy of nondisease, Mendelian and complex disease genes. The number of genes belong to 19 different phylostrata is plotted. Original phylostrata obtained from the study of (Domazet-Loso and Tautz 2008) were combined into super-phylostrata in bold. The species tree was reproduced with permission from Domazet-Loso and Tautz. (*B*) Number of genes in six combined age groups.





Figure S3. Number of genes in three disease gene sets. Note that the numbers given here are before genes were mapped onto the PPI network.



Figure S4. Interactions between neighboring nodes of broker and non-broker proteins. In each subfigure, the protein itself and linkers between the protein and its neighboring nodes are removed. Only the linkers between neighboring nodes are plotted. (**A**) SUMO4 and (**B**) PRKCZ are two examples of broker and disease genes; (**C**) PCBP1 and (**D**) BMS1 are two examples of non-broker and nondisease genes.





Figure S5. Network metrics as a function of gene age: Mendelian disease genes versus nondisease genes.

6



Figure S6. Network metrics as a function of gene age: complex disease genes versus nondisease genes.



Figure S7. Network metrics as a function of gene age: GWAS genes versus nondisease genes.



Figure S8. Distribution of age of disease and nondisease genes.

Table S1. Named and unnamed genes in disease and nondisease gene sets. Note that the numbers given here are before filtering and before genes were mapped onto the PPI network.

	Nondisease	Mendelian	Complex	GWAS
Named	7706	641	758	585
Unnamed	419	6	19	3

## References

- Blekhman, R., O. Man, L. Herrmann, A. R. Boyko, A. Indap, C. Kosiol, C. D. Bustamante, K. M. Teshima, and M. Przeworski. 2008. Natural Selection on Genes that Underlie Human Disease Susceptibility. Curr Biol 18:883-889.
- Bossi, A., and B. Lehner. 2009. Tissue specificity and the human protein interaction network. Mol Syst Biol **5**:260.
- Domazet-Loso, T., and D. Tautz. 2008. An ancient evolutionary origin of genes associated with human genetic diseases. Mol Biol Evol **25**:2699-2707.