



OPEN

Internetwork connectivity of molecular networks across species of life

Tarun Mahajan¹ & Roy D. Dar^{1,2,3,4}

Molecular interactions are studied as independent networks in systems biology. However, molecular networks do not exist independently of each other. In a network of networks approach (called multiplex), we study the joint organization of transcriptional regulatory network (TRN) and protein–protein interaction (PPI) network. We find that TRN and PPI are non-randomly coupled across five different eukaryotic species. Gene degrees in TRN (number of downstream genes) are positively correlated with protein degrees in PPI (number of interacting protein partners). Gene–gene and protein–protein interactions in TRN and PPI, respectively, also non-randomly overlap. These design principles are conserved across the five eukaryotic species. Robustness of the TRN–PPI multiplex is dependent on this coupling. Functionally important genes and proteins, such as essential, disease-related and those interacting with pathogen proteins, are preferentially situated in important parts of the human multiplex with highly overlapping interactions. We unveil the multiplex architecture of TRN and PPI. Multiplex architecture may thus define a general framework for studying molecular networks. This approach may uncover the building blocks of the hierarchical organization of molecular interactions.

Biological functions and characteristics are consequences of complex interactions between numerous components¹. These components can be molecules such as DNA, RNA, proteins and other small molecules or larger units such as cells, tissues, whole organisms or entire ecosystems. These interactions are organized into a hierarchy of networks. Networks at different levels of this hierarchy have been studied extensively. For instance, at the subcellular level, transcriptional regulatory networks (TRN) model protein–DNA interactions^{1–12}, protein–protein interaction (PPI) networks capture physical interactions between proteins^{6,13–26} and metabolic networks map interactions between the set of biochemical reactions in an organism^{1,27–29}. Analysis of individual network layers has answered important biological questions ranging from organization of gene expression^{5,8,29–31}, predicting phenotype from molecular interaction networks^{16,24}, to understanding disease biology^{32–36}.

However, biological networks do not function in isolation. These networks comprise of different types of interactions and even interact with other networks^{1,37}. For instance, TRN and PPI networks interact with each other. Proteins are translated from genes in accordance with the regulatory program encoded in the TRN. These translated proteins interact with each other in the PPI layer. Transcription factor proteins interact with other proteins in the PPI layer and also regulate downstream genes in the TRN network. Further, PPI networks can also encode different kinds of physical interactions between proteins, such as the ones revealed by Yeast Two-Hybrid (Y2H) binary, Affinity Purification (AP) protein complexes, synthetic lethality, dosage lethality, genetic interactions, etc.¹⁴. Such multilayer networks (comprising multiple networks) can be interdependent when different network layers interact with each other to form a network of networks (NON) architecture³⁸. For instance, the interaction between TRN and PPI networks forms an interdependent NON (Fig. 1). Alternatively, multilayer networks can be multiplex with different networks, which encode distinct types of interactions between the same molecular species such as the different types of PPI interactions.

Until recently, network science has focused largely on the study of individual biological networks. Even some of the studies that worked with multiple networks aggregated or integrated the different networks and did not consider a multilayer approach^{39–42}. This could partly be attributed to the fact that multilayer networks have gained popularity only in recent years, especially in statistical physics³⁸. Now, extensive work has been done to

¹Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ²Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ³Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ⁴Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ✉email: tarunm3@illinois.edu; roydar@illinois.edu

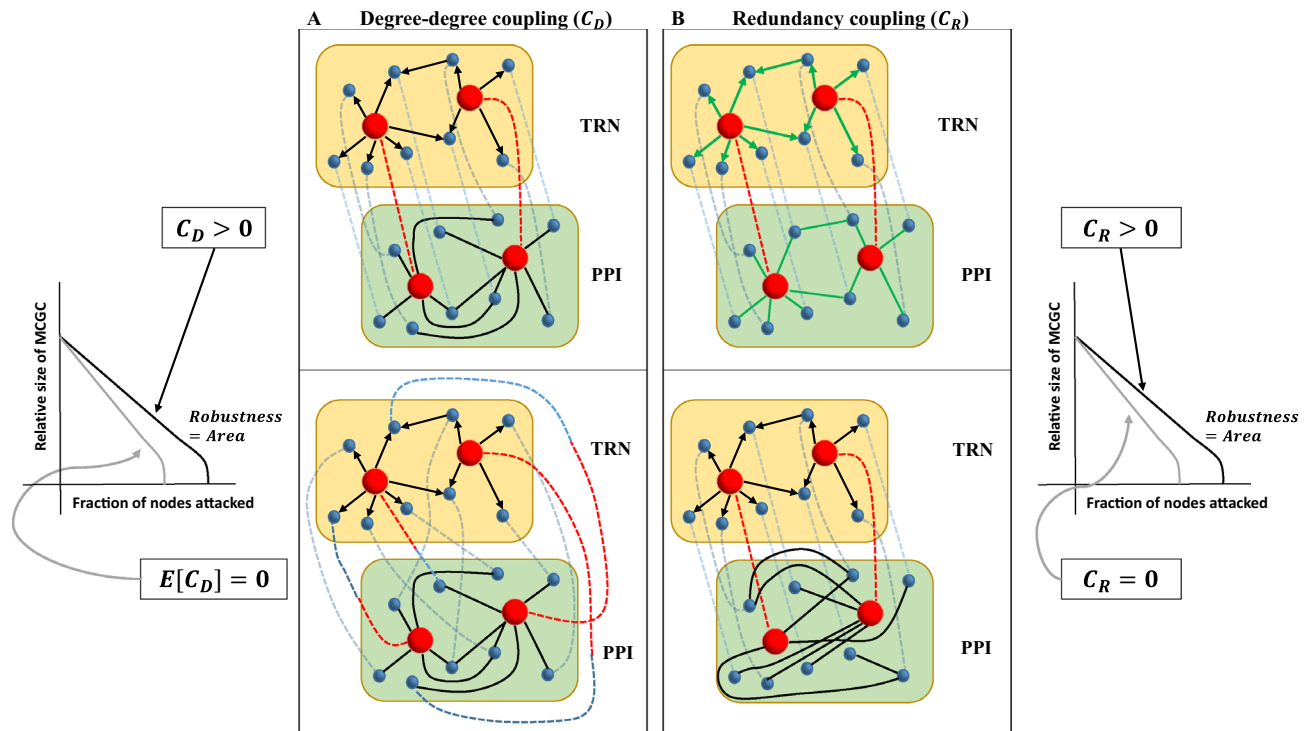


Figure 1. Degree–degree coupling and redundancy as potential modulators of robustness in molecular multiplexes. Degree–degree coupling (C_D) and redundancy coupling (C_R) can assume any value in the multiplex of TRN (yellow layer) and PPI (green layer) networks across species. **(A)** (Right, Top) For $C_D > 0$, highly connected genes in TRN (red spheres) are more likely to produce hubs or highly connected proteins in PPI (red spheres), while sparsely connected genes in TRN (blue spheres) are highly likely to produce spokes or sparsely connected proteins in PPI (blue spheres). (Right, Bottom) For $C_D = 0$, TRN and PPI will be uncoupled in the multiplex, and the association between genes and proteins would be randomized. Here, highly connected genes produce spoke proteins and vice-versa sparsely connected genes produce hub proteins. (Left) Based on theoretical studies, C_D is expected to be positively correlated with robustness. Robustness is quantified by area under the attack curve for the Mutually Connected Giant Component (MCGC) (“Methods”). **(B)** (Left, Top) For $C_R > 0$, there will be non-zero number of edges which are simultaneously present in both TRN and PPI (edges marked green). (Left, Bottom) For $C_R = 0$, TRN and PPI will not have common edges (no green edges). (Right) Based on theoretical studies, C_R is expected to be positively correlated with robustness. Black directed edges represent regulation of downstream genes by transcription factors in TRN. Black undirected edges represent protein–protein interactions in PPI. Dashed edges represent correspondence between a gene and its corresponding protein.

study robustness properties of multilayer networks^{43–53}. Counter-intuitively, interdependent networks are more fragile to random failure than independent individual networks⁴³. Real interdependent networks mitigate this vulnerability by means of specific intra- and interlayer degree–degree correlation or coupling^{46,51}. For a given TRN–PPI interdependent (or multiplex) network (Fig. 1A), degree–degree coupling (C_D) is quantified as the correlation between the connectivity of a protein in the PPI network, K , and the connectivity of its corresponding gene in the TRN, either in-degree (k_{in} , number of regulations incident on a gene from upstream transcription factors), out-degree (k_{out} , number of downstream genes regulated by a transcription factor), or total degree ($k = k_{in} + k_{out}$). In this case, C_D can be negative, positive, or zero. Particularly, positive C_D makes the multiplex robust to attack (Fig. 1A)^{54,55}. With positive C_D , hub nodes are likely to be hub nodes in all the network layers (Fig. 1A, top). For negative C_D , hub nodes in one layer are dependent on spokes in the other network layer. For zero C_D , hubs and spokes are randomly dependent on each other across network layers (Fig. 1B, bottom). Here, we investigate C_D across species and assess whether it is positive, negative, or uncoupled, and if there exists a global trend across various species.

Edge overlap or redundancy between network layers also mitigates vulnerability in interdependent networks⁵⁶. Two genes in the TRN network have an interaction between them if one is the transcription factor for the other. While in PPI networks, interaction between two proteins depicts physical or functional interaction between these proteins. We define multiplex redundancy as similarity in interactions between the TRN and PPI networks. We quantify redundancy (C_R) by counting the number of common interactions simultaneously in both the TRN and PPI networks. If TRN and PPI networks are represented as graphs, then C_R can be measured by counting the number of edges which are simultaneously present in both the graphs (Fig. 1B and “Methods”). Higher redundancy makes the multiplex more robust to attack (Fig. 1B).

We study the multilayer network of TRN and PPI networks in nine different species, namely *H. pylori*, *M. tuberculosis*, *E. coli*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *A. thaliana*, *M. musculus* and *H. sapiens*, spanning two domains of life (bacteria and eukaryotes). We collected TRN networks from nine different sources, and 16 different PPI networks from five different sources (see “Methods” and Supplementary Table S1). Two of the PPI sources are publicly curated databases—BioGRID¹⁴ and HINT⁵⁷. Interlayer connectivity is defined by one-to-one correspondence between a gene and its corresponding protein. Therefore, this multilayer network can be reduced to an equivalent multiplex network⁴³. Henceforth, we call the TRN–PPI multilayer network a TRN–PPI multiplex. Based on quality control, five (*S. cerevisiae*, *C. elegans*, *D. melanogaster*, *M. musculus* and *H. sapiens*) of the nine species were used for further analysis; multiplexes with visually continuous percolation curves representing second-order like behavior were studied. Here, we show that for species TRN–PPI multiplexes, positive C_D increases robustness to targeted attack on the genes and proteins. Further, increasing C_R also increases robustness. We find a trade-off between robustness and independence. Independent multiplexes with no degree–degree coupling and redundancy are highly vulnerable, while positively degree–degree coupled and highly redundant multiplexes are highly robust. We show that this trade-off exists for different species individually. Multiplex coupling is also correlated with the distribution of functionally important genes and proteins such as essential, disease and pathogen-interacting genes and proteins. These functionally important genes are selectively situated in redundant and essential parts of the multiplex and, consequently, are vulnerable. Interlayer degree–degree coupling and redundancy offer design mechanisms for tuning robustness in molecular multiplex networks.

Results

TRN–PPI multiplex is non-randomly coupled across species. We study coupling between TRN and PPI networks using two multiplex properties—degree–degree coupling (C_D) and redundancy coupling (C_R). C_D is quantified by Pearson’s or Spearman’s rank correlation between PPI degree, K , and TRN out-degree, k_{out} (Fig. 2A–D, “Methods”). We quantify C_R by counting the total number of interactions present simultaneously in both TRN and PPI networks (Fig. 1B, “Methods”). We find that C_D is significantly positive across different eukaryotes (Fig. 2A–D and Supplementary Figure S1); C_D values are statistically significant as evident from the p-values (two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) and z-scores for comparison against the null model (Fig. 2A–D and Supplementary Figures S19 and S20). C_R is also non-randomly positive across different eukaryotes (Fig. 2E–H and Supplementary Figure S2). Node-specific (for each gene–protein pair in the multiplex) C_R values are more long-tailed compared to a randomly shuffled null model. Shuffled null model is generated by randomly shuffling labels on genes in TRN, while keeping protein labels fixed in PPI.

Multiplex degree–degree and redundancy couplings modulate robustness. We study the relationship between multiplex degree–degree (C_D) and redundancy (C_R) couplings and robustness (R) across species and for individual species. We use a previously reported formalism to study topological robustness of the TRN–PPI multiplex under targeted attack on its nodes^{43,58} (“Methods”). Robustness is related to the size of the mutually connected giant component (MCGC). MCGC is defined as the largest connected component between both layers of the multiplex (“Methods”). Buldyrev et al.⁴³ and Kleinberg et al.⁵⁸ track the size of MCGC under attack to quantify robustness. We specifically focus on targeted attack. Under targeted attack, at each step of the attack, gene–protein pairs are removed in decreasing order of multiplex degree, $K_{mult}(i) = \max(K(i), k_{out}(i))$ ⁵⁸, where $K_{mult}(i)$ is the multiplex degree for gene–protein pair i , $K(i)$ is the degree of protein i in the PPI network and $k_{out}(i)$ is the out-degree of gene i in the TRN network (“Methods”). Absolute robustness is then measured by tracking the relative size of MCGC (MCGC divided by number of gene–protein pairs in the multiplex) as we successively remove gene–protein pairs from the multiplex. Figure 3 and Supplementary Figure S3 show the relative size of MCGC as a function of the fraction of gene–protein pairs removed during targeted attack for all the species. We call the curve for MCGC the “attack curve”. Absolute robustness is quantified by area under the attack curve (we will call this area *RobustArea*) (“Methods”). Large *RobustArea* implies large robustness for a given multiplex, and vice versa small *RobustArea* means low robustness. Cohen’s d ⁵⁹ is used to quantify effect size for robustness by comparing *RobustArea* for a given multiplex against an appropriate null model (“Methods”). This quantity is used as an estimate of robustness (R) in this work. We use a *Zero-Coupling-Zero-Redundancy null model* (see Supplementary methods). Under this null model, we generate multiplexes with C_D and C_R fixed to zero.

Attack curves for different eukaryotic species are shown in Fig. 3. Visually, we see that organismal *RobustArea* values are larger than the null model on average (except for yeast with HINT PPI network). This is quantified in Fig. 4A, where C_D and R are positively correlated across different eukaryotic species with Spearman’s correlation coefficient of 0.68 (p-value = 0.044, two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$), and C_R and R are positively correlated across different eukaryotic species with Spearman’s correlation coefficient of 0.72 (p-value = 0.037, two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$). We also quantify the dependence of R on C_D and C_R for individual species (Fig. 4B,C). For each species, we sample a subset of the TRN–PPI multiplex such that the sampled multiplex has specific values of C_D and C_R (see Supplementary Methods). This sampling is repeated 1000 times. Attack curves and *RobustArea* are then computed for the sampled multiplexes. Therefore, for a given combination of C_D and C_R , we get a distribution of *RobustArea*, and R can be calculated. For each species, we explore C_D and C_R values over a grid (Fig. 4B). Changing both C_D and C_R increases R (Fig. 4B and Supplementary Figure S6). For all the species, multiplexes with the highest C_D and C_R values exhibit maximum R (Fig. 4B and Supplementary Figure S6). Further, for a fixed value of C_R , R increases with C_D . Similarly, for fixed C_D , R also increases with C_R . The effect of C_R is stronger than C_D , which is evident for the human multiplex, where R is high for high C_R irrespective of C_D . The independent effect of C_D and C_R on R is shown in Fig. 4C.

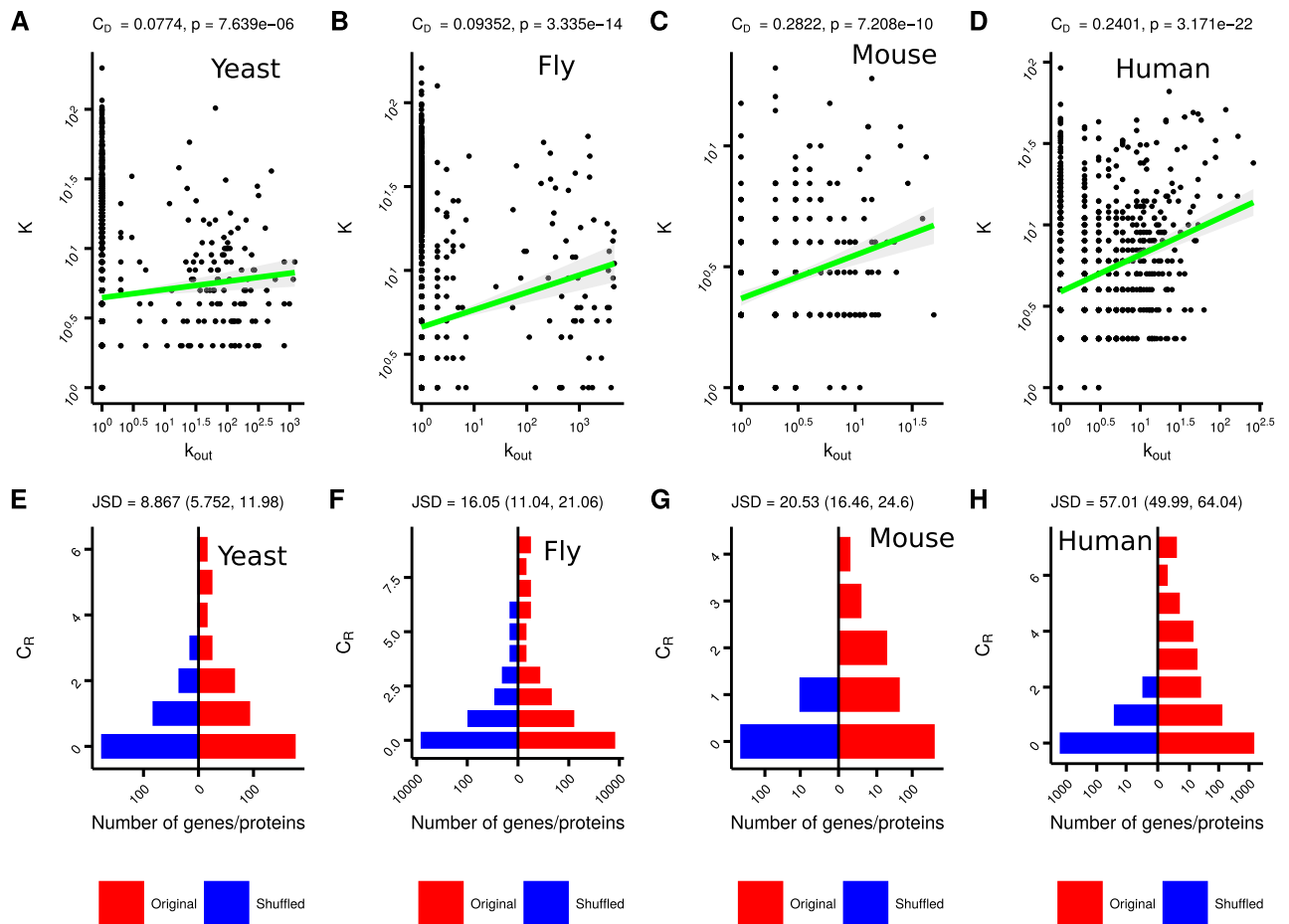


Figure 2. TRN–PPI multiplex is coupled across species. Scatter plot for degree (K) of proteins in the PPI network versus out-degree (k_{out}) of genes in the TRN for eukaryotes, (A) yeast, (B) fly, (C) mouse and (D) human. Both K and k_{out} values have been log-transformed after adding 1. The PPI networks are from the BioGRID database (“Methods”). Linear interpolated fits between K and k_{out} are also shown (green line) with 95% confidence region shaded in gray. Degree–degree coupling (C_D) values and corresponding p-values (two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) are also shown. Distribution of redundancy (C_R) for gene–protein pairs for species multiplex and the randomly shuffled null model for eukaryotes, (E) yeast, (F) fly, (G) mouse and (H) human. For each gene–protein pair, C_R is quantified by the number of redundant edges incident on that gene–protein pair. Shuffled null model is generated by randomly shuffling labels on genes in TRN, while keeping protein labels fixed in PPI. Jensen Shannon divergence (JSD) between distributions of C_R in organismal and shuffled multiplexes is also given in (E–H).

Across species, we find that C_D , C_R and R are positively correlated with the number of gene–proteins in the multiplex (Supplementary Figure S4). To assess whether across species correlation between C_D (C_R) and R is simply an artifact of the difference between the sizes of the multiplexes or not, we sample two subsets of different sizes for yeast, fly and human (Supplementary Figures S16–S18). For each of these species, we find that the larger sized subset has a larger robustness while keeping C_R and C_D fixed (Supplementary Figure S16). However, despite this dependence of R on multiplex size, dependence on C_D and C_R can still be assessed by comparing the two subsets (Supplementary Figures S17 and S18). This suggests that the correlations with R seen in Fig. 4A,C are indeed because of C_D and C_R in addition to the dependence on the number of gene–protein pairs in the multiplex.

We also considered a configuration null model–*Multiplex-Configuration* (see Supplementary Methods)– to assess multiplex robustness. Briefly, under this null model, gene and protein degrees in TRN and PPI, respectively, are fixed, while edges are randomly shuffled. Additionally, the one-to-one correspondence between genes in TRN and proteins in PPI is fixed as well. *Multiplex-Configuration* preserves C_D , while randomly shuffling C_R . With the *Multiplex-Configuration* model, robustness of species multiplex is less pronounced compared to the *Zero-Coupling-Zero-Redundancy* model (see Supplementary Figures S21 and S22); two of the multiplexes (*C. elegans* HINT and yeast HINT) are even less robust than the null model. However, these results do not contradict our previous conclusions regarding TRN–PPI multiplexes. *Zero-Coupling-Zero-Redundancy* and *Multiplex-Configuration* null models answer different questions. The former investigates, given the TRN and PPI networks, whether the one-to-one correspondence between genes in TRN and proteins in PPI creates non-random coupling between the two network layers. Whereas, the latter null model quantifies the impact of TRN and PPI network structures

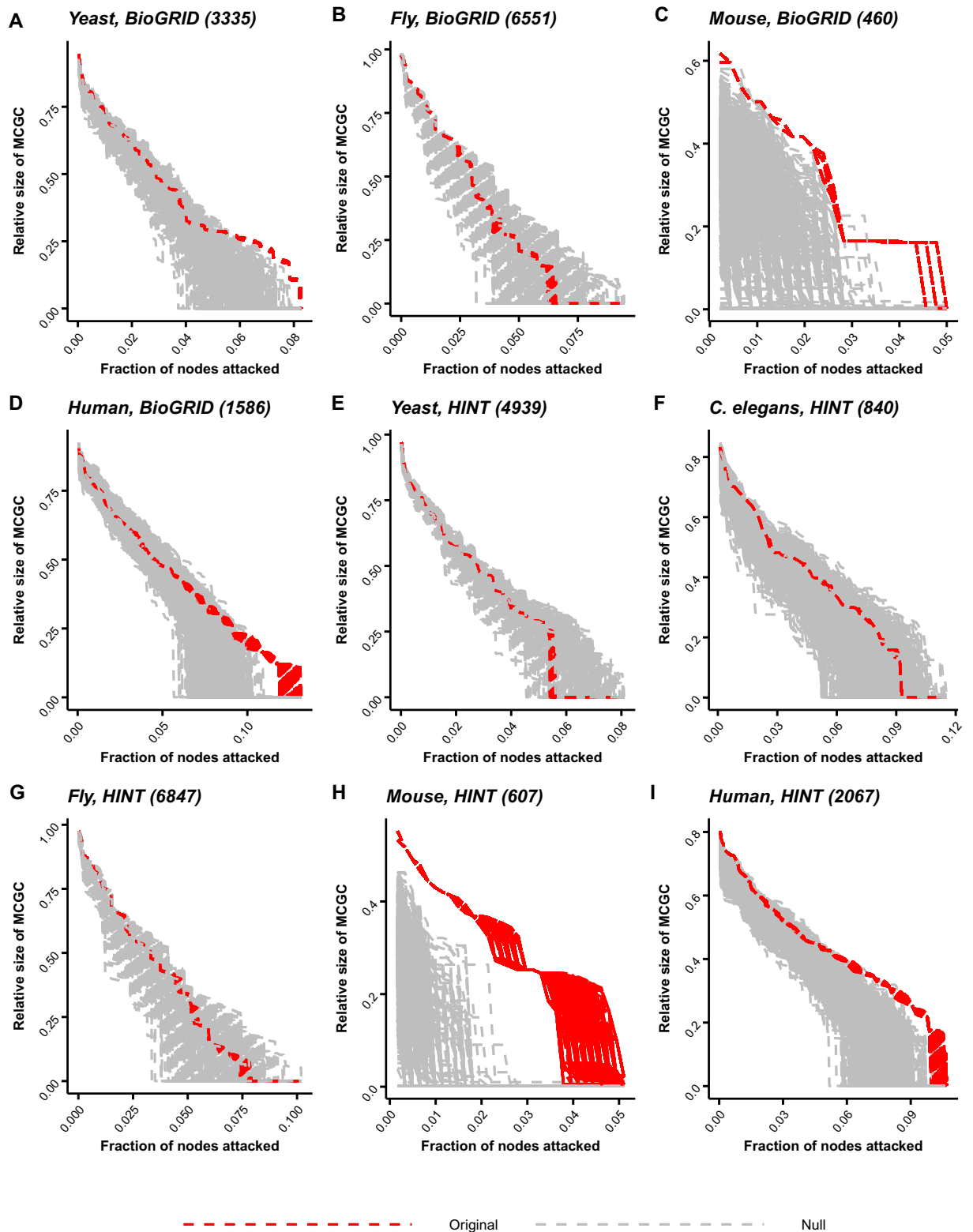


Figure 3. Multiplex attack curves for eukaryotes. Relative size of the mutually connected giant component (MCGC) is plotted as a function of the fraction of gene–protein pairs attacked and removed from the multiplex (“Methods”). Attack curves are shown for five species using PPI networks from either BioGRID or HINT databases (“Methods”); database for a given panel are annotated next to the species name. Along with the attack curves for the species (red), attack curves for the *Zero-Coupling–Zero-Redundancy* (Supplementary Methods) null model are also shown (gray). Under this null model, multiplexes have no degree–degree coupling and no redundancy. On average, species attack curves are more robust than the null model. Robustness is quantified by area under the curve. Attack on the multiplex (organism or null) is repeated 1000 times.

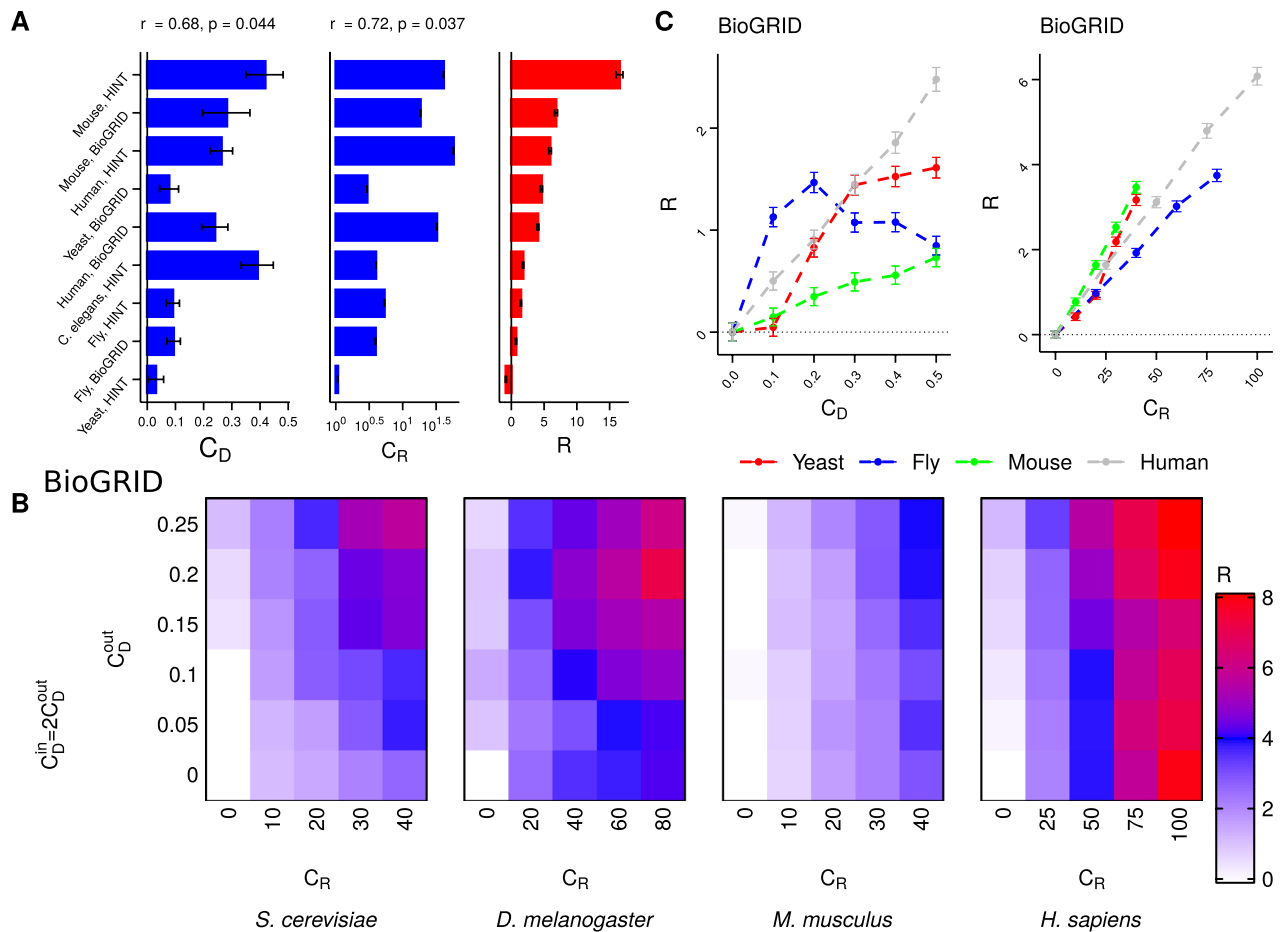


Figure 4. Degree-degree and redundancy couplings modulate multiplex robustness. (A) Robustness to targeted attack (R) (right) is positively correlated with degree-degree coupling (C_D) (left) across species with Spearman's rank correlation coefficient of 0.6803 (p-value = 0.04372). Robustness to targeted attack (R) (right) is also positively correlated with redundancy coupling (C_R) (center) across species with Spearman's rank correlation coefficient of 0.7167 (p-value = 0.03687). C_R is computed as a z-score against the null model (“Methods”) (B) We sample a subset of gene–protein pairs from the species multiplexes (sizes for the subsets are: *S. cerevisiae*-1000, *D. melanogaster*-2000, *M. musculus*-300, *H. sapiens*-500) with specific C_D and C_R values. We repeat the sampling 100 times. We explore C_D and C_R over a grid. For each point over the 2D grid, the heatmap shows the robustness (R) value. R is computed by comparing *RobustArea* of any point over the grid against the lower-left point of the grid (with $C_D = 0, C_R = 0$). For each point over the grid, for each of the sampled subsets, targeted attack is performed. Mean R values are shown here. Lower and upper 95% confidence interval (CI) values are shown in Supplementary Figure S5. C_D^{in-out} : degree-degree coupling between k_{out} and K , C_D^{in} : degree-degree coupling between k_{in} and K . (C) (Left) Dependence of robustness on degree-degree coupling while redundancy is fixed. For each species curve, R is calculated by comparison against the sampled multiplex with $C_D = 0$. (Right) Dependence of robustness on redundancy coupling while degree-degree coupling is fixed to the value in the full multiplex. For each species curve, R is calculated by comparison against the sampled multiplex with $C_R = 0$. In all the panels, error bars show 95% CIs. In panel (B), BioGRID PPI networks are used; results with HINT PPI networks are shown in Supplementary Figure S6B. For panel C (left), C_R is set equal to the number of redundant edges in the multiplex.

on redundancy coupling. In this work, we specifically focus on the coupling between TRN and PPI created by the one-to-one correspondence between their nodes, and not on the impact of individual network structures.

Essential genes and proteins are essential for multiplex robustness. We collected essential and non-essential genes for three species (yeast, fly and human) from the Online GEne Essentiality (OGEE) database^{60,61}. To gauge importance of essential genes for the multiplex, we selectively attack essential genes and proteins in decreasing order of multiplex degree. Partial attack curves are generated by successively attacking essential genes in decreasing order of multiplex degree. The attack process is halted once all the genes in the set of essential genes have been removed (“Methods”). *RobustArea* is quantified by computing area under such partial attack curves. R is calculated by comparison against three random null models—**Random Degree Preserving** (*RanDP*), **Random Degree Preserving-Redundancy Zero** (*RanDP-RZ*) and **Random Degree Preserving-no**

C_D ($RanDP-noC_D$). Steps to generate random subsets from these models are given in Supplementary Methods. Briefly, $RanDP$ generates subsets by matching network degrees in the random subsets to essential gene–protein pairs in the multiplex. $RanDP-RZ$ further ensures that C_R is zero in the sampled subsets. Both $RanDP$ and $RanDP-RZ$ preserve C_D . On the other hand, $RanDP-noC_D$ matches degree distributions rather than individual gene or protein degrees; this does not preserve C_D . $RanDP-noC_D$ perfectly matches all the degree distributions to essential genes and proteins (Supplementary Figures S38–S40). However, $RanDP$ (Supplementary Figures S32–S34) and $RanDP-RZ$ (Supplementary Figures S35–S37) do not perfectly match all the degree distributions.

For all the null models, essential genes and proteins are more vulnerable than random genes and proteins for yeast and human multiplexes (Fig. 5A, Supplementary Figures S41 and S42). Fly essential genes are more vulnerable only against the $RanDP-RZ$ model (Fig. 5A). Vulnerability of essential genes in a species multiplex is greater than dictated by either TRN or PPI networks. We establish this by performing targeted attack on TRN and PPI networks and comparing against attack on the multiplex (Supplementary Figures S23–S31). For the human multiplex, essential genes are topologically important in PPI and TRN networks as well (Supplementary Figures S23–S28). In the fly multiplex, essential genes are more and less vulnerable than random genes and proteins in TRN and PPI, respectively (Supplementary Figures S23–S28). Essential genes and proteins in the yeast multiplex are equally/less and more vulnerable than random genes and proteins in TRN and PPI, respectively (Supplementary Figures S23–S28). However, comparing TRN and PPI attack curves against null model attack curves for the multiplex shows that attack on the multiplex is more lethal than attack on the individual networks (Supplementary Figures S29–S31). Moreover, yeast and human essential genes are situated in highly topologically important parts of the multiplex, which is evident by higher vulnerability compared to random genes and proteins (Supplementary Figures S29–S31). Fly essential genes and proteins appear topologically more important than random genes and proteins only against the $RanDP-RZ$ null model (Supplementary Figures S29–S31). Collectively, these results show that essential genes and proteins are topologically essential in species multiplex, and this importance is not trivially dependent on the independent relevance in TRN and PPI.

In the human multiplex, higher C_R of essential genes compared to random genes co-occurs with lower R (Fig. 5A, Supplementary Figures S41 and S42). This suggests that redundancy might play a role in the topological importance of essential genes. We also study the impact of C_R on robustness by sampling subsets of genes and proteins (size 100) in the human multiplex (Supplementary Figure S15). As the redundancy of the sampled genes and proteins increases, robustness decreases against a random set of genes. Thus, redundancy might control selective placement of a subset of genes and proteins in important parts of the multiplex. For yeast and fly, such correlation between C_R and R is only seen against the $RanDP-RZ$ model (Fig. 5A). This implies that C_R might not be the only property controlling topological importance of essential genes.

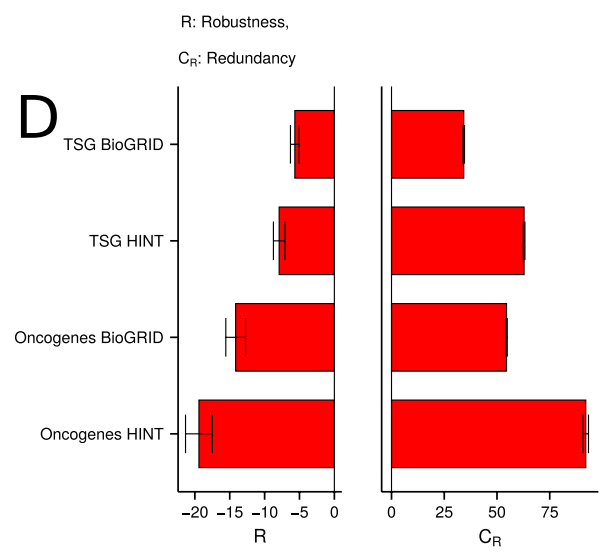
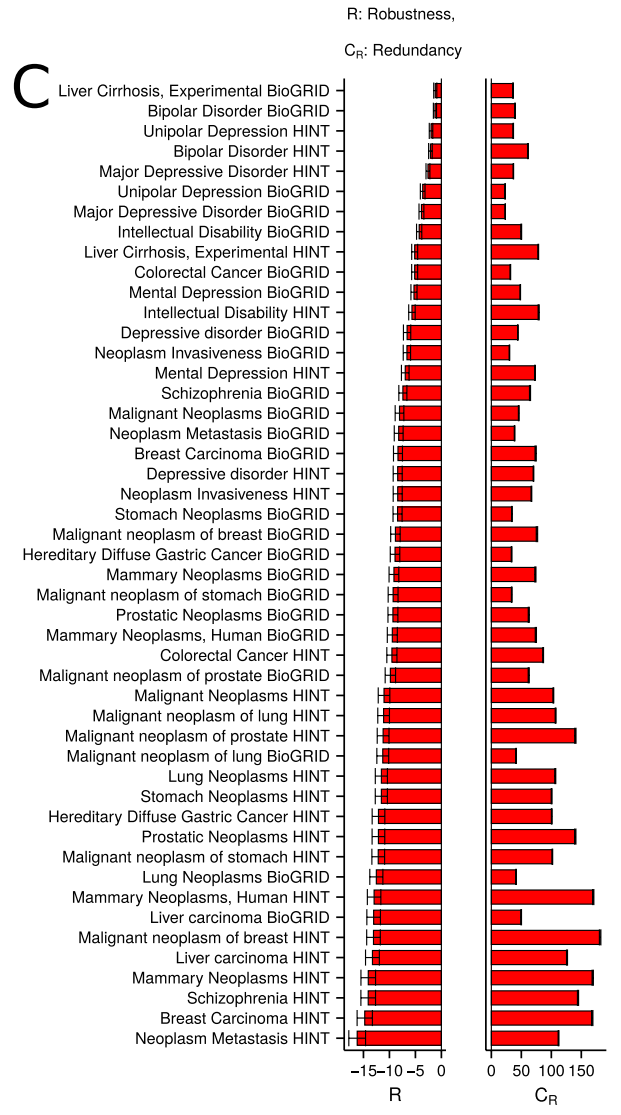
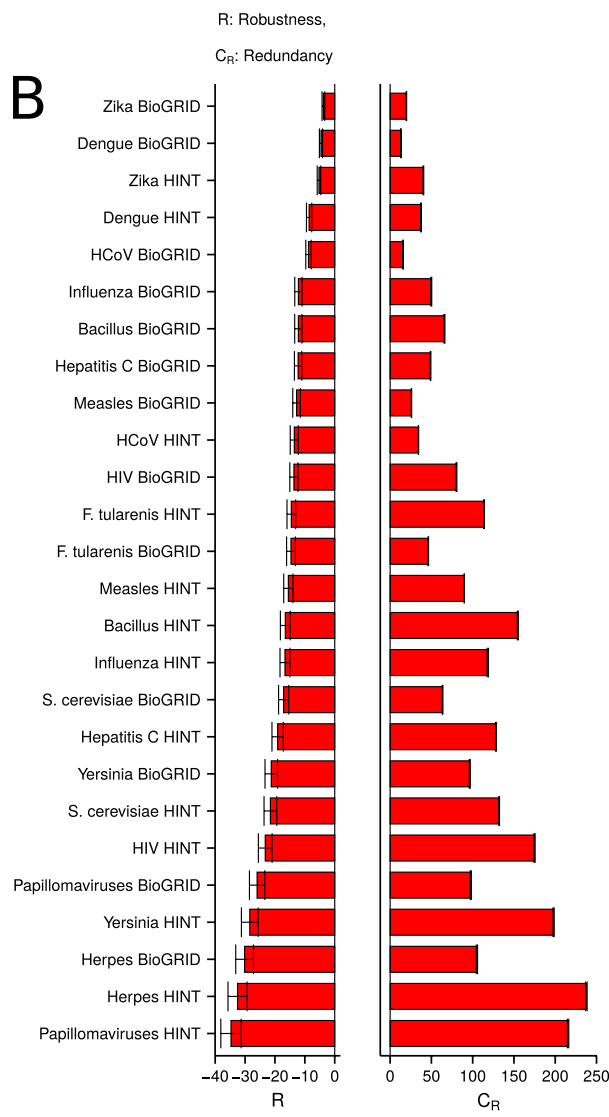
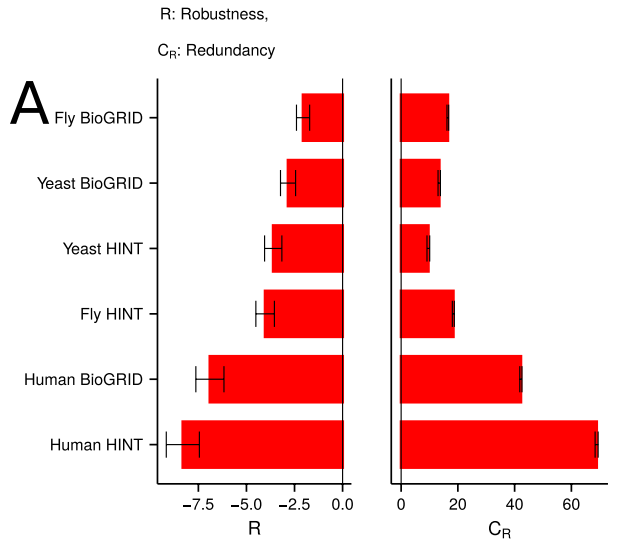
We conclude that attacking essential genes breaks the multiplex faster than attacking a random set of genes, which shows that essential genes and proteins are situated in a highly important part of the multiplex.

Pathogen- and disease-related genes and proteins are situated in essential parts of the human multiplex.

Pathogen. We collected human–pathogen protein–protein interaction data for 13 different pathogens. Data for 12 of these 13 pathogens was collected from a publicly available database, HPIDB 3.0^{62,63}. This is a curated database which currently contains 69,787 unique protein interactions between 66 host and 668 pathogen species. We studied interactions for 12 different human pathogens from HPIDB 3.0 (Fig. 5B). Besides these 12 pathogens, we also included human–pathogen protein interactions for various human coronaviruses (HCoVs). We collected a list of 119 human proteins which interact with various HCoVs⁶⁴. Therefore, in total we have 13 pathogens in our analysis.

Similar to essential genes, we assess topological relevance of pathogen-related genes and proteins using $RanDP$, $RanDP-RZ$ and $RanDP-noC_D$ null models. For all the pathogens, except Zika, targeted attack on the pathogen-related genes and proteins makes the multiplex highly vulnerable against all the null models (Fig. 5B and Supplementary Figures S41 and S42). As with essential genes, we also gauged the topological relevance of pathogen-related genes and proteins in TRN and PPI independently. Pathogen-related genes are highly vulnerable to attack and topologically essential in PPI (Supplementary Figures S26–S28). In TRN, majority of the pathogens exhibit a similar behavior (Supplementary Figures S23–S25). Similar to essential genes, pathogen-related genes are more vulnerable in the multiplex than in TRN or PPI independently (Supplementary Figures S29–S31).

For all the pathogens with high vulnerability to targeted attack, this vulnerability co-occurs with higher C_R for the pathogen-related genes and proteins compared to a random set of genes (Fig. 5B and Supplementary Figures S41 and S42). Further, given our simulations (Supplementary Figure S15), this suggests that higher redundancy makes the pathogen-related genes and proteins highly important for the human multiplex. R and C_R for pathogen-related genes and proteins are negatively correlated with Spearman's rank correlation of -0.82 (p-value = 2.025×10^{-6} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$), -0.87 (p-value = 1.873×10^{-6} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$) and -0.86 (p-value = 1.898×10^{-6} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$) against $RanDP$, $RanDP-RZ$ and $RanDP-noC_D$ null models, respectively (Supplementary Figures S7–S9). Even after controlling for the different number of pathogen-related gene–protein pairs for different pathogens, R and C_R are negatively correlated with Spearman's rank correlation of -0.71 (p-value = 6.706×10^{-5} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$), -0.85 (p-value = 1.905×10^{-6} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$) and -0.85 (p-value = 1.908×10^{-6} , two-tailed z-test using Fisher's z-transformation for $\alpha = 0.05$) against $RanDP$, $RanDP-RZ$ and $RanDP-noC_D$ null models, respectively (Supplementary Figures S7–S9). In agreement with this conclusion, pathogen-related genes and proteins are significantly enriched in the human multiplex (Supplementary Figure S13).



◀ **Figure 5.** Functionally important genes and proteins are redundant and essential. **(A)** (Left) Essential genes and proteins have lower robustness (R) to targeted attack across species against *RanDP-RZ* null model. (Right) Lower R is accompanied by higher redundancy (C_R) for essential genes and proteins against *RanDP-RZ* null model. Panels **(B–D)** show results for the human multiplex. **(B)** (Left) Pathogen-related genes and proteins have lower R to targeted attack across pathogens against *RanDP-RZ* null model. (Right) Lower R is accompanied by higher C_R for pathogen genes and proteins against *RanDP-RZ* null model. **(C)** (Left) Disease-related genes and proteins have lower R to targeted attack across diseases against *RanDP-RZ* null model. (Right) Lower R is accompanied by higher C_R for disease genes and proteins against *RanDP-RZ* null model. **(D)** (Left) Oncogenes (tumor suppressor genes (TSGs)) and proteins have lower R to targeted attack against *RanDP-RZ* null model. (Right) Lower R is accompanied by higher C_R for oncogenes (TSGs) and proteins against *RanDP-RZ* null model. In all the panels, database used for PPI networks is annotated on the y-axis (“Methods”). In all the panels, relative robustness (R) is measured against *RanDP-RZ* null model (“Methods”), and C_R values are calculated as the difference in redundancy against *RanDP-RZ* null model. Results against *RanDP* and *RanDP-noC_D* null models are shown in Supplementary Figures S42 and S41, respectively. Error bars show 95% CIs. For all the panels, C_R is calculated as the mean difference in the number of redundant edges between the species’ multiplex and the null model.

Disease. We collected disease-related genes from a publicly available database, DisGeNET^{65–67}. The current version (v6.0) contains gene-disease associations between 17,549 genes and 24,166 diseases, disorders, traits, and clinical or abnormal human phenotypes. We collected disease-gene associations for diseases which have at least 100 genes (with the HINT PPI network) in the human multiplex considered in this work. After this filtering, we retain 24 diseases in our analysis (Fig. 5C).

We assess topological relevance of disease-related genes and proteins using *RanDP*, *RanDP-RZ* and *RanDP-noC_D* null models. For most of the diseases, targeted attack on the disease-related genes and proteins makes the multiplex highly vulnerable against all the null models (Fig. 5C and Supplementary Figures S41 and S42). We also gauged the topological relevance of disease-related genes and proteins in TRN and PPI independently. Disease-related genes are highly vulnerable to attack and topologically essential in PPI (Supplementary Figures S26–S28). In TRN, a majority of the diseases exhibit a similar behavior (Supplementary Figures S23–S25). Similar to essential and pathogen-related genes, disease-related genes are more vulnerable in the multiplex than in TRN or PPI independently (Supplementary Figures S29–S31).

For all the diseases with high vulnerability to targeted attack, this vulnerability co-occurs with higher C_R for the disease-related genes and proteins compared to a random set of genes (Fig. 5C and Supplementary Figures S41 and S42). Further, given our simulations (Supplementary Figure S15), this suggests that higher redundancy makes the disease-related genes and proteins highly important for the human multiplex. R and C_R for disease-related genes are negatively correlated with Spearman’s rank correlation of -0.30 (p-value = 0.03986, two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$), -0.7129 (p-value = 6.898×10^{-8} , two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) and -0.608 (p-value = 7.29×10^{-6} , two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) against *RanDP*, *RanDP-RZ* and *RanDP-noC_D* null models, respectively (Supplementary Figures S10–S12). Even after controlling for the different number of disease-related gene–proteins pairs for different diseases, R and C_R are negatively correlated with Spearman’s rank correlation of -0.35 (p-value = 0.01475, two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$), -0.68 (p-value = 3.131×10^{-7} , two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) and -0.56 (p-value = 4.589×10^{-5} , two-tailed z-test using Fisher’s z-transformation for $\alpha = 0.05$) against *RanDP*, *RanDP-RZ* and *RanDP-noC_D* null models, respectively (Supplementary Figures S10–S12). In agreement with this conclusion, disease-related genes and proteins are significantly enriched in the human multiplex (Supplementary Figure S13).

We also collected the set of genes that contain mutations which have been causally implicated in cancer from the Network of Cancer Genes (NCG) database⁶⁸. NCG also includes information on whether a given cancer gene is an oncogene or a tumor suppressor gene (TSG). As before, we conduct robustness analysis on the set of oncogenes and TSG against the three null models. We find that targeted attack on oncogenes (or TSGs) makes the multiplex highly vulnerable (Fig. 5D and Supplementary Figures S41 and S42). These sets are more vulnerable in the multiplex than in TRN or PPI independently (Supplementary Figures S23–S31). Multiplex vulnerability co-occurs with higher C_R for the set of oncogenes (or TSGs) (Fig. 5D and Supplementary Figures S41 and S42).

Discussion

Recently, robustness properties of different biological multiplex and multilayer networks have been studied. These include brain networks^{51,58}, multiplex of PPI interactions⁵⁸, and a TRN-metabolic multilayer network⁶⁹. However, these studies have limitations, which are described next. Kleineberg et al.⁵⁸ do not draw broader conclusions about the organization of PPI multiplex across different species. Further, the different layers encode different types of interactions between proteins. This framework does not capture interaction between different types of molecules. Their conclusions are based on a generative model of network growth. This makes the results contingent on the accuracy of the generative model. This generative model is based on geometric principles and does not incorporate biological motivations or mechanisms. Klosik et al.⁶⁹ study robustness under random failure of a TRN-metabolic multilayer network. The study only focuses on *E. coli* and there are no species wide comparisons. Moreover, they do not study the dependence of robustness on degree–degree coupling and redundancy. Another recent study focuses on the interdependent or multiplex network of TRN–PPI-metabolic networks in human⁷⁰. Liu et al. 2019⁷⁰ show that this multiplex is more robust than an uncoupled or shuffled multiplex. They also showed that essential and cancer genes are preferentially arranged in essential parts of the

multiplex. However, they do not study the dependence of robustness on multiplex properties. Further, there is no cross-species analysis.

This study bridges the gap between theoretical developments and sub-cellular multilayer networks of molecular interactions. The central goal of our work is to investigate the organization and traits of molecular multiplexes. We focus our attention on the multiplex of TRN and PPI networks across five different eukaryotes. Our analysis spans five different TRN and 9 different PPI networks. We show that degree–degree coupling and redundancy are universal principles that shape robustness of the multiplex. Both are independent modulators of robustness. Though maximum robustness is achieved for a degree–degree coupling of 1 and a completely redundant multiplex, the observed species multiplexes have low absolute degree–degree coupling and redundancy. This suggests that robustness is not the only evolutionary pressure shaping the TRN–PPI multiplex. Independence might be a countering force to robustness. One possible explanation for low absolute degree–degree coupling, redundancy and hence robustness could be an inability to tune degree–degree coupling and redundancy independently. Redundancy and degree–degree coupling are positively correlated across species (Supplementary Figure S14). Multiplexes in nature might be tuning degree–degree coupling and redundancy in unison. Therefore, increasing robustness would increase redundancy as well. If redundancy were high, both TRN and PPI layers would be encoding similar interactions and the amount of unique information captured by the multiplex would be low⁷¹. Species multiplexes might have an upper bound on redundancy which could explain the low absolute values for robustness.

Robustness, independence and redundancy are only some of the pressures which might affect the structure of TRN–PPI multiplex across the domains of life. Other topological factors might possibly be involved. For instance, theoretical studies have established that controllability⁷² and navigability⁷³ both depend on multiplex structure. Further, these results have been confirmed in macroscale networks^{72,73}. For multiplex networks, with one-to-one correspondence between the nodes in the two layers, controllability decreases with increasing degree–degree coupling⁷². This means that the TRN–PPI multiplex might become less controllable at high degree–degree coupling, and more genes and proteins will need to be controlled to steer the multiplex towards a desired state. Navigability is negatively affected by redundancy as well. As the number of overlapping edges, and hence redundancy, increases, navigability decreases⁷³. Navigability is quantified by two different metrics; maximum entropy of trajectories explored by a random walker over the multiplex, and uniformity in the steady state probability distribution of node occupation under random walks over the multiplex. Maximum entropy decreases and probability distribution of node occupation becomes more heterogeneous with increasing edge overlap. At high redundancy or edge overlap, low maximum entropy will mean that a random walker can only explore a limited set of trajectories, and highly heterogeneous steady-state distributions will lead to unbiased occupancy among nodes. Therefore, controllability and navigability might exert countering pressures to robustness in shaping the structure of the TRN–PPI multiplex. Disassortative mixing is another important property of molecular networks⁷⁴. Individual network disassortativity might interact with multiplex coupling to create higher order effects, where degrees of neighbors in individual networks might be coupled in the multiplex. Such higher order coupling may have additional impact on robustness and other multiplex properties.

We have identified degree–degree coupling and redundancy as two modulators of TRN–PPI multiplex robustness across five different eukaryotes. These modulators can potentially be tuned to control robustness in naturally existing TRN–PPI multiplexes. Further, we can even custom design synthetic TRN–PPI multiplexes to have desired robustness values. For instance, if robustness is the desired property for a set of genes, the multiplex could be rewired such that protein hubs are also highly regulated transcriptionally. On the other hand, if independence between TRN and PPI layers is the desired behavior, degree–degree coupling and redundancy can be reduced synthetically. In principle, similar ideas can be extended to multiplexes comprising different types of molecular species, for example, protein coding mRNA, miRNA and protein-binding mRNA. The results of this study can be easily extended to other molecular multiplexes and can inform the design of novel multiplexes with different molecular species to achieve a desired biological function.

Besides the global design principles for multiplex organization, we have also shown that functionally important genes and proteins have a distinct distribution over the TRN–PPI multiplex. Essential, disease- and pathogen-related genes and proteins are preferentially situated in essential parts of the multiplex. This topological placement is dictated by redundancy. Attack on these functionally important genes quickly dismantles the multiplex. For diseases and pathogens, this suggests that these diseases and pathogens might have evolved with the human multiplex and preferentially interact with the vulnerable genes and proteins. Thus, multiplex framework can be useful in the study of disease evolution. Network analysis has previously been used for repurposing existing drugs⁷⁵. We believe that our multiplex approach might help in better identification of drug targets, since a multiplex better captures the complexity of the underlying molecular networks. Therefore, multiplex framework might have application in network medicine.

One limitation of any network analysis of molecular interactions is incomplete data. This problem is further compounded due to partial overlap between the observed TRN and PPI networks. However, it has been shown previously that if the size of the incomplete network is above a certain threshold such that a giant component exists, incomplete networks are representative of the complete network⁷⁶. This suggests that our analysis is representative of the complete species multiplexes. Addition of more genes and proteins in the multiplex might change the specific values of C_D , C_R and R , however the general dependence between robustness and degree–degree coupling and redundancy will hold. Further, we have not incorporated information about isoform proteins in this study. However, it is straightforward to include such information. In the presence of isoforms, the correspondence between TRN genes and PPI proteins will be one-to-many rather than one-to-one.

Quantifying structure and topology in complex biological networks has been actively researched within network biology. Design principles include the universality of scale-free networks^{1,77} (from metabolic networks across species²⁸ and gene regulatory networks^{5,8,13,30,78,79}, to power grids and the internet⁸⁰), lethal deletions

in the hubs of yeast¹⁶, disassortative mixing in molecular networks⁷⁴, and the existence of sub-modules and reoccurring network motifs^{9,12}. Network biology has only recently investigated the influence of multilayered multiplex networks in comparison to single network layers in isolation^{51,58,69–71,81}. This study contributes to the understanding of internetwork connectivity in layered molecular interaction networks. It is the first to compare TRN–PPI multiplexes across species. We discover global trends across species with degree–degree coupling and edge redundancy positively correlated with increased robustness. Robustness is explored in the context of the TRN–PPI multiplex and is proposed as one of the selective pressures by which evolution has shaped internetwork connectivity, degree–degree coupling and redundancy. The design principles presented here may be useful for the future design and understanding of multiplex networks and to improve efficacy for targeting specific gene subgroups, e.g. in disease. This research presents a multiplex framework for additional investigations of design principles in interlayered biological networks.

Methods

Data. Networks. We compiled network data for nine different species. These nine species span two domains of life, namely bacteria and eukaryotes. There are three bacteria—*H. pylori*, *M. tuberculosis* and *E. coli*—and six eukaryotes—*S. cerevisiae*, *C. elegans*, *D. melanogaster*, *A. thaliana*, *M. musculus* and *H. sapiens*. For the eukaryotes, we collected three datasets—one TRN and two PPI networks. Among the bacteria, *E. coli* also has three datasets (one TRN and two independent PPI networks), while *H. pylori* and *M. tuberculosis* have one TRN and one PPI network each. These datasets have been collected from diverse sources (see Supplementary Table S1). We use PPI data from multiple published sources—species-specific publications^{15,18,24}, BioGRID database¹⁴ and HINT database⁵⁷. Different experimental methods uncover different information about PPI networks²⁶. Therefore, we only use PPI networks inferred from Yeast two-hybrid (Y2H) experiments⁸². Y2H infers binary protein–protein interactions and is a prominent strategy for identifying protein–protein interactions⁸³. BioGRID and HINT do not have data for *H. pylori* and *M. tuberculosis*. PPI networks for these bacteria were collected from individual publications, Häuser et al.¹⁵ and Wang et al.²⁴ respectively. *E. coli* only exists in HINT. We include another published PPI network for *E. coli*¹⁸. A consolidated database of TRN networks across species does not exist. Therefore, we collected protein–DNA interactions from different publications. References for all the species are given in Supplementary Table S1. Available TRN and PPI networks are incomplete. Consequently, they only contain a fraction of the total number of possible genes and proteins in the genome and proteome, respectively. Further, TRN and PPI networks used in this study have different numbers of genes and proteins (see Supplementary Table S1, Supplementary Information Additional File 2). For a given species, we have only considered genes and proteins which are present in both TRN and PPI networks in our analysis.

The following characteristics of the TRN and PPI networks used in this study are included in Supplementary Table S1—Number of genes and proteins, % proteome coverage in the TRN–PPI multiplex (fraction of the total proteome covered in the multiplex), number of network edges (edges represent connections between genes and proteins in TRN and PPI networks respectively), average degrees (average K in PPI and average k_{in} or k_{out} in TRN) and size of the Largest Connected Component (LCC) (subset of genes/proteins in TRN/PPI where every gene/protein is reachable from every other gene/protein) are shown for all nine species.

Essential genes. List of essential genes was collected for three species (yeast, fly and human) from the Online GENE Essentiality (OGEE) database^{60,61}. The database has gene essentiality information on 48 species.

Pathogen-related genes. We collected human–pathogen protein–protein interaction data for 12 different pathogens from a publicly available database HPIDB 3.0^{62,63}. This curated database contains 69,787 unique protein interactions between 66 host and 668 pathogen species. Human–pathogen protein interactions for various human coronaviruses (HCoV)s were collected from a recently published paper⁶⁴. In total, we analyzed pathogen-related gene–protein pairs for 13 pathogens.

Disease-related genes. We collected disease-related genes from a publicly available database DisGeNET^{65–67}. The current version (v6.0) contains gene–disease associations between 17,549 genes and 24,166 diseases, disorders, traits, and clinical or abnormal human phenotypes. We collected disease–gene associations for diseases which have at least 100 genes in the human multiplex considered in this work.

Oncogenes and tumor suppressor genes. We collected the set of genes which contain mutations which have been causally implicated in cancer from the Network of Cancer Genes (NCG)⁶⁸. NCG also includes information on whether a given cancer gene is an oncogene or TSG.

Multiplex formulation of transcriptional regulatory and protein–protein interaction networks. TRN and PPI networks are modeled as interdependent networks (Fig. 1A). TRN layer encodes the transcriptional program for producing proteins from genes. The proteins translated from the TRN layer participate in protein–protein interactions in the PPI layer. There is one-to-one correspondence between genes and proteins in the TRN and PPI network layers. This specific configuration of interdependent networks can be reduced to a multiplex network⁴³, and we can apply the framework developed by Buldyrev et al.⁴³.

We use graph theory to model and analyze TRN–PPI multiplex in this work. TRN and PPI networks are modeled as graphs with nodes representing genes and proteins respectively (Fig. 1A). Connections between nodes are represented by edges. PPI edges are undirected. Edges in TRN have directionality—transcription factors have edges emanating from them, while downstream genes have incoming edges. The connectivity pattern of edges

is quantified by the concept of degree at each node. In PPI networks, degree (K) is the number of edges incident on a protein. For TRN, in-degree (k_{in}) is the number of transcription factors upstream of a gene, and out-degree (k_{out}) is the number of genes downstream of a transcription factor.

Since TRN and PPI layers have different coverage of the genome and proteome (see Supplementary Table S1), all analysis was done with genes and proteins present in both TRN and PPI networks.

Quantifying multiplex coupling. *Degree–degree coupling.* We quantify degree–degree coupling (C_D) using either Pearson’s correlation or Spearman’s rank correlation coefficient (Eq. 1).

$$C_D = \text{cor}(k_{out}, K), \quad (1)$$

where $\text{cor}()$ is the sample Pearson’s correlation or Spearman’s rank correlation coefficient, k_{out} is the TRN out-degree and K is the PPI degree.

Redundancy coupling. Redundancy coupling (C_R) is quantified by the number of edges simultaneously present in TRN and PPI. Assume that G^1 and G^2 are graphs representing TRN and PPI networks respectively, and V^1 and V^2 are the corresponding vertex sets. Let $E(G^1)$ and $E(G^2)$ be the edge sets for G^1 and G^2 respectively. The elements of the edge sets are vertex pairs. For instance, $(V_i^1, V_j^1) \in E(G^1)$ means that gene i is a transcription factor regulator for gene j , $(V_i^2, V_j^2) \in E(G^2)$ means that proteins i and j interact with each other. Interactions common between TRN and PPI can be mathematically represented by the number of common edges between G^1 and G^2 (Eq. 2).

$$\text{Edges}_{12} = |\{e \mid e \in E(G^1), e \in E(G^2)\}|, \quad (2)$$

where Edges_{12} is the number of edges common between G^1 and G^2 and e represents an edge either in G^1 or G^2 . In Fig. 2, we compute node-specific redundancy coupling. Here, C_R for each gene–protein is equal to the number of redundant edges incident on that gene–protein pair. C_R is either calculated as a z-score, which is computed as $C_R = \frac{\text{Edges}_{12} - \text{mean}(\text{Edges}_{12}^{\text{null}})}{\text{sd}(\text{Edges}_{12}^{\text{null}})}$, where $\text{Edges}_{12}^{\text{null}}$ is the number of redundant edges in a null model, or $C_R = \text{Edges}_{12}$ or $C_R = \text{mean}(\text{Edges}_{12}) - \text{mean}(\text{Edges}_{12}^{\text{null}})$. The definition of C_R used is specified in each figure’s caption.

Multiplex robustness. *Quantifying multiplex robustness.* We use MCGC to quantify multiplex robustness⁴³. MCGC is the set of genes and proteins which are simultaneously connected in both the network layers—every gene/protein in MCGC is reachable from every other gene/protein in MCGC. MCGC is computed by finding the intersection between the largest connected components (LCCs) of the TRN and PPI network layers. To quantify response to targeted attack, we track the size of the largest MCGC at each step of the attack. We simulate attack on the multiplex via the following algorithm.

1. Compute multiplex degree for all gene–protein pairs in the multiplex. Multiplex degree is defined as $K_{\text{mult}}(i) = \max(K(i), k_{out}(i))$ ⁵⁸, where $K_{\text{mult}}(i)$ is the multiplex degree for gene–protein pair i , $K(i)$ is the degree of protein i in the PPI network and $k_{out}(i)$ is the out-degree of gene i in the TRN network. Order multiplex degrees into a list of gene–protein pairs, D , arranged in decreasing order of multiplex degree.
2. At step L of the attack, remove the L th gene–protein pair in D . Removing a gene (protein) from TRN (PPI) layer may lead to the failure of dependent proteins (genes) in the PPI (TRN) layer. This failure may progress recursively, affecting more nodes in the multiplex. This process is called a cascade of failures⁴³.
3. After removing the attacked gene–protein pair and other failed dependent nodes at step L (cascade of failures), find the LCC in either TRN or PPI layer. At this stage, MCGC coincides with the LCC. Compute the size of MCGC. For computing size of MCGC, the TRN network is converted to an undirected version. Therefore, we calculate “weak” MCGC, where weak refers to the undirected nature of TRN.
4. Repeat steps 2 and 3 of this algorithm until MCGC breaks down.

This algorithm will generate a sequence of values, which give the trajectory of the MCGC as the multiplex is successively attacked. If we plot this trajectory as a function of the fraction of gene–protein pairs removed from the multiplex, robustness to attack can be assessed from either the area under the curve or the critical number of nodes removed for which the MCGC is fragmented⁴⁶. We use area under the curve (*RobustArea*) as the measure for multiplex robustness. Thus, *RobustArea* is given as

$$\text{RobustArea} = \int_0^1 \text{RMCGC}(f) \, df, \quad (3)$$

where *RobustArea* $\in (0, 1]$ is the area under the curve, f is the fraction of gene–protein pairs removed during the attack, and $\text{RMCGC}(f)$ is the size of MCGC relative to the total number of gene–protein pairs in the multiplex (n) after f fraction of gene–protein pairs have been removed from the multiplex. *RobustArea* quantifies absolute robustness. Relative robustness (R) is computed by comparing *RobustArea* against a null model. Cohen’s d is used to compute effect size of relative robustness (Eq. 4).

$$R = \frac{\text{mean}(\text{RobustArea}_{\text{obs}}) - \text{mean}(\text{RobustArea}_{\text{null}})}{\sqrt{\frac{\text{var}(\text{RobustArea}_{\text{obs}}) + \text{var}(\text{RobustArea}_{\text{null}})}{2}}}, \quad (4)$$

where R is the relative robustness, $RobustArea_{obs}$ and $RobustArea_{null}$ are the $RobustArea$ values for the observed multiplex and null model respectively and $mean()$ and $var()$ are the mean and variance functions respectively. We have assumed that $RobustArea_{obs}$ and $RobustArea_{null}$ have the same number of samples.

Since the attack is stochastic, given multiple nodes can have the same multiplex degree, we repeat targeted attack multiple times. For the species multiplexes, we repeat the attack either 1000 or 100 times.

Robustness for partial attack curves. For Fig. 5, we estimate robustness for a subset of functionally important gene–protein pairs and compare that against a random set of gene–protein pairs in the multiplex. Here, we explain the strategy to conduct such a comparison. Assume that $S = \{S_1, S_2, \dots, S_M\}$ is a collection of sets of gene–protein pairs in a multiplex. Here M is the total number of sets. Sets S_i , $i \in \{1, 2, \dots, M\}$, can be mutually exclusive or not. Let M_{min} be the size of the smallest set in S . For an equitable comparison, we randomly sample (under an appropriate model) M_{min} number of gene–protein pairs from all the subsets, except for the smallest subset. We sample each subset 100 times. For each sampled version of a subset $S_i \in S$, we attack the gene–protein pairs in S_i in decreasing order of multiplex degree for gene–protein pairs in that set, using the attack algorithm explained previously. We stop the attack once M_{min} number of gene–protein pairs have been removed from the multiplex. We also perform a similar partial attack on a set of randomly selected gene–protein pairs (under an appropriate null model). The size of the random set is set equal to M_{min} . Robustness can be calculated for each subset based on the obtained partial attack curves. Relative robustness for each subset is calculated by comparing $RobustArea$ for that subset against the random set.

Data availability

All TRN and PPI networks are provided as an R programming language⁸⁴ data object (“NetworkMultiplex.RData” in Supplementary Dataset). Data for all the pathogens and diseases are also provided as R programming language data objects in Supplementary Dataset. All the analysis was performed in the R programming language⁸⁴. Custom scripts for reproducing Figs. 2, 3, 4, 5 are provided in Supplementary Dataset.

Received: 18 September 2020; Accepted: 23 December 2020

Published online: 13 January 2021

References

- Barabási, A.-L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101 (2004).
- Abdulrehman, D. *et al.* YEASTRACT: Providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface. *Nucleic Acids Res.* **39**, D136–D140 (2010).
- Blais, A. & Dynlacht, B. D. Constructing transcriptional regulatory networks. *Genes Dev.* **19**, 1499–1511 (2005).
- Deplancke, B. *et al.* A gene-centered *C. elegans* protein–DNA interaction network. *Cell* **125**, 1193–1205 (2006).
- Guelzim, N., Bottani, S., Bourguin, P. & Képès, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* **31**, 60 (2002).
- Han, J.-D.J. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88 (2004).
- Jin, J. *et al.* An Arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.* **32**, 1767–1773 (2015).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Reece-Hoyes, J. S. *et al.* A compendium of *Caenorhabditis elegans* regulatory transcription factors: A resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110 (2005).
- Sandmann, T. *et al.* A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436–449 (2007).
- Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64 (2002).
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14**, 283–291 (2004).
- Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).
- Häuser, R. *et al.* A second-generation protein–protein interaction network of *Helicobacter pylori*. *Mol. Cell. Proteom.* **13**, 1318–1329 (2014).
- Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41 (2001).
- Murali, T. *et al.* DroID 2011: A comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.* **39**, D736–D743 (2010).
- Rajagopala, S. V. *et al.* The binary protein–protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32**, 285 (2014).
- Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173 (2005).
- Schwikowski, B., Uetz, P. & Fields, S. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **18**, 1257 (2000).
- Stelzl, U. *et al.* A human protein–protein interaction network: A resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
- Szklarczyk, D. *et al.* STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2014).
- Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. Global protein function prediction from protein–protein interaction networks. *Nat. Biotechnol.* **21**, 697 (2003).
- Wang, Y. *et al.* Global protein–protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J. Proteome Res.* **9**, 6665–6677 (2010).
- Yook, S.-H., Oltvai, Z. N. & Barabási, A.-L. Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928–942 (2004).
- Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
- Guimera, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895 (2005).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651 (2000).
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190 (2002).

30. Babu, M. M., Teichmann, S. A. & Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.* **358**, 614–633 (2006).
31. Luscombe, N. M. *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308 (2004).
32. Ay, M., Goh, K.-I., Cusick, M. E., Barabási, A.-L. & Vidal, M. Drug–target network. *Nat. Biotechnol.* **25**, 1119–1127 (2007).
33. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56 (2011).
34. Goh, K.-I. *et al.* The human disease network. *Proc. Nat. Acad. Sci.* **104**, 8685–8690 (2007).
35. Hopkins, A. L. Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682 (2008).
36. Zhou, X., Menche, J., Barabási, A.-L. & Sharma, A. Human symptoms–disease network. *Nat. Commun.* **5**, 4212 (2014).
37. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499 (2002).
38. Kivelä, M. *et al.* Multilayer networks. *J. Complex Netw.* **2**, 203–271 (2014).
39. Ames, R. M., MacPherson, J. I., Pinney, J. W., Lovell, S. C. & Robertson, D. L. Modular biological function is most effectively captured by combining molecular interaction data types. *PLoS One* **8**, e62670 (2013).
40. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
41. Padi, M. & Quackenbush, J. Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst. Biol.* **9**, 80 (2015).
42. Yeager-Lotem, E. *et al.* Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proc. Nat. Acad. Sci.* **101**, 5934–5939 (2004).
43. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025 (2010).
44. Dong, G., Gao, J., Tian, L., Du, R. & He, Y. Percolation of partially interdependent networks under targeted attack. *Phys. Rev. E* **85**, 016112 (2012).
45. Lee, K.-M., Kim, J. Y., Cho, W.-K., Goh, K.-I. & Kim, I. M. Correlated multiplexity and connectivity of multiplex random networks. *New J. Phys.* **14**, 033027 (2012).
46. Liu, X., Stanley, H. E. & Gao, J. Breakdown of interdependent directed networks. *Proc. Nat. Acad. Sci.* **113**, 1138–1143 (2016).
47. Parshani, R., Buldyrev, S. V. & Havlin, S. Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.* **105**, 048701 (2010).
48. Parshani, R., Rozenblat, C., Ietri, D., Ducruet, C. & Havlin, S. Inter-similarity between coupled networks. *EPL (Europhys. Lett.)* **92**, 68002 (2011).
49. Radicchi, F. Percolation in real interdependent networks. *Nat. Phys.* **11**, 597 (2015).
50. Radicchi, F. & Arenas, A. Abrupt transition in the structural formation of interconnected networks. *Nat. Phys.* **9**, 717 (2013).
51. Reis, S. D. *et al.* Avoiding catastrophic failure in correlated networks of networks. *Nat. Phys.* **10**, 762 (2014).
52. Son, S.-W., Bizhani, G., Christensen, C., Grassberger, P. & Paczuski, M. Percolation theory on interdependent networks based on epidemic spreading. *EPL (Europhys. Lett.)* **97**, 16006 (2012).
53. Zhou, D. *et al.* Simultaneous first- and second-order percolation transitions in interdependent networks. *Phys. Rev. E* **90**, 012803 (2014).
54. Bianconi, G. *Multilayer Networks: Structure and Function* (Oxford University Press, Oxford, 2018).
55. Min, B., Do Yi, S., Lee, K. M. & Goh, K. I. Network robustness of multiplex networks with interlayer degree correlations. *Phys. Rev. E* **89**, 042811 (2014).
56. Cellai, D., López, E., Zhou, J., Gleeson, J. P. & Bianconi, G. Percolation in multiplex networks with overlap. *Phys. Rev. E* **88**, 052811 (2013).
57. Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* **6**, 92 (2012).
58. Kleineberg, K.-K., Buzna, L., Papadopoulos, F., Boguñá, M. & Serrano, M. A. Geometric correlations mitigate the extreme vulnerability of multiplex networks against targeted attacks. *Phys. Rev. Lett.* **118**, 218301. <https://doi.org/10.1103/PhysRevLett.118.218301> (2017).
59. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, 2013).
60. Chen, W.-H., Lu, G., Chen, X., Zhao, X.-M. & Bork, P. OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic acids research gkw1013* (2016).
61. Chen, W.-H., Minguez, P., Lercher, M. J. & Bork, P. OGEE: An online gene essentiality database. *Nucleic Acids Res.* **40**, D901–D906 (2011).
62. Kumar, R. & Nanduri, B. HPIDB—A unified resource for host–pathogen interactions. *BMC Bioinform.* **11**, S16. <https://doi.org/10.1186/1471-2105-11-S6-S16> (2010).
63. Ammari, M. G., Gresham, C. R., McCarthy, F. M. & Nanduri, B. Hpidb 2.0: A curated database for host–pathogen interactions. *Database* **2016** (2016).
64. Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell Discov.* **6**, 1–18 (2020).
65. Piñero, J. *et al.* Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**, (2015).
66. Piñero, J. *et al.* Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* gkw943 (2016).
67. Piñero, J. *et al.* The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2020).
68. Repana, D. *et al.* The network of cancer genes (ncg): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol.* **20**, 1 (2019).
69. Klosik, D. F., Grimbs, A., Bornholdt, S. & Hütt, M.-T. The interdependent network of gene regulation and metabolism is robust where it needs to be. *Nat. Commun.* **8**, 534 (2017).
70. Liu, X. *et al.* Robustness and lethality in multilayer biological molecular networks. *bioRxiv* 818963 (2019).
71. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **6**, 6864 (2015).
72. Nie, S., Wang, X. & Wang, B. Effect of degree correlation on exact controllability of multiplex networks. *Phys. A* **436**, 98–102 (2015).
73. Battiston, F., Nicosia, V. & Latora, V. Efficient exploration of multiplex networks. *New J. Phys.* **18**, 043035 (2016).
74. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
75. Guney, E., Menche, J., Vidal, M. & Barabási, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 1–13 (2016).
76. Menche, J. *et al.* Uncovering disease–disease relationships through the incomplete interactome. *Science* **347**, (2015).
77. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).
78. Faith, J. J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
79. Thieffry, D., Huerta, A. M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: A global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays* **20**, 433–440 (1998).
80. Barabási, A.-L. & Bonabeau, E. Scale-free networks. *Sci. Am.* **288**, 60–69 (2003).
81. Zheng, W., Wang, D. & Zou, X. Control of multilayer biological networks and applied to target identification of complex diseases. *BMC Bioinform.* **20**, 271 (2019).

82. Fields, S. & Song, O.-K. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245 (1989).
83. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10**, 2763–2788 (2009).
84. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2013).

Acknowledgements

R. D. D. acknowledges support by NSF CAREER (1943740). We are grateful to the members of the Dar lab for fruitful discussions on the manuscript.

Author contributions

R.D.D. and T.M. conceived the analytical and computational work. T.M. carried out the computational work. R.D.D. and T.M. analyzed and interpreted the data. R.D.D. and T.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80745-9>.

Correspondence and requests for materials should be addressed to T.M. or R.D.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021