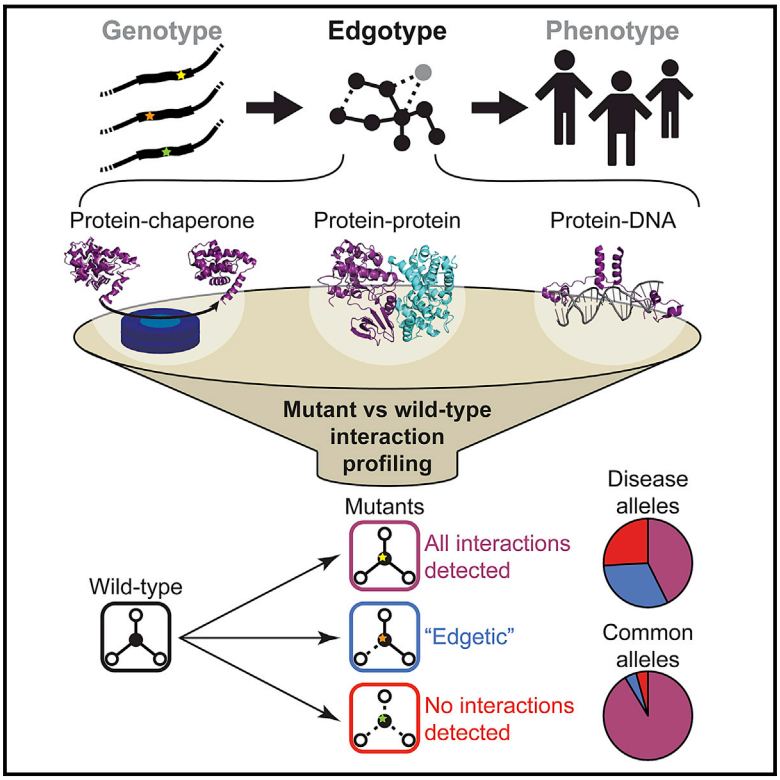
Resource

Cell

Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders

Graphical Abstract



Authors

Nidhi Sahni, Song Yi, ..., Susan Lindquist, Marc Vidal

Correspondence

lindquist@wi.mit.edu (S.L.), marc_vidal@dfci.harvard.edu (M.V.)

In Brief

A large-scale characterization of disease mutations reveals surprisingly widespread, yet specific perturbations in macromolecular interactions. Different mutations in the same gene lead to different interaction profiles, often resulting in distinct disease phenotypes.

Highlights

- Most missense disease mutations appear not to impair protein folding or stability
- Interaction profiling helps distinguish disease mutations from non-disease variants
- Distinct interaction perturbations underlie distinct disease phenotypes
- Integrative interaction networks enhance genotype-to-phenotype understanding

Sahni et al., 2015, Cell *161*, 647–660
April 23, 2015
http://dx.doi.org/10.1016/j.cell.2015.04.013

CellPress

Cell

Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders

Nidhi Sahni,[1,2,3,21] Song Yi,[1,2,3,21] Mikko Taipale,[4,21] Juan I. Fuxman Bass,[5,21] Jasmin Coulombe-Huntington,[6,21] Fan Yang,[1,2,7,8,9] Jian Peng,[10] Jochen Weile,[1,2,7,8,9] Georgios I. Karras,[4] Yang Wang,[1,2,3] István A. Kovács,[1,2,11] Atanas Kamburov,[2,3] Irina Krykbaeva,[4] Mandy H. Lam,[8] George Tucker,[10] Vikram Khurana,[4] Amitabh Sharma,[1,2,11] Yang-Yu Liu,[2,11] Nozomu Yachie,[1,2,7,8,9] Quan Zhong,[2,3] Yun Shen,[1,2,3] Alexandre Palagi,[2,3] Adriana San-Miguel,[2,3] Changyu Fan,[1,2,3] Dawit Balcha,[1,2,3] Amelie Dricot,[1,2,3] Daniel M. Jordan,[12,13] Jennifer M. Walsh,[2,3] Akash A. Shah,[2,3] Xinping Yang,[2,3] Ani K. Stoyanova,[2,3] Alex Leighton,[10] Michael A. Calderwood,[1,2,3] Yves Jacob,[2,3,14] Michael E. Cusick,[1,2,3] Kourosh Salehi-Ashtiani,[2,3] Luke J. Whitesell,[4,15] Shamil Sunyaev,[12,16] Bonnie Berger,[10,17] Albert-László Barabási,[1,2,11,18] Benoit Charloteaux,[1,2,3] David E. Hill,[1,2,3] Tong Hao,[1,2,3] Frederick P. Roth,[1,2,7,8,9,19,22] Yu Xia,[2,6,22] Albertha J.M. Walhout,[2,5,22] Susan Lindquist,[4,15,20,22,*] and Marc Vidal[1,2,3,22,*]

[1]Genomic Analysis of Network Perturbations Center of Excellence in Genomic Science (CEGS), Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[4]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[5]Program in Systems Biology, Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA
[6]Department of Bioengineering, Faculty of Engineering, McGill University, Montreal, QC H3A 0C3, Canada
[7]Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada
[8]Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
[9]Lunenfeld-Tanenbaum Research Institute, Mt. Sinai Hospital, Toronto, ON M5G 1X5, Canada
[10]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[11]Center for Complex Network Research (CCNR) and Departments of Physics, Biology and Computer Science, Northeastern University, Boston, MA 02115, USA
[12]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[13]Program in Biophysics, Harvard University, Cambridge, MA 02139, USA
[14]Département de Virologie, Unité de Génétique Moléculaire des Virus ARN (GMVR), Institut Pasteur, UMR3569, Centre National de la Recherche Scientifique, and Université Paris Diderot, Paris, France
[15]Department of Biology and Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[16]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[17]Department of Mathematics and Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[18]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
[19]Canadian Institute for Advanced Research, Toronto, ON M5G 1Z8, Canada
[20]Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[21]Co-first author
[22]Co-senior author
*Correspondence: lindquist@wi.mit.edu (S.L.), marc_vidal@dfci.harvard.edu (M.V.)
http://dx.doi.org/10.1016/j.cell.2015.04.013

SUMMARY

How disease-associated mutations impair protein activities in the context of biological networks remains mostly undetermined. Although a few renowned alleles are well characterized, functional information is missing for over 100,000 disease-associated variants. Here we functionally profile several thousand missense mutations across a spectrum of Mendelian disorders using various interaction assays. The majority of disease-associated alleles exhibit wild-type chaperone binding profiles, suggesting they preserve protein folding or stability. While common variants from healthy individuals rarely affect interactions, two-thirds of disease-associated alleles perturb protein-protein interactions, with half corresponding to "edgetic" alleles affecting only a subset of interactions while leaving most other interactions unperturbed. With transcription factors, many alleles that leave protein-protein interactions intact affect DNA binding. Different mutations in the same gene leading to different interaction profiles often result in distinct disease phenotypes. Thus disease-associated alleles that perturb distinct protein activities rather than grossly affecting folding and stability are relatively widespread.

INTRODUCTION

Over a hundred thousand genetic variants have been identified across a large number of Mendelian disorders (Amberger et al., 2011), complex traits (Hindorff et al., 2009), and cancer types

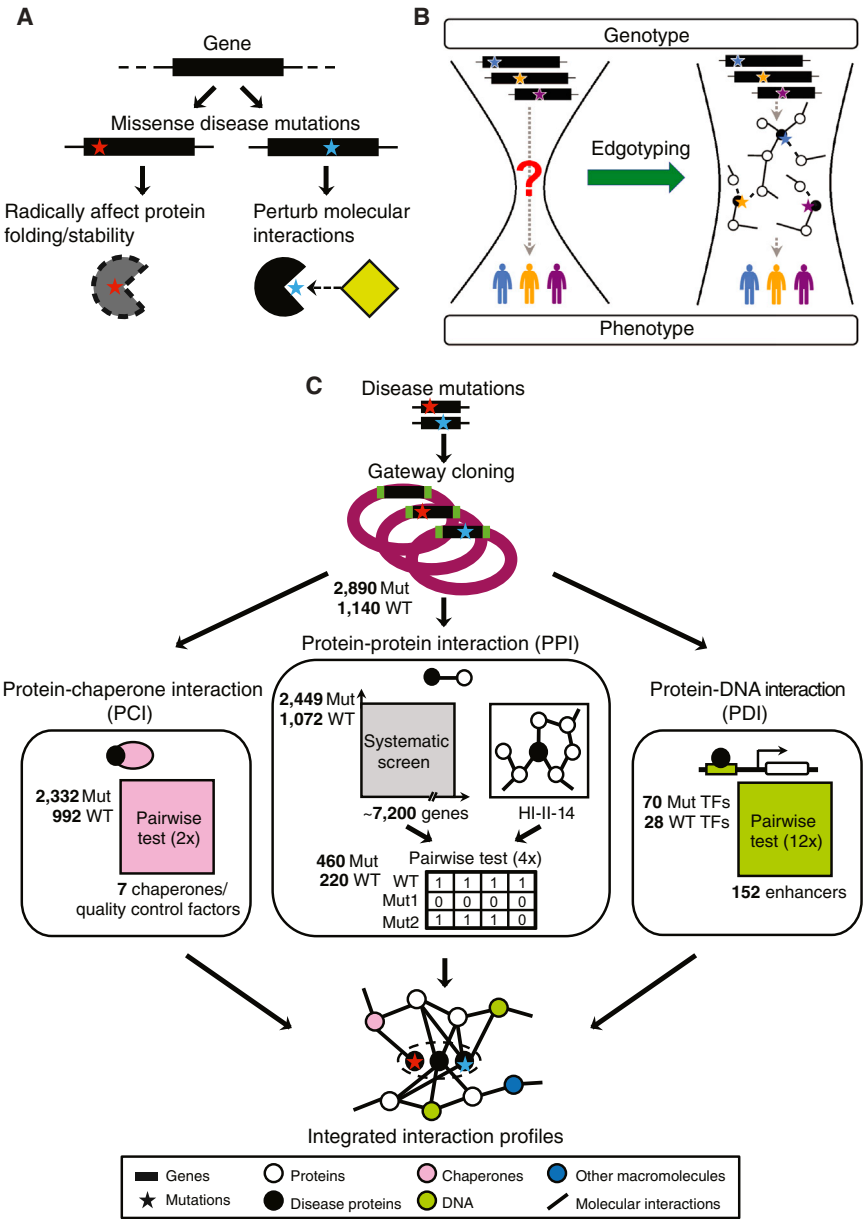


Figure 1. Systematic Characterization of Human Disease Missense Mutations

(A) Two possible effects of missense disease mutations: protein folding/stability changes and molecular interaction perturbations.

(B) Understanding mutational effects by edgotyping links genotype to phenotype. Solid and dashed lines represent retained and perturbed interactions, respectively.

(C) Experimental pipeline for characterizing alterations of molecular interactions, including protein-chaperone (PCI), protein-protein (PPI) and protein-DNA (PDI) interactions. WT: Wild-type, Mut: mutation. TF: transcription factor. "1," detected PPI; "0," not detected PPI. Dashed oval: variants in the same gene. See also Figure S1.

(Chin et al., 2011). However, many fundamental questions regarding genotype-phenotype relationships remain unresolved (Vidal et al., 2011). One critical challenge is to distinguish causal disease mutations from non-pathogenic polymorphisms. Even when causal mutations are identified, the functional consequence of such mutations is often elusive (Sahni et al., 2013).

Genotypic information alone rarely elucidates the mechanistic insights pertaining to disease pathogenesis. Although genotype-phenotype relationships can be modeled under the assumption that most disease-associated mutations lead to complete loss of protein function, e.g., through radical changes such as protein misfolding and instability (Subramanian and Kumar, 2006) (Figure 1A), the reality is often more complex, as in the case of mutations affecting the same gene but giving rise to clinically distinguishable diseases (Zhong et al., 2009). In addition, since genes and gene products do not function in isolation but interact with each other in the context of interactome networks (Vidal et al., 2011), it is likely that many diseases result from perturbations of such complex networks (Goh et al., 2007).

Missense mutations are among the most common sequence alterations in Mendelian disorders, accounting for more than half of all reported mutations in the Human Gene Mutation Database (HGMD) (Stenson et al., 2014). In principle, missense mutations may have no functional consequences, disrupt the three-dimensional structure of the corresponding protein, or exert specific effects on particular molecular or biochemical interactions (Figure 1A), such as protein-protein interactions (PPIs), protein-DNA interactions (PDIs), or enzyme-substrate interactions, while leaving all other functional properties unperturbed. We previously reported that a considerable portion of Mendelian disease mutations could indeed be predicted computationally to cause interaction-specific, or "edgetic," perturbations (Zhong et al., 2009). However, only a small number of genes and associated mutations were experimentally tested in that study, and the extent to which disease mutations globally lead to interaction perturbations remains to be determined.

Here we describe a multi-pronged approach to systematically decipher molecular interaction perturbations associated with missense mutations. Since chaperones and associated quality control factors (QCFs) can salvage unstable proteins by assisting with folding, and an increase in protein-chaperone interactions (PCIs) has been observed for a number of disease mutants (Whitesell and Lindquist, 2005), our systematic approach begins with characterizing PCIs for large numbers of disease-associated alleles, followed by systematic measurements of PPI and

PDI profile changes caused by mutations, a strategy referred to as "edgotyping" (Figure 1B).

We provide evidence for widespread interaction perturbations across a broad spectrum of human Mendelian disorders. Our results suggest that interaction profiling helps distinguish disease-causing mutations from common variants. Furthermore, the integration of different types of molecular interactions expands our ability to understand complex genotype-phenotype relationships.

RESULTS

Human Mutation ORFeome Version 1.1

To globally characterize disease-associated alleles, we selected mutations associated with a wide range of disorders, including cancer susceptibility and heart, respiratory, and neurological diseases. We retrieved from HGMD (Stenson et al., 2014) a list of ~16,400 mutations affecting over 1,200 genes for which we have a wild-type (WT) open-reading frame (ORF) clone in our human "ORFeome" collection (Yang et al., 2011) and selected up to four mutations per gene (Figure 1C; Tables S1A and S1B; Extended Experimental Procedures). Using properties related to RNA abundance, GO annotation, and protein domains (Extended Experimental Procedures), we verified there is no significant bias between our selected genes and the rest of the human genome or all genes represented in HGMD (Figures S1B–S1G).

Altogether, we cloned and sequence-verified 2,890 human mutant ORFs (hmORFs), each harboring a single nucleotide change that results in an amino acid change relative to the corresponding WT ORF of 1,140 genes. To our knowledge, this human mutation ORFeome version 1.1 resource (hmORFeome1.1; Figure S1A) is the most extensive human mutation collection reported to date.

Disease Mutations and Protein Folding and Stability

Using enhanced binding to a chaperone as an indicator of protein instability or misfolding, we examined how disease mutations impact protein folding and disposition. We determined the extent to which hmORF-encoded proteins and their WT counterparts interact with QCFs using a quantitative high-throughput LUMIER assay (Taipale et al., 2012; Taipale et al., 2014) (Figure 1C and Table S2A). We selected the following QCFs based on their broad specificity (Taipale et al., 2014): (1) the cytoplasmic chaperones HSP90 and HSC70, (2) their co-chaperones BAG2 and CHIP/STUB1, (3) the proteasomal regulatory subunit PSMD2 (formerly known as RPN1), and (4) the ER chaperones GRP78/BIP and GRP94 (Extended Experimental Procedures). We did not survey mitochondrial chaperones since only ~7% of disease-associated gene products are predicted to localize solely in mitochondria (Huntley et al., 2015).

Increased interaction between a QCF and mutant or WT protein, as measured by the LUMIER assay, indicates a mutation-induced perturbation in conformational stability, often associated with compromised or complete loss of function (Taipale et al., 2012). The interaction profiles of most mutant proteins correlated with their WT counterparts. However, compared to a background control set, we observed a significant enrichment of mutant alleles showing increased interaction with QCFs (Figures 2A–2H and S2A) but little or no enrichment for decreased interaction (Figures 2A and S2B; Extended Experimental Procedures). The interaction profiles of mutant proteins with the different cytoplasmic QCFs were highly correlated, distinct from those with ER factors (Figure 2I). These results highlight the coordination and specificity of cellular quality control pathways. Altogether ~28% of the tested alleles exhibited increased binding to at least one of the seven QCFs tested. Although this fraction is likely a conservatively low estimate due to limited assay sensitivity, the strong correlation between chaperone interaction profiles (Figure 2I) suggests that the estimate would not increase substantially by assaying more chaperones. We validated several mutant-specific interactions with endogenous chaperones by co-immunoprecipitation followed by western blot, corroborating the results obtained with the LUMIER assay (Figure 2J).

We next estimated protein abundance using semiquantitative ELISA, which provides a proxy for steady-state protein stability. Although the expression levels of mutant alleles correlated with their WT counterparts (Figure S2C), mutant proteins exhibiting enhanced interactions with cytoplasmic, but not ER, chaperones were detected at lower steady-state levels than their WT counterparts ($p < 1.0 \times 10^{-4}$; Figure 3A). This is possibly a result of retention in the ER of mutant proteins that would normally be secreted and therefore not be detected by an assay that captures intracellular proteins. Interestingly, recessive alleles exhibited lower protein abundance levels and increased binding with QCFs compared with proteins encoded by dominant alleles (Figures S2D and S2E). This is consistent with the hypothesis that recessive mutations are more likely to result in loss-of-function phenotypes than dominant mutations (Lesage and Brice, 2009).

To gain insight into the structural properties of mutant proteins that exhibit increased binding to QCFs, we assessed the impact of different disease mutations on predicted protein structures. The disease alleles associated with increased binding to QCFs corresponded significantly more often to mutations of residues buried in the core of the protein (Figure 3B and Table S1C), and less often to mutations in intrinsically disordered regions (Figure 3C) when compared to mutant proteins with no change in binding. Next, we estimated the relative "deleteriousness" associated with distinct genetic mutations using PolyPhen-2 algorithm (Adzhubei et al., 2010). Deleterious mutations predicted by PolyPhen were significantly enriched in alleles that exhibited increased binding to QCFs (Figure S2F).

Previous studies suggested that increased chaperone binding reflects a change in protein stability (Falsone et al., 2004; Taipale et al., 2012). To provide further evidence for this, we assessed protein stability in cellular lysates by measuring solubility in a cellular thermal shift assay (CeTSA). We found that the majority (5 of 6) of mutant proteins with increased chaperone binding also exhibited decreased stability as measured by CeTSA (Figures S3A–S3D). In addition, computational predictions by the FoldX program (Schymkowitz et al., 2005) suggest that mutant proteins with increased binding to QCFs are likely to be significantly less stable than their WT counterpart (Figure 3D and Table S2B). Taken together, experimental and computational analyses

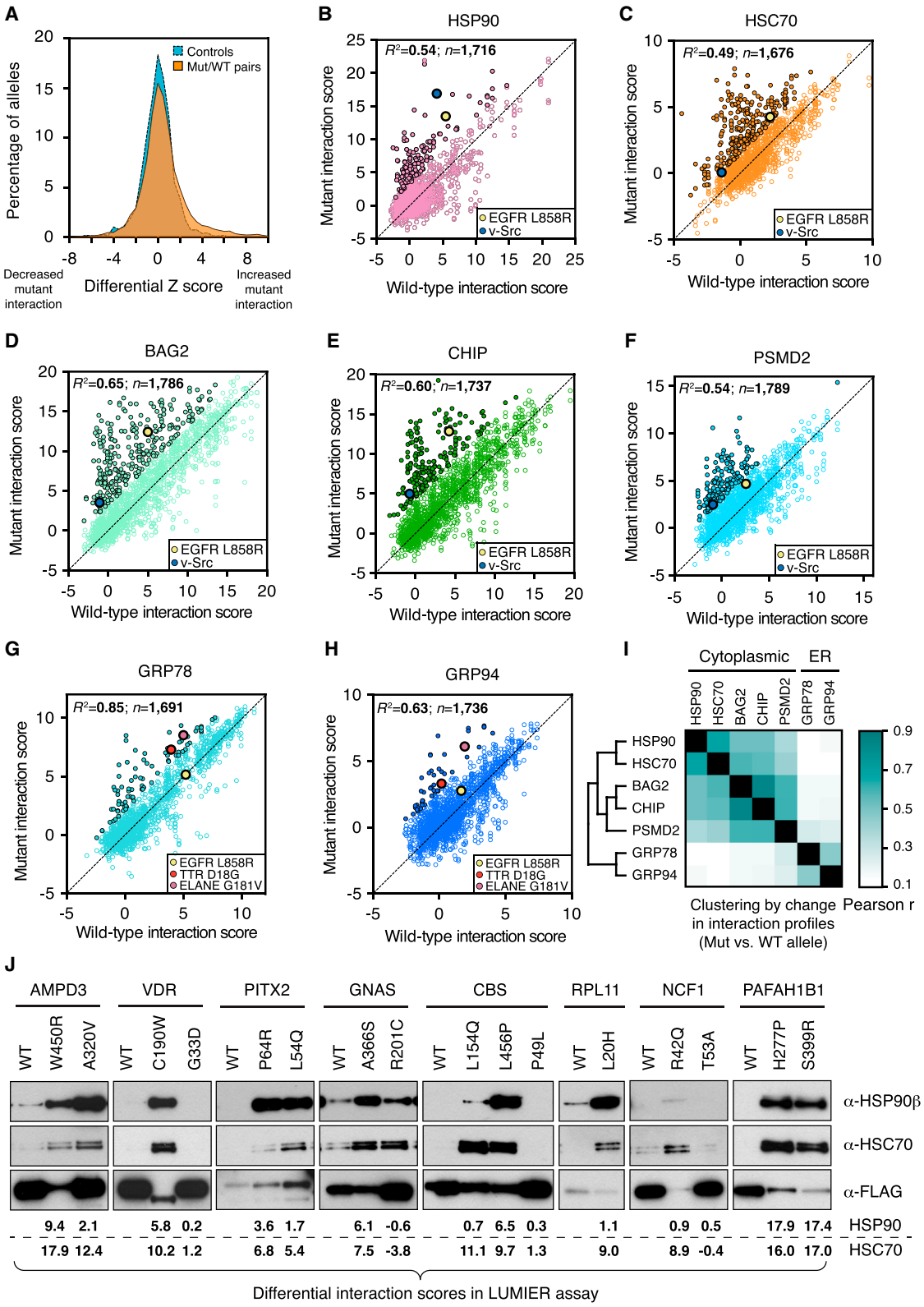


Figure 2. Most Disease Missense Mutations Do Not Impair Protein Folding or Stability

(A) Differential Z score distributions in LUMIER assay. Normalized differential Z scores are calculated as the difference in chaperone binding between all mutant/WT pairs expressed at detectable levels (n = 12,131). Non-expressed pairs serve as controls (n = 1,567).

(legend continued on next page)

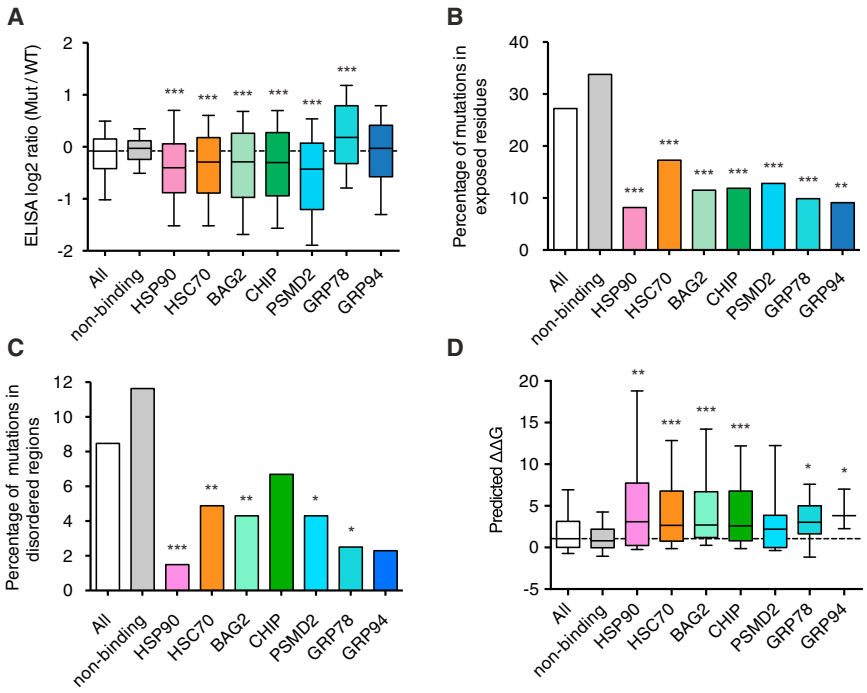


Figure 3. Mutant Proteins with Enhanced Binding to QCFs Are Likely to Be Unstable

(A) Protein expression levels measured by ELISA. x axis shows all tested mutants (All), mutants with no change (non-binding) or an increase in binding to QCFs.

(B) Solvent accessibility of mutant proteins.

(C) Disorder analysis of mutant proteins.

(D) Stability predictions by FoldX. ΔΔG, free energy change.

Dashed line (A and D) represents the median of all mutants. p values (A) and (D) by one-sided Wilcoxon rank sum test; (B) and (C) by Chi-square test. For n values, see Table S7B. *p < 0.05; **p < 0.01; ***p < 0.001. See also Figures S2 and S3.

suggest that mutant proteins with enhanced binding to QCFs have a destabilized protein structure.

Our quantitative survey of allele-specific interactions estimates that the majority of missense disease mutations do not dramatically impact protein structure or folding (Tables S1D and S2). Therefore, they may exert their deleterious effects through other mechanisms such as perturbation of molecular interactions.

Disease Mutations and PPI Perturbations

In principle, the effects of missense disease mutations on molecular interactions (Zhong et al., 2009), or "edgotype" (Sahni et al., 2013), could range from no apparent detectable change in interactions ("quasi-WT"), to specific loss of some interaction(s) ("edgetic"), to an apparent complete loss of interactions ("quasi-null") (Figure 4A). To systematically characterize PPI perturbations associated with disease mutations and identify potential gain of interactions, we used the yeast two-hybrid (Y2H) interaction assay followed by a stringent validation assay. After autoactivator removal, we screened 2,449 mutant proteins and their 1,072 corresponding WT proteins for interactions with proteins encoded by the ~7,200 ORFs in the human ORFeome v1.1 (Rual et al., 2004). Mutant and WT proteins were then tested pair-wise against all partners found both in these Y2H screens and in our human interactome map HI-II-14 (Rolland et al., 2014) (Figure 1C). Altogether, we obtained interaction profiles for 460 mutant proteins and their 220 WT counterparts and found 521 perturbed interactions out of 1,316 PPIs (Table S3A).

To validate these results, we used the orthogonal in vivo *Gaussia princeps* luciferase protein complementation assay (GPCA) performed in human 293T cells (Cassonnet et al., 2011) (Table S3B). Unperturbed interactions were recovered at a rate statistically indistinguishable from that of a well-documented positive reference set (PRS), similar to the interactions of the WT alleles (Braun et al., 2009; Venkatesan et al., 2009). Perturbed interactions were recovered at a rate as low as a negative control "random reference set" (RRS) (Figures 4B and S4A), demonstrating the high quality of the identified perturbations induced by disease mutations.

To analyze global and topological characteristics of gene products with edgetic, quasi-null, or quasi-WT mutations, we used the human interactome map HI-II-14 (Rolland et al., 2014). According to the studied network properties (betweenness, k-core centrality, degree, closeness), the nodes (genes) examined in our edgotyping study appear unbiased, in that their topological properties are statistically indistinguishable from other genes in the network (Figures S4B–S4F). Interestingly, we found that the genes carrying edgetic mutations tend to be more central than either non-edgetic genes or the rest of the network (Table S4).

Out of a total of 197 mutations, corresponding to 89 WT proteins with two or more interaction partners, our interaction profiling identified 26% as quasi-null alleles, 31% edgetic and 43% quasi-WT (Figure 4C and Table S3C). We also analyzed

(B–H) Interaction scatter plots for 2,332 disease alleles. Alleles were assayed for interaction with QCFs HSP90 (B), HSC70 (C), BAG2 (D), CHIP (E), PMSD2 (F), GRP78 (G), and GRP94 (H). EGFR L858R and v-Src can interact with HSP90 (Shimamura et al., 2005; Taipale et al., 2012), and TTR D18G and ELANE G181V can interact with GRP78 (Köllner et al., 2006; Sörgjerd et al., 2006); hence used as controls. Filled circles with black border represent significantly increased chaperone binding. Correlations by Pearson coefficient of determination, R^2.

(I) Clustering analysis based on chaperone interaction profile similarity.

(J) Validation by co-immunoprecipitation (co-IP). LUMIER scores are shown below the blots.

See also Figure S2.

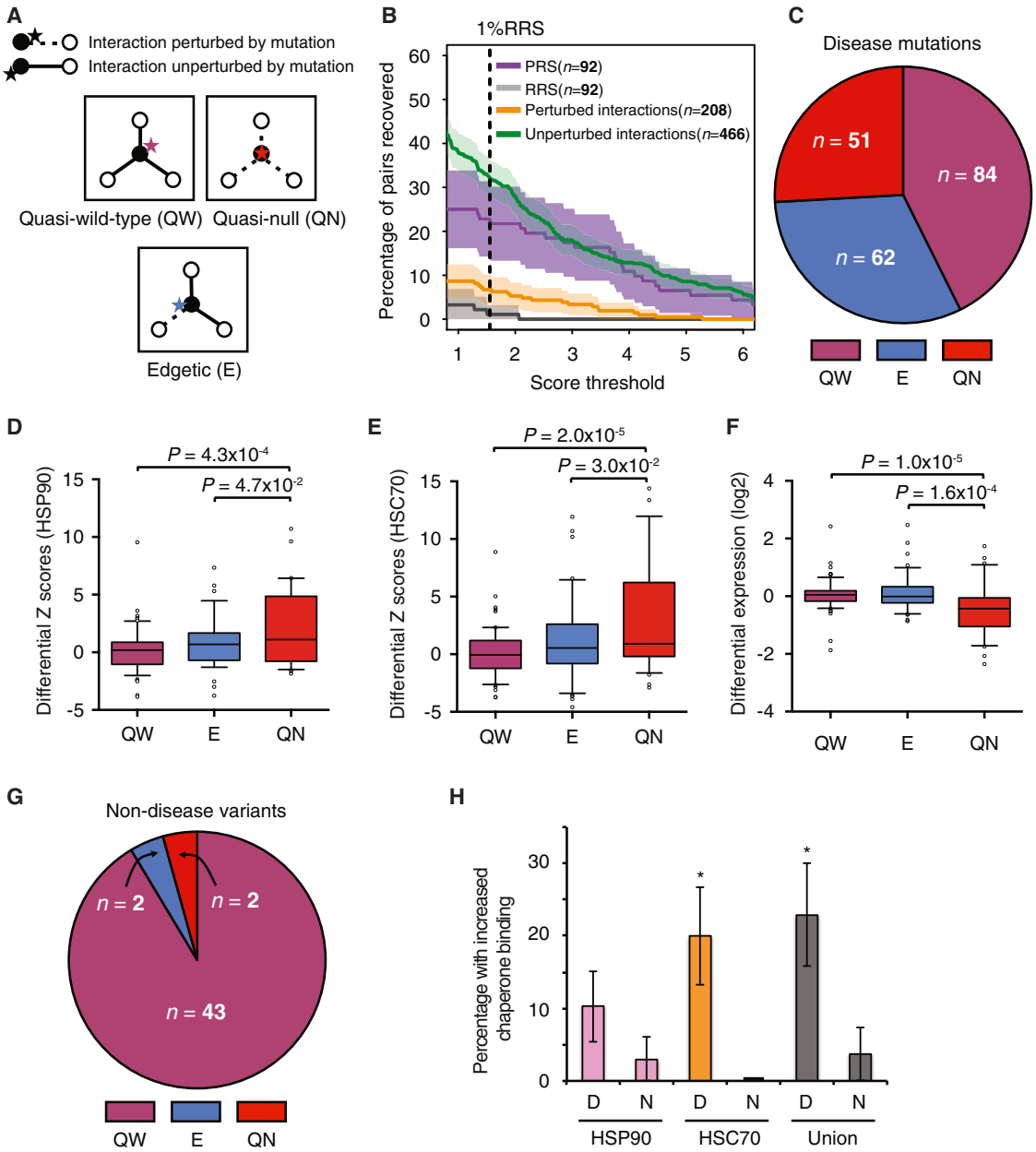


Figure 4. Interaction Perturbation Profiles Distinguish Disease Mutations from Non-Disease Variants

(A) Three classes of PPI profiles (edgotypes) for mutations.

(B) Percentage of protein pairs recovered in GPCA at increasing score thresholds. Shading indicates SE of the proportion.

(C) Distribution of different edgotype classes for disease mutations.

(D and E) Differential LUMIER interaction scores among different edgotype classes, for binding to HSP90 (D) and HSC70 (E). p values by one-sided unpaired t test.

(F) Differential expression among different edgotype classes (ELISA log2 ratio of mutant over WT). QW: n = 75, E: n = 49, QN: n = 42. p values by one-sided Wilcoxon rank sum test.

(G) Distribution of different edgotype classes for non-disease variants.

(H) Increased binding to HSP90, HSC70, or either (Union) for non-disease (N) or disease (D) variant proteins. p values by one-sided Fisher's exact test. Error bars indicate SE of the proportion. *p < 0.05.

See also Figures S4 and S5.

disease mutations annotated by ClinVar (Landrum et al., 2014) and found the distribution of quasi-null, edgetic, and quasi-WT alleles was statistically indistinguishable from that of HGMD (Figure S4G). We only identified two mutations that conferred PPI gains, suggesting that gain of interactions may be a rare event in human disease.

Protein Folding and Expression Levels of Edgetic Mutations

Differences between edgotype classes could be due to protein folding and/or relative expression levels. Quasi-null proteins associated significantly more with cytoplasmic, but not ER, chaperones, whereas edgetic and quasi-WT proteins did not significantly change their chaperone association (Figures 4D–4E, and S5A–S5E). Quasi-null proteins appeared to be poorly expressed, while edgetic and quasi-WT proteins were expressed at levels similar to those of their WT controls (Figure 4F). We validated several mutant-chaperone interactions and expression profiles by co-immunoprecipitation with endogenous chaperones, followed by western blot (Figure S5F). All tested quasi-null proteins exhibited more binding to HSP90 and HSC70, although they were expressed at lower levels than their WT controls. However, the edgetic TAT-P220S protein and the quasi-WT NCF2-R395W protein did not show any detectable chaperone association. Among mutant proteins with no change in chaperone binding, edgetic (28%) and quasi-WT (57%) proteins comprised the majority, while quasi-null proteins comprised a significantly lower percentage (15%) (Figure S5G). Altogether, these results suggest that quasi-null proteins are more often unstable/misfolded and diminished in their steady-state expression levels. In contrast, edgetic and quasi-WT proteins likely exhibit normal folding and expression levels, further supporting the idea that they may cause disease through interaction perturbations or other mechanisms rather than simple loss of protein function.

Disease-Causing Mutations Versus Common Variants

Genome-wide association studies have identified hundreds of loci linked to particular disorders. However, these loci often contain several genes and multiple variants, making it challenging to distinguish causal mutations from non-pathogenic variants. We observed previously that among binary interactions found by WT proteins, disease-causing alleles were more likely to perturb interactions than non-disease variants (Rolland et al., 2014). We further investigated both disease-causing alleles from HGMD and common variants identified in healthy individuals from diverse geographical sites (1000 Genomes Project Consortium, 2012) (Table S1A) with respect to the edgetic character and chaperone binding of their protein products. Interaction profiling showed that only a small fraction of non-disease alleles lost interactions (8%, Figure 4G), a 7-fold reduction relative to disease mutations (57%; $p = 1.7 \times 10^{-9}$; Figure 4C). In addition, non-disease alleles on average did not alter chaperone association (Table S2A), a characteristic distinct from disease mutations annotated by HGMD (Figure 4H) or ClinVar (Figure S5H). Together, interaction perturbations can help distinguish disease-associated alleles from non-disease alleles.

To assess the predictive power of edgotyping to identify disease-causing mutations, we determined its precision and sensitivity in classifying an allele as causal based on interaction perturbation profiles. As a "gold standard" for causal alleles, we used a set of mutations annotated in HGMD as disease-causing ("DM" in Table S1A). As a negative control, we used a set of alleles most likely not associated with disease. We observed that 96% (105 of 109) of the alleles found to perturb interactions (E or QN) were disease-causing (Figure S6A). Conversely, 61% (105 of 172) of disease-causing mutations annotated by HGMD were interaction-perturbing (Figure S6B). Together, our prediction achieved a precision (96%) and sensitivity (61%) significantly higher than random expectation. It is possible that current incompleteness of interaction network maps might limit the power of edgotyping to properly classify disease-causing mutations. To evaluate this possibility, we performed a down-sampling analysis and found negligible effect on mutation classification over a broad range of network sizes (Figure S6C).

Edgetic Mutations and Interaction Interfaces

To explore edgotypes from a structural point-of-view, we assessed the possible impact of distinct classes of mutations on protein function using PolyPhen-2 analysis (Adzhubei et al., 2010). Interaction-perturbing mutations are significantly more often predicted to be deleterious than non-interaction-perturbing mutations (Figure 5A). We next investigated whether mutations from the different classes might differ in evolutionary conservation, based on the presumption that conservation of amino acid residues is a property that generally reflects functionality (1000 Genomes Project Consortium, 2012; Subramanian and Kumar, 2006; Sunyaev, 2012). The residues affected by interaction-perturbing mutations are significantly more conserved across species compared to non-interaction-perturbing mutations (Figure S6D). However, PolyPhen and conservation analysis could not distinguish between edgetic and quasi-null mutations within the interaction-perturbing group.

Given that structural domains often mediate protein interactions, different classes of mutation might vary in their locations relative to protein domains. Interaction-perturbing mutations are indeed significantly enriched within structural domains compared to non-interaction-perturbing alleles (Figure 5B and Table S1C). In addition to structural domains, intrinsically disordered regions and linear motifs could also play a role in mediating PPIs. However, we found interaction-perturbing disease alleles to be depleted in intrinsically disordered regions (Figure S6E), and occurring in linear motifs as frequently as non-perturbing alleles (Figure S6F). These results suggest that mutations perturbing PPIs are preferentially located within structural domains. Nevertheless, none of the above properties could reliably predict whether a mutation would give rise to an edgetic or quasi-null PPI effect.

We next investigated whether edgetic and quasi-null mutations differ in their physical location within three-dimensional protein structures (Zhong et al., 2009). Edgetic mutations are significantly more enriched in structurally exposed residues compared to quasi-null mutations (Figure 5C). Consistently, edgetic mutations do not tend to cause a change in hydrophobicity, a destabilizing feature that generally disrupts protein function (Balasubramanian et al., 2005), while quasi-null mutations often lead to a decrease in hydrophobicity (Figure S6G).

We also investigated whether or not edgetic mutations are more frequently located at an interface that supports interaction with a partner protein. Starting from all available structures of co-crystal complexes in the Protein Data Bank (PDB) involving a disease gene product, we determined the relative location of each mutated residue within these structures (Extended Experimental

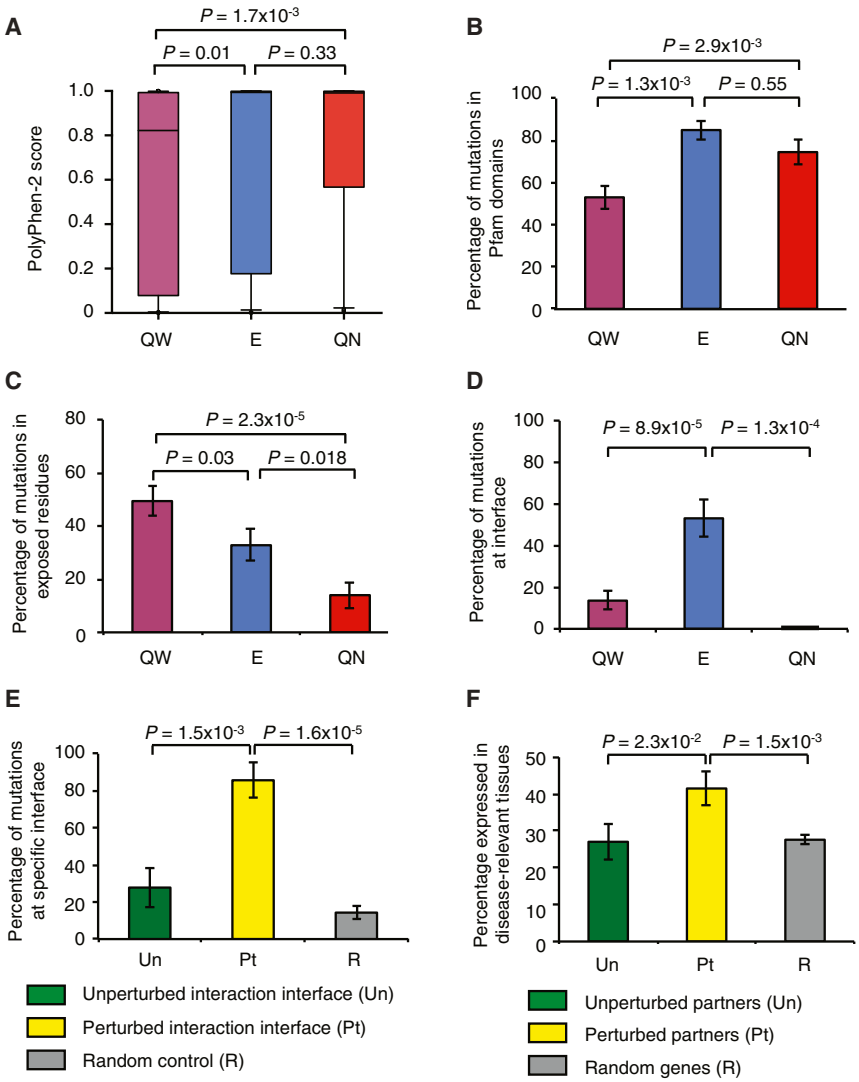


Figure 5. Edgetic Mutations Perturb Interaction Interfaces with Protein Partners Expressed in Disease-Relevant Tissue

(A) PolyPhen-2 scores for mutations in different edgotype classes. p values by one-sided Wilcoxon rank sum test.

(B) Percentage of mutations within Pfam domains. p values by one-sided position-shuffling test.

(C) Percentage of mutations in exposed residues. QW: n = 83; E: n = 61; QN: n = 50.

(D) Percentage of mutations at PPI interfaces. QW: n = 59; E: n = 32; QN: n = 16.

(E) Percentage of interfacial mutations for perturbed (n = 14) and unperturbed (n = 18) interactions, compared to random mutations.

(F) Percentage of perturbed (n = 118) and unperturbed (n = 85) interactors expressed in disease relevant tissues. Thirty random genes from RNA-Seq dataset are assessed for each disease gene. p values from (C) to (F) by one-sided Fisher's exact test. Error bars (B) to (F), SE of the proportion. See also Figure S5.

Procedures and Table S5A). In contrast to quasi-null mutations, edgetic mutations are significantly enriched at interaction interfaces identified from the corresponding co-crystal structures (Figure 5D). Notably, edgetic mutations also exhibit a significant tendency to reside at interaction interfaces with the perturbed partners, as compared to unperturbed partners or random controls (Figure 5E). These results suggest that edgetic mutations are preferentially located at PPI interfaces, perturbing the corresponding interaction.

Edgetic Mutations Perturb Interactions with Protein Partners Expressed in Disease-Relevant Tissues

We hypothesized that protein interaction partners perturbed by edgetic mutations are likely to function together within the tissue known to be affected by the relevant disease. To test this, we compared gene expression patterns for perturbed and unperturbed partners in disease-relevant tissues using RNA-seq data from the Illumina Human Body Map 2.0 project. Perturbed interactors exhibit a striking tendency to be expressed in disease-relevant tissues compared with unperturbed interactors or random genes (Figures 5F and S6H; Table S5B). These results indicate that disease mutations most often perturb interactions that are functionally relevant in the particular tissue(s) affected by a specific disease.

Distinct Interaction Perturbations May Underlie Diverse Disease Phenotypes

Our edgotyping model suggests that different mutations in the same gene may result in different, pleiotropic phenotypic outcomes through perturbation of distinct interactions (Figure 6A). To test this, we compared mutation edgotype classes and the resulting disease phenotypes. Among pleiotropic genes associated with two or more diseases, mutant alleles associated with different disease manifestations were more likely to exhibit different edgotype classes of perturbed PPI profiles (Table S5C).

This is exemplified by mutations in *TPM3*, which encodes slow muscle alpha-tropomyosin. Three *TPM3* edgetic mutations L100M, R168G, and R245G are associated with fiber-type disproportion myopathy through an unknown mechanism (Adzhubei et al., 2010; Clarke et al., 2008) (Figure 6B). These edgetic mutations perturb five of the ten interaction partners of the WT gene product. The majority of perturbed partners are expressed in muscle, the tissue most relevant to this disease (Figure 6C). One of the disrupted interactions is the interaction between TPM3 and troponin, which was shown to be vital for the transduction of calcium-induced signals required for muscle contraction (Gunning et al., 1990). Two other perturbed interactors, HSF2, involved in myotube regeneration (McArdle et al., 2006), and CCHCR1, required for cytoskeleton organization (Tervaniemi

et al., 2012), could also be of disease relevance. In contrast to these edgetic mutations, the quasi-WT mutation M9R causes a different disease, nemaline myopathy. M9R might affect actin binding, thus leading to the formation of abnormal nemaline rods (Laing et al., 1995).

The possible disease relevance of our approach was further illustrated by edgetic mutations in the gene *EFHC1*, mutations in which can cause epilepsy. One perturbed partner, ZBED1, plays a role in a major cell proliferation pathway affected by *EFHC1* knockouts (Yamashita et al., 2007), while another perturbed interactor, TCF4, is required for neuronal differentiation (Flora et al., 2007) (Figure 6D).

We next reasoned that mutations perturbing a greater number of interactions would be likely to have a larger impact on protein function, and hence result in more severe phenotypic effects. We used the age of disease onset as a proxy for severity and determined whether an increase in the fraction of interactions lost correlated with an increase in severity for each pair of mutations causing the same disease (as annotated by HGMD) (Figure 6E and Table S5D). We found that mutations perturbing more PPIs were associated with an earlier age of disease onset significantly more often than random expectation (Figure 6E). Although computational predictions based on PolyPhen-2 were able to distinguish between interaction-perturbing versus non-perturbing alleles (Figure 5A), they did not perform as well as our approach in predicting disease severity (Figure S6I). This limitation is consistent with the inability of PolyPhen-2 to distinguish between edgetic and quasi-null mutations (Figure 5A).

Protein-DNA Interactions

We hypothesized that mutations for which no PPI perturbation has yet been detected likely cause changes in other types of molecular interactions. As a proof-of-concept, we examined the effect of disease mutations on protein-DNA interactions (PDIs) between human transcription factors (TFs) (Reece-Hoyes et al., 2011a) and developmental enhancers (Fuxman Bass et al., 2015). Our hmORFeome1.1 mutant library contains 70 TF ORFs altogether harboring 173 mutations (Table S6A). A primary screen using enhanced yeast-one hybrid (eY1H) assays (Reece-Hoyes et al., 2011b) identified PDIs between 152 enhancers (Visel et al., 2007) and 28 WT TFs (Figure 1C and Extended Experimental Procedures). We then performed pairwise assays to compare the PDIs of mutant TFs and their WT counterparts in eY1H assays (Table S6B).

Using systematic PDI profiling, we determined edgotype classes for 58 mutations in 22 TFs that bound at least two enhancers. We identified 38% of the mutations as quasi-null, 43% as edgetic (loss or gain of interaction), and 19% as quasi-WT (Figure 7A). More than 80% of TF missense disease mutations tested either abrogated DNA binding or caused partial change of PDIs. Interestingly, almost half of the mutations are edgetic, challenging the assumption that TF mutations that affect DNA binding do so in a similar fashion across their targets. Among these, a significant fraction of mutations exhibit gain of PDIs, likely because these mutations cause a reduction in DNA-binding specificity and allow greater promiscuity in target recognition.

Given that TFs interact with their DNA targets through DNA-binding domains (DBDs), we assessed whether disease mutations perturbing PDIs are enriched within DBDs. Mutations within versus outside DBDs exhibited strikingly different PDI perturbation patterns ($p = 1.1 \times 10^{-3}$; Figure 7B and Table S6C). Among quasi-null mutations, the proportion of mutations within DBDs was ~10-fold higher than outside DBD regions. These results confirm that most PDI perturbing mutations reside within the DBDs of proteins, further supporting the quality and validity of our PDI perturbation data.

Mutations within the same TF that cause different PDI changes would affect the expression of different targets, resulting in different diseases. We examined disease-causing TF mutations in pleiotropic genes associated with two or more diseases. Mutations with different PDI edgotype classes were likely to be associated with different clinical manifestations (Figure 7C), consistent with our results for PPI perturbations (Figure 6A).

Of the disease mutations for which both PPI and PDI data were available, about half did not perturb any PPIs (Figure 7D). Interestingly, for ~80% of these we did identify PDI perturbations. For instance, mutations in the TGF-β-induced transcription factor *TGIF1* cause holoprosencephaly (Gripp et al., 2000). While the two mutant variants S28C and P63R are still able to bind their protein partners CTBP1 and CTBP2 (quasi-WT for PPI), both mutations completely abrogated the ability of TGIF1 to bind any of the tested DNA targets (quasi-null for PDI) (Figure S7A). Clearly, integrating different types of molecular interactions will enhance our ability to understand specific mechanisms that underlie many genetic disorders.

To gain further insights into alternative molecular interaction perturbations, we computationally examined the effect of disease mutations on protein-chemical interactions (Reva et al., 2011). We found that the frequency with which disease mutations are at protein-chemical interfaces is significantly higher than that of non-disease variants (Figure S7B). In addition, disease mutations that perturb PPIs have no discernable tendency to locate at protein-chemical interfaces (Figure S7C), suggesting that protein-protein and protein-chemical interfaces do not tend to overlap. Interestingly, ~13% of PPI non-perturbing mutations are located at protein-chemical interfaces, supporting the conclusion that these mutations could cause disease through perturbation of alternative types of molecular interactions.

We combined computational predictions and interaction profiling to optimize our performance in disease mutation stratification. Although computational methods such as PolyPhen-2 could predict interaction-perturbing alleles as deleterious (Figure 5A), they fail to explain many disease-causing mutations, and misclassify them as "benign" (Figure S7D). Among these misclassified mutations, ~50% could be explained by molecular interaction perturbations (PCI, PPI, or PDI). For instance, the S140F mutation in *PKP2* encoding the adhesion protein plakophilin leads to arrhythmogenic right ventricular dysplasia (Gerull et al., 2004). While PolyPhen-2 predicts S140F as benign, the S140F mutant exhibited increased binding to the chaperones HSC70 and BAG2, and lost all the PPIs of the WT protein (Table S7A). All together, existing computational methods alone fail to precisely predict disease causality. Examining different types of molecular interaction perturbations is critical for a full comprehension of disease-causing mutations in human.

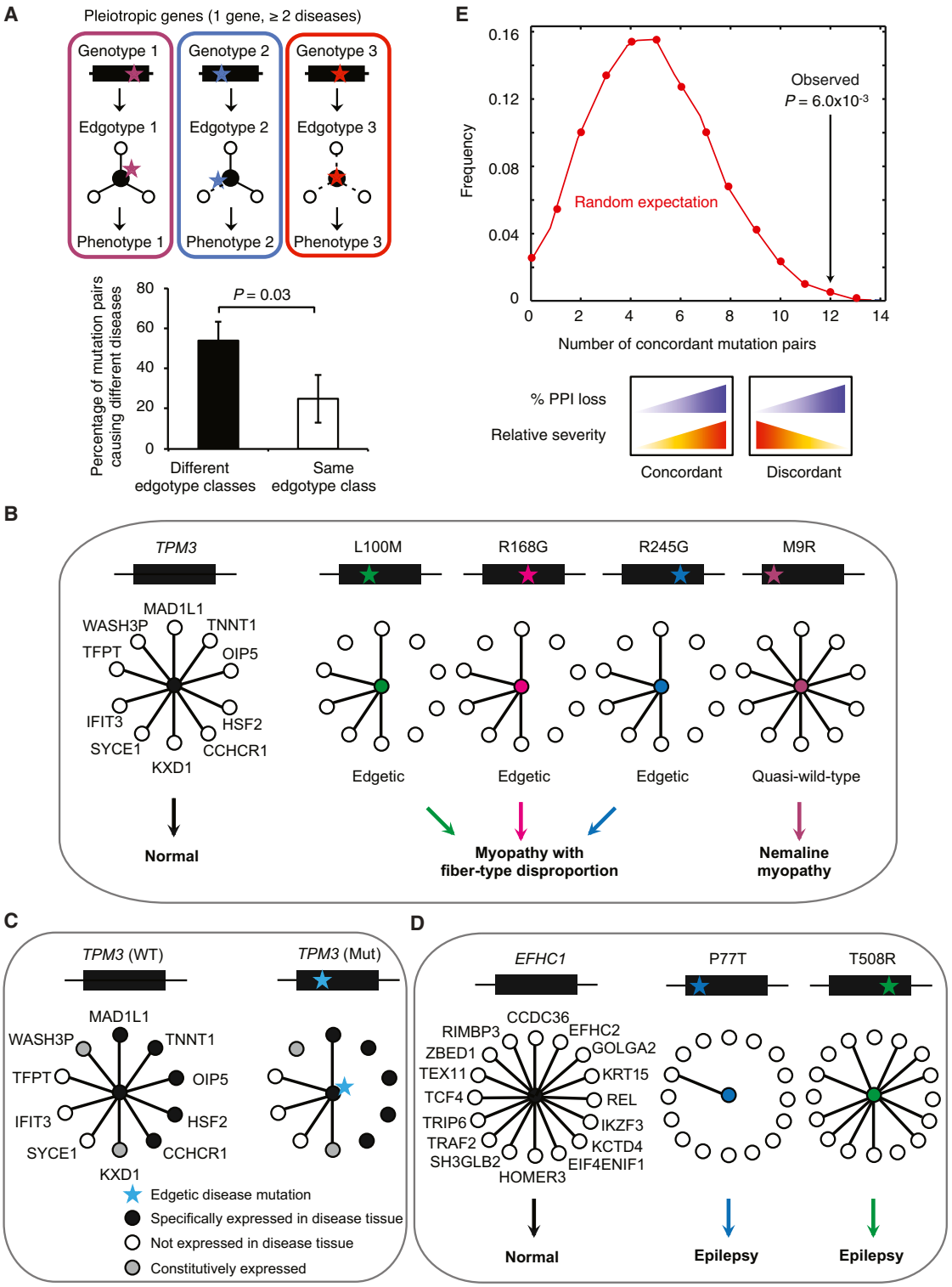


Figure 6. Heterogeneous Genetic Mutations Give Rise to Diverse Disease Outcomes through Distinct Interaction Perturbations

(A) Schematic of pleiotropic disease outcomes resulting from distinct interaction patterns (edgotypes) caused by distinct mutations. Percentage of mutation pairs causing different diseases out of all pairs with different or the same edgotype classes is shown. n = 52. Error bars, SE of the proportion. p values by one-sided Fisher's exact test.

(B) Example of edgotyping four disease mutations in the pleiotropic gene *TPM3*.

(legend continued on next page)

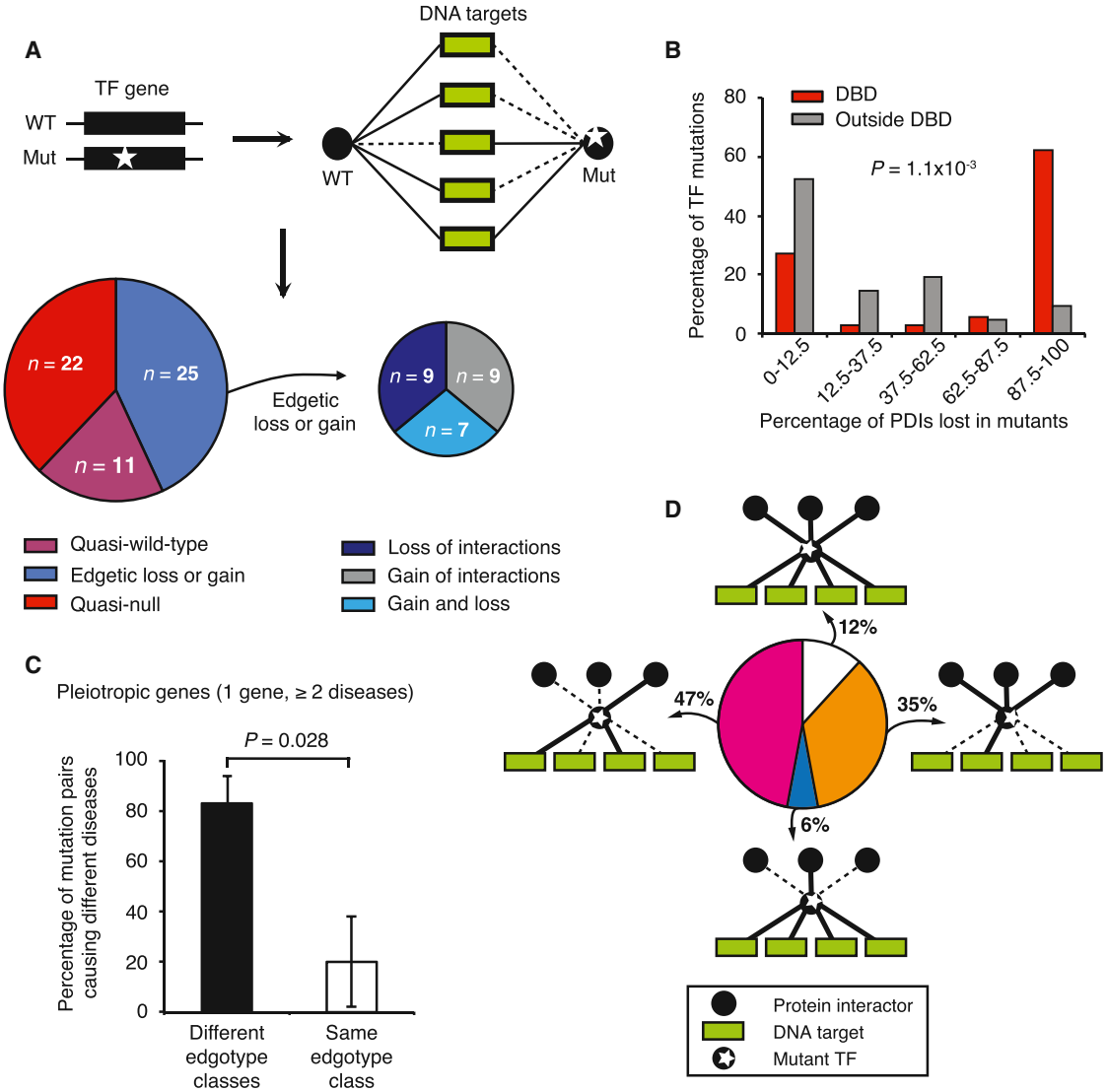


Figure 7. Integration of Protein-Protein and Protein-DNA Interaction Perturbations

(A) PDI edgotype distribution for disease mutations in 22 TFs that bind to more than one enhancer.

(B) Histogram showing percentage of mutations within and outside DBDs as a function of the percentage of PDI loss. Numbers on x axis indicate bin range. p values by one-sided Wilcoxon rank sum test.

(C) Percentage of TF mutation pairs that cause different diseases out of all pairs with different or the same PDI edgotype classes (n = 17). Error bars, SE of the proportion. p values by one-sided Fisher's exact test.

(D) PPI-PDI integration enables mutation characterization at higher resolution. Percentage of mutations is shown for: PPI and PDI unperturbed; PPI unperturbed and PDI perturbed; PPI perturbed and PDI unperturbed; and PPI and PDI perturbed in the integrated network.

See also Figure S7.

DISCUSSION

In this systematic characterization of mutations across various human Mendelian disorders, we have found surprisingly widespread disease-specific perturbations of macromolecular interactions. Approximately 60% of disease-associated missense mutations perturb PPIs, among which half result in complete loss of interactions, generally caused by protein misfolding and

(C) Most perturbed partners of TPM3 are expressed in the disease-relevant tissue.

(D) Edgetic mutations in *EFHC1* perturb epilepsy-related protein partners.

(E) Correlation between the fraction of PPI perturbation and age of onset for mutation pairs causing the same disease. p values by comparing the observed value to 100,000 random controls (n = 13; Extended Experimental Procedures).

See also Figure S6.

impaired expression, and the other half lead to edgetic perturbations. Importantly, different mutations in the same gene frequently result in different interaction perturbation profiles. This strongly suggests that the "edgotype" of a mutation represents a fundamental link between genotype and phenotype.

Our systematic edgotyping strategy provides a practical approach to classifying candidate disease alleles emerging from genome-wide association studies and from sporadic and somatic mutation sequencing approaches. Edgotyping achieves a high precision in identifying candidate disease-causing mutations based on the interaction perturbations relative to WT alleles (Figure S6A). However, the overall sensitivity of an edgotyping approach is compromised due to the false negative rate inherent to the assays used. We expect that a significant fraction of variants currently viewed as non-interaction-perturbing (quasi-WT) will eventually be proven to be edgetic and possibly cause disease. This circumstance likely arises from the incomplete nature of current human interactome network maps (Rolland et al., 2014). Nevertheless, because edgetic mutations cannot become quasi-WT or quasi-null even as interactome maps improve, our estimate of edgetic mutations already provides a reliable minimum lower bound for their frequency.

An alternative possibility is that quasi-WT mutations affect disease phenotypes through perturbation of different types of molecular interactions. Biological signaling is regulated at multiple levels, and various types of molecular interactions are involved (Sahni et al., 2013) as we have shown for PPI and PDI networks. In addition, protein-RNA (Lee et al., 2006) and protein-metabolite (Carpten et al., 2007) interactions have also been shown to be involved in disease. Perturbations of these alternative interaction networks will undoubtedly result in distinct disease consequences. One can envision that integration of additional types of interaction perturbation information with computational predictions will be necessary for a complete understanding of the cellular networks governing a particular disease state (Figure S7D). As a major benefit, perturbed interactions spotlight specific targets and pathways that are altered in a patient-specific context. This type of information could provide a much-needed guide in efforts to developing better diagnostic tools and more personalized medical treatments.

EXPERIMENTAL PROCEDURES

Using ORFs in the human ORFeome v8.1 collection as template, we PCR amplified the two DNA fragments flanking the mutations, followed by a fusion PCR to stitch the fragments together. The resulting fusion ORFs harboring the mutations were Gateway cloned into the Donor vector pDONR223 to derive Entry clones (Rual et al., 2004), which were subsequently verified by next-generation sequencing (Yang et al., 2011).

Interaction with chaperones and other QCFs was performed using a quantitative LUMIER assay (Taipale et al., 2012; Taipale et al., 2014). All wild-type and mutant allele clones were transferred via Gateway recombination into a mammalian expression vector containing a C-terminal 3xFLAG-V5 tag. Stable HEK293T cell lines expressing luciferase-QCF fusion proteins were generated by lentiviral infection, and plasmids carrying wild-type and disease mutation alleles were transfected into the stable HEK293T lines (Taipale et al., 2012). Following capture of FLAG-tagged proteins, luminescence was measured to determine QCF-target interaction. Following luminescence measurement, FLAG-tagged mutant and wild-type proteins were detected as described (Taipale et al., 2012).

We performed a binary protein-protein interaction screen for all mutant and wild-type alleles as baits against ~7,200 human prey proteins (Rual et al., 2004). The identified interactions were combined with the known pairs cataloged by the human binary interaction dataset HI-II-14 (Rolland et al., 2014). All first-pass pairs from the primary Y2H screens were subjected to pairwise testing in which all interactors of any allele of a gene were then tested against all alleles of that gene. The resulting verified protein-protein interaction profiles of disease mutants were compared with their wild-type counterparts. We validated perturbed and unperturbed interactions from mutation-mediated interaction perturbation data ("edgotyping" data) using an orthogonal in vivo *Gaussia princeps* luciferase protein complementation assay (GPCA). Human HEK293T cells were co-transfected with each construct expressing complementary fragments of the *Gaussia* luciferase fused in frame with the tested protein pairs and luciferase activity was measured as described (Cassonnet et al., 2011).

An enhanced yeast one-hybrid (eY1H) assay was used to detect binary protein-DNA interactions (PDIs) between a DNA bait and a protein prey (Reece-Hoyes et al., 2011a; Reece-Hoyes et al., 2011b). DNA baits corresponding to human enhancers were retrieved from the Vista Enhancer Browser (http://enhancer.lbl.gov) (Visel et al., 2007). Protein preys were a set of TFs for which mutant clones are available in our human mutation ORFeome version 1.1. We performed pairwise eY1H assays of an arrayed collection of TF preys comprising all the wild-type TFs and their mutant clones against 152 available enhancer baits.

Disease-causing mutations were annotated by HGMD, and the deleteriousness of amino acid substitutions was predicted by PolyPhen-2 program (Adzhubei et al., 2010). For structural features, distinct mutations were compared with respect to protein domains from the Pfam database, and interaction interfaces on co-crystal structures from PDB. Tissue-specific gene expression was analyzed with normalized RNA-seq data from Human Body Map 2.0 (GSE30611). Network properties analyzed included betweenness centrality, k-core centrality, degree, and closeness centrality (de Nooy et al., 2005).

Full details are provided in the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.04.013.

AUTHOR CONTRIBUTIONS

M.V., S.L., L.J.W., D.E.H., and K.S.-A. conceived the project. N.S., S.Y., M.T., and J.I.F.B. designed and performed experiments, with help from G.I.K., I.K., M.H.L., Q.Z., A.P., D.B., A.D., J.M.W., A.A.S., X.Y., A.K.S., and Y.J. J.C.-H., F.Y., J.P., J.W., Y.W., I.A.K., and T.H. performed computational analyses with contributions from N.S., S.Y., M.T., J.I.F.B., A.K., G.T., V.K., A.S., Y.-Y.L., Y.S., A.S.-M., C.F., D.M.J., A.L., and B.C. V.K., Y.J., N.Y., M.E.C., M.A.C, S.S., B.B., L.J.W., B.C., and D.E.H. provided constructive feedback. M.A.C., S.S., B.B., D.E.H., A.-L.B., T.H., F.P.R., Y.X., A.J.M.W., S.L., and M.V. supervised research and provided critical advice on the study. N.S., S.Y., M.T., J.I.F.B., J.C.-H., M.A.C., B.C., D.E.H., F.P.R., Y.X., A.J.M.W., S.L., and M.V. wrote the manuscript, with contributions from other co-authors.

ACKNOWLEDGMENTS

We thank the members of the DFCI Center for Cancer Systems Biology (CCSB) for valuable discussions and acknowledge A.A. Chen, M. Koeva, and E. Guney for helpful suggestions. This work was supported by NHGRI (P50HG004233 to M.V., F.P.R. and A.-L.B.; RC4HG006066 to M.V., T.H., D.E.H., K.S.-A., L.J.W., and S.L.; and R01HG001715 to M.V., D.E.H., and F.P.R.), NIGMS (GM082971 to A.J.M.W.), NSF (CCF-1219007 to Y.X.), NSERC (RGPIN-2014-03892 to Y.X.), and the Krembil Foundation, a Canada Excellence Research Chair, an Ontario Research Fund–Research Excellence Award awarded to F.P.R. J.I.F.B. is supported by a Pew Latin American Fellowship. A.K. is supported by an EMBO Long-Term Fellowship. S.L. is an investigator

of the Howard Hughes Medical Institute. M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation).

Received: December 15, 2014
Revised: March 5, 2015
Accepted: April 6, 2015
Published: April 23, 2015

REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

Amberger, J., Bocchini, C., and Hamosh, A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM). Hum. Mutat. *32*, 564–567.

Balasubramanian, S., Xia, Y., Freinkman, E., and Gerstein, M. (2005). Sequence variation in G-protein-coupled receptors: analysis of single nucleotide polymorphisms. Nucleic Acids Res. *33*, 1710–1721.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., de Smet, A.-S., et al. (2009). An experimentally derived confidence score for binary protein-protein interactions. Nat. Methods *6*, 91–97.

Carpten, J.D., Faber, A.L., Horn, C., Donoho, G.P., Briggs, S.L., Robbins, C.M., Hostetter, G., Boguslawski, S., Moses, T.Y., Savage, S., et al. (2007). A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. Nature *448*, 439–444.

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G., et al. (2011). Benchmarking a luciferase complementation assay for detecting protein complexes. Nat. Methods *8*, 990–992.

Chin, L., Hahn, W.C., Getz, G., and Meyerson, M. (2011). Making sense of cancer genomic data. Genes Dev. *25*, 534–555.

Clarke, N.F., Kolski, H., Dye, D.E., Lim, E., Smith, R.L., Patel, R., Fahey, M.C., Bellance, R., Romero, N.B., Johnson, E.S., et al. (2008). Mutations in TPM3 are a common cause of congenital fiber type disproportion. Ann. Neurol. *63*, 329–337.

de Nooy, W., Mrvar, A., and Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek (Cambridge University Press).

Falsone, S.F., Leptihn, S., Osterauer, A., Haslbeck, M., and Buchner, J. (2004). Oncogenic mutations reduce the stability of SRC kinase. J. Mol. Biol. *344*, 281–291.

Flora, A., Garcia, J.J., Thaller, C., and Zoghbi, H.Y. (2007). The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc. Natl. Acad. Sci. USA *104*, 15382–15387.

Fuxman Bass, J.I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D.E., Vidal, M., and Walhout, A.J.M. (2015). Human Gene-Centered Transcription Factor Networks for Enhancers and Disease Variants. Cell *161*, this issue, 661–673.

Gerull, B., Heuser, A., Wichter, T., Paul, M., Basson, C.T., McDermott, D.A., Lerman, B.B., Markowitz, S.M., Ellinor, P.T., MacRae, C.A., et al. (2004). Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. Nat. Genet. *36*, 1162–1164.

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabási, A.L. (2007). The human disease network. Proc. Natl. Acad. Sci. USA *104*, 8685–8690.

Gripp, K.W., Wotton, D., Edwards, M.C., Roessler, E., Ades, L., Meinecke, P., Richieri-Costa, A., Zackai, E.H., Massagué, J., Muenke, M., and Elledge, S.J. (2000). Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. Nat. Genet. *25*, 205–208.

Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. *320*, 369–387.

Gunning, P., Gordon, M., Wade, R., Gahlmann, R., Lin, C.S., and Hardeman, E. (1990). Differential control of tropomyosin mRNA levels during myogenesis suggests the existence of an isoform competition-autoregulatory compensation control mechanism. Dev. Biol. *138*, 443–453.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.

Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J., and O'Donovan, C. (2015). The GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res. *43*, D1057–D1063.

Köllner, I., Sodeik, B., Schreek, S., Heyn, H., von Neuhoff, N., Germeshausen, M., Zeidler, C., Krüger, M., Schlegelberger, B., Welte, K., and Beger, C. (2006). Mutations in neutrophil elastase causing congenital neutropenia lead to cytoplasmic protein accumulation and induction of the unfolded protein response. Blood *108*, 493–500.

Laing, N.G., Wilton, S.D., Akkari, P.A., Dorosz, S., Boundy, K., Kneebone, C., Blumbergs, P., White, S., Watkins, H., Love, D.R., et al. (1995). A mutation in the alpha tropomyosin gene TPM3 associated with autosomal dominant nemaline myopathy. Nat. Genet. *9*, 75–79.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985.

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO J. *25*, 522–532.

Lesage, S., and Brice, A. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. Hum. Mol. Genet. *18* (R1), R48–R59.

McArdle, A., Broome, C.S., Kayani, A.C., Tully, M.D., Close, G.L., Vasilaki, A., and Jackson, M.J. (2006). HSF expression in skeletal muscle during myogenesis: implications for failed regeneration in old mice. Exp. Gerontol. *41*, 497–500.

Reece-Hoyes, J.S., Barutcu, A.R., McCord, R.P., Jeong, J.S., Jiang, L., MacWilliams, A., Yang, X., Salehi-Ashtiani, K., Hill, D.E., Blackshaw, S., et al. (2011a). Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. Nat. Methods *8*, 1050–1052.

Reece-Hoyes, J.S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., Pesyna, C., Dekker, J., Myers, C.L., and Walhout, A.J. (2011b). Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. Nat. Methods *8*, 1059–1064.

Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. *39*, e118.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O., et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res. *14* (10B), 2128–2135.

Sahni, N., Yi, S., Zhong, Q., Jailkhani, N., Charloteaux, B., Cusick, M.E., and Vidal, M. (2013). Edgotype: a fundamental link between genotype and phenotype. Curr. Opin. Genet. Dev. *23*, 649–657.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Res. *33*, W382–W388.

Shimamura, T., Lowell, A.M., Engelman, J.A., and Shapiro, G.I. (2005). Epidermal growth factor receptors harboring kinase domain mutations

associate with the heat shock protein 90 chaperone and are destabilized following exposure to geldanamycins. Cancer Res. *65*, 6401–6408.

Sörgjerd, K., Ghafouri, B., Jonsson, B.H., Kelly, J.W., Blond, S.Y., and Hammarström, P. (2006). Retention of misfolded mutant transthyretin by the chaperone BiP/GRP78 mitigates amyloidogenesis. J. Mol. Biol. *356*, 469–482.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum. Genet. *133*, 1–9.

Subramanian, S., and Kumar, S. (2006). Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. BMC Genomics *7*, 306.

Sunyaev, S.R. (2012). Inferring causality and functional significance of human coding DNA variants. Hum. Mol. Genet. *21* (R1), R10–R17.

Taipale, M., Krykbaeva, I., Koeva, M., Kayatekin, C., Westover, K.D., Karras, G.I., and Lindquist, S. (2012). Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. Cell *150*, 987–1001.

Taipale, M., Tucker, G., Peng, J., Krykbaeva, I., Lin, Z.Y., Larsen, B., Choi, H., Berger, B., Gingras, A.C., and Lindquist, S. (2014). A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. Cell *158*, 434–448.

Tervaniemi, M.H., Siitonen, H.A., Söderhäll, C., Minhas, G., Vuola, J., Tiala, I., Sormunen, R., Samuelsson, L., Suomela, S., Kere, J., and Elomaa, O. (2012). Centrosomal localization of the psoriasis candidate gene product, CCHCR1, supports a role in cytoskeletal organization. PLoS ONE *7*, e49920.

Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., et al. (2009). An empirical framework for binary interactome mapping. Nat. Methods *6*, 83–90.

Vidal, M., Cusick, M.E., and Barabási, A.L. (2011). Interactome networks and human disease. Cell *144*, 986–998.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92.

Whitesell, L., and Lindquist, S.L. (2005). HSP90 and the chaperoning of cancer. Nat. Rev. Cancer *5*, 761–772.

Yamashita, D., Sano, Y., Adachi, Y., Okamoto, Y., Osada, H., Takahashi, T., Yamaguchi, T., Osumi, T., and Hirose, F. (2007). hDREF regulates cell proliferation and expression of ribosomal protein genes. Mol. Cell. Biol. *27*, 2003–2013.

Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat. Methods *8*, 659–661.

Zhong, Q., Simonis, N., Li, Q.R., Charloteaux, B., Heuze, F., Klitgord, N., Tam, S., Yu, H., Venkatesan, K., Mou, D., et al. (2009). Edgetic perturbation models of human inherited disorders. Mol. Syst. Biol. *5*, 321.