A Context-Specific, Literature-Supported Framework for Validating Stress Response Models in Mammals

Bradley Frishman, University of Florida (frishmanbradley@ufl.edu); Jorge Gonzalez, Embry-Riddle Aeronautical University; Valery Forbes, Florida Atlantic University

Abstract

Computational models of stress responses can highlight candidate genes underlying physiological adaptation, but their utility depends on rigorous validation. Using existing biological databases, we tested a novel approach to identify and group differentially expressed genes (DEGs) from RNA-seq data. The approach was used to build a neural network to model the epigenetic response of human cells to temperature stress. A Key-Response, Treatment Specific, Support, and Noisy group of genes were identified as representing the Principal Response of the species to the stress. The Support Group was also suspected to represent housekeeping genes based on variability patterns.

To validate these assumptions, we built protein–protein interaction (PPI) networks using the Human Protein Atlas and STRING-db, incorporating both direct and second-order connections. Crucially, second-order connections were restricted to those made via DEGs, ensuring that connectivity reflected condition-specific stress responses rather than generic hubs.

Across two conditions, >75% of Principal Response genes assembled into LCCs significantly larger than random networks ($p<0.00005$). Support Group genes also showed strong connectivity and enriched overlap with a housekeeping gene database, supporting their distinct classification. STRING-db confirmed PPI enrichment but produced less reliable results than our DEG-restricted framework.

Overall, this study demonstrates that the model identifies biologically meaningful, interconnected stress-response networks. By emphasizing DEG-restricted second-order connections, our framework addresses key limitations of context-free enrichment methods and advances validation strategies for computational models of gene regulation, with implications for stress physiology and personalized medicine.

Introduction

This project focused on validating a computationally-derived stress response model in mammals, with the goal of assessing whether predicted gene networks reflect biologically meaningful interactions. Computational models of gene regulation are increasingly applied to study how organisms respond to environmental challenges, yet their credibility depends on rigorous validation against empirical data. A central obstacle is that existing protein-protein interaction (PPI) databases are not necessarily comprehensive nor specific to the cellular contexts of interest. Additionally, in our case, a limited number of replicates (n=5 human samples) limits the usefulness of traditional statistical validation methods. To address these challenges, we developed a context-specific network validation framework that integrates biological literature and second-order gene interactions, providing a more targeted alternative to enrichment tools such as those found on STRING-db.org.

The model tested here sought to identify the genes responsible for eliciting a principal response to temperature stress that is characteristic to a species. We began with RNA-seq data from five human samples exposed to both heat stress (41°C) and cold stress (32°C) conditions. The model identified differentially expressed genes (DEGs) using a novel approach due to the small sample size (n=5) of humans. The model, without relying on mean fold difference or statistical significance, defined DEGs as genes with an activity after treatment greater than at

baseline (Gonzalez et al., 2025). For a detailed explanation on how DEGs were identified, see definition 4.2 in Gonzalez et al., 2025; the goal of this study is not to describe the model itself but rather to develop and demonstrate a strategy for validating its assumptions—and those of other similar models. Nevertheless, the model employed a neural network with one hidden layer that relied on comparing baseline interindividual variation in RNA transcript count to post-stress interindividual variation in RNA transcript count. Of the 9825 candidate human genes, 2810 were identified to be DEGs in the 32º condition and 2679 in the 41º condition. Among the DEGs, the model categorized the extreme response genes, those which are thought to be most important in driving a response, into four groups (Figure 1). The Key-Response Group, Treatment-Specific Group, and Noisy Group (denoted as g1, g2, and g4, respectively) were believed to be part of a principal response to temperature stress characteristic of the species, while the final group was hypothesized to represent Support Genes (denoted as g3) based on their patterns of inter-individual variability. Support genes are thought to be well-regulated, like the Principal Response genes, but also likely to be involved in many processes for many species. We define Support Genes as genes which are constantly expressed but at fluctuating levels due to the fact that they are involved in different processes under different circumstances. Classically, Housekeeping Genes have been defined as genes that are consistently expressed across different cell types and conditions because they are required for essential cellular functions such as metabolism, transcription, and structural maintenance. We suspected that a significant amount of Support Genes were also Housekeeping Genes.

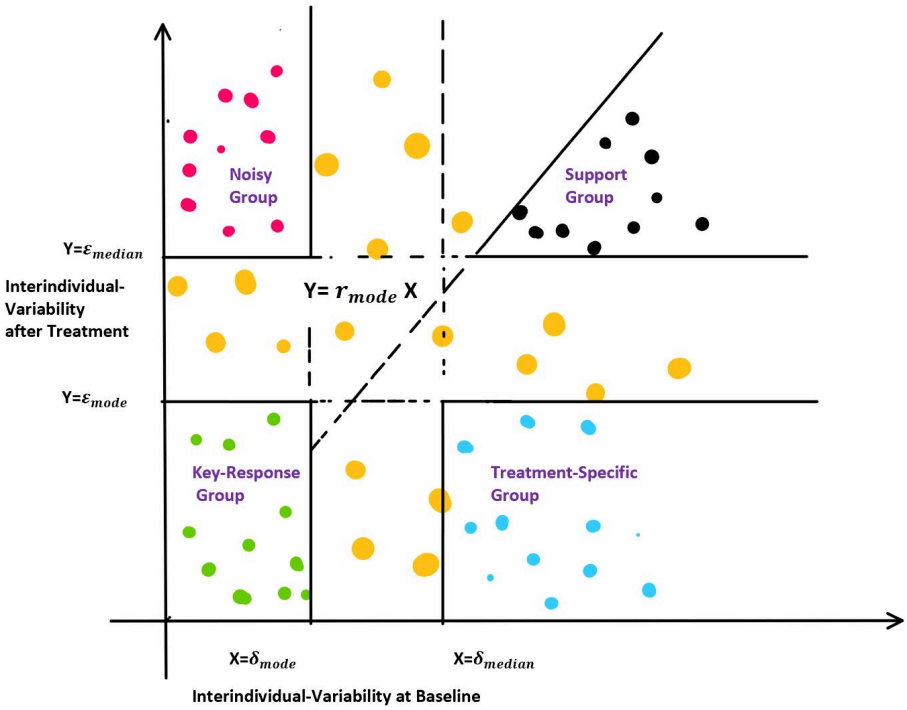


Figure 1. Graphical representation of the groups from the model evaluated in this study. Certain differentially expressed genes (DEGs) were classified according to their relative interindividual-variability at baseline and after treatment. Genes located at the extremes of the distribution depicted were categorized into four groups: Noisy, Key-Response, Treatment-Specific, and Support (Gonzalez et al., 2025).

Methods

Definitions:

1. First order connection: A direct model-gene-to-model-gene protein-protein interaction. Protein interactions were obtained from the Human Protein Atlas (PA) "Interaction" section, which aggregates data from four major sources: IntAct (high and medium confidence physical associations), BioGRID (multivalidated physical interactions), BioPlex (interactions with >75% probability), and

OpenCell (significant physical interactions). We included any known interaction reported in at least one of these four major sources.

2. Second order connection: is one in which two model-identified genes are connected via a literature-validated interaction with the same differentially expressed gene (as defined by the model). In other words, a bridge between model genes A and B would be formed if they shared an interaction with the same DEG. Note that Principal Response Genes ('model genes') were themselves DEGs and therefore could act as intermediates for the purpose of second order connections.

Network Validation Methods:

1. Public resources including the Human Protein Atlas and STRING-db were used to identify known interactions among genes found in the model-produced sets. Housekeeping and Reference Transcript Atlas's online database would be used to verify the status of genes in the Support Group, which were previously hypothesized to be housekeeping. Protein-protein interaction network construction included both first-order direct gene connection and second-order connections. As stated previously, second-order interactions were only considered if the intermediate (bridge) gene was itself also a DEG. The largest connected component (LCC) was chosen as the primary measure for validating the model, as we assumed it better reflects mechanistic coherence than edge counts. The network test statistic used should reflect the goal of the model being tested. In our case, the model attempts to precisely find the mechanism of the highest importance (referred to in Gonzalez et al., 2025 as the group of epigenetic drivers which make up the high order response of the species to the treatment). If the model truly identifies the genes responsible for the principal temperature-stress response mechanism (as it sought to do), then those genes should *not* appear as isolated entities (see Figure 3). Instead, they would ideally

form a cohesive mechanism, and the best evidence of such a mechanism is the presence of consistent and thorough protein–protein interactions documented in the literature. We aimed to determine whether the predicted Principal Response Genes would assemble into a network that resembles a biologically meaningful mechanism. We expect the model-selected genes to have a greater LCC than would be expected by random chance, thereby validating the use of interindividual variability (as the model did) in identifying the core mechanism of response of the species.

Degree-preserving randomizations (using 20 bins) provided null distributions against which observed connectivity could be compared. For each temperature condition, two networks were built: one from the genes found in the Key-Response, Treatment-Specific, and Noisy groups (Principal Response Genes) and a second from the Support Group genes on their own.

2. When considering the networks formed by both first and second order interactions, edges were counted identically, regardless of the type (first or second order). The algorithm also collapsed multiple interactions between the two same genes and formed a simple (no parallel edges), undirected graph (Figure 2).

3. In order to gain a better understanding of how significant a small observed difference in the LCC size may be, we performed degree-shifted reference scenarios. For each temperature condition, we first computed the degree distributions of the Principle Response genes, defined as the number of known interactions per gene. We then generated four modified degree distributions. In the conservative upshift, only the lowest non-zero degree bin was moved up by one. That is, we found the lowest degree bin and moved all the genes counted in that bin to the next higher bin. In the aggressive upshift, the entire degree distribution is shifted one bin to the right. The same logic applies for the converse. In the conservative *downshift*, we again moved

only the (number of) genes in the lowest bin down into an even lower degree bin. In the aggressive downshift, we shifted the entire distribution down by one bin. For each scenario, we repeatedly sampled background genes (this refers to the 9825 candidate genes) to match the target degree distribution as closely as possible, built PPI networks, and measured LCC.

4. String-db. was also used to generate an PPI enrichment p-value for the genesets. The web database version 12.0 was used (Szklarczyk et al., 2023) . For the analysis, the whole genome was used as the statistical background since our background list of 9825 was too large to accommodate. Default settings were used: full STRING network (the edges indicate both functional and physical protein associations) composed of interactions from textmining, experiments, databases, co-expression, neighborhood, gene fusion, and co-occurrence; medium confidence (0.400) minimum required interaction score.

Figure 2. Graphical Explanation of the Networks created in this study to validate the stress-response model.

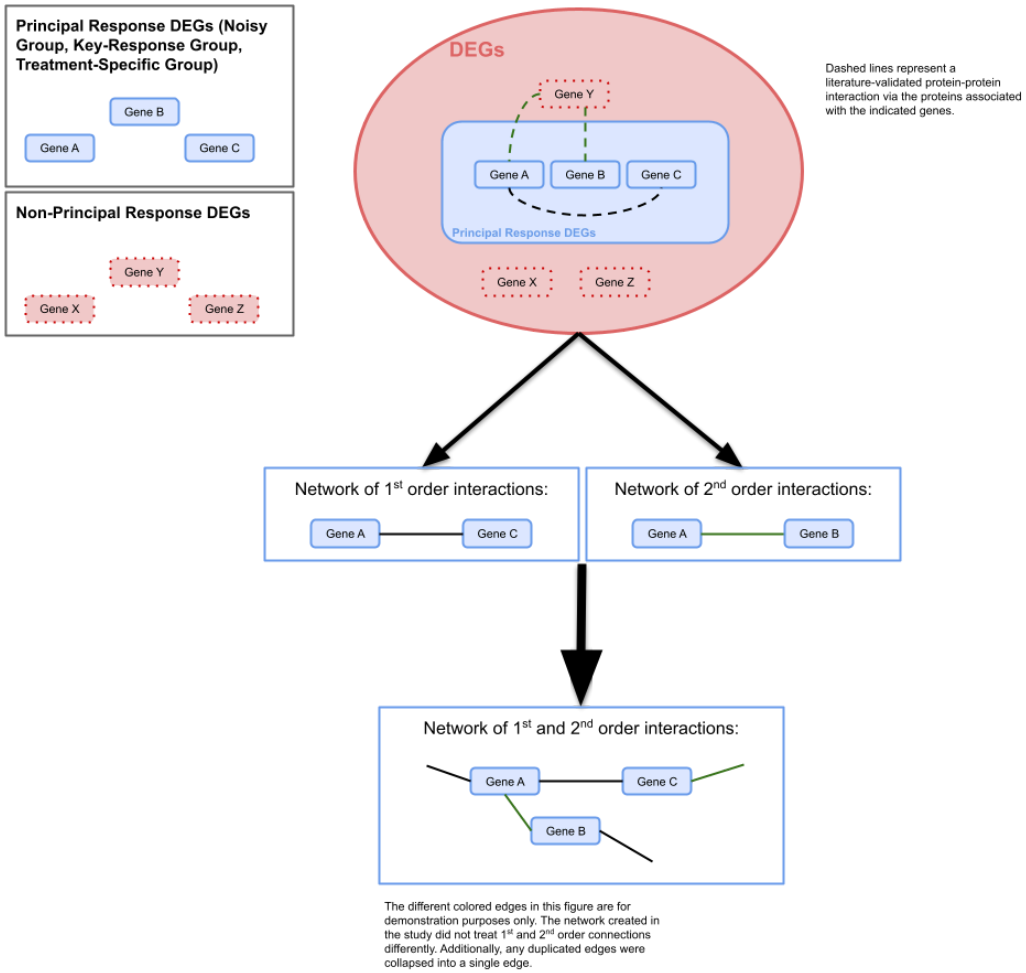
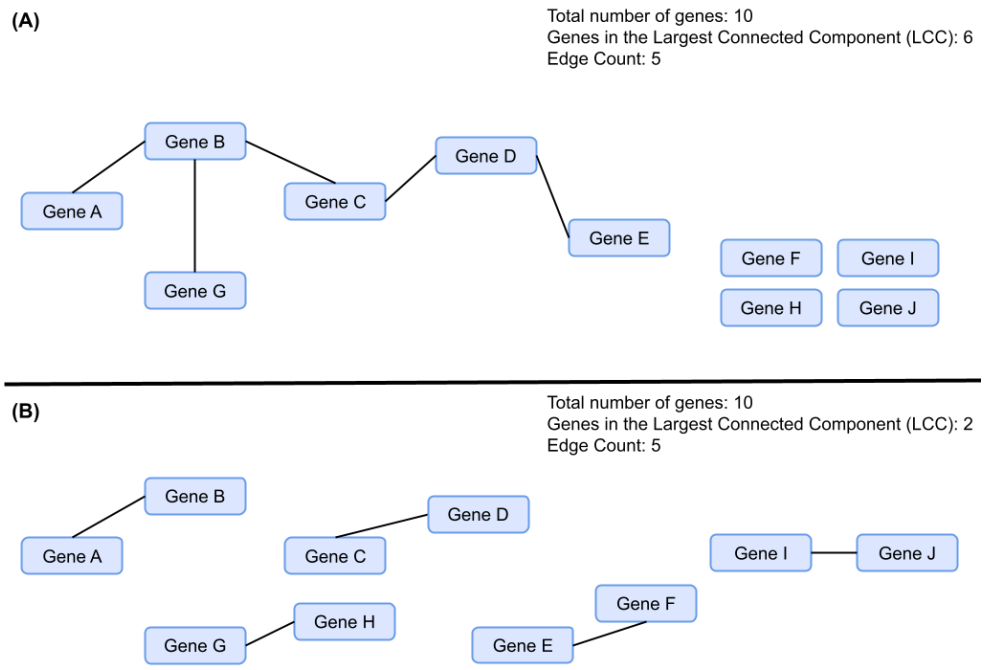


Figure 3. Gene networks consist of the same number of genes with different makeups.



We posit that (A) is more reflective of a driving epigenetic mechanism to the stress response than (B), despite their equal edge counts.

Housekeeping Gene Validation:

As previously mentioned, we suspected group 3/ Support Group genes to be enriched for Housekeeping Genes. To test this, in addition to performing the network analysis described above for solely these genes, we also cross-referenced a known housekeeping gene database. We used the Housekeeping and Reference Transcript Atlas (version 1.0), a web database which mined high-quality human RNA-seq data sets (Hounkpe et al., 2020).

We also compared basic statistics on the degree distributions (known interactions identified as described above) Principle Response Group genes and the Support Group genes.

Results

Model-identified Principle Response Group genes formed significantly more interconnected networks than expected by chance. At both temperature conditions, more than 75 percent of g1,g2,g4 genes were incorporated into the LCC, with p-values below 0.00005 when compared to randomized networks (Table 1).

The Support Genes group, g3, also showed strong evidence of biological validity. First, using the same network-construction and null comparison applied to other groups, the LCCs of g3 gene sets were found to be significantly larger than one would expect by chance (Table 2). Additionally, between 27 and 34 percent of these genes overlapped with an established housekeeping database, a much higher proportion than observed for g1,g2,g4 gene sets and the overall background prevalence (about 22% of all 9825 genes were in the HK database) (Table 3). We found that g3 genes tended to have a much higher degree compared to any other set of genes (Figure 4). Compared to the other groups, Support Group genes are known to be slightly more connected, with more genes having 50+ known interactions. This suggests this subgroup may include more hub-like genes. Similarly, the randomly generated files with genes that followed a comparable degree distribution as g3 tended to highly overlap with the housekeeping database. Furthermore, unlike the g1,g2, and g4 network, g3 gene networks had a significantly larger than expected edge count in their LCC (Table 2). Note that this grouping was created based on analyzing interindividual variability, completely independent from degree, but the results indicate the chosen genes are more well-connected than those in the other sets. In other words, results indicate that the tested model may be a reliable way to identify high-degree genes, which in turn are likely housekeeping genes.

STRING-db analysis confirmed that g1,g2,g4 model genes were enriched for protein-protein interactions with p-values of 5.86e-05 (32º condition) and 1.3e-08 (41º condition). The results from using STRING-db's PPI network builder tended to be more significant relative to randomly generated gene sets as well. Notably, the randomly generated sets (produced using a degree-preserving method), often showed highly inconsistent PPI enrichment significance despite consisting of ostensibly random genes.

Table 1. Results from the given network of genes that the model identified to be a part of the principle stress response (Principle Response Group: groups 1,2, or 4).

	32º	**41º**
Number of Model Genes Total	182	204
Number of Genes with ≥1 known interactions in PA	167 (91.76%)	186 (91.18%)
Number of Genes in Largest Connected Component [1st & 2nd Order]	134 (80.24%) Mean of random files: 125.55 P-value: 0.00000	141 (75.81%) Mean of random files: 134.75 P-value: 0.00002
Number of Genes in Largest Connected Component [2nd Order ONLY]	133 (79.64%) Mean of random files: 122.80 P-value: 0.00000	140 (75.27%) Mean of random files: 131.65 P-value: 0.00000

Table 2: Results for Support Group genes (g3) which were suspected to be enriched with housekeeping genes.

	32º	**41º**
Number of Genes Total	67	53
Number of Genes with ≥1 known interactions in PA	62 (92.54%)	46 (86.79%)
Number of Genes in **Largest Connected Component** (1 & 2nd Order)	57 (91.94%) Mean of random files: 51.45 P-value: 0.00000	37 (69.81%) Mean of random files: 35.90 P-value: 0.06494
Number of Genes in **Largest Connected Component** (2nd Order ONLY)	57 (91.94%) Mean of random files: 49.45 P-value: 0.00000	36 (67.92%) Mean of random files: 34.80 P-value: 0.06376
Number of **Edges in LCC** (1st & 2nd order)	476 Mean of random files: 313.35 P-value: 0.00000	168 Mean of random files: 139.15 P-value: 0.00000
Number of **Edges in LCC** (2nd order only)	469 Mean of random files: 311.80 P-value: 0.00000	165 Mean of random files: 142.00 P-value: 0.00000
Number of Edges in LCC Per # of Genes in LCC (2nd order only)	8.23 Other Groups in Model: 5.21	4.58 Other Groups in Model: 5.25

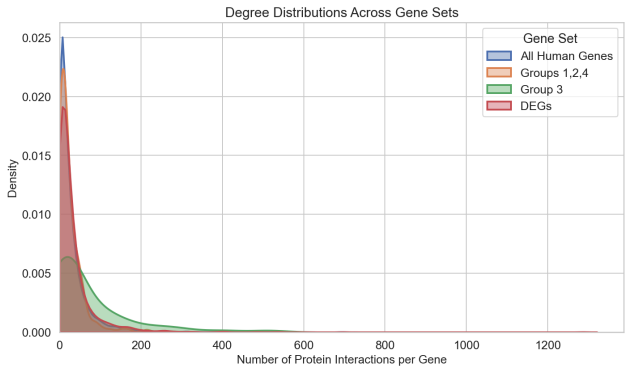
Table 3. Prevalence of model-selected Support Group genes (g3) in Housekeeping & Reference Transcript Atlas

	32º	41º

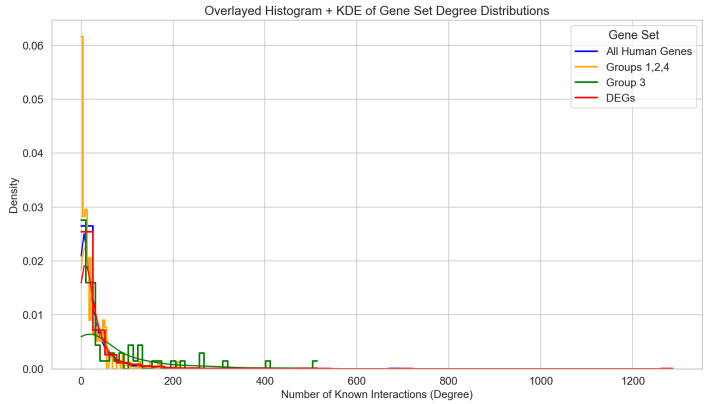
Percent of Support Group Genes found in Housekeeping Database	26.9%	34.0%
Percent of Key-Response, Treatment-Specific, and Noisy Group Genes found in Housekeeping Database	11.0%	11.76%
Mean Percent of Random* Genes Found in Housekeeping Database	27.31%	26.60%
Percent of All† Genes Found in Housekeeping Database	21.77%	

*Random files generated using degree preserving randomization of the corresponding g3 geneset.
†Background list of ~9800 genes from which DEGs were identified

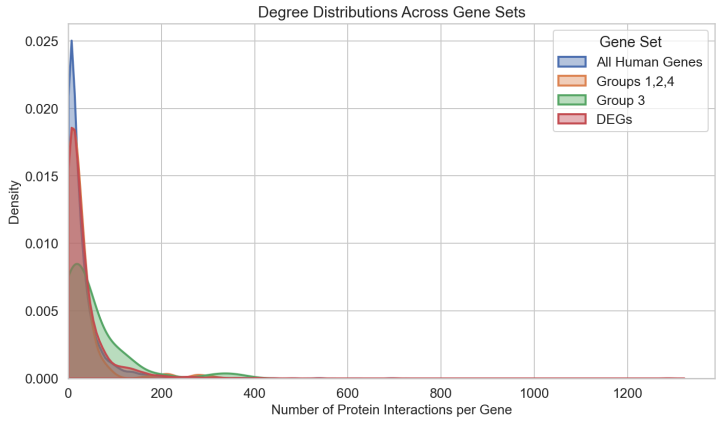
Figure 4. Analysis of Model Selected Group 3 Genes Number of Known Interactions (Degree)



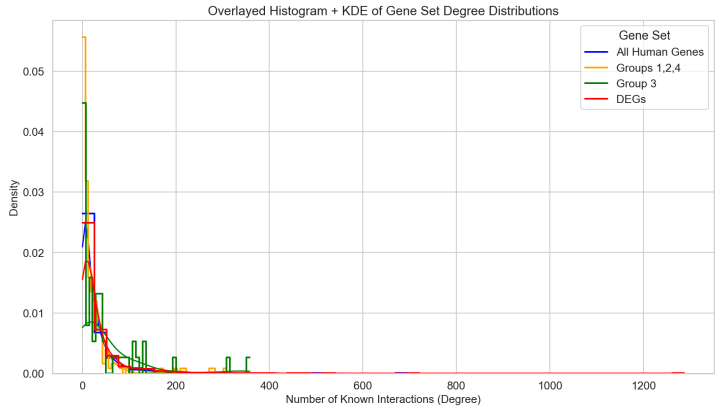
(a)



(b)



(c)



(d)

(e)

	All 9825 Human Genes	DEGs	Groups 1,2,4	**Group 3 (Suspected Housekeeping)**
Median	13.0	16.0	11.0	**22.0**

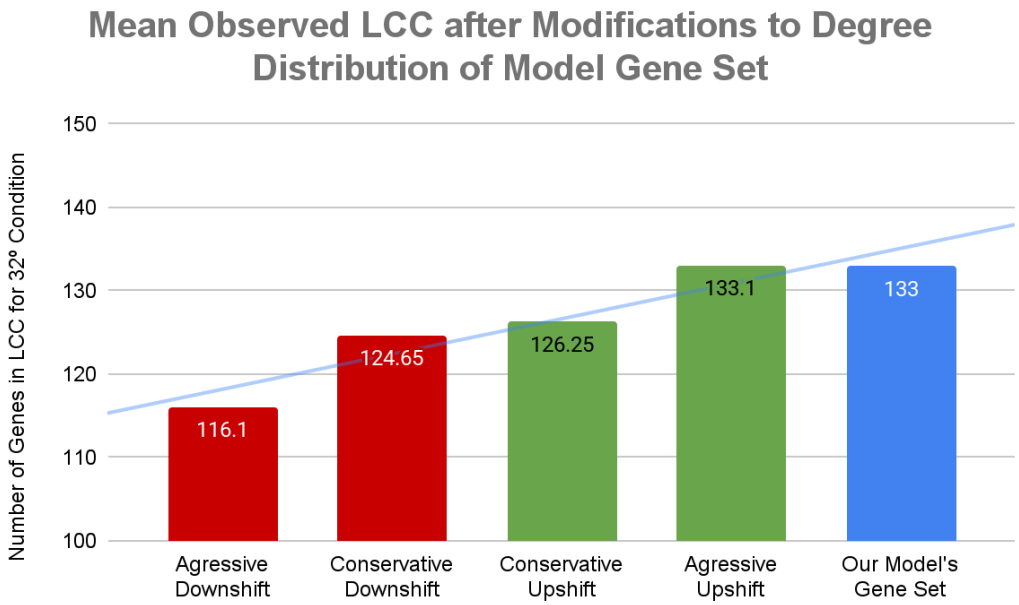
Mean	28.12	32.67	24.34	**48.68**
Std Dev	48.98	56.22	44.77	**72.21**

a. 32º. Kernel Density Estimation (KDE) with the x-axis representing the number of known interactors for each gene, based on HumanProteinAtlas data. The Y-axis ("Density") reflects the relative frequency (via a kernel density estimate) of genes with a given number of interactions within each group. Normalized so the area under each curve equals 1 (not absolute counts).
b. 32º. Histograms and KDE.
c. 41º. KDE with the x-axis representing the number of known interactors for each gene, based on ProteinAtlas data. The Y-axis ("Density") reflects the relative frequency (via a kernel density estimate) of genes with a given number of interactions within each group. Normalized so the area under each curve equals 1 (not absolute counts).
d. 41º. Histograms and KDE.
e. Statistical Analysis of the Varying Degree Distributions

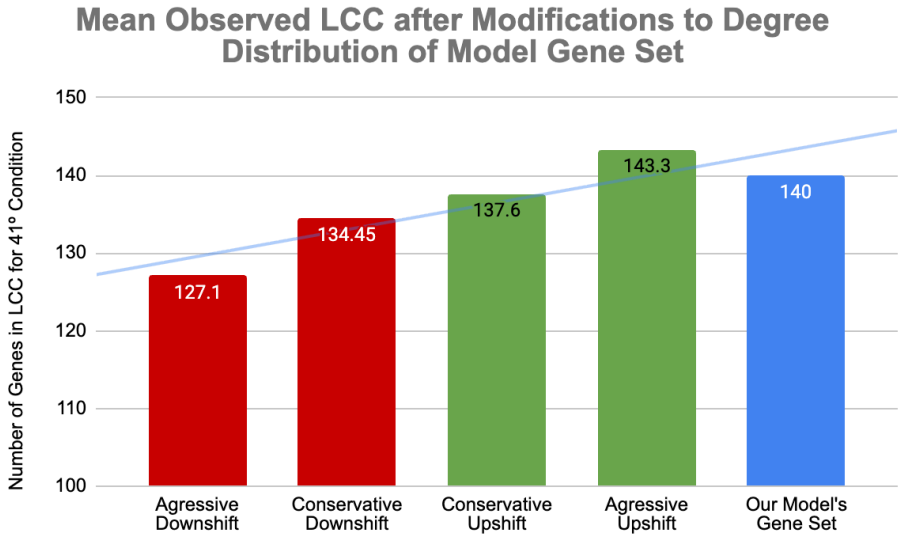
LCC Change per Distribution Shift:

After performing degree-shifted reference scenarios, we found that the genes selected to be used in the model tended to have a larger LCC even when compared to a gene set with more known interactions (Fig. 5). The purpose of this analysis is to bound the LCC and to gain insight into how sensitive it is to conservative or aggressive changes in degree distribution. Using this information, we can better assess how significant the apparent differences between LCCs are.

Figure 5. LCC Change After Distribution Shifts



(A)



(B)

(A) represents the results from the 31º condition, (B) represents the results from the 41º condition.

Discussion

Overall, this study shows that the tested stress response model identifies gene networks that are more biologically interconnected than would occur by random chance, underscoring the validity of its predictions. By restricting validation to second-order connections through DEGs, the approach enhanced specificity, reduced data bias, and provided clearer evidence of context-dependent biological relevance than traditional enrichment analyses. Most importantly, the observation that model-identified stress genes assemble into a strongly connected network supports the novel computational model's effectiveness. If the selected genes take part in a species' primary response to a particular stress, they must act together as part of a mechanism. That single mechanism (or network) should be complete, ideally composed of most to all of the selected genes. The literature-supported connectivity we observed is precisely the hallmark of such a mechanism.The findings illustrate the utility of tailoring network validation frameworks to experimental conditions rather than relying on broad, genome-wide baselines.

Comparison with an Existing Enrichment Analysis Method:

We assess that the results from STRING-db are less reliable than our own network building process for the following reasons. First, STRING's enrichment p-values are calculated using the entire genome of up to 20,000 genes as the background (genes which can be used in the randomly generated gene sets), whereas our analysis restricts comparison to the smaller list of 9825 genes that were actually studied. Using a relevant and restricted universal gene set for randomization creates a more appropriate null distribution. Second, our approach only incorporates second-order connections that are made through differentially expressed genes, directly tying the network structure to the experimental condition of interest; STRING does not employ this specificity. Third, while more statistically involved than our calculations, STRING

evaluates enrichment primarily by edge count, while we focus on the size of the largest connected component. As stated, we argue LCC is a more biologically meaningful measure of a functional mechanism, at least for our use case due to the goal of the tested model. These factors may explain why the enrichment results from STRING-db showed no consistent pattern across random gene sets, in contrast to the stability of our validation approach.

Support Group (Suspected Housekeeping) Genes:

Housekeeping genes and the proteins they encode have been shown to exhibit high degree in protein-protein interaction networks (Lin, Liu, & Hwang, 2009). There are several reasons why: First, because they encode proteins required for universal and continuous cellular processes such as transcription, translation, and energy metabolism, they function within complex molecular pathways which would require numerous physical interactions (Joshi et al., 2022). Their broad expression across all cell types also increases both true interaction opportunities and the observed connectivity due to higher cellular abundance. In other words, since these genes are found everywhere, they have more places to make interactions and are more likely to be well-studied. Finally, HK genes are typically evolutionarily ancient and conserved, thus they have had more time to accumulate interactions with other proteins (Joshi et al., 2022; Eisenberg & Levanon, 2003). These reasons likely underlie our findings that the high-degree Support Group genes (and their degree-comparable random gene set counterparts) were more likely to be found in the published housekeeping gene database. Note that the Support Group was created based on analyzing interindividual variability, completely independent from degree, but the results indicate these genes are more well-connected than the rest of the genesets. Support genes were the set of genes that have high inter-individual variability both before and after treatment (Gonzalez et al., 2025, definition 2.1.3). Gonzalez et al. assert that these are still

well-regulated genes in the sense that their behaviors are consistent before and after treatment. It was assumed that their variability remained consistently large because they were connected to many biological processes. Our results substantiate this assumption. They indicate that we may be able to define high-degree/housekeeping genes in terms of connectivity and in terms of variability since they have a similar likelihood of being found in the known HK database.

Future Direction

Future work would build on these results by validating individual model predictions against the biological literature and distinguishing between positive and negative forms of protein-protein interactions. Expanding the framework to other computational models of gene regulation under stress and other stressors will further test its generalizability. In the long term, this work contributes to establishing more reliable approaches for integrating computational modeling with experimental biology, which has implications for stress physiology, disease modeling, and personalized medicine.

Works Cited

Eisenberg, E., & Levanon, E. Y. (2003). Preferential attachment in the Protein Network Evolution. Physical Review Letters, 91(13). https://doi.org/10.1103/physrevlett.91.138701

Gonzalez, J., Genereux, D., Crouse, K., Hindle, A., Karlsson, E., Madelaire, C., & Forbes, V. (2025). New approaches to discovering epigenetic rules of homeostasis in diverse mammal species. 10.13140/RG.2.2.34754.64968.

Hounkpe, B. W., Chenou, F., de Lima, F., & De Paula, E. V. (2020, July 14). HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Research, 49(D1), D947-D955. https://doi.org/10.1093/nar/gkaa609

Joshi, C. J., Ke, W., Drangowska-Way, A., O'Rourke, E. J., & Lewis, N. E. (2022). What are housekeeping genes?. PLoS computational biology, 18(7), e1010295. https://doi.org/10.1371/journal.pcbi.1010295

Lin, Wh., Liu, Wc. & Hwang, Mj. Topological and organizational properties of the products of house-keeping and tissue-specific genes in protein-protein interaction networks. BMC Syst Biol 3, 32 (2009). https://doi.org/10.1186/1752-0509-3-32

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic acids research, 51(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000