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

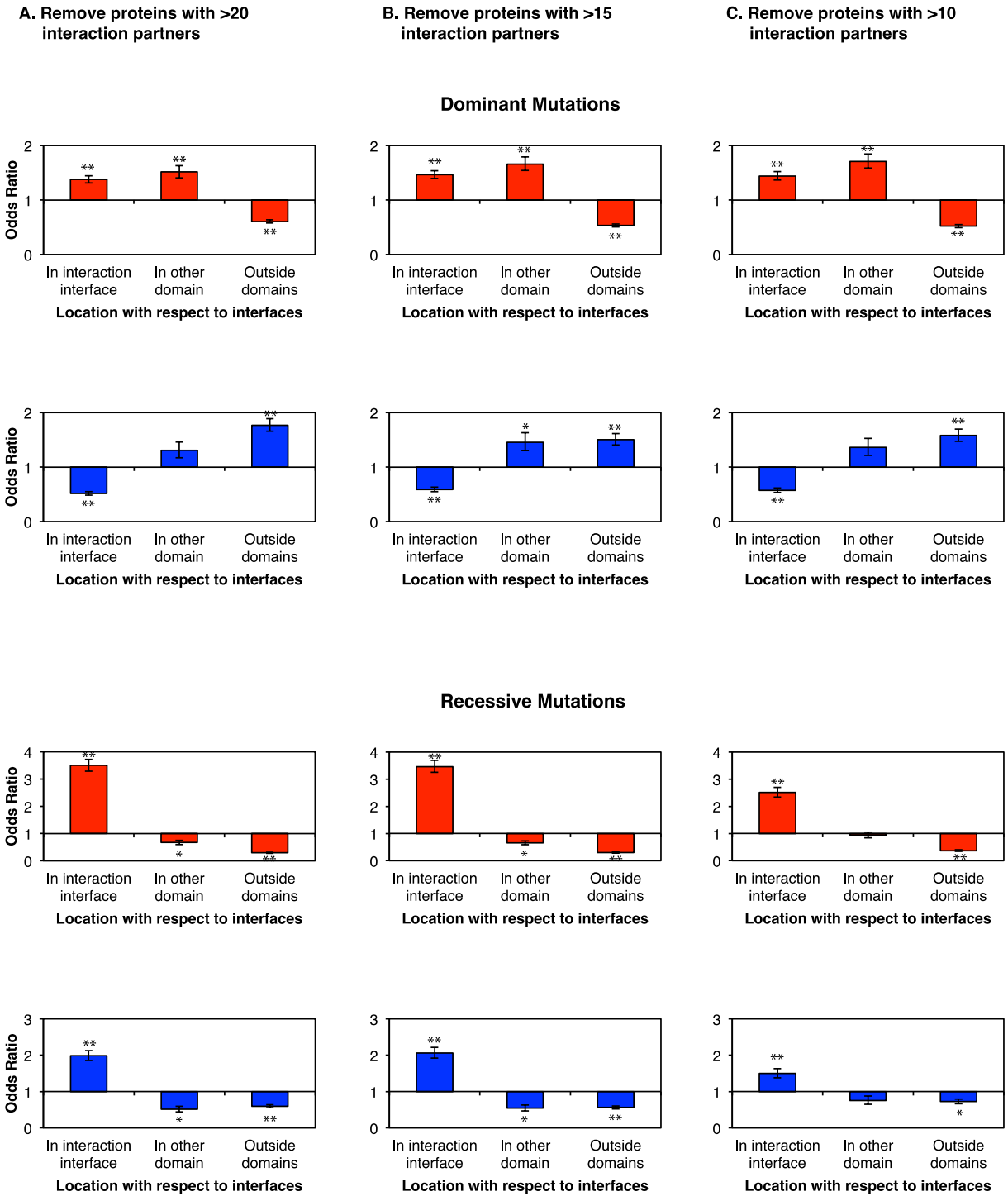


Figure S1. Distribution of Recessive and Dominant Disease Mutations with Respect to Interaction Interfaces after Removal of Protein Hubs

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.

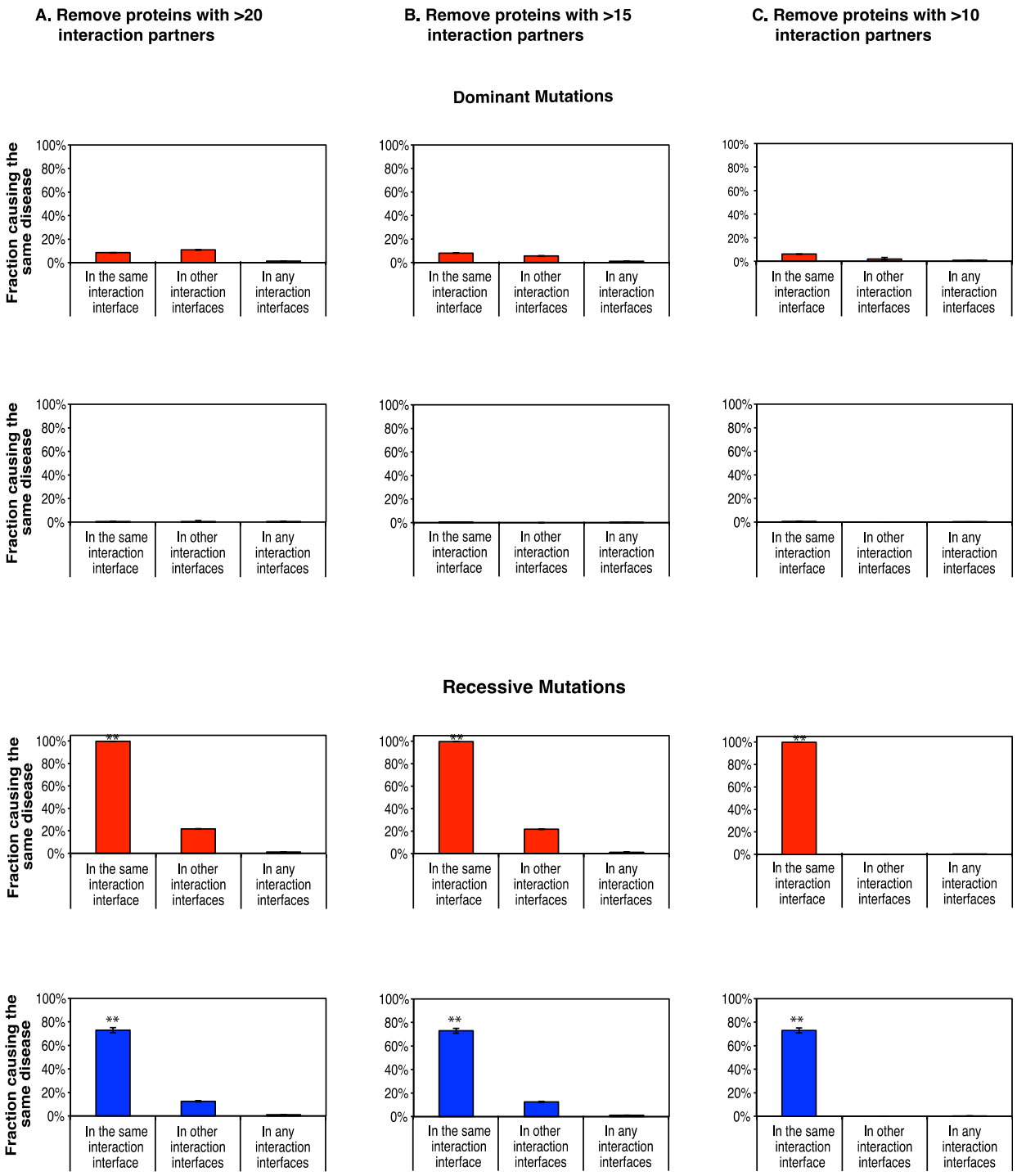


Figure S2. Analysis of Locus Heterogeneity among Dominant and Recessive Disease Mutations after Removal of Protein Hubs

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

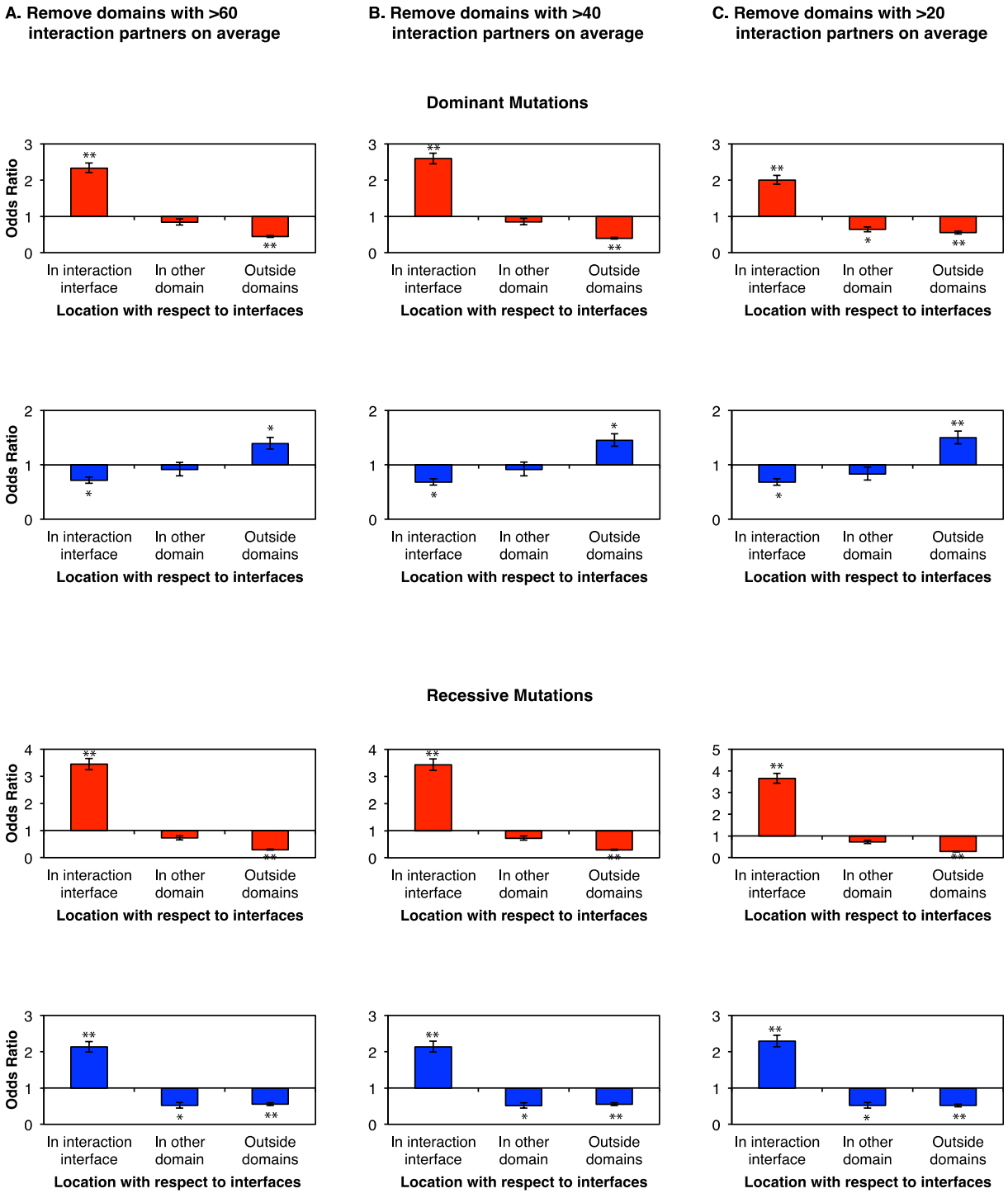


Figure S3. Distribution of Recessive and Dominant Disease Mutations with Respect to Interaction Interfaces after Removal of Domain Hubs

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.

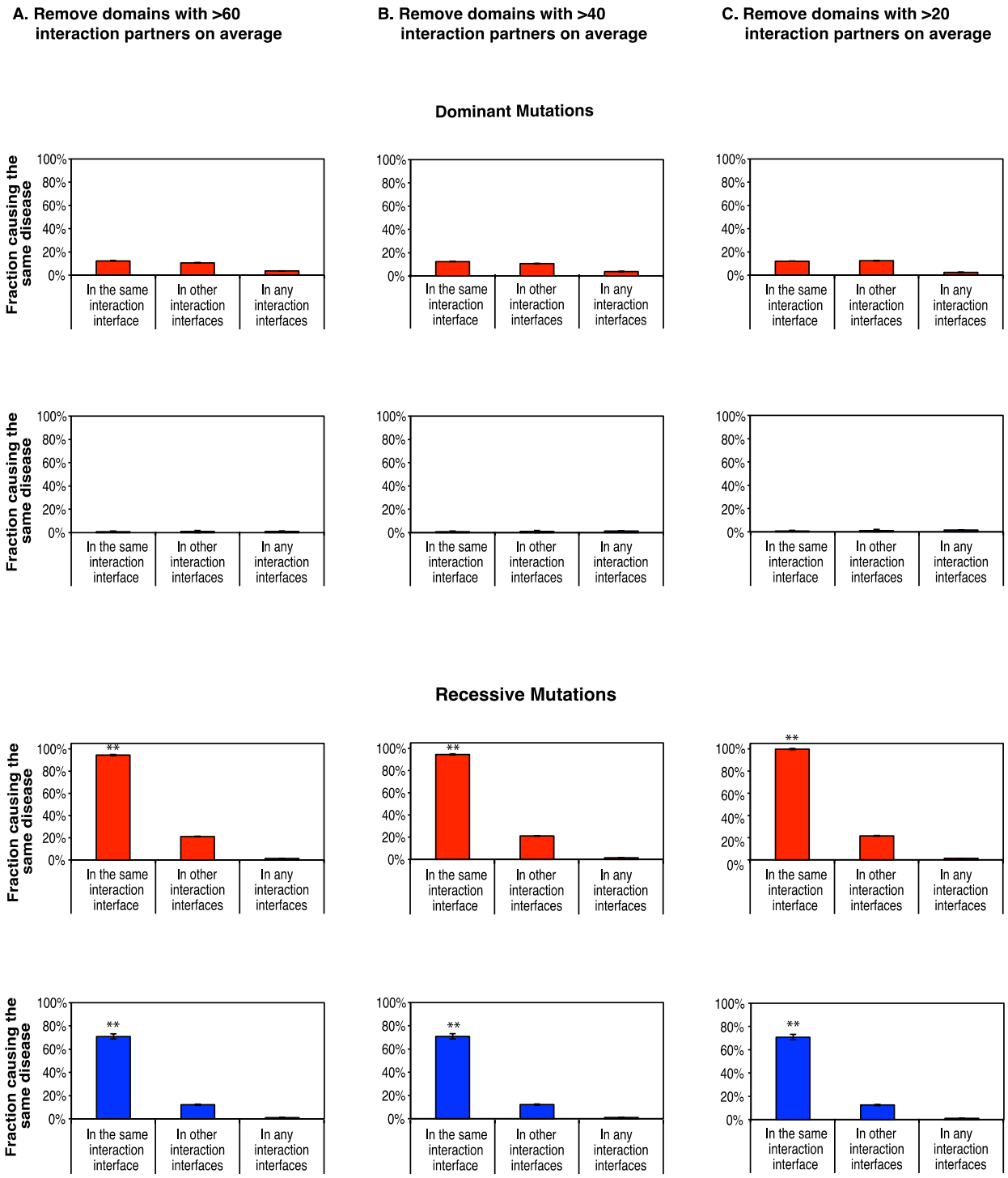
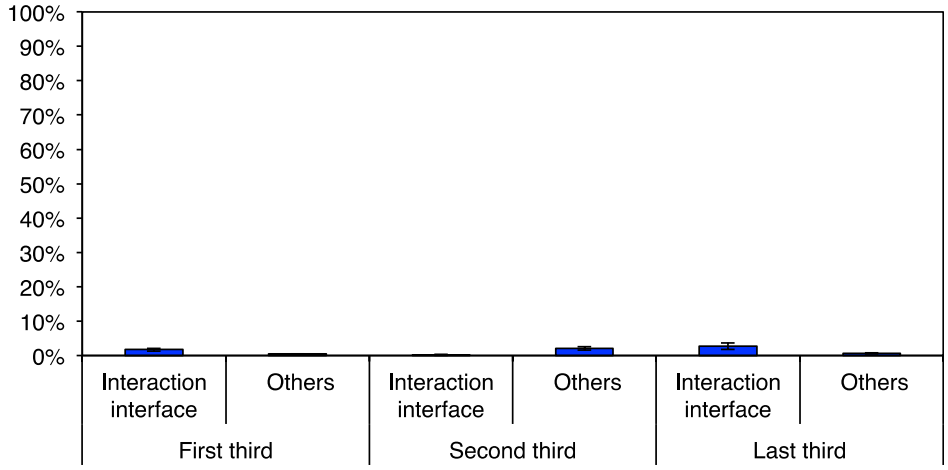


Figure S4. Analysis of Locus Heterogeneity among Dominant and Recessive Disease Mutations after Removal of Domain Hubs

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

A



B

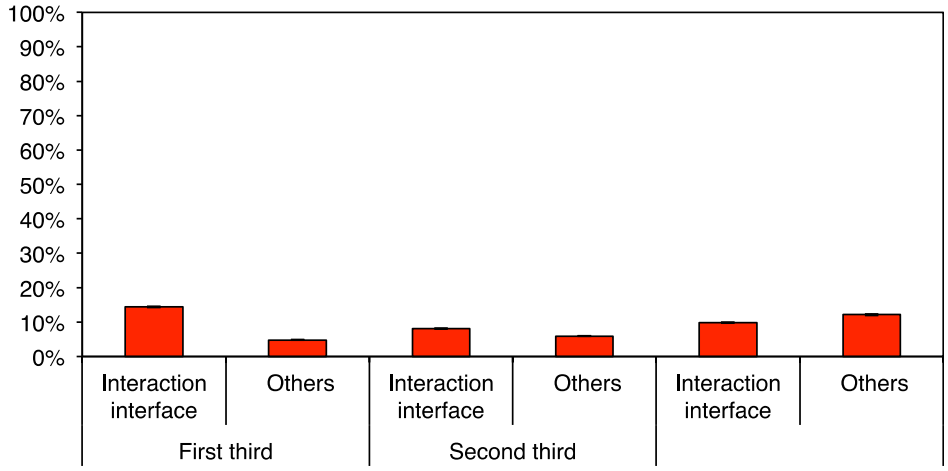


Figure S5. Analysis of the Effect of Mutation Location on Locus Heterogeneity of Dominant Disease Mutations

(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.

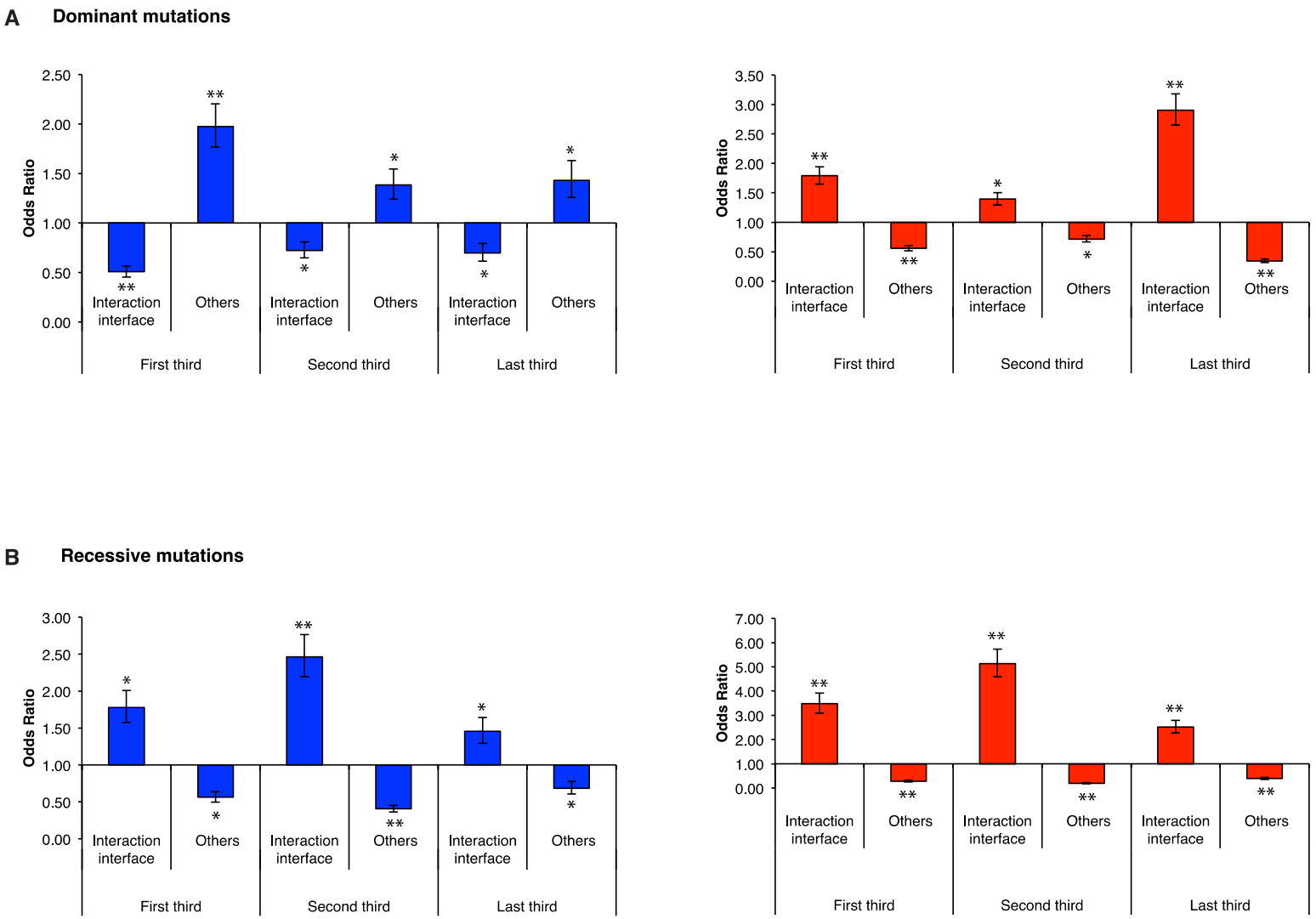
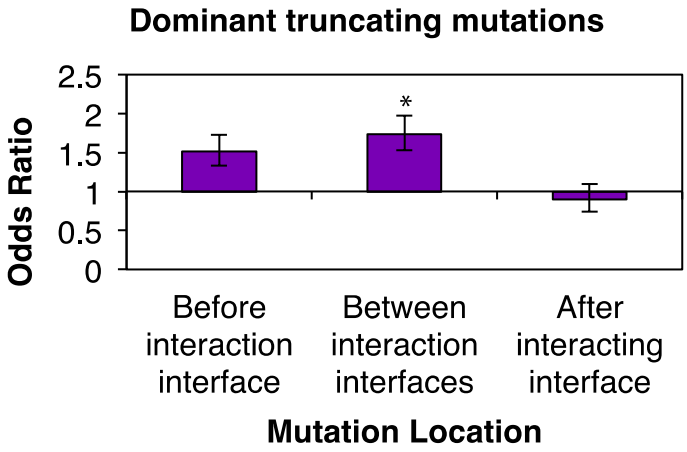


Figure S6. Effect of Mutation Location Relative to the N Terminus on the Mutation Distribution Patterns

(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.

A



B

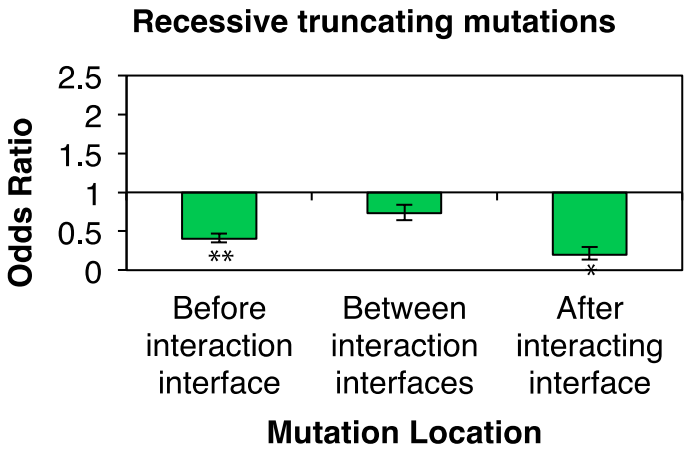


Figure S7. Distribution of Recessive and Dominant Truncating Mutations outside of Interaction Interfaces

(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.

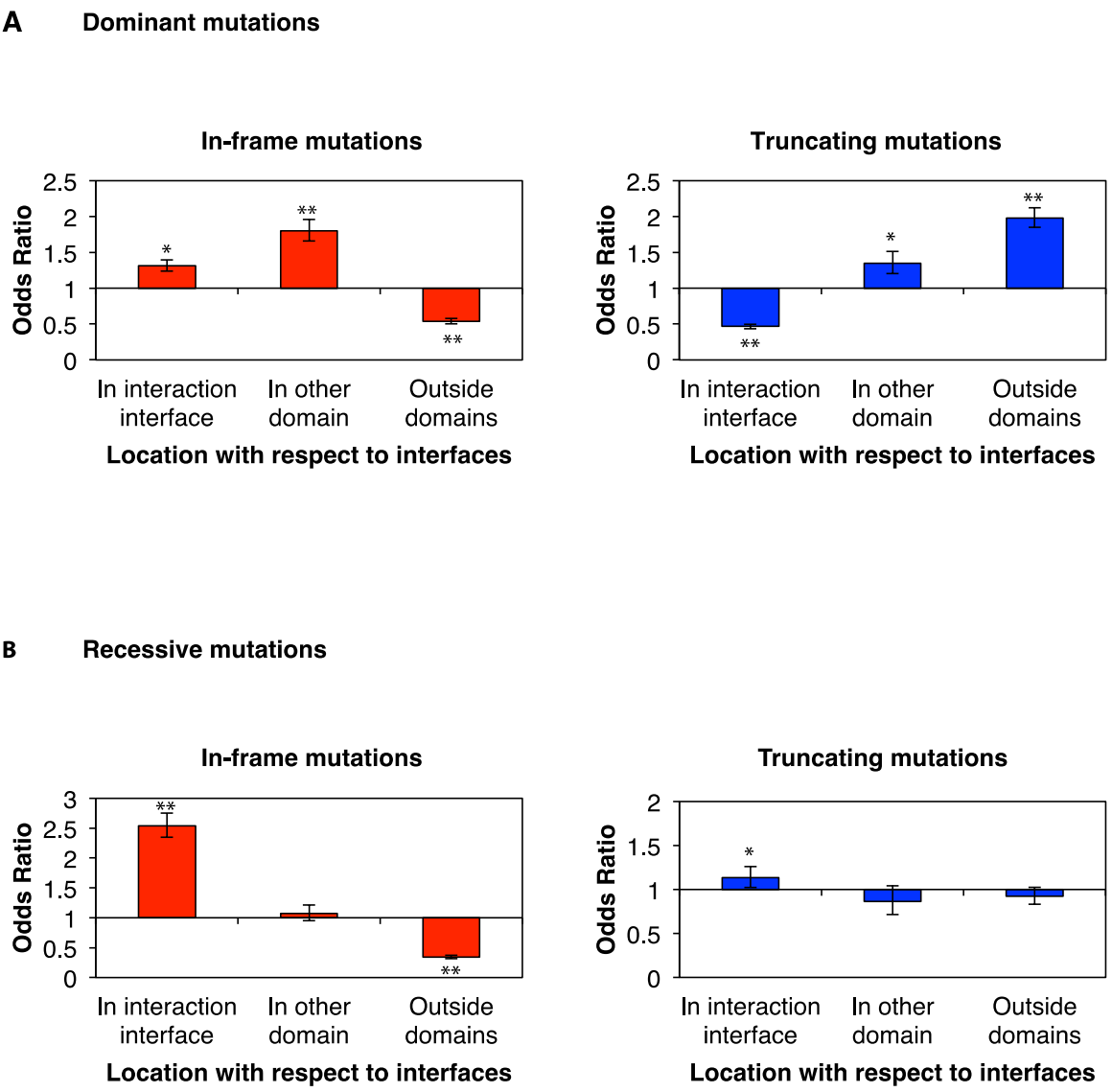


Figure S8. Distribution of Recessive and Dominant Disease Mutations from HGMD with Respect to Interaction Interfaces

(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.

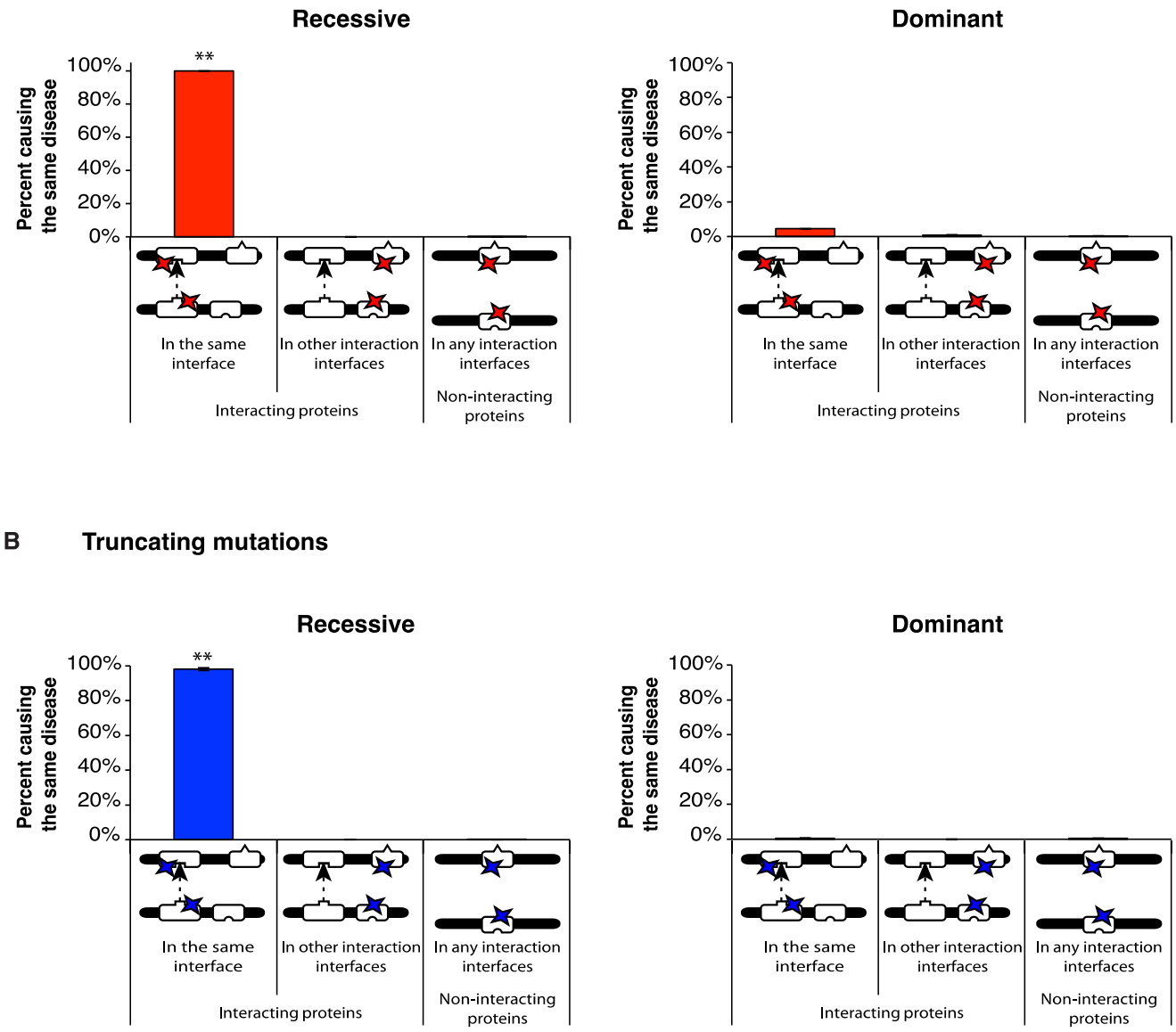


Figure S9. Analysis of Locus Heterogeneity among Dominant and Recessive HGMD Disease Mutations

(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 ±SE. **$p < 10^{-20}$. The p values are calculated using cumulative binomial tests.

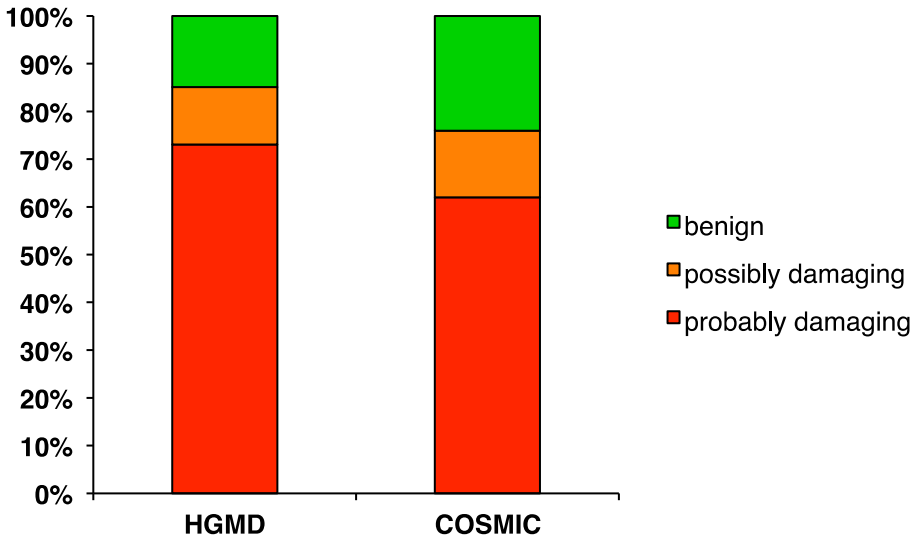


Figure S10. PolyPhen-2 Predictions on the Missense Mutations Used for the Analyses (HumVar Model)

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

Average Degree	Clustering Coefficient	Characteristic Path Length	Diameter
2.73	0.216	8.7	27

Table S2. Sample Sizes Used in the Calculations

Figure 2. Distribution of Dominant and Recessive Mutations

Location of Mutations	Dominant		Recessive	
	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 3A. Locus Heterogeneity of Recessive In-Frame Mutations

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	2145	296
	Different interfaces	3158	11882
Non-interacting genes	Any interfaces	69832	5257750

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

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	7138	63539
	Different interfaces	2490	21066
Non-interacting genes	Any interfaces	154193	8273199

Figure 3B. Locus Heterogeneity of Recessive Truncating Mutations

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	307	197
	Different interfaces	484	3569
Non-interacting genes	Any interfaces	20010	1758000

Figure 3B. Locus Heterogeneity of Dominant Truncating Mutations

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	6	1282
	Different interfaces	1	168
Non-interacting genes	Any interfaces	4076	864499

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

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	1909	7275
	Different interfaces	772	5711
Non-interacting genes	Any interfaces	25823	1233426

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

Location of Mutation Pairs on the Interacting Proteins		Same Disease	Different Diseases
Interacting genes	Same interface	2360	22805
	Different interfaces	938	8014
Non-interacting genes	Any 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

Figure 5A. Locus Heterogeneity of Truncating Recessive Mutations in Thirds

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

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

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

Figure 6A. Enrichment of Truncating Mutations in between Interacting Interfaces

	Dominant	**Recessive**
Between interacting interfaces	302	265