The American Journal of Human Genetics, Volume 93

## **Supplemental Data**

## **Dissecting Disease Inheritance Modes**

### in a Three-Dimensional Protein Network

### Challenges the "Guilt-by-Association" Principle

Yu Guo, Xiaomu Wei, Jishnu Das, Andrew Grimson, Steven M. Lipkin, Andrew G. Clark, and Haiyuan Yu

A. Remove proteins with >20 interaction partners B. Remove proteins with >15 interaction partners

#### C. Remove proteins with >10 interaction partners



# Figure S1. Distribution of Recessive and Dominant Disease Mutations with Respect to Interaction Interfaces after Removal of Protein Hubs

Error bars represent 95% confidence intervals of odds ratios.  $**p < 10^{-20}$ ,  $*p < 10^{-10}$ . The p values are calculated using Z-tests for log odds ratio. Red: in-frame mutations. Blue: truncating mutations.

A. Remove proteins with >20 interaction partners B. Remove proteins with >15 interaction partners

#### C. Remove proteins with >10 interaction partners

#### **Dominant Mutations**





| 100% - |   |                                       |                                     |
|--------|---|---------------------------------------|-------------------------------------|
| 80% -  |   |                                       |                                     |
| 60% -  |   |                                       |                                     |
| 40% -  |   |                                       |                                     |
| 20% -  |   |                                       |                                     |
| 0% 1   |   |                                       |                                     |
|        | In the same<br>interaction<br>interface | In other<br>interaction<br>interfaces | In any<br>interaction<br>interfaces |





| 100% |   |                                       |                                     |
|------|---|---------------------------------------|-------------------------------------|
| 80%- |   |                                       |                                     |
| 60%- |   |                                       |                                     |
| 40%- |   |                                       |                                     |
| 20%- |   |                                       |                                     |
| 0% - |   |                                       |                                     |
|      | In the same<br>interaction<br>interface | In other<br>interaction<br>interfaces | In any<br>interaction<br>interfaces |

#### **Recessive Mutations**









# Figure S2. Analysis of Locus Heterogeneity among Dominant and Recessive Disease Mutations after Removal of Protein Hubs

Error bars represent ±SE.  $**p < 10^{-20}$ . The p values are calculated using cumulative binomial tests. Red: in-frame mutations. Blue: truncating mutations.

A. Remove domains with >60 interaction partners on average B. Remove domains with >40 interaction partners on average C. Remove domains with >20 interaction partners on average

Dominant Mutations



## Figure S3. Distribution of Recessive and Dominant Disease Mutations with Respect to Interaction Interfaces after Removal of Domain Hubs

Error bars represent 95% confidence intervals of odds ratios.  $**p < 10^{-20}$ ,  $*p < 10^{-10}$ . The p values are calculated using Z-tests for log odds ratio. Red: in-frame mutations. Blue: truncating mutations. Note that the enrichment of dominant mutations in other domains decreased after the removal of domain hubs, suggesting that this enrichment might be due to over-represented domains in the 3D protein interactome network.

A. Remove domains with >60 interaction partners on average B. Remove domains with >40 interaction partners on average

**Dominant Mutations** 

C. Remove domains with >20 interaction partners on average

#### 100% Fraction causing the same disease 80% 60% 40% 20% 0% In other In the same In any interaction interaction interaction interface interfaces interfaces







| 100% |  |      |
|------|--|------|
| 80%  |  |      |
| 60%· |  |      |
| 40%  |  |      |
| 20%  |  |      |
| 0%·  |  | <br> |
|      |  |      |

| 0% -   | In the same<br>interaction | In other<br>interaction | In any<br>interaction |
|--------|----------------------------|-------------------------|-----------------------|
| 20%    |                            |                         |                       |
| 40% -  |                            |                         |                       |
| 60%·   |                            |                         |                       |
| 80% •  |                            |                         |                       |
| 100 /0 |                            |                         |                       |

**Recessive Mutations** 









Figure S4. Analysis of Locus Heterogeneity among Dominant and Recessive Disease Mutations after Removal of Domain Hubs

Error bars represent ±SE.  $**p < 10^{-20}$ . The p values are calculated using cumulative binomial tests. Red: in-frame mutations. Blue: truncating mutations.



В



# Figure S5. Analysis of the Effect of Mutation Location on Locus Heterogeneity of Dominant Disease Mutations

(A) Percentage of dominant truncating mutations located on different parts of the protein that cause the same disease with mutations on its interaction partner.

(B) Percentage of dominant in-frame mutations located on different parts of the protein that cause the same disease with mutations on its interaction partner.

#### A Dominant mutations



#### **B** Recessive mutations



#### Figure S6. Effect of Mutation Location Relative to the N Terminus on the Mutation Distribution Patterns

(A) Odds ratios of the distributions of dominant truncating (left) and in-frame (right) mutations on different locations of proteins.

(B) Odds ratios of the distributions of recessive truncating (left) and in-frame (right) mutations on different locations of proteins. Error bars represent 95% confidence intervals of odds ratios.  $**p < 10^{-20}$ ,  $*p < 10^{-5}$ . The p values are calculated using Z-tests for log odds ratio.



В



# Figure S7. Distribution of Recessive and Dominant Truncating Mutations outside of Interaction Interfaces

(A) Odds ratios of the distribution of dominant truncating mutations outside of interaction interfaces.

(B) Odds ratios of the distribution of recessive truncating mutations outside of interaction interfaces. Error bars represent 95% confidence intervals of odds ratios.  $**p < 10^{-20}$ ,  $*p < 10^{-10}$ . The p values are calculated using Z-tests for log odds ratio.

#### A Dominant mutations



**B** Recessive mutations



## Figure S8. Distribution of Recessive and Dominant Disease Mutations from HGMD with Respect to Interaction Interfaces

(A) Odds ratios of the distributions of dominant in-frame (left) and truncating (right) mutations on different locations of proteins.

(B) Odds ratios of the distribution of recessive in-frame (left) and truncating (right) mutations on different locations of proteins. Error bars represent 95% confidence intervals of odds ratios.  $**p < 10^{-20}$ , \*p < 0.05. The p values are calculated using Z-tests for log odds ratio.

#### A In-frame mutations



#### B Truncating mutations



## Figure S9. Analysis of Locus Heterogeneity among Dominant and Recessive HGMD Disease Mutations

(A) Percentage of recessive (left) or dominant (right) in-frame mutation pairs on two different proteins causing the same disease.

(B) Percentage of recessive (left) or dominant (right) truncating mutation pairs on two different proteins causing the same disease. Error bars represent  $\pm$ SE. \*\*p < 10<sup>-20</sup>. The p values are calculated using cumulative binomial tests.



Figure S10. PolyPhen-2 Predictions on the Missense Mutations Used for the Analyses (HumVar Model)

Table S1. Network Statistics of the 3D Protein Interactome Network

| Average Degree | Clustering Coefficient | <b>Characteristic</b> | Path | Diameter |
|----------------|------------------------|-----------------------|------|----------|
|                |                        | Length                |      |          |
| 2.73           | 0.216                  | 8.7                   |      | 27       |

### Table S2. Sample Sizes Used in the Calculations

| Looding of Mutations      | Dominant |            | Recessive |            |  |
|---------------------------|----------|------------|-----------|------------|--|
| Location of Withauons     | In-Frame | Truncating | In-Frame  | Truncating |  |
| In interaction interfaces | 4421     | 1398       | 3620      | 2119       |  |
| In other domains          | 764      | 355        | 410       | 188        |  |
| Outside domains           | 2339     | 2013       | 1122      | 1243       |  |

**Figure 2. Distribution of Dominant and Recessive Mutations** 

### Figure 3A. Locus Heterogeneity of Recessive In-Frame Mutations

| Location of Mutation Pa<br>Interacting Proteins | airs on the             | Same Disease | Different Diseases |
|---|-------------------------|--------------|--------------------|
| <b>•</b>  | Same interface          | 2145         | 296                |
| interacting genes                               | Different<br>interfaces | 3158         | 11882              |
| Non-interacting genes                           | Any<br>interfaces       | 69832        | 5257750            |

### Figure 3A. Locus Heterogeneity of Dominant In-Frame Mutations

| Location of Mutation Pa<br>Interacting Proteins | airs on the          | Same Disease | Different Diseases |
|---|----------------------|--------------|--------------------|
| Interacting cones                               | Same interface       | 7138         | 63539              |
| interacting genes                               | Different interfaces | 2490         | 21066              |
| Non-interacting genes                           | Any<br>interfaces    | 154193       | 8273199            |

### Figure 3B. Locus Heterogeneity of Recessive Truncating Mutations

| Location of Mutation Pa<br>Interacting Proteins | airs on the          | Same Disease | Different Diseases |
|---|----------------------|--------------|--------------------|
| Interacting gapos                               | Same interface       | 307          | 197                |
| Interacting genes                               | Different interfaces | 484          | 3569               |
| Non-interacting genes                           | Any<br>interfaces    | 20010        | 1758000            |

|          | -           |       |           |         |        |         |     |         |          |       |
|----------|-------------|-------|-----------|---------|--------|---------|-----|---------|----------|-------|
| Figure   | 2D          | Loona | Hotomog   | amaitre | of D   | aminant | Two | oting   | Mutati   | 0.000 |
| righte   | <b>.D</b> . | LOCHS | neieroy   | enenv   | 01 170 | линани  |     | 'ALLIN' | VIIIAII  | ons.  |
| - igui v | ·           | Locus | THE COLOR | enerey  | ~ ~    |         |     |         | TITTTTTT |       |

| Location of Mutation P<br>Interacting Proteins | airs on the          | Same Disease | Different Diseases |
|--|----------------------|--------------|--------------------|
| Interacting cones                              | Same interface       | 6            | 1282               |
| Interacting genes                              | Different interfaces | 1            | 168                |
| Non-interacting genes                          | Any<br>interfaces    | 4076         | 864499             |

| Location of Mutation Pa<br>Interacting Proteins | airs on the          | Same Disease | Different Diseases |
|---|----------------------|--------------|--------------------|
| Interacting gapos                               | Same interface       | 1909         | 7275               |
| Interacting genes                               | Different interfaces | 772          | 5711               |
| Non-interacting genes                           | Any<br>interfaces    | 25823        | 1233426            |

Figure 4A. Locus Heterogeneity of Haploinsufficient (HI) In-Frame Mutations

### Figure 4A. Locus Heterogeneity of non-HI In-Frame Mutations

| Location of Mutation Pairs on the<br>Interacting Proteins |                      | Same Disease | Different Diseases |
|---|----------------------|--------------|--------------------|
| Interacting genes   | Same interface       | 2360         | 22805              |
|   | Different interfaces | 938          | 8014               |
| Non-interacting genes                                     | Any<br>interfaces    | 42080        | 2902891            |

### Figure 4B. Distribution of HI vs. non-HI Truncating Mutations

| Location of Mutations     | HI  | non-HI |
|---------------------------|-----|--------|
| In interaction interfaces | 516 | 882    |
| In other domains          | 113 | 242    |
| Outside domains           | 401 | 1611   |

| Elemente 54 Leane | TI at a ma man aiter | of Trans oction of | Decerting          | Mutationa in Thinda   |
|-------------------|----------------------|--------------------|--------------------|-----------------------|
| rigure 5A. Locus  | s Heterogeneity (    | of fruncating      | <b>Kecessive</b> . | viutations in 1 nirus |

| Location of Mutation Pairs on the<br>Interacting Proteins |                                | Same Disease | Different Diseases |
|---|--------------------------------|--------------|--------------------|
| First third   | In<br>interaction<br>interface | 170          | 150                |
|   | Others                         | 277          | 1828               |
| Second third  | In<br>interaction<br>interface | 242          | 114                |
|   | Others                         | 160          | 1127               |
| Last third  | In<br>interaction<br>interface | 208          | 182                |
|   | Others                         | 115          | 367                |

| Location of Mutation Pairs on the<br>Interacting Proteins |                                | Same Disease | Different Diseases |
|---|--------------------------------|--------------|--------------------|
| First third   | In<br>interaction<br>interface | 936          | 103                |
|   | Others                         | 1429         | 2448               |
| Second third  | In<br>interaction<br>interface | 2193         | 240                |
|   | Others                         | 540          | 1481               |
| Last third  | In<br>interaction<br>interface | 1525         | 309                |
|   | Others                         | 432          | 614                |

Figure 5B. Locus Heterogeneity of In-Frame Recessive Mutations in Thirds

| Figure 6A. Enrichment of Truncating Mutations in between Interacting Interf | aces |
|---|------|
|---|------|

| Figure 6A. Enrichment of Truncating Mutations in b |          |           |  |
|--|----------|-----------|--|
|  | Dominant | Recessive |  |
| Between interacting interfaces                     | 302      | 265       |  |