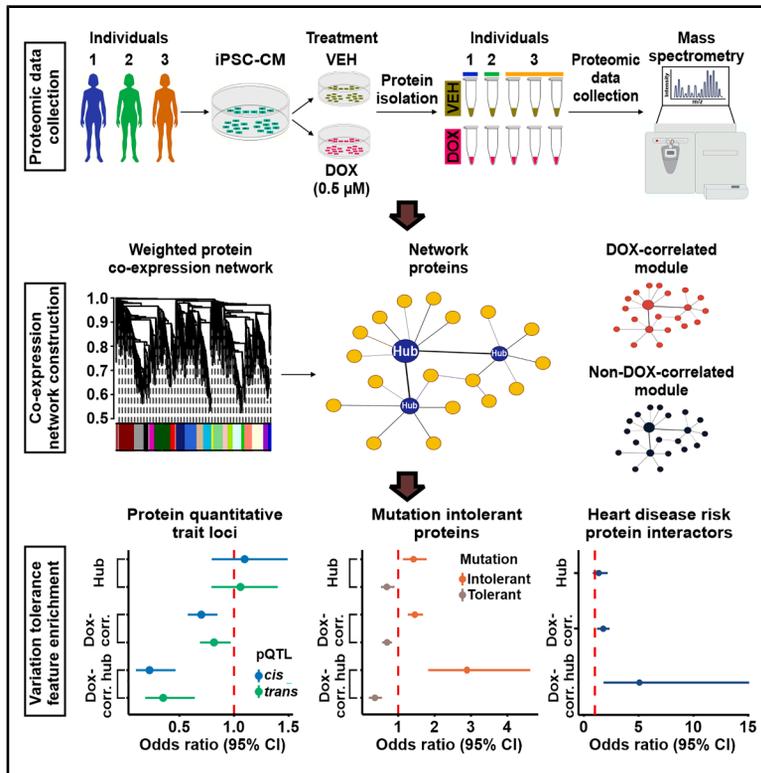


DNA-damage-associated protein co-expression network in cardiomyocytes informs on tolerance to genetic variation and disease

Graphical abstract



Authors

Omar Darrel Johnson, Sayan Paul,
José Angel Gutiérrez,
William Kent Russell, Michelle Claire Ward

Correspondence

miward@utmb.edu

In brief

Molecular network; Integrative aspects of cell biology; Proteomics

Highlights

- DNA damage affects protein co-expression in cardiomyocytes
- Damage-responsive hub proteins are depleted for protein quantitative trait loci
- Damage-responsive hubs are enriched for mutation-intolerant proteins
- Damage-responsive hubs are enriched for heart disease risk protein interactors



Article

DNA-damage-associated protein co-expression network in cardiomyocytes informs on tolerance to genetic variation and disease

Omar Darrel Johnson,^{1,2,5} Sayan Paul,^{3,5} José Angel Gutiérrez,³ William Kent Russell,³ and Michelle Claire Ward^{3,4,6,7,*}¹Biochemistry, Cellular and Molecular Biology Graduate Program, University of Texas Medical Branch, Galveston, TX 77555, USA²MD-PhD Combined Degree Program, University of Texas Medical Branch, Galveston, TX 77555, USA³Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555, USA⁴Senior author⁵These authors contributed equally⁶X(formerly Twitter): @michelle_c_ward⁷Lead contact*Correspondence: miward@utmb.edu<https://doi.org/10.1016/j.isci.2025.112474>

SUMMARY

Cardiovascular disease (CVD) is associated with genetic variants and environmental factors. A consequence of multiple risk factors is DNA damage. To examine how DNA damage influences the cardiomyocyte proteome and its relationship to CVD risk, we treated human induced pluripotent stem cell (iPSC)-derived cardiomyocytes with the DNA-damaging agent doxorubicin (DOX). A network constructed from 4,178 proteins reveals 12 co-expressed modules with 403 hub proteins. Five modules correlate with DOX and associate with RNA processing, chromatin regulation, and metabolism. DOX-correlated hub proteins are depleted for proteins that vary in expression across individuals due to genetic variation but are enriched for proteins encoded by loss-of-function intolerant genes. While not enriched for known CVD risk proteins, DOX-correlated hub proteins are enriched for the physical protein interactors of CVD risk proteins. These data demonstrate that protein connectivity in DNA-damage-associated modules influences the tolerance to genetic variation and supports the use of dynamic networks to explore complex traits.

INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of mortality globally.¹ CVD risk factors including age, diagnosis of metabolic disease, and treatment with chemotherapeutics can result in oxidative stress and apoptosis, leading to accumulated DNA damage and activation of the DNA damage response in the heart.² The amount of DNA damage in the myocardium, as determined by the DNA double-strand break (DSB) marker γ H2AX, is predictive of heart failure.³ Persistent DNA damage in cardiac cell types can lead to disease and has been detected among some of the most prevalent CVDs, including heart failure and atrial fibrillation.⁴ The myocardium is composed of cardiomyocytes that produce the contractile force of the heart necessary to circulate oxygenated blood throughout the body, and therefore damage to these cells can lead to cardiac dysfunction. Adult human cardiomyocytes are also particularly susceptible to DNA damage, given that they are post-mitotic and unable to regenerate.⁴ This means that DNA damage, induced through DSBs, can only be repaired through error-prone non-homologous end joining, unlike proliferative cell types that can also repair DNA through homologous recombination.⁵

Doxorubicin (DOX) is an effective anthracycline chemotherapeutic that can adversely induce cardiac dysfunction through

the formation of DSBs in cardiomyocytes.⁶ This is primarily mediated through its interaction with the DNA topology regulator topoisomerase II (TOP2). Physiologically, the TOP2A and TOP2B isoforms resolve torsional stress in DNA by DSBs; however, in the presence of DOX, TOP2 is trapped on DNA where it generates DNA lesions.⁷ The predominant TOP2 isoform expressed in the heart is TOP2B. TOP2B has been shown to mediate the cardiotoxic effects of DOX in *in vivo* animal models and *in vitro* human disease models.^{8,9} While DOX can lead to cellular effects through mechanisms including the generation of ROS, at clinically tolerated sub-micromolar doses, DSBs induced through interactions with TOP2B are the main contributors to DOX-induced cardiotoxicity.¹⁰ Clinically, DOX-induced cardiac dysfunction is not a unique pathology but shares characteristics of multiple CVDs.⁶ For example, 9% of individuals receiving DOX exhibit reductions in their left ventricular ejection fractions within values that would constitute heart failure.^{11,12} Similarly, treatment with DOX increases the risk for electrophysiologic dysfunction and atrial fibrillation by 10-fold and is associated with other related clinically measurable phenotypes, such as an increased QT-interval.^{13,14} These pathologies overlap with those influenced by DNA damage and are impacted by genetic risk.^{4,15}

Genome-wide association studies (GWASs) have identified hundreds of risk loci associated with complex CVDs including



atrial fibrillation, heart failure, and clinical cardiovascular phenotypes, highlighting the genetic component of CVD.¹⁵ Although GWASs identify genetic risk loci that can be mapped to genes that implicate putative regulatory effects, they do not explain the molecular mechanisms of the diseases that they associate with.^{16,17} This ultimately impedes our understanding of the effect of genetic variation on CVD and its applicability to identify potential drug targets.¹⁷

One approach to understand a complex disease phenotype is to construct networks based on molecular phenotypes such as global mRNA or protein expression levels. For example, while transcriptome profiling of human brain regions has provided insight into cell-type specificity for the risk for neuropsychiatric diseases,¹⁸ it has been shown that expression networks differ at the mRNA and protein level and that many complex disease phenotypes are observed only at the proteome level.^{19–21} Targeted studies investigating the protein interactomes of proteins encoded in GWAS loci have identified convergent points in the interaction network for autism spectrum disorder and coronary artery disease, highlighting disease-relevant biology.^{22,23} Complex disease networks are likely to exert tissue- and cell-type-specific effects. Indeed, proteins relevant to late onset Alzheimer disease are localized within glial cells in the brain.²¹ Similarly, the appropriate cellular context, such as cell type and state, is important for understanding the basis of CVD. Clinically observed CVDs such as DOX-induced cardiotoxicity can be recapitulated in human induced pluripotent stem cell (iPSC)-derived cardiomyocyte models,²⁴ allowing for the study of disease-relevant states such as exposure to DOX and hypoxia to understand CVD risk.^{25–27} However, a protein network has not been generated in this context to understand the interplay between a cellular stressor relevant to CVD and proteins encoded by genes that are implicated in complex CVDs.

We therefore designed a study to determine the effects of the DNA-damaging drug, DOX, on the proteome of human cardiomyocytes. We differentiated cardiomyocytes from iPSCs from three healthy individuals, treated them with a sub-lethal dose of DOX, and measured global protein expression levels. We construct a protein expression network consisting of co-expression modules correlated with DNA-damaging treatment, define the tolerance of network components to genetic variation, and localize CVD risk proteins and their interactors within the network.

RESULTS

iPSC-CM proteome resembles the heart ventricle proteome and is affected by DNA damage

We differentiated iPSCs from three healthy female donors into cardiomyocytes (iPSC-CMs) using biphasic WNT modulation (Figure 1A and Table S1). iPSC-CMs were metabolically selected and matured for 27 days post-differentiation initiation (See STAR Methods).²⁶ Flow cytometry analysis of two individuals indicated high-purity cultures with a median of 97% of cells expressing cardiac troponin T (Figure S1).

To determine the effect of DNA damage on the cardiomyocyte proteome, we treated iPSC-CMs with a TOP2B-inhibiting, clinically relevant concentration of DOX (0.5 μ M) and a water vehicle

control (VEH) for 24 h. This dose of DOX causes minimal cell death in iPSC-CMs but induces thousands of mRNA expression changes for genes in pathways related to p53 signaling, base excision repair, and DNA replication.²⁸ We confirmed the DNA-damaging effects of DOX under these conditions by assaying the expression of the DNA DSB marker γ H2AX (Figure 1B). DOX-treated cardiomyocytes have significantly higher γ H2AX expression compared to VEH-treated cardiomyocytes (Figure 1C; 90% vs. 7%; t test; $p < 0.05$). To account for technical variability in the drug treatment and proteomic data collection, the DOX and VEH treatment in iPSC-CMs from one individual was replicated three times, resulting in a total of 10 samples across individuals and treatments.

Global protein expression data were inferred from peptide identification and quantification using data-independent acquisition mass spectrometry (DIA; see STAR Methods). Peptides were mapped to 4,261 proteins present in at least one sample (Data S1). Four non-human proteins and proteins that were present in less than half of the samples were filtered out. To enable construction of as complete a network as possible, we used the remaining 4,178 proteins to impute protein abundance data for the 246 proteins with missing data using a feature-clustering-based imputation method commonly used for proteomics data (Figure S2A).^{29,30} On average, imputed proteins are less abundant than proteins present in all 10 samples (median \log_2 abundance all = 19.83, median \log_2 imputed abundance = 16.98; Figure S2B). We took advantage of the technical replicates to remove unwanted variation in the data (see STAR Methods).³¹ After correction, principal-component analysis reveals that PC1, which accounts for 32% of variation in the data, associates with drug treatment, whereas PC2, accounting for 29% of the variation in the data, associates with individual (Figures S2C and S2D). Similarly, when comparing all pairwise sample correlations, the data primarily separate into two clusters corresponding to DOX and VEH treatment (Figure 1D).

To gain insight into the utility of our iPSC-CM proteome data for understanding the effect of DNA damage on the heart, we correlated the median expression of our set of proteins with the expression of proteins measured in 26 postmortem human tissues from hundreds of individuals.²⁰ The iPSC-CM proteome is most similar to the proteome from heart left ventricle and atrial appendage, followed by skeletal muscle (Figure 1E). iPSC-CMs express cardiac-specific and cardiac-abundant proteins, including myosin heavy chain 7 (MYH7), myosin heavy chain 6 (MYH6), and troponin I (TNNI3; Figure 1F). Together, these findings affirm that the proteome of our iPSC-CMs closely resembles ventricular and atrial tissue, includes key cardiomyocyte proteins, and demonstrates the influence of DNA damage as the main contributor to variation of protein abundance values within our experimental model.

Network analysis reveals DOX-correlated modules, response proteins, and hub proteins

In order to identify sets of co-expressed proteins within our data, we utilized weighted correlation network analysis (WGCNA).³² WGCNA assumes the co-expression network follows a scale-free topology, where few nodes (proteins) have high connectivity and many nodes have low connectivity, and

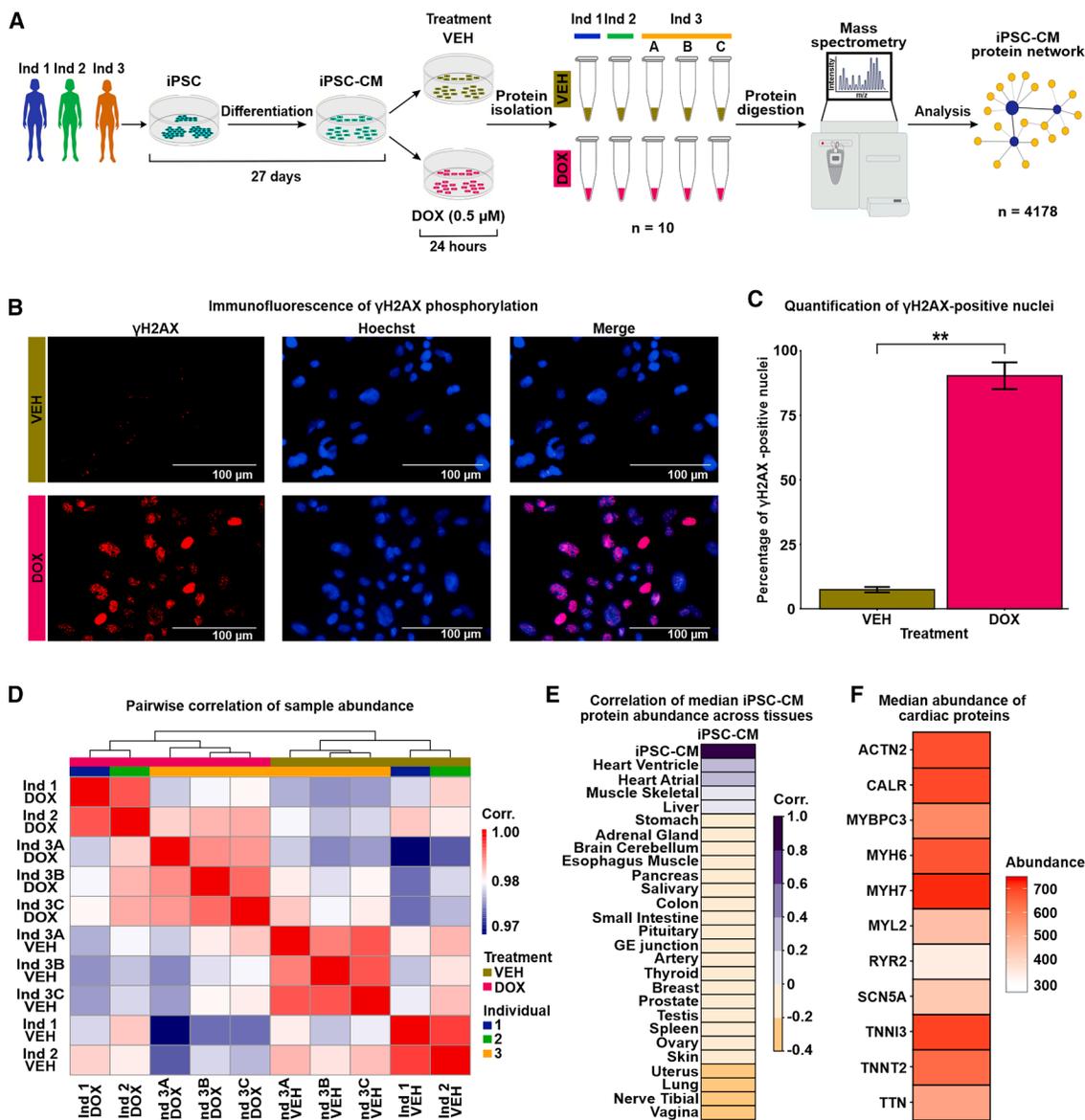


Figure 1. iPSC-CM protein samples cluster by DOX treatment and most closely resemble the heart ventricle proteome

(A) Flowchart representing the study design. iPSCs from three individuals (Ind), Ind 1 (blue), Ind 2 (green), and Ind 3 (orange), were differentiated into cardiomyocytes (iPSC-CMs). iPSC-CMs were exposed to 0.5 μ M doxorubicin (DOX) or a vehicle control (VEH) for 24 h. The treatment was replicated in Ind 3 three times, yielding 10 total samples. Peptides were isolated and quantified by mass spectrometry, allowing the construction of an iPSC-CM network from 4,178 proteins.

(B) Immunostaining of the DNA damage marker, γ H2AX, and Hoechst nuclear stain in VEH- and DOX-treated iPSC-CMs. Scale bar: 100 μ m

(C) Percentage of VEH- and DOX-treated iPSC-CMs that stain positive for γ H2AX. Data representative of treatment experiments from three individuals. Data are presented as mean \pm SD. Asterisk represents a statistically significant change in γ H2AX expression (** $p < 0.01$).

(D) Pairwise Pearson correlation of the median protein abundance across all 10 samples.

(E) Pearson correlation of the median iPSC-CM protein abundance for all proteins across all experimental samples to the median abundance of those proteins across different human postmortem tissues.²⁰

(F) Median protein abundance across experimental samples of select proteins known to be elevated in heart tissue in comparison to other tissue types.

requires a soft power threshold to determine the weights of edges connecting nodes. We regressed total network connectivity on the connectivity frequency distribution at different power thresholds and determined that the lowest power threshold that yielded the best fit to a scale-free network (0.71 regression

coefficient) was 20 (Figure S3A). The scale-free fit is further supported by decreasing connectivities at higher power thresholds, where we find that at a power threshold of 20, our network has a median connectivity score of 53.6, mean connectivity of 65.4, and maximum connectivity of 216 (Figure S3B). Using these

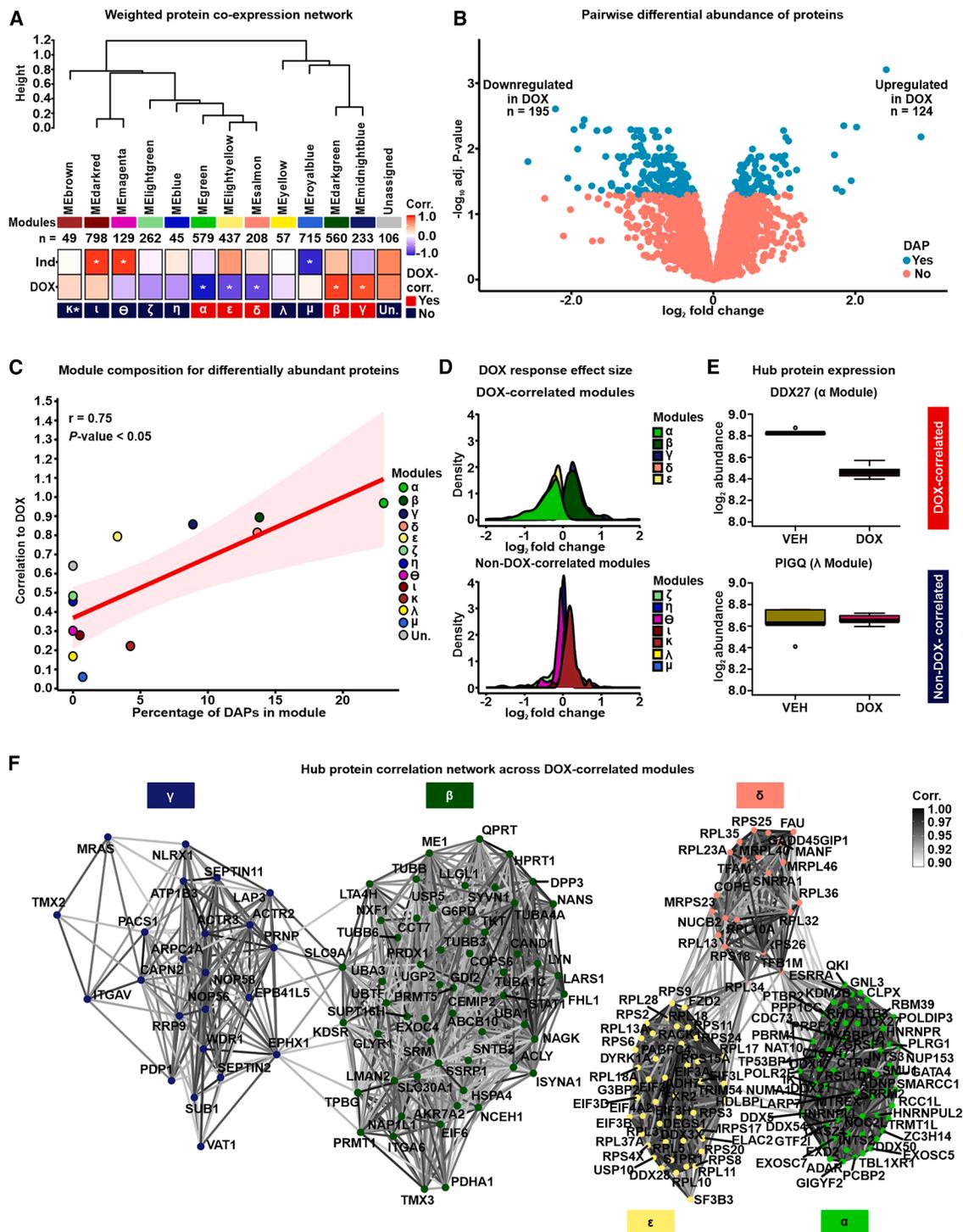


Figure 2. Network analysis of the iPSC-CM proteome identifies protein co-expression modules correlated with DOX treatment

(A) Hierarchical clustering of 12 co-expression module eigen (ME) proteins based on their Pearson correlation. Height represents the dissimilarity between ME proteins ($1 - \text{corr.}$). Each module is represented by a color and the number of proteins in the module specified. The "Unassigned" module, shaded gray, includes proteins that cannot be represented by one of the 12 ME proteins. The correlation between each ME protein and the known biological variables: Individual (Ind) and DOX treatment is shown. Asterisk represents a significant correlation between the ME protein and the trait ($*p < 0.01$). Modules are designated by Greek letters in order of decreasing correlation with DOX and summarized as DOX-correlated modules (red) and non-DOX-correlated modules (dark blue).

(B) Volcano plot representing proteins that are differentially abundant (DAPs; adjusted $p < 0.05$; blue) and not differentially abundant (salmon) between VEH- and DOX-treated iPSC-CMs.

(legend continued on next page)

criteria, we generated a network comprising 21 co-expressed modules with at least 40 proteins per module (Figure S3C), summarized by 21 eigen proteins (Figure S3D). All combinations of proteins in the co-expression network are summarized as pairwise correlations (Data S2). To better distinguish individual modules, we merged modules with highly correlated eigen proteins (Pearson correlation >0.85) and overlapping gene ontologies, which collapsed the data into 12 distinct modules (Figures 2A and S3E–S3G). These modules contain between 45 and 798 proteins (Table S2) and represent diverse gene ontologies (Table S3). One hundred six proteins (2.5% of all expressed proteins) do not fit the profile of any module and were categorized as “unassigned” (Table S2). These proteins are not co-expressed and have lower intra-modular connectivity (kIN) than the median kIN across assigned modules (normalized kIN unassigned = 0.015, normalized kIN assigned = 0.06; Wilcoxon rank-sum test; $p < 0.05$). This set of proteins is not enriched for any gene ontology. Imputed proteins are distributed across modules representing 1.6%–16% of all proteins within a module. We associated each module with the two biological features of our data, namely DOX treatment and Individual (IND). To do so, we measured the correlation between each module’s eigen protein and either IND or DOX treatment (Figure 2A). We named each of the 12 modules based on the order of their absolute correlation with DOX treatment: α , β , γ , δ , ϵ , ζ , η , θ , ι , κ , λ , and μ . Eigen proteins from the α , β , γ , ϵ , and δ modules exhibited significant correlations with DOX treatment (Pearson correlation; $p < 0.01$) and were therefore categorized as DOX-correlated modules. Modules α , δ , and ϵ have strong negative DOX correlations of -0.97 , -0.81 , and -0.79 , whereas modules β and γ have similar positive correlations of 0.89 and 0.86, respectively. In contrast, eigen proteins from the μ , ι , and θ modules are significantly correlated with IND. This delineation indicates that 5 of the 12 co-expressed modules in our iPSC-CM protein abundance correlation network are specifically associated with DOX treatment.

To support the categorization of five modules as DOX-correlated, we performed an independent pairwise differential abundance (DA) test. Using this approach, we identified 319 DA proteins among the 4,178 evaluated proteins (adjusted $p < 0.05$; Table S2; see STAR Methods). One hundred twenty-four DA proteins exhibit increased abundance in the DOX-treated iPSC-CMs, whereas 195 DA proteins show decreased abundance (Figure 2B). The percentage of DA proteins in a module correlates with the module’s correlation to DOX ($\rho = 0.75$; $p < 0.05$; Figure 2C). Similarly, proteins within DOX-correlated modules tend to have a greater response to DOX treatment as measured by their log fold change, compared to proteins in non-DOX-correlated modules (Figure 2D).

In order to test the robustness of our approach for identifying DOX-responsive proteins, we also acquired protein measurements of the same samples by data-dependent acquisition (DDA) on the mass spectrometer (see Methods). We identified 4,501 proteins using this method. All proteins identified by DIA were also identified by DDA, and the abundance of the 3,027 proteins present across all samples in both datasets is highly correlated ($\rho = 0.75$; $p < 0.001$; Figure S4). Following imputation and unwanted variance removal of the DDA data, we identified 223 DA proteins among 3,954 proteins. The effect size of the response to DOX treatment among proteins included in both acquisition methods is correlated (Pearson correlation coefficient = 0.35; $p < 0.001$; Figure S5), and the proportion of DA proteins is similar (7% for DDA and 8% for DIA). These results suggest that protein abundance changes in response to DOX identified by DIA are replicated when abundance data are collected through data-dependent acquisition.

We identified a set of proteins with a high level of intra-modular connectivity that are likely to play a central role in the biological processes associated with each module. These 403 hub proteins have the highest correlation with the module eigen proteins and the highest intra-modular connectivity (kIN) score (top 10% of all network proteins; see STAR Methods; Figure S6 and Table S2). Hub proteins predominantly reflect the module’s collective response to DOX or VEH treatment. For example, DDX27 in the α module is downregulated in response to DOX treatment, whereas PIGQ in the λ module shows no difference in abundance in response to DOX (Figure 2E). We then focused on the relationship between hub proteins across the five DOX-correlated modules ($n = 202$). The resulting network revealed not only strong intra-modular connections but also significant inter-modular correlations among proteins with a similar direction of effect in response to DOX treatment, indicating potential roles in inter-modular overlap for biological processes (Figure 2F).

Module proteins differ in their tissue specificity and cellular localization

Having identified modules of co-expressed proteins that are correlated with DOX treatment, we next sought to investigate the properties associated with each module. To determine the specificity of module proteins to heart tissue, we utilized data from the Human Protein Atlas (HPA) and the Genotype-Tissue Expression (GTEx) projects.^{20,33} These resources provide extensive gene and protein expression measurements across various tissues. The HPA database highlights 419 genes with elevated expression in heart tissue, defined as at least a 4-fold higher mRNA level in the heart compared to the average in other tissues. We detect 277 proteins corresponding to heart-elevated genes in our data. Heart-elevated proteins constitute only a small

(C) Correlation between the proportion of DAPs in a module and the correlation of the module to DOX. The best-fit line together with 95% confidence interval is shown.

(D) Distribution of effect sizes of response to DOX treatment (\log_2 fold change from pairwise differential abundance model) for the five DOX-correlated and seven non-DOX-correlated modules.

(E) Examples of hub protein abundance values in VEH- and DOX-treated iPSC-CMs in a DOX-correlated module (α ; DDX27) and a non-DOX-correlated module (λ ; PIGQ).

(F) DOX-correlated hub protein co-expression correlation network, where nodes are hub proteins and edges represent the weighted correlation between them. Connections among DOX-correlated proteins with a correlation of ≥ 0.9 are depicted for visualization.

proportion of proteins in each module regardless of DOX-correlation status (median all = 4.5%; range all = 0%–8%; Figure S7A). For a quantitative analysis of tissue specificity of DOX-correlated module proteins, we obtained tissue specificity (TS) scores for proteins in heart left ventricle tissue from GTEX. For each module in the network, module TS scores were compared to the mutually exclusive set of all other network proteins (network TS median = 0.33). We find significant deviations in TS scores from DOX-correlated modules α , β , γ , and δ , as well as non-DOX-correlated modules κ and ι (Wilcoxon rank-sum test; $p < 0.05$). Modules β , δ , κ , and ι have higher TS scores, whereas α and γ have lower TS scores (Figure S7B). While the module with the strongest correlation to DOX, α , contains proteins that are the least specific to heart ventricle tissue (TS score = -0.47), there is no correlation between tissue specificity and correlation to DOX ($\rho = 0.18$; Figure S7C). Both DOX-correlated proteins and DOX-correlated hub proteins are neither enriched nor depleted in the set of 419 elevated heart proteins compared to non-DOX-correlated proteins and hub proteins (Figure S7D). These findings suggest that while there may be heterogeneous tissue-specific effects of proteins at the modular level, proteins that are central to the DNA damage response in cardiomyocytes are not specific to the heart ventricle.

We next asked whether proteins within each module are restricted to specific intracellular or extracellular locations. We first investigated the localization of module proteins across four broad categories: intracellular, membrane-bound, plasma-detected, and secreted proteins as defined by HPA. We find that DOX-correlated modules α , β , and δ contain proteins that are enriched in the intracellular category compared to proteins not contained within each of these modules (Fisher's exact test; odds ratio [OR] = 1.8, 1.3, and 1.5, respectively; $p < 0.05$; Figure S8A). Module γ is the only DOX-correlated module enriched for membrane proteins (OR = 1.4; $p < 0.05$), along with non-DOX-correlated modules η , λ , and μ . Plasma-detected proteins are only enriched in module β (OR = 1.61; $p < 0.05$), whereas secreted proteins are only enriched in module μ (OR = 2.3; $p < 0.05$). The ϵ module is the only DOX-correlated module not enriched for any category. These results suggest that the cellular localization of proteins differs across modules. We then asked whether the sub-cellular localization of intracellular proteins differs across modules using annotation data from the UniProt database.³⁴ Network proteins are generally distributed across multiple sub-cellular organelles including the sarcomere, nucleus, mitochondrion, lysosome, Golgi apparatus, endoplasmic reticulum, cytoskeleton, cytoplasm, cell membrane, and autophagosome (Figure S8B). Nuclear proteins are enriched in modules α , β , and ϵ (Fisher's exact test; $p < 0.05$). Modules α and ϵ also share enrichment for cell membrane, cytoplasm, endoplasmic reticulum, and mitochondrial and lysosomal proteins. Module δ is enriched for mitochondrial proteins, whereas γ is enriched for endoplasmic reticulum and cell membrane proteins.

DOX-correlated module proteins are enriched for distinct biological processes

To elucidate the broad functional roles of proteins within the DOX-correlated modules, we tested whether proteins associated with distinct biological processes are enriched in each

module (see STAR Methods). We found that the α , β , δ , and ϵ modules show enrichment for various biological processes (Fisher's exact test; adjusted $p < 0.05$; Figures S9 and S10). The α module is enriched for 160 unique processes related to the DNA damage response, gene regulation via RNA splicing and metabolism, as well as chromatin regulating processes such as histone methylation and acetylation (Table S3). The β module is uniquely enriched for 35 processes related to protein localization within the nucleus and metabolism of carbohydrates, organophosphates, and oxoacids. Proteins in the δ module are uniquely enriched for 13 processes related to mitochondrial gene expression, translation, and ATP synthesis. The ϵ module is uniquely enriched for 26 processes related to DNA replication and chromatin assembly. There are no processes significantly enriched within the γ module; however, the most represented processes include glutamine family amino acid catabolic processes and ion homeostasis. Biological processes shared across modules include cytoplasmic translation (enriched in δ and ϵ), post-transcriptional regulation of gene expression (α and ϵ), and ribosome and ribonucleoprotein complex biosynthesis (α , ϵ , and δ). Biological process enrichment testing using only hub proteins of each module identified similar enrichment patterns to those in the complete module (for example, RNA metabolism and processing in the α module), indicating that the set of module hub proteins captures the core set of biological processes most representative of the entire module. Together, these data show a diversity of processes enriched among DOX-correlated modules.

We then asked whether proteins in each module are enriched for specific protein families consistent with their distinct biological processes. The α module is enriched for the splicing factor SR family and Rnase PH family (Fisher's exact test; adjusted $p < 0.05$; Figure S11), whereas β is enriched for the TCP-1 chaperonin family, and δ is enriched for signaling proteins in the 14-3-3 family. Domain-enriched modules therefore generally corroborate the biological process enrichment analysis results.

We selected five key protein categories to investigate further based on our module-specific functional enrichment results (Figure S12A; see STAR Methods).^{20,35–37} The α module is uniquely enriched for transcription factors (Fisher's exact test; OR = 5.6; adjusted $p < 0.05$; Figure S12B), particularly the homeodomain, GATA, and C2H2 Zinc Finger families that include the GATA4, MEIS1, and ZNF629 transcription factors (Figure S12C). Mutations in GATA4 are associated with atrial septal defects, arrhythmia, and a reduced capacity for the cardiac hypertrophic response³⁸ and have been implicated in DOX-induced cardiotoxicity.³⁹ Both the α and ϵ modules show enrichment for stress granule components and RNA-binding proteins (adjusted $p < 0.05$; Figure S12A). RNA-binding proteins related to multiple post-transcriptional processes including splicing, translation, and mRNA stability are enriched (Figure S12D) and include proteins such as QKI, YBX3, and GNL3 (Figure S12E). QKI regulates pre-mRNA splicing, export of mRNAs from the nucleus, protein translation, and mRNA stability⁴⁰ and is implicated in cardiomyocyte calcium dynamics and contractility,⁴¹ cardiomyopathies,⁴⁰ and attenuation of DOX-induced cardiotoxicity.⁴² The β module uniquely exhibits enrichment for enzymes (OR = 1.9; adjusted $p < 0.05$; Figure S12A) including glutathione metabolism, fatty

acid activation, and drug metabolism (Figure S12F). The most DOX-responsive enzymes include NDUFB7, a member of the δ module and critical component of the mitochondrial membrane respiratory chain complex I (Figure S12G). These results are consistent with the biological processes enriched in each module.

Hub proteins in DOX-correlated modules are depleted for pQTLs

We next focused our attention on the properties of the 403 hub proteins within the network. We were particularly interested in understanding the tolerance of these central proteins to physiological genetic variation. To do so, we investigated proteins whose expression level varies across individuals depending on the genotype of an associated SNP, i.e., protein quantitative trait loci (pQTLs; Figure 3A). Prior research focused on serum proteins has shown that highly connected proteins central to co-expression networks are typically depleted for pQTLs.⁴³ However, how network centrality interacts with the dynamic molecular response to perturbation to shape the distribution of pQTLs is not known. We therefore obtained a set of pQTLs identified in human plasma from thousands of individuals¹⁷ and overlapped these with iPSC-CM network proteins (Table S2). We first asked whether pQTL proteins associated with either *cis*- or *trans*-SNPs are enriched among hub proteins. Hub proteins are neither enriched nor depleted for pQTLs (Fisher's exact test; OR = 1.1; 95% confidence interval [CI] = 0.8–1.5 for *cis*-pQTLs and OR = 1.0; 95% CI = 0.8–1.4 for *trans*-pQTLs; Figure 3B). We then asked whether pQTL proteins are enriched among proteins that respond to DNA damage. DOX-correlated proteins are modestly depleted for pQTL proteins associated with either *cis*- (OR = 0.7; 95% CI = 0.6–0.8; $p < 0.05$; Figure 3B) or *trans*-SNPs (OR = 0.8; 95% CI = 0.7–1.0; $p < 0.05$, Figure 3B). However, DOX-correlated hub proteins show an even more pronounced depletion for pQTL proteins mapped to *cis*- (OR = 0.2; 95% CI = 0.1–0.5; $p < 0.05$; Figure 3B) or *trans*-SNPs (OR = 0.4; 95% CI = 0.2–0.6; $p < 0.05$; Figure 3B).

Next, we investigated those pQTLs that correspond to proteins expressed within our network and asked whether the SNP effect size of the pQTLs differed across network components. We assigned each pQTL protein to the *cis*- or *trans*-SNP with the greatest effect size. As expected, the median effect size for *cis*-pQTL proteins in the network is greater than that for *trans*-pQTLs (0.2 vs. 0.1). There is no difference in the *cis*- or *trans*-pQTL effect sizes between DOX-correlated and non-DOX-correlated proteins (Figure 3C). However, DOX-correlated hub proteins have lower *cis*- and *trans*-pQTL effect sizes than hub proteins that are not correlated with DOX (Wilcoxon rank-sum test; $p < 0.05$; Figure 3D).

Given the depletion of pQTLs among DOX-correlated hub proteins and not the total set of hub proteins, we next asked if there is a relationship between the intra-modular connectivity of a protein and the probability of that protein being a pQTL that is influenced by DOX. We assigned proteins into two groups based on their DOX-correlation status, stratified proteins within each group into deciles based on their connectivity, and calculated the percentage of pQTL proteins in each decile. In the non-DOX-correlated group, the decile with the highest connectivity

shows the greatest percentage of pQTL proteins (21%), whereas the decile with the lowest connectivity shows the lowest percentage of pQTL proteins (9%). Across deciles, there is a general upward trend in the percentage of pQTL proteins as intra-modular connectivity increases (Figure 3E). In the DOX-correlated group, the highest connectivity decile shows a reduction in the percentage of pQTL proteins (8%) relative to the lowest connectivity decile (16%) and more variability for the percentage of pQTL proteins across deciles. Therefore, the percentage of pQTL proteins across different deciles showed opposite trends depending on their correlation to DOX. The decrease in pQTL proteins with the highest connectivity among DOX-correlated proteins suggests that hub proteins correlated to DOX treatment are under stronger evolutionary constraints, leading to reduced genetic variation in these highly connected proteins.

We also tested for enrichment of pQTLs across modules. The α module, with the highest correlation to DOX, is depleted for pQTLs (Fisher's exact test; OR = 0.6; CI = 0.5–0.8; adjusted $p < 0.05$; Figure S13). Conversely, the μ module, with the lowest correlation to DOX, is the only module enriched for pQTLs (OR = 1.4; CI = 1.1–1.7; adjusted $p < 0.05$) and is one of three modules correlated with Individual. There are no modules correlated with both Individual and DOX, which supports differences in tolerance to variation across these protein sets.

In summary, our analysis reveals that DOX-correlated hub proteins are less likely to be associated with pQTLs, indicating their likelihood to play an essential role in maintaining network stability and function under stress conditions.

DOX-correlated hub proteins are enriched for loss-of-function-intolerant proteins

Given that DOX-correlated hub proteins are depleted for proteins that vary in their expression across healthy individuals, we next asked about the tolerance of these proteins to variation more broadly. We utilized several genetic tolerance scores to assess the functional impact and evolutionary constraint of genes encoding proteins within specific modules. First, we considered the probability of each protein in a module being haploinsufficient (pHaplo), which estimates whether a single copy of a gene can maintain normal function, or triplosensitive (pTriplo), which estimates the risks of gene dose increases, that can be equally detrimental. Genes with a high pHaplo score (≥ 0.86) or a high pTriplo score (≥ 0.94) are deemed haploinsufficient or triplosensitive, respectively, and likely precipitate health consequences.⁴⁴ We compared the gene dose scores of DOX-correlated module proteins against the broader network (see STAR Methods). The α and ϵ modules have higher pHaplo scores compared to network proteins, whereas δ module proteins have lower pHaplo scores (Wilcoxon rank-sum test; $p < 0.05$; Figure S14A). The α module also shows increased pTriplo scores compared to the network average, whereas the δ module's pTriplo scores are lower ($p < 0.05$; Figure S14B). These results show that the most DOX-correlated module, α , is most sensitive to gene dosage changes.

Next, we examined the tolerance of module proteins to mutations that reduce or eliminate protein function. Genes with a high probability (≥ 0.9) of loss-of-function intolerance (pLI) are considered loss-of-function intolerant, meaning they cannot withstand

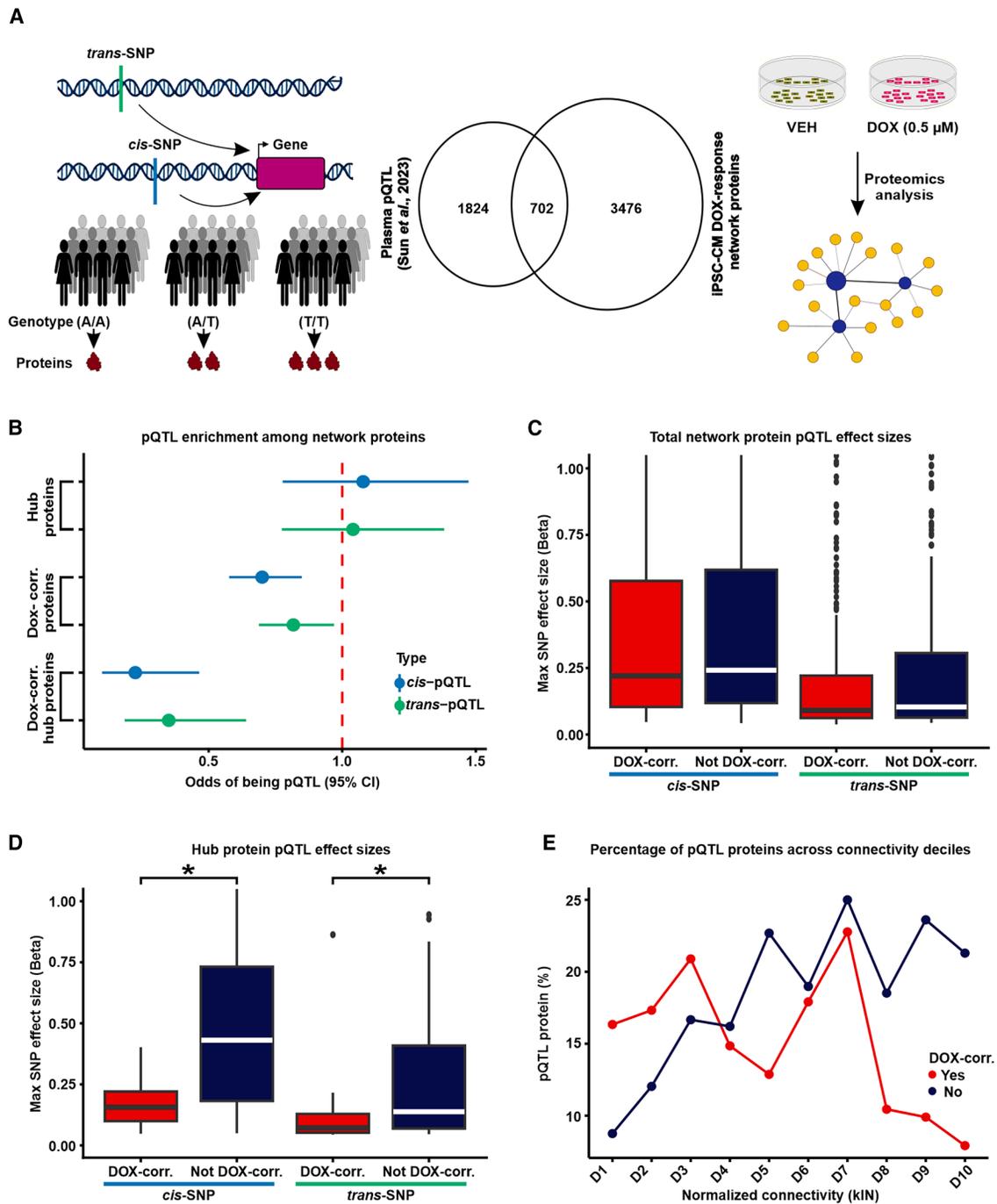


Figure 3. DOX-correlated hub proteins are depleted for protein quantitative trait loci

(A) Schematic representing the rationale behind the test design. Protein expression can be influenced by *cis*- or *trans*-SNPs, resulting in differential abundance of proteins across individuals with different genotypes. We integrated existing protein quantitative trait loci (pQTL) data from plasma samples¹⁷ with our DOX-treated iPSC-CM protein network.

(B) Enrichment of *cis*- (blue) and *trans*-pQTLs (green) among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins.

(C) Maximum pQTL effect sizes for *cis*- and *trans*-pQTLs among all proteins in DOX-correlated modules and non-DOX-correlated modules.

(D) Maximum pQTL effect sizes for *cis*- and *trans*-pQTLs among all hub proteins in DOX-correlated modules and non-DOX-correlated modules. Asterisk represents a significant difference in effect sizes between DOX-correlated and non-DOX-correlated modules ($p < 0.05$).

(E) Percentage of pQTL proteins across connectivity deciles for normalized kIN for proteins in DOX-correlated (red) and non-DOX-correlated (blue) modules. Deciles are ordered by increasing kIN scores.

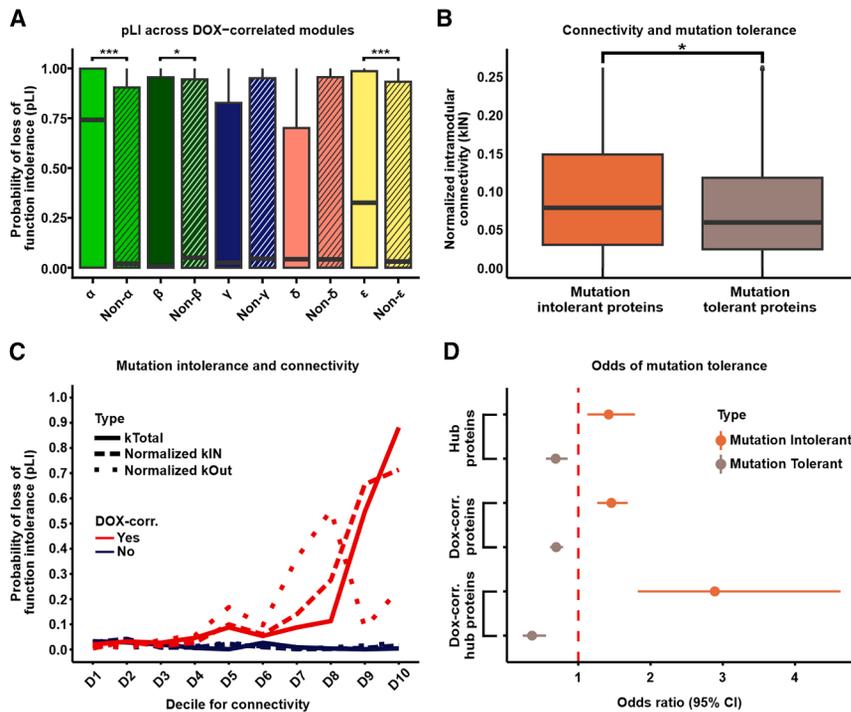


Figure 4. DOX-correlated hub proteins are enriched for loss-of-function-intolerant proteins

(A) Probability of loss-of-function intolerance (pLI) for all proteins in each DOX-correlated module (x) and all proteins outside of the module (Non-x). Asterisk represents a significant difference between x and Non-x (* $p < 0.05$; *** $p < 0.001$).

(B) Connectivity (normalized kIN) among mutation-tolerant (pLI ≤ 0.1) and mutation-intolerant (pLI ≥ 0.9) proteins. Asterisk represents a significant difference in scores between module-specific proteins and all proteins outside of modules (* $p < 0.05$).

(C) Median pLI scores across proteins in connectivity deciles (kTotal (solid line), normalized kIN (dashed line), and normalized kOut (dotted line)) for proteins in DOX-correlated modules (red) and non-DOX-correlated modules (blue). Deciles are ordered by increasing connectivity scores.

(D) Enrichment of mutation-intolerant (orange) and mutation-tolerant (brown) proteins among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins.

loss-of-function mutations without significant phenotypic impact. Conversely, genes with a low probability (≤ 0.1) can tolerate loss-of-function mutations with minimal phenotypic consequences and are considered loss-of-function tolerant.^{44,45} Both α and ϵ module proteins exhibit significantly higher pLI scores than proteins outside these modules (Wilcoxon rank-sum test; $p < 0.05$; Figure 4A and Table S2), indicating that they are less tolerant to genetic variation. In contrast, the β module is characterized by lower pLI scores ($p < 0.05$; Figure 4A), suggesting these proteins can better tolerate loss-of-function mutations. These results emphasize that the α module contains proteins that are highly intolerant to pathogenic variation.

To test the relationship between modular connectivity and pLI across the network, we identified the set of loss-of-function-tolerant and loss-of-function-intolerant proteins. We find that loss-of-function-intolerant proteins have a higher level of intramodular connectivity (normalized kIN) compared to loss-of-function-tolerant proteins (Wilcoxon rank-sum test; $p < 0.05$; Figure 4B). To understand the relationship between different types of connectivity and DOX-correlation, we next considered the spectrum of tolerance to mutation across all proteins in the network. First, we classified all proteins into two groups based on their DOX-correlation status. We then generated deciles for intra-modular connectivity (kIN), inter-modular connectivity (kOut), and total network connectivity (kTotal) for both DOX-correlated and non-DOX-correlated proteins. For DOX-correlated proteins, there is a marked increase in pLI scores in the highest kTotal and kIN deciles (Figure 4C). Conversely, pLI scores for kOut appeared more variable at higher connectivity deciles. In comparison to DOX-correlated proteins, median pLI scores for non-DOX-correlated proteins remain consistently low across deciles for all measures of connectivity. These find-

ings highlight that DOX-correlated proteins with the highest total connectivity, primarily driven by intra-modular connectivity, are essential for cellular function under DNA damage conditions and are under strong evolutionary constraints to maintain their functional integrity. In contrast, non-DOX-correlated proteins do not exhibit a relationship between connectivity and mutation intolerance, suggesting a less critical role in maintaining network stability.

We next asked whether loss-of-function-intolerant proteins are enriched among proteins central to the protein network given that there are increased pLI scores in the highest decile for intramodular connectivity. DOX-correlated proteins are modestly enriched for proteins whose encoding genes are loss-of-function intolerant (Fisher's exact test; OR = 1.5; CI = 1.3–1.7; $p < 0.05$; Figure 4D) and depleted for proteins that are tolerant to mutation (OR = 0.7; CI = 0.6–0.8; $p < 0.05$). Hub proteins are modestly enriched for loss-of-function-intolerant proteins (OR = 1.4; CI = 1.1–1.8; $p < 0.05$) and depleted for mutation-tolerant proteins (OR = 0.7; CI = 0.6–0.9; $p < 0.05$). However, DOX-correlated hub proteins are highly enriched for mutation-intolerant proteins (OR = 2.9; CI = 1.8–4.6; $p < 0.05$) and depleted for mutation-tolerant proteins (OR = 0.4; CI = 0.2–0.6; $p < 0.05$; Figure 4D). Similarly, Individual-correlated module proteins are more tolerant of loss-of-function mutations than DOX-correlated module proteins, and modules correlated with neither Individual nor DOX (Wilcoxon rank-sum test; $p < 0.05$, Figure S14C). These data demonstrate that DOX-correlated hub proteins are likely to be the most critical proteins to the DNA damage response network.

DOX-correlated modules contain cardiovascular-disease-associated proteins

DOX treatment in cancer patients is associated with increased risk for atrial fibrillation (AF) and heart failure (HF).^{46–49} We

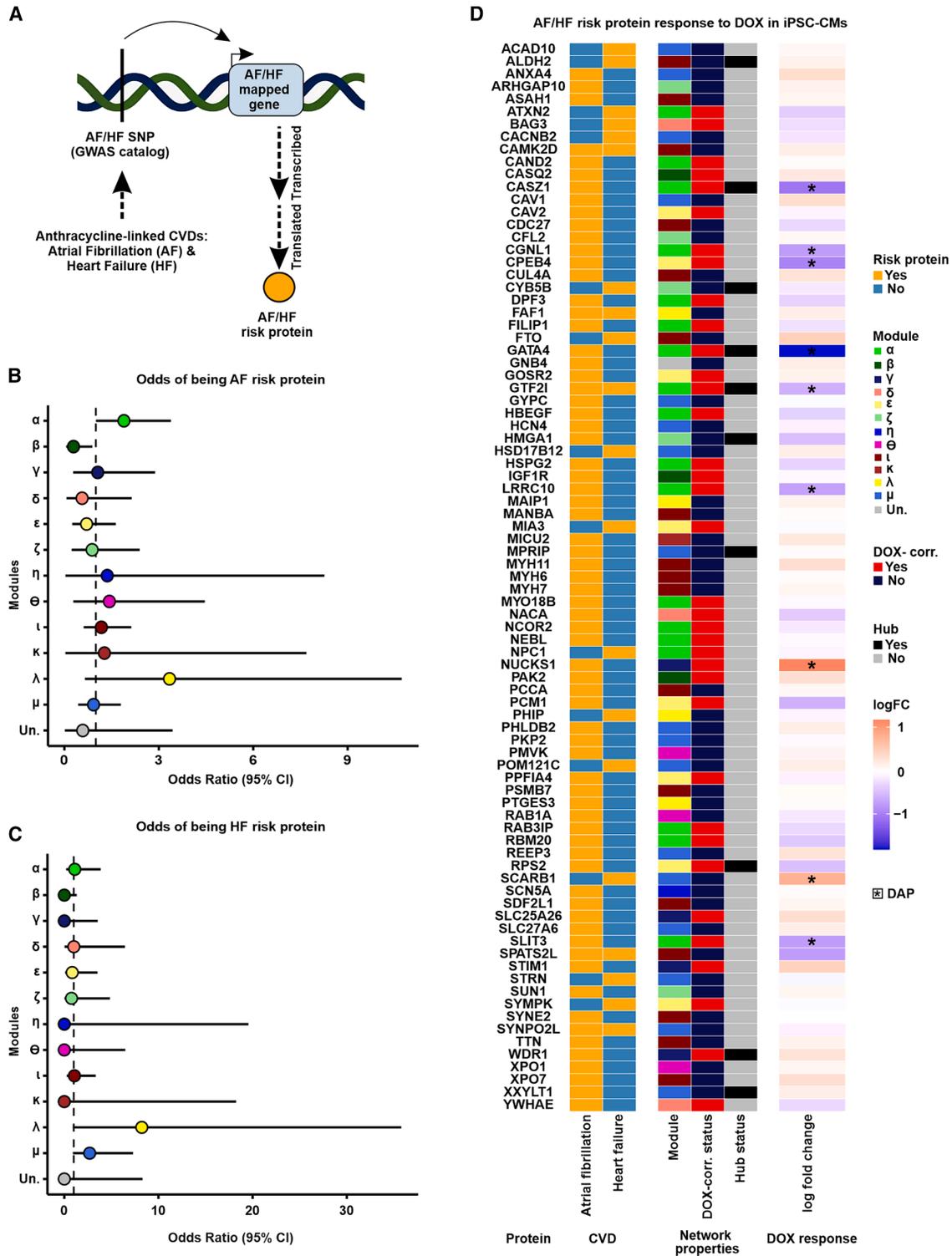


Figure 5. DOX-correlated modules contain proteins associated with risk for atrial fibrillation

(A) Schematic representing how risk proteins for the anthracycline-associated CVD traits atrial fibrillation (AF) and heart failure (HF; “heart disease” GWAS catalog ontology class EFO_0003777) are obtained. AF- and HF-associated SNPs (AF/HF-SNP) from GWAS are mapped to nearby genes (AF/HF mapped gene) and converted to the corresponding protein identifier (AF/HF risk protein).

(legend continued on next page)

therefore obtained data for genetic loci associated with these two CVDs from the GWAS catalog⁵⁰ and asked whether these loci are near-expressed proteins belonging to DOX-correlated modules (Figure 5A). Seventy mapped AF genes and 20 mapped HF genes are expressed as proteins in our iPSC-CM data. The α module shows nominal enrichment for AF proteins (OR = 1.89; CI = 1.0–3.4; nominal $p < 0.05$; Figure 5B) but is not enriched for HF proteins (Figure 5C). No other DOX-correlated module shows evidence for enrichment with either trait. We next investigated the response of individual disease-associated proteins to DOX; 31 of 70 expressed AF risk genes and 6 of 20 expressed HF risk genes are within DOX-correlated modules (Figure 5D). Many AF risk genes in the α module, including CASZ1, CGNL1, GATA4, GTF2I, LRRRC10, and SLIT3, are also DAPs that are downregulated in response to DOX. GTF2I is the only DOX-correlated protein that is associated with both AF and HF. NUCKS1 is an AF-associated DAP in the γ module and is therefore upregulated in response to DOX. Transcription factors CASZ1, GATA4, and GTF2I are the only risk proteins for AF or HF in DOX-correlated modules that are both DAPs and hub proteins. These results show that several CVD risk proteins respond to DNA damage.

DOX-correlated hub proteins are enriched for physical protein-protein interactors of CVD-associated proteins

To understand how proteins in our network correspond to CVD more generally, we obtained GWAS data for 84 CVD traits available in the GWAS catalog. We assessed the relative likelihood for key network proteins to be contained in the set of proteins mapped to CVD (Figure 6A). CVD-associated proteins are neither enriched nor depleted among hub proteins, DOX-correlated proteins, or DOX-correlated hub proteins (Figure 6B), suggesting that they do not contribute to the key features of our protein network. These results are in line with the observation that DOX-correlated hub proteins are depleted for pQTLs and are intolerant to genetic mutation.

We next reasoned that proteins that are known to physically interact with CVD-associated proteins may provide a mechanism through which DNA damage associates with CVD risk. We therefore identified pairwise interactions between all 4,178 network proteins (STRINGdb; confidence score ≥ 0.9) and generated a total physical protein-protein interaction (PPI) network ($n = 13,750$ edges). We extracted the subnetwork containing CVD proteins ($n = 213$) and their interactors ($n = 798$; $n = 909$ edges; Data S2 & Table S4). We then asked whether CVD proteins and their interactors differ in their tolerance to mutation. CVD proteins have lower pLI scores than proteins they interact with (0.07 vs. 0.69; Wilcoxon rank-sum test; $p < 0.05$; Figure 6C). To test the robustness of these results, we analyzed 10,000 randomly generated subnetworks from the total network that maintained a similar degree distribution as the CVD protein network (See STAR Methods). The enrichment p value of CVD-associated proteins and their interactors is lower than

the 5th percentile of the random distribution ($p = 0.00001$ vs. $p = 0.006$), and the median pLI score difference between our test sets (0.63) is within the 95th (0.64) and 96th (0.62) percentiles of the randomly generated networks. These findings suggest that genes encoding CVD risk proteins are less likely to be essential compared to their interacting partners.

We next asked about the relative likelihood for hub proteins in the network to be contained in the set of CVD protein interactors. We first calculated the frequency of protein interactors falling in the same module. We find that the proportion of interacting proteins expressed in the same module is higher for interactions where one of the proteins is a hub protein compared to interactions that do not include a hub protein (Wilcoxon rank-sum test; $p < 0.05$; Figure S15). Although hub proteins are more likely to be co-expressed in the same module with their physical protein interactors, hub proteins are not enriched among CVD protein interactors (Figure 6D). However, CVD protein interactors are enriched for DOX-correlated proteins in comparison to proteins that are not DOX-correlated (Fisher's exact test; OR = 1.5; CI = 1.1–2.1; $p < 0.05$; Figure 6D). We further find that CVD protein interactors are enriched for DOX-correlated hub proteins in comparison to hub proteins that are not DOX-correlated (OR = 5.1; CI = 1.8–15.1; $p < 0.05$; Figure 6D). We similarly observe a significant enrichment of CVD protein interactors among DOX-correlated hub proteins when identifying interactors with a lower stringency score ($n = 49,020$ interactions; STRINGdb confidence score ≥ 0.4 ; OR = 2.3; $p < 0.05$). These data suggest that the pathogenicity of CVD variants identified by GWASs might not only be a consequence of a direct effect on a protein's function but also indirectly through interactions with essential proteins that are highly correlated to the DNA damage response in cardiomyocytes. We therefore illustrate the CVD protein and CVD protein interactor network together with our experimentally derived DOX-correlation and hub protein status annotations (Figures 6E and S16).

We posit that the DNA-damage-associated CVD protein network in cardiomyocytes can also be used to reduce the search space for druggable targets that might have the greatest impact on influencing various CVD phenotypes. We therefore also annotated each network protein by whether they are a target of an FDA-approved drug (Table S2).³³ We identified 79 proteins in DOX-correlated modules that are druggable, including the chromatin-modifying enzymes HDAC1, HDAC2, and HDAC3. Seven of the druggable DOX-correlated proteins are CVD proteins (FADS1, FADS2, FINC, IGF1R, AT2B1, RL3, and TRXR1), and five have elevated expression in heart tissue (RZR2, ADT1, LDHB, ODO1, and AAPK2), thereby opening further avenues of investigation.

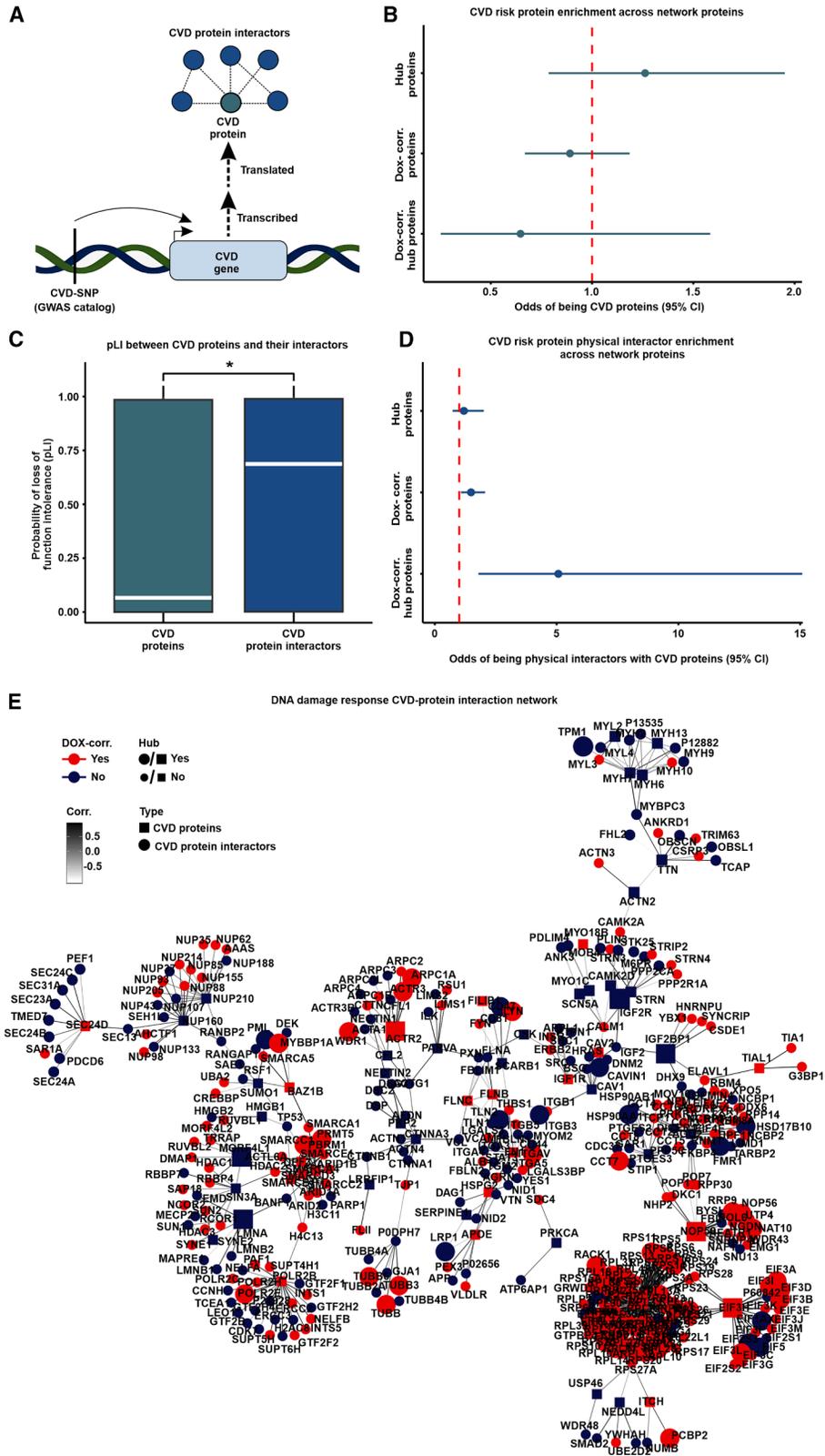
DISCUSSION

Many genetic loci have been associated with CVD. While the genes in these loci can be inferred to play a role in disease

(B) Enrichment of 70 expressed atrial fibrillation risk proteins across network modules.

(C) Enrichment of 20 expressed heart failure risk proteins across network modules.

(D) Response of AF and HF risk proteins to DOX. A positive fold change indicates proteins that increase in expression in response to DOX. Asterisk represents differentially abundant proteins (DAPs; adjusted $p < 0.05$).



(legend on next page)

risk, the mechanisms behind these associations is often unclear. We hypothesized that a relevant environmental factor may provide insight. DNA damage is a ubiquitous stressor that is both implicated in and predictive of CVD.³ It can be induced through exogenous factors including through administration of drugs used in the treatment of cancer. To understand how DNA damage in cardiomyocytes influences CVD risk, we used an *in vitro* model of iPSC-CMs from multiple individuals to study the effects of DOX on the proteome. We constructed a co-expression network and identified many proteins that change in their abundance in response to DOX and characterized their network properties and tolerance to variation.

DOX induces changes to the cardiomyocyte proteome relevant to anthracycline-induced cardiotoxicity

The effects of genetic variation and environmental perturbations on mediating risk for complex disease have most often been assayed through the transcriptome. This includes studies investigating the influence of DOX on cardiomyocytes.^{24,27,28} Here, we measured the effects of DOX on the cardiomyocyte proteome using both a pairwise differential abundance test and a co-expression network correlated with DOX treatment. We identified 319 differentially abundant proteins (8% of the proteome) and determined that 5 out of 12 co-expressed modules are correlated with DOX treatment. A previous study that assayed the proteomic response to anthracyclines, including DOX, using both human microtissues and tissues from heart failure patients identified four anthracycline-associated hub proteins.⁵¹ We find two of these proteins, BAG3 and hub protein CAND1, in DOX-correlated modules in cardiomyocytes, indicating the relevance of our cardiomyocyte model. In addition, our study generated protein data for thousands more proteins, implicating many more proteins in DOX toxicity. Conversely, we found little overlap between our data and DOX-treated rodent cardiac proteomic data from rat cardiomyocytes⁵² and rat heart tissue,⁵³ suggesting species specificity in DOX responses.

Our network analysis identified the α module as the most DOX-correlated co-expression module. The 579 proteins in this module decrease in their abundance following DOX treatment and are enriched for proteins in gene regulatory processes and DNA damage repair. The most enriched processes relate to RNA processing and splicing. The associated proteins likely contribute to the large-scale splicing changes that have previously been identified following DOX treatment.²⁷ Processes related to chromatin organization and histone modifications are also enriched, in line with previous work in a murine DOX toxicity

model indicating effects on histone eviction and histone-modifying enzymes.⁵⁴ This module is also enriched for transcription factors including GATA4, CTCF, ZNF629, ESSRA, GATAD2B, ZNF346, YBX3, GTF2I, and GATAD2A, indicating potential drivers of the gene expression changes observed due to DOX treatment in heart cells.^{38,39,55–65} Many of these processes are important across cell types, and we correspondingly observe decreased heart tissue specificity for proteins in this module. Indeed, DOX can lead to neurotoxicity, hepatotoxicity, and nephrotoxicity in addition to cardiotoxicity.^{66–68} Conversely the DOX-correlated δ module includes proteins with higher heart specificity than proteins in other modules and is enriched for processes related to mitochondrial functions such as oxidative phosphorylation. Proteins in this module may thus contribute to some of the *in vivo* tissue-specific effects of DOX on the heart.

The target of DOX is TOP2, expressed in the heart as both the α and β isoforms. It is TOP2B that is thought to mediate the cardiotoxic effects of anthracycline chemotherapeutics such as DOX. While we do not detect a change in TOP2B protein abundance in response to DOX, it is co-expressed in the ϵ module, which is DOX-correlated and consists of proteins downregulated in response to DOX. Notably, ~75% of proteins shown to physically interact with TOP2B (10/13 expressed in iPSC-CMs), including CTCF, are present in DOX-correlated modules α and ϵ and exhibit high connectivity.⁶⁹ GWASs have identified 10 risk loci associated with anthracycline-induced cardiotoxicity.^{70–72} We find three proteins encoded by genes mapped to these risk loci expressed in our network, where two are localized in the ϵ module. These proteins include POLRMT and RPL7. Decreased abundance of POLRMT, a DNA-directed RNA polymerase located in mitochondria, and RPL7, a component of the large ribosomal subunit, suggest that these proteins concordantly lead to decreased mitochondrial transcription and translation due to DOX treatment. Together, the proteins that we identify as responding to DOX in cardiomyocytes have been implicated in DOX-induced cardiotoxicity through molecular and genetic approaches.

While our study was designed to understand general principles of DOX treatment on cardiomyocytes using three individuals, our data suggest that there may be individual-specific effects on protein abundance and response to DOX and that studies including more individuals are warranted.

DOX-induced protein expression changes are relevant to CVD

Our network approach allowed us to intersect sets of DOX-correlated proteins with proteins implicated in risk for complex CVDs.

Figure 6. DOX-correlated hub proteins are enriched for physical protein interactors of CVD risk proteins

(A) Schematic representing the rationale behind the test design. CVD-associated SNPs (CVD-SNP) are mapped to nearby genes (CVD gene) that are translated into proteins (CVD proteins) that may physically interact with other proteins (CVD protein interactors).

(B) Enrichment of CVD risk proteins among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins.

(C) pLI score distribution of CVD risk proteins and CVD risk protein interactors. Asterisk represents a significant difference in the scores between protein groups (* $p < 0.05$).

(D) Enrichment of CVD risk protein interactors among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins.

(E) Protein-protein interaction network for CVD proteins (square) and CVD protein interactors (circle) expressed within the co-expression network. Color represents if the protein is in a DOX-correlated module (red) or a non-DOX-correlated module (blue). Edges represent weighted correlation within the co-expression network. Node size indicates if a protein is a hub (large icon) or not a hub (small icon) protein. A CVD protein subnetwork containing the most highly connected proteins is presented here with the full CVD protein network displayed in Figure S16.

However, there are hundreds of proteins within each DOX-correlated module and hundreds of DAPs. Prioritization of proteins as diagnostic and therapeutic targets therefore remains a challenge. We provide annotations to subset module proteins by including DOX response effect size and direction, hub status, pLI score, druggability, and association with CVD, which may help in target selection for proteins most likely to modulate the cardiomyocyte response to DNA damage (Table S2). For example, the application of these filters uncovered highly responsive proteins that are central to the DOX response network, such as downregulated proteins GATA4, QKI, and TFAM, which have been shown to directly improve DOX-induced cardiotoxicity when their expression is increased.^{39,73,74} We highlight some proteins below that may contribute to a CVD trait and inform on cardioprotective drug development.

The α module contains proteins encoded by genes in loci associated with atrial fibrillation. These proteins include GATA4, LRRRC10, RBM20, SLIT3, CAND2, GTF2I, HSPG2, FILIP1, MYO18B, CASZ1, and DPF3 that are downregulated in response to DOX treatment. Many of these proteins show evidence for independently impacting atrial fibrillation risk, as well as playing an essential role in cardiac development.^{38,39,55–65,75} The co-expression of these proteins involved in the genetic risk for atrial fibrillation indicates that DNA damage may mediate atrial fibrillation by collectively reducing the abundance of these critical proteins. Atrial fibrillation can be a cause or consequence of many complex CVD traits,⁷⁶ highlighting the importance of these proteins to disease.

The β module, containing proteins with increased abundance due to DOX treatment, shows enrichment for both metabolic processes as well as processes involved in the translocation of enzymatic proteins to the nucleus, uniquely placing it at the intersection between the proteomic and metabolomic response to DNA damage. This includes the hub protein PRDX1, which has been shown to translocate to the nucleus upon DNA double-strand breakage and clear damage-induced nuclear reactive oxygen species and γ H2AX.⁷⁷ Similarly, four TCA cycle proteins help prevent DOX-mediated cellular damage by translocating from the mitochondria to the nucleus upon DNA damage.⁷⁸ Three of these proteins, PDH-E1, MDH-2, and CS, are co-expressed in the β module. Therefore, the β module may identify proteins for future studies investigating how the nuclear translocation of enzymatic proteins can attenuate DOX-induced cardiotoxicity.

Highly connected DNA-damage-associated proteins are intolerant to mutation

Our network analysis identified not only modules of co-expressed proteins but also subsets of highly connected hub proteins belonging to DOX-correlated and non-DOX-correlated modules. This allowed us to investigate the relationship between DNA-damage-associated connectivity and tolerance of these proteins to physiological and pathological variation. We find that DNA-damage-associated hub proteins are not only depleted for pQTLs but also enriched for loss-of-function-intolerant proteins, indicating the importance of these proteins. Considering the spectrum of connectivity, beyond the highly connected hub proteins, revealed increasing pLI values with increasing protein connectivity for DOX-correlated module proteins. pLI values remain low

across a range of connectivities for non-DOX-correlated module proteins. The trend for DOX-correlated proteins was observable in both total connectivity and intra-modular connectivity, but not inter-modular connectivity, suggesting that intolerance to mutation in the network is centered around DOX-correlated modules and their related biological processes. The opposite trend is true for pQTLs, where enrichment tends to decrease with increased connectivity of DOX-correlated proteins. It has previously been shown that pQTLs are depleted among hub proteins in a steady-state network.⁴³ Here, we show that DOX-correlated hub proteins are depleted for *cis*- and *trans*-pQTLs compared to non-DOX-correlated hub proteins, and those that are pQTLs have lower effect sizes in the DOX-correlated set. These results suggest that the proteins most central to the DNA damage response are constrained in their expression with little variation across individuals and support the observation that these proteins are intolerant to loss-of-function mutations. We note that the set of unassigned proteins are neither enriched nor depleted in mutation-intolerant genes or pQTLs, suggesting that as a group these do not have distinct genetic tolerance properties. The differential relationship between connectivity, genetic influence, and pLI emphasizes the evolutionary constraint of DOX-correlated hub proteins by purifying selection to minimize potential disruptions in their expression.

Highly connected DNA-damage-associated proteins influence CVD risk proteins through protein-protein interactions

While CVD risk proteins are not enriched among hub proteins or DOX-correlated proteins, physical protein interactors of CVD risk proteins are enriched among DOX-correlated proteins and DOX-correlated hub proteins in particular. This suggests that DNA damage can affect the expression of proteins that interact with proteins implicated in CVD risk, hinting at the mechanism through which some of the disease-associated loci might exert their effects. For example, DOX-correlated hub proteins PBRM1 and SMARCC1 are both chromatin regulators that interact with atrial-fibrillation-associated protein DPF3.⁶⁵ These findings are also corroborated by the observation that GWAS risk proteins are generally very tolerant to mutation, but their physical PPIs are not. However, there are exceptions to this trend such as is observed in atrial fibrillation, where GATA4, GTF2I, and CASZ1 are both CVD proteins and DOX-correlated hub proteins. These data support the notion that genetic variation could contribute to CVD phenotypes by altering the stability and functionality of regulatory proteins that are central to the proteomic DNA damage response network through physical protein interactions. Therefore, our network pinpoints CVD risk proteins that are highly connected to proteins central to the cardiomyocyte DNA damage response that can be prioritized for cell-type-specific co-immunoprecipitation studies that have proved informative for understanding mechanisms behind genes implicated in coronary artery disease and autism spectrum disorder.^{22,23}

Considering the DNA damage response network in the context of the omnigenic model

The omnigenic model for complex phenotypes posits that thousands of genes expressed in disease-relevant cell types can

influence complex traits, whereby core genes have a direct impact on the phenotype, while peripheral genes influence it indirectly through distant gene interactions.⁷⁹ In an omnigenic architecture, the majority of the heritability influencing complex traits with polygenic architecture is distributed across peripheral genes and involves extensive regulatory networks connecting them to core genes.^{79,80} In practice, core genes are likely identified on a graded scale rather than a binary classification, where heritability decreases as the degree of separation from core genes is reduced. The use of co-expression networks to investigate the omnigenic model provides a powerful approach to untangle the complex interactions specified in this framework.

Co-expression networks across different conditions or tissues can use connectivity as a method to identify core and peripheral genes and specify the distribution of heritability across degrees of separation. For example, Mähler et al., used a transcriptional co-expression network from *Populus tremula* leaf buds to demonstrate that eQTL genes are predominantly located at the network's periphery and that connectivity is inversely correlated with eQTL effect sizes, implying that core genes (hub genes) within modules experience strong selective pressures.⁸¹ The purifying selection predominantly acting on core genes implies an evolutionary conservation that possibly underscores their fundamental biological roles. This finding is also consistent with the notion that damage to core genes by loss-of-function mutations tends to have strong effects on disease risk.^{79,81} Fóthi et al. evaluated the omnigenic model's application to autism spectrum disorder using brain-specific gene interaction networks and found that autism gene clusters are significantly more connected to each other and the peripheral genes in brain-related tissues than in non-brain-related tissues.⁸² These data support the notion that disease-relevant tissues are the appropriate context for assessing omnigenic architectures to better understand complex traits. Hartl et al. used this foundation to generate a co-expression network derived from gene expression profiles across 12 brain regions to contextualize the functional pathways of risk genes for multiple neuropsychiatric diseases.¹⁸ Despite the omnigenic model's suggestion that disease risk is mediated by a small number of core genes indirectly influenced by peripheral genes, the study indicated that gene effects are distributed more continuously across the networks rather than being separated into distinct core and peripheral categories.

The omnigenic model suggests that the connections between peripheral and core genes vary by trait and include transcriptional networks, post-translational modifications, and protein-protein interactions^{79,83} and that genetic variants influencing disease may affect expression in specific cell types or conditions.⁷⁹ The aforementioned studies all utilize steady-state expression data to construct their networks, whereas an omnigenic architecture relevant to disease may emerge under the conditions of specific cell stressors that drive the purifying selection pressure that core genes are placed under. The results obtained for DOX-correlated modules resemble what would be predicted by an omnigenic architecture, where there is a negative correlation between the influence of genetic variation and network connectivity. However, the opposite is true for proteins that are not correlated to DOX. This suggests that the core-periphery struc-

ture may develop in response to selective pressures. Our dynamic network differentiates proteins that are actively responding to stress from those that are not. While modules derived from steady-state networks can enrich for various biological processes, the connectivity observed at steady state may not accurately reflect gene relationships under disease-relevant conditions. Instead, the core-periphery structure may be dynamic, shifting according to different cellular states. If true, it means that networks constructed in response to various stressors might establish a connectivity profile that pinpoints unique or shared core-peripheral relationships to different cell states. Therefore, networks generated from cell models that capture the dynamic molecular responses to selective pressures or disease-associated stimuli may be needed to more accurately understand how the core-periphery structure described in the omnigenic model emerges in the transition from non-disease to disease states.

In summary, there are no studies integrating proteomic data from cardiomyocytes subjected to DNA damage with measures of genetic tolerance to variation and disease. Here, we profiled protein abundance in cardiomyocytes treated with DOX across multiple individuals. We found that the level of protein connectivity in DNA-damage-associated co-expression modules influences the tolerance to genetic variation. We believe that the data and analysis presented here will be a resource for further studies into the mechanistic effects of DNA damage on the cardiomyocyte proteome and DOX-induced cardiotoxicity, as well as for studies investigating the architecture of complex traits in response to perturbation.

Limitations of the study

We generate cardiomyocytes through directed differentiation of iPSCs as it allows us to include multiple individuals and use these cardiomyocytes in carefully controlled experiments where we can treat the same batch of cells with DOX and measure their protein abundance. However, it is possible that our *in vitro* system may not fully recapitulate the *in vivo* molecular profile. We also selected a single, sub-lethal, dose of DOX to study the primary effects of DNA damage on the proteome. It is possible that different doses of DOX would apply different selective pressures on the proteome and identify different response proteins and mutation tolerance. However, we believe our results using a clinically relevant dose of a widely used chemotherapeutic agent allow us to provide useful insights. Similarly, the response to DNA damage may be temporally dynamic and hence our results may not extrapolate to shorter or longer exposure times.

While mass spectrometry is a powerful tool for measuring protein abundance, it has inherent limitations that can affect the comprehensiveness and accuracy of the data acquired. For example, it may not detect low-abundance proteins effectively, leading to an incomplete representation of the proteome. Further, while protein abundance is a critical metric, it is not the only relevant measure when assessing protein function and cellular responses. Other important protein metrics include post-translational modifications such as phosphorylation, ubiquitination, and glycosylation, which can significantly alter protein function, stability, localization, and interactions. Future studies that integrate more comprehensive transcriptional and

proteomic networks that capture multiple time points and greater protein coverage may enhance the findings from similarly designed studies.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Dr. Michelle Claire Ward (miward@utmb.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: mass spectrometry data generated in this study are available in the MassIVE database (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>). The accession number is listed in the [key resources table](#).
- Code: all original code has been deposited at GitHub and is publicly available as of the date of publication. The link is listed in the [key resources table](#). Custom analysis scripts were generated using workflow.⁸⁴
- Other items: any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank all members of the Ward Lab for helpful discussions. We thank Dr. Kelly A. Frazer and the University of California San Diego for providing the iPSC lines through the iPSCORE resource. We acknowledge the Wellcome Trust Sanger Institute as the source of the WTSli048-A human induced pluripotent cell line, which was generated under the human induced pluripotent stem cell initiative funded by a grant from the Wellcome Trust and Medical Research Council, supported by the Wellcome Trust (WT098051) and the NIHR/Wellcome Trust Clinical Research Facility, and acknowledge Life Science Technologies Corporation as the provider of Cytotune. We thank the Mass Spectrometry Facility at the University of Texas Medical Branch for performing the mass spectrometry experiments, and the Flow Cytometry and Cell Sorting Core Facility for access to flow cytometers. This work was funded by a Cancer Prevention & Research Institute of Texas (<https://www.cpr.it.state.tx.us/>) Recruitment of First-Time Faculty Award (RR190110) to M.C.W., the National Institutes of Health grant R35GM150459 to M.C.W., and CPRIT grant RP190682 to W.K.R. in partial support of the UTMB Mass Spectrometry Facility. O.D.J. was supported by a Jeane B. Kempner Predoctoral Fellowship administered through UTMB (<http://www.kempnerfund.org/>).

AUTHOR CONTRIBUTIONS

M.C.W. conceived and designed the study. S.P., J.A.G., and W.K.R. performed the experiments and generated the data. O.D.J., S.P., and M.C.W. analyzed the data. O.D.J. and S.P. visualized the data. O.D.J. and M.C.W. wrote the paper with input from all authors. M.C.W. supervised the project. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Human induced pluripotent stem cell (iPSC) lines

- Ethical approvals and consent
- Sex and gender consideration
- Sample size and allocation to experimental groups

METHOD DETAILS

- iPSC culture
- Cardiomyocyte differentiation
- iPSC-CM purity determination using flow cytometry
- Drug treatment of iPSC-CMs
- γ H2AX immunofluorescence staining and quantification of DNA double-strand breaks
- Protein isolation and quantification
- Protein digestion
- Nanoflow liquid chromatography mass spectrometry
- NanoLC MS/MS analysis for DDA
- NanoLC MS/MS analysis for DIA
- Database searching for DDA proteins
- Database searching for DIA proteins
- Abundance matrix filtering and imputation
- Removal of unwanted technical variation and normalization
- Comparison of the iPSC-CM proteome with the proteome across human tissues
- WGCNA network construction
- Identification of hub proteins
- Module correlation to DOX treatment and individual
- Hub protein correlation network visualization in DOX-correlated modules
- Linear modeling to identify differentially abundant proteins
- Comparison of the DIA and DDA datasets
- Comparison of proteins elevated in heart tissue across modules
- Comparison of heart ventricle tissue specificity scores across modules
- Cellular localization of module proteins
- Biological process enrichment across modules
- Functional categorization of module proteins
- Protein family enrichment amongst module proteins
- pQTL protein enrichment
- Comparison of pQTL SNP effect sizes
- Proportion of pQTLs across network protein connectivity deciles
- pLI, pHaplo and pTripto comparisons across modules
- Comparisons between pLI scores and network connectivity
- CVD GWAS enrichment testing
- CVD risk protein interactor enrichment testing
- pLI comparison between CVD proteins and CVD protein interactors
- PPI network construction

QUANTIFICATION AND STATISTICAL ANALYSIS

ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2025.112474>.

Received: September 5, 2024

Revised: December 8, 2024

Accepted: April 15, 2025

Published: April 18, 2025

REFERENCES

1. Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., et al. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021.

2. Nikfarjam, S., and Singh, K.K. (2023). DNA damage response signaling: A common link between cancer and cardiovascular diseases. *Cancer Med.* *12*, 4380–4404.
3. Dai, Z., Ko, T., Fujita, K., Nomura, S., Uemura, Y., Onoue, K., Hamano, M., Katoh, M., Yamada, S., Katagiri, M., et al. (2024). Myocardial DNA damage predicts heart failure outcome in various underlying diseases. *Heart Fail.* *12*, 648–661.
4. Wu, L., Sowers, J.R., Zhang, Y., and Ren, J. (2023). Targeting DNA damage response in cardiovascular diseases: from pathophysiology to therapeutic implications. *Cardiovasc. Res.* *119*, 691–709.
5. Fugger, K., and West, S.C. (2016). Keeping homologous recombination in check. *Cell Res.* *26*, 397–398.
6. Linders, A.N., Dias, I.B., López Fernández, T., Tocchetti, C.G., Bomer, N., and Van der Meer, P. (2024). A review of the pathophysiological mechanisms of doxorubicin-induced cardiotoxicity and aging. *npj Aging* *10*, 9.
7. Yang, F., Teves, S.S., Kemp, C.J., and Henikoff, S. (2014). Doxorubicin, DNA torsion, and chromatin dynamics. *Biochim. Biophys. Acta* *1845*, 84–89.
8. Zhang, S., Liu, X., Bawa-Khalife, T., Lu, L.-S., Lyu, Y.L., Liu, L.F., and Yeh, E.T.H. (2012). Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nat. Med.* *18*, 1639–1642.
9. Mailet, A., Tan, K., Chai, X., Sadananda, S.N., Mehta, A., Ooi, J., Hayden, M.R., Pouladi, M.A., Ghosh, S., Shim, W., and Brunham, L.R. (2016). Modeling doxorubicin-induced cardiotoxicity in human pluripotent stem cell derived-cardiomyocytes. *Sci. Rep.* *6*, 25333.
10. Minotti, G., Menna, P., Salvatorelli, E., Cairo, G., and Gianni, L. (2004). Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol. Rev.* *56*, 185–229.
11. van der Zanden, S.Y., Qiao, X., and Neefjes, J. (2021). New insights into the activities and toxicities of the old anticancer drug doxorubicin. *FEBS J.* *288*, 6095–6111.
12. Cardinale, D., Colombo, A., Bacchiani, G., Tedeschi, I., Meroni, C.A., Veglia, F., Civelli, M., Lamantia, G., Colombo, N., Curigliano, G., et al. (2015). Early detection of anthracycline cardiotoxicity and improvement with heart failure therapy. *Circulation* *131*, 1981–1988.
13. Kilickap, S., Barista, I., Akgul, E., Aytemir, K., Aksoy, S., and Tekuzman, G. (2007). Early and late arrhythmogenic effects of doxorubicin. *South. Med. J.* *100*, 262–265.
14. Go, A.S., Hylek, E.M., Phillips, K.A., Chang, Y., Henault, L.E., Selby, J.V., and Singer, D.E. (2001). Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA* *285*, 2370–2375.
15. Walsh, R., Jurgens, S.J., Erdmann, J., and Bezzina, C.R. (2023). Genome-wide association studies of cardiovascular disease. *Physiol. Rev.* *103*, 2039–2055.
16. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* *20*, 467–484.
17. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.-H., Richardson, T. G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S.G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* *622*, 329–338.
18. Hartl, C.L., Ramaswami, G., Pembroke, W.G., Muller, S., Pintacuda, G., Saha, A., Parsana, P., Battle, A., Lage, K., and Geschwind, D.H. (2021). Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nat. Neurosci.* *24*, 1313–1323.
19. Johnson, E.C.B., Carter, E.K., Dammer, E.B., Duong, D.M., Gerasimov, E.S., Liu, Y., Liu, J., Betarbet, R., Ping, L., Yin, L., et al. (2022). Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat. Neurosci.* *25*, 213–225.
20. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., Dong, G., Fang, H., Robinson, A.E., and GTEx Consortium; and Snyder, M.P. (2020). A quantitative proteome map of the human body. *Cell* *183*, 269–283.e19.
21. Seyfried, N.T., Dammer, E.B., Swarup, V., Nandakumar, D., Duong, D.M., Yin, L., Deng, Q., Nguyen, T., Hales, C.M., Wingo, T., et al. (2017). A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell Syst.* *4*, 60–72.e4.
22. Pintacuda, G., Hsu, Y.-H.H., Tsafou, K., Li, K.W., Martin, J.M., Riseman, J., Biagini, J.C., Ching, J.K.T., Mena, D., Gonzalez-Lozano, M.A., et al. (2023). Protein interaction studies in human induced neurons indicate convergent biology underlying autism spectrum disorders. *Cell Genom.* *3*, 100250.
23. Zhu, Q.M., Hsu, Y.-H.H., Lassen, F.H., MacDonald, B.T., Stead, S., Malolepsza, E., Kim, A., Li, T., Mizoguchi, T., Schenone, M., et al. (2024). Protein interaction networks in the vasculature prioritize genes and pathways underlying coronary artery disease. *Commun. Biol.* *7*, 87.
24. BurrIDGE, P.W., Li, Y.F., Matsa, E., Wu, H., Ong, S.-G., Sharma, A., Holmström, A., Chang, A.C., Coronado, M.J., Ebert, A.D., et al. (2016). Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nat. Med.* *22*, 547–556.
25. Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, C.A., Alazizi, A., Pazokitoroudi, A., Sankaraman, S., Wen, X., et al. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *Elife* *10*, e67077.
26. Ward, M.C., Banovich, N.E., Sarkar, A., Stephens, M., and Gilad, Y. (2021). Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *Elife* *10*, e57345.
27. Knowles, D.A., Burrows, C.K., Blischak, J.D., Patterson, K.M., Serie, D. J., Norton, N., Ober, C., Pritchard, J.K., and Gilad, Y. (2018). Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *Elife* *7*, e33480.
28. Matthews, E.R., Johnson, O.D., Horn, K.J., Gutiérrez, J.A., Powell, S.R., and Ward, M.C. (2024). Anthracyclines induce cardiotoxicity through a shared gene expression response signature. *PLoS Genet.* *20*, e1011164.
29. Harris, L., Fondrie, W.E., Oh, S., and Noble, W.S. (2023). Evaluating Proteomics Imputation Methods with Improved Criteria. *J. Proteome Res.* *22*, 3427–3438. <https://doi.org/10.1021/acs.jproteome.3c00205>.
30. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., and Botstein, D. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* *17*, 520–525.
31. Gagnon-Bartsch, J.A., and Speed, T.P. (2012). Using control gene to correct for unwanted variation in microarray data. *Biostatistics* *13*, 539–552.
32. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* *9*, 559–613.
33. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjösted, E., and Asplund, A. (2015). Tissue-based map of the human proteome. *Science* *347*, 1260419.
34. UniProt Consortium (2023). the universal protein knowledgebase in 2023. *Nucleic Acids Res.* *51*, D523–D531.
35. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. *Cell* *172*, 650–665.
36. Nunes, C., Mestre, I., Marcelo, A., Koppenol, R., Matos, C.A., and Nóbrega, C. (2019). MSGP: the first database of the protein components of the mammalian stress granules. *Database* *2019*, baz031.
37. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature* *583*, 711–719.

38. Fan, D., Pang, S., Chen, J., Shan, J., Cheng, Q., and Yan, B. (2021). Identification and functional study of GATA4 gene regulatory variants in atrial septal defects. *BMC Cardiovasc. Disord.* *21*, 321.
39. Kobayashi, S., Volden, P., Timm, D., Mao, K., Xu, X., and Liang, Q. (2010). Transcription factor GATA4 inhibits doxorubicin-induced autophagy and cardiomyocyte death. *J. Biol. Chem.* *285*, 793–804.
40. Chen, X., Liu, Y., Xu, C., Ba, L., Liu, Z., Li, X., Huang, J., Simpson, E., Gao, H., Cao, D., et al. (2021). QKI is a critical pre-mRNA alternative splicing regulator of cardiac myofibrillogenesis and contractile function. *Nat. Commun.* *12*, 89.
41. Montañés-Agudo, P., Aufiero, S., Schepers, E.N., van der Made, I., Cócera-Ortega, L., Ernault, A.C., Richard, S., Kuster, D.W.D., Christoffels, V.M., Pinto, Y.M., and Creemers, E.E. (2023). The RNA-binding protein QKI governs a muscle-specific alternative splicing program that shapes the contractile function of cardiomyocytes. *Cardiovasc. Res.* *119*, 1161–1174.
42. Yan, F., Liu, R., Zhuang, X., Li, R., Shi, H., and Gao, X. (2021). Salidroside attenuates doxorubicin-induced cardiac dysfunction partially through activation of QKI/FoxO1 pathway. *J. Cardiovasc. Transl. Res.* *14*, 355–364.
43. Gudjonsson, A., Gudmundsdottir, V., Axelsson, G.T., Gudmundsson, E. F., Jonsson, B.G., Launer, L.J., Lamb, J.R., Jennings, L.L., Aspelund, T., Emilsson, V., and Gudnason, V. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat. Commun.* *13*, 480. <https://doi.org/10.1038/s41467-021-27850-z>.
44. Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L., Morley, T., Niestroj, L.-M., Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map of the human genome. *Cell* *185*, 3041–3055.e25.
45. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285–291. <https://doi.org/10.1038/nature19057>.
46. Yang, H., Bhoo-Pathy, N., Brand, J.S., Hedayati, E., Grassmann, F., Zeng, E., Bergh, J., Bian, W., Ludvigsson, J.F., Hall, P., and Czene, K. (2022). Risk of heart disease following treatment for breast cancer - results from a population-based cohort study. *Elife* *11*, e71562. <https://doi.org/10.7554/eLife.71562>.
47. Buza, V., Rajagopalan, B., and Curtis, A.B. (2017). Cancer Treatment-Induced Arrhythmias: Focus on Chemotherapy and Targeted Therapies. *Circ. Arrhythm. Electrophysiol.* *10*, e005443. <https://doi.org/10.1161/CIRCEP.117.005443>.
48. Larsen, C.M., Garcia Arango, M., Dasari, H., Arciniegas Calle, M., Adjei, E., Rico Mesa, J., Scott, C.G., Thompson, C.A., Cerhan, J.R., Haddad, T. C., et al. (2023). Association of Anthracycline With Heart Failure in Patients Treated for Breast Cancer or Lymphoma, 1985-2010. *JAMA Netw. Open* *6*, e2254669. <https://doi.org/10.1001/jamanetworkopen.2022.54669>.
49. Dean, Y.E., Dahshan, H., Motawea, K.R., Khalifa, Z., Tanas, Y., Rakha, I., Hasan, W., Kishk, M., Mahmoud, A., Elsayed, A., et al. (2023). Anthracyclines and the risk of arrhythmias: A systematic review and meta-analysis. *Medicine (Baltim.)* *102*, e35770. <https://doi.org/10.1097/MD.00000000000035770>.
50. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* *51*, D977–D985.
51. Nguyen, N., Souza, T., Verheijen, M.C.T., Gmuender, H., Selevsek, N., Schlapbach, R., Kleinjans, J., and Jennen, D. (2021). Translational proteomics analysis of anthracycline-induced cardiotoxicity from cardiac microtissues to human heart biopsies. *Front. Genet.* *12*, 695625.
52. Kumar, S.N., Konorev, E.A., Aggarwal, D., and Kalyanaraman, B. (2011). Analysis of proteome changes in doxorubicin-treated adult rat cardiomyocyte. *J. Proteomics* *74*, 683–697.
53. Yuan, Y., Fan, S., Shu, L., Huang, W., Xie, L., Bi, C., Yu, H., Wang, Y., and Li, Y. (2020). Exploration the mechanism of doxorubicin-induced heart failure in rats by integration of proteomics and metabolomics data. *Front. Pharmacol.* *11*, 600561.
54. Pang, B., Qiao, X., Janssen, L., Velds, A., Groothuis, T., Kerkhoven, R., Nieuwland, M., Ovaa, H., Rottenberg, S., Van Tellingen, O., et al. (2013). Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat. Commun.* *4*, 1908.
55. Chiamvimonvat, N., and Song, L.S. (2018). LRRC10 (Leucine-Rich Repeat Containing Protein 10) and REEP5 (Receptor Accessory Protein 5) as Novel Regulators of Cardiac Excitation-Contraction Coupling Structure and Function. *Am Heart Assoc* *7*, e008260.
56. Fochi, S., Lorenzi, P., Galasso, M., Stefani, C., Trabetti, E., Zipeto, D., and Romanelli, M.G. (2020). The emerging role of the RBM20 and PTBP1 ribonucleoproteins in heart development and cardiovascular diseases. *Genes* *11*, 402.
57. Liu, X., Li, B., Wang, S., Zhang, E., Schultz, M., Touma, M., Monteiro Da Rocha, A., Evans, S.M., Eichmann, A., and Herron, T. (2024). Stromal Cell-Slit3/Cardiomyocyte-Robo1 Axis Regulates Pressure Overload-Induced Cardiac Hypertrophy. *Circ. Res.* *134*, 913–930.
58. Feghaly, J., Zakka, P., London, B., MacRae, C.A., and Refaat, M.M. (2018). Genetics of atrial fibrillation. *J. Am. Heart Assoc.* *7*, e009884.
59. Fijalkowska, I., Sharma, D., Bult, C.J., and Danoff, S.K. (2010). Expression of the transcription factor, TFII-I, during post-implantation mouse embryonic development. *BMC Res. Notes* *3*, 203–208.
60. Johnson, B.B., Cosson, M.-V., Tsansizi, L.I., Holmes, T.L., Gilmore, T., Hampton, K., Song, O.-R., Vo, N.T.N., Nasir, A., Chabronova, A., et al. (2024). Perlecan (HSPG2) promotes structural, contractile, and metabolic development of human cardiomyocytes. *Cell Rep.* *43*, 113668.
61. Grande, V., Schuld, J., van der Ven, P.F.M., Gruss, O.J., and Fürst, D. O. (2023). Filamin-A-interacting protein 1 (FILIP1) is a dual compartment protein linking myofibrils and microtubules during myogenic differentiation and upon mechanical stress. *Cell Tissue Res.* *393*, 133–147.
62. Latham, S.L., Weiß, N., Schwanke, K., Thiel, C., Croucher, D.R., Zweigerdt, R., Manstein, D.J., and Taft, M.H. (2020). Myosin-18B regulates higher-order organization of the cardiac sarcomere through thin filament cross-linking and thick filament dynamics. *Cell Rep.* *32*, 108090.
63. Malfatti, E., Böhm, J., Lacène, E., Beuvin, M., Romero, N.B., Laporte, J., and Laporte, J. (2015). A premature stop codon in MYO18B is associated with severe nemaline myopathy with cardiomyopathy. *J. Neuromuscul. Dis.* *2*, 219–227.
64. Liu, Z., Li, W., Ma, X., Ding, N., Spallotta, F., Southon, E., Tessarollo, L., Gaetano, C., Mukoyama, Y.-s., and Thiele, C.J. (2014). Essential role of the zinc finger transcription factor Casz1 for mammalian cardiac morphogenesis and development. *J. Biol. Chem.* *289*, 29801–29816.
65. Lange, M., Kaynak, B., Forster, U.B., Tönjes, M., Fischer, J.J., Grimm, C., Schlesinger, J., Just, S., Dunkel, I., Krueger, T., et al. (2008). Regulation of muscle development by DPFF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* *22*, 2370–2384.
66. Damodar, G., Smitha, T., Gopinath, S., Vijayakumar, S., and Rao, Y. (2014). An evaluation of hepatotoxicity in breast cancer patients receiving injection Doxorubicin. *Ann. Med. Health Sci. Res.* *4*, 74–79. <https://doi.org/10.4103/2141-9248.126619>.
67. Lahoti, T.S., Patel, D., Thekkemadom, V., Beckett, R., and Ray, S.D. (2012). Doxorubicin-induced in vivo nephrotoxicity involves oxidative stress-mediated multiple pro- and anti-apoptotic signaling pathways. *Curr. Neurovasc. Res.* *9*, 282–295. <https://doi.org/10.2174/156720212803503636>.

68. Kesler, S.R., and Blayney, D.W. (2016). Neurotoxic Effects of Anthracycline- vs Nonanthracycline-Based Chemotherapy on Cognition in Breast Cancer Survivors. *JAMA Oncol.* 2, 185–192. <https://doi.org/10.1001/jamaoncol.2015.4333>.
69. Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M. V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., and Young, E.J. (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* 17, 1–22.
70. Park, B., Sim, S.H., Lee, K.S., Kim, H.J., and Park, I.H. (2020). Genome-wide association study of genetic variants related to anthracycline-induced cardiotoxicity in early breast cancer. *Cancer Sci.* 111, 2579–2587.
71. Aminkeng, F., Bhavsar, A.P., Visscher, H., Rassekh, S.R., Li, Y., Lee, J. W., Brunham, L.R., Caron, H.N., van Dalen, E.C., Kremer, L.C., et al. (2015). A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. *Nat. Genet.* 47, 1079–1084.
72. Velasco-Ruiz, A., Nuñez-Torres, R., Pita, G., Wildiers, H., Lambrechts, D., Hatse, S., Delombaerde, D., Van Brussel, T., Alonso, M.R., Alvarez, N., et al. (2021). POLRMT as a novel susceptibility gene for cardiotoxicity in epirubicin treatment of breast cancer patients. *Pharmaceutics* 13, 1942.
73. Gupta, S.K., Garg, A., Bär, C., Chatterjee, S., Foinquinos, A., Milting, H., Streckfuß-Bömeke, K., Fiedler, J., and Thum, T. (2018). Quaking Inhibits Doxorubicin-Mediated Cardiotoxicity Through Regulation of Cardiac Circular RNA Expression. *Circ. Res.* 122, 246–254. <https://doi.org/10.1161/CIRCRESAHA.117.311335>.
74. Li, J., Wang, P.Y., Long, N.A., Zhuang, J., Springer, D.A., Zou, J., Lin, Y., Bleck, C.K.E., Park, J.H., Kang, J.G., and Hwang, P.M. (2019). p53 prevents doxorubicin cardiotoxicity independently of its prototypical tumor suppressor activities. *Proc. Natl. Acad. Sci. USA* 116, 19626–19634. <https://doi.org/10.1073/pnas.1904979116>.
75. Roselli, C., Chaffin, M.D., Weng, L.C., Aeschbacher, S., Ahlberg, G., Albert, C.M., Almgren, P., Alonso, A., Anderson, C.D., Aragam, K.G., et al. (2018). Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* 50, 1225–1233. <https://doi.org/10.1038/s41588-018-0133-9>.
76. Anter, E., Jessup, M., and Callans, D.J. (2009). Atrial fibrillation and heart failure: treatment considerations for a dual epidemic. *Circulation* 119, 2516–2525.
77. Moreton, A., Kourtis, S., Gañez Zapater, A., Calabrò, C., Espinar Calvo, M.L., Fontaine, F., Darai, E., Abad Cortel, E., Block, S., Pascual-Reguant, L., et al. (2023). A metabolic map of the DNA damage response identifies PRDX1 in the control of nuclear ROS scavenging and aspartate availability. *Mol. Syst. Biol.* 19, e11267.
78. Srivastava, S., Gajwani, P., Jousma, J., Miyamoto, H., Kwon, Y., Jana, A., Toth, P.T., Yan, G., Ong, S.-G., and Rehman, J. (2023). Nuclear translocation of mitochondrial dehydrogenases as an adaptive cardioprotective mechanism. *Nat. Commun.* 14, 4360.
79. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186.
80. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans effects on gene expression can drive omnigenic inheritance. *Cell* 177, 1022–1034.e6.
81. Mähler, N., Wang, J., Terebieniec, B.K., Ingvarsson, P.K., Street, N.R., and Hvidsten, T.R. (2017). Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* 13, e1006402.
82. Fóthi, Á., Pintér, C., Pollner, P., and Lőrincz, A. (2022). Peripheral gene interactions define interpretable clusters of core ASD genes in a network-based investigation of the omnigenic theory. *npj Syst. Biol. Appl.* 8, 28.
83. Ratnakumar, A., Weinhold, N., Mar, J.C., and Riaz, N. (2020). Protein-Protein interactions uncover candidate ‘core genes’ within omnigenic disease networks. *PLoS Genet.* 16, e1008903.
84. Blischak, J.D., Carbonetto, P., and Stephens, M. (2019). Creating and sharing reproducible research code the workflow way. *F1000Research* 8, 1749.
85. Panopoulos, A.D., D’Antonio, M., Benaglio, P., Williams, R., Hashem, S. I., Schuldt, B.M., DeBoever, C., Arias, A.D., Garcia, M., Nelson, B.C., et al. (2017). iPSCORE: a resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Rep.* 8, 1086–1100.
86. Streeter, I., Harrison, P.W., Faulconbridge, A., The HipSci Consortium; Flicek, P., Parkinson, H., and Clarke, L. (2017). The human-induced pluripotent stem cell initiative—data resources for cellular genetics. *Nucleic Acids Res.* 45, D691–D697.
87. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
88. Hermann, G.K., Russell, W.K., Garg, N.J., and Yin, Y.W. (2021). Poly (ADP-ribose) polymerase 1 regulates mitochondrial DNA repair in an NAD-dependent manner. *J. Biol. Chem.* 296, 100309.
89. Searle, B.C., Swearingen, K.E., Barnes, C.A., Schmidt, T., Gessulat, S., Küster, B., and Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* 11, 1548. <https://doi.org/10.1038/s41467-020-15346-1>.
90. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 30, 918–920. <https://doi.org/10.1038/nbt.2377>.
91. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* 14, 513–520. <https://doi.org/10.1038/nmeth.4256>.
92. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. <https://doi.org/10.1038/s41592-019-0638-x>.
93. Gagnon-Bartsch, J.A., and Speed, T.P. (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics* 13, 539–552.
94. Almeida-Silva, F., and Venancio, T.M. (2022). BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct. Integr. Genomics* 22, 131–136.
95. R Core Team (2024). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).
96. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695. <https://igraph.org>.
97. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43, e47.
98. Molania, R., Gagnon-Bartsch, J.A., Dobrovic, A., and Speed, T.P. (2019). A new normalization for Nanostring nCounter gene expression data. *Nucleic Acids Res.* 47, 6073–6083.
99. Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al.; Gene Ontology Consortium (2023). The Gene Ontology knowledgebase in 2023. *Genetics* 224, iyad031.

100. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141.
101. Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., Genome Aggregation Database Consortium; Rehm, H. L., MacArthur, D.G., and O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Hum. Mutat.* 43, 1012–1030.
102. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646.
103. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
104. Pedersen, T. (2024). tidygraph: A Tidy API for Graph Manipulation. R package version 1.3.1.9000. <https://CRAN.R-project.org/package=tidygraph>.
105. Pedersen, T. (2024). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 2.2.1.9000. <https://CRAN.R-project.org/package=ggraph>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|---|---|
| Antibodies | | |
| Anti-Cardiac Troponin T antibody | BD Biosciences | Cat# 564767; RRID: AB_2738939 |
| Phospho-Histone H2A.X (Ser139) Rabbit mAb | Thermo Fisher Scientific | Cat# NC1602516; RRID: AB_2898141 |
| Donkey anti-Rabbit Alexa Fluor 594 antibody | Invitrogen | Cat# A-21207; RRID: AB_141637 |
| Chemicals, peptides, and recombinant proteins | | |
| mTESR1 | Stem Cell Technology | Catalog No. 85850 |
| Penicillin/Streptomycin | Corning | Catalog No. 30-002-CI |
| Matrigel hESC-qualified Matrix | Corning | Catalog No. 354277 |
| CHIR99021 trihydrochloride | Tocris Bioscience | Catalog No. 4953 |
| RPMI 1640 | Corning | Catalog No. 15-040-CM |
| B-27 minus insulin | Thermo Fisher Scientific | Catalog No. A1895601 |
| GlutaMAX | Thermo Fisher Scientific | Catalog No. 35050-061 |
| Wnt-C59 | Tocris Bioscience | Catalog No. 5148 |
| RPMI without glucose | Thermo Fisher Scientific | Catalog No. 11879 |
| L-Ascorbic acid 2-phosphate sesquimagnesium salt | Santa Cruz Biotechnology | Catalog No. sc228390 |
| Human Recombinant Albumin | Sigma-Aldrich | Catalog No. A0237 |
| HEPES (L+)Lactic acid sodium | Sigma-Aldrich | Catalog No. L7022 |
| Trypsin-EDTA | Corning | Catalog No. 25-053 CI |
| DMEM without glucose | Thermo Fisher Scientific | Catalog No. A1443001 |
| FBS | Genemate | Catalog No. S1200-500 |
| Galactose | Sigma-Aldrich | Catalog No. G5388 |
| Sodium pyruvate | Thermo Fisher Scientific | Catalog No. 11360-070 |
| FOXP3/Transcription Factor Staining Buffer Set | Thermo Fisher Scientific | Catalog No. 00-5523 |
| autoMACS® Buffer | Miltenyi Biotec | Catalog No. 130-091-221 |
| Doxorubicin | Sigma-Aldrich | Catalog No. D1515 |
| Hoechst 33342 nucleic acid stain | Thermo Fisher Scientific | Catalog No. PI62249 |
| Tris(2-carboxyethyl) phosphine | Thermo Fisher Scientific | Catalog No. 77720 |
| Trypsin | Promega | Catalog No. V5280 |
| Critical commercial assays | | |
| Zombie Violet Fixable Viability Kit | BioLegend | Catalog No. 423113 |
| BCA Protein Assay kit | Thermo Fisher Scientific | Catalog No. 23227 |
| Nanoflow liquid chromatography-tandem mass spectrometry and analysis | UltiMate 3000 RSLCnano, Dionex | N/A |
| Quantitative Fluorometric Peptide Assay | ThermoFisher Scientific | Catalog No. 23290 |
| Deposited data | | |
| Mass Spectrometry data | MassIVE database (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) | MassIVE: MSV000094446 |
| Code | GitHub | GitHub: https://github.com/mward-lab/Johnson_DNA_damage_DOX_2025 |
| Experimental models: Cell lines | | |
| Human iPSCORE iPSC line: UCSD143i-87-1 | University of California San Diego & the WiCell Research Institute | https://www.wicell.org/home/stem-cells/catalog-of-stem-cell-lines/ucsd143i-87-1.cmsx?closable=true |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|--|---|
| Human iPSCORE iPSC line: UCSD131i-77-1 | University of California San Diego & the WiCell Research Institute | https://www.wicell.org/home/stem-cells/catalog-of-stem-cell-lines/ucsd131i-77-1.cmsx?closable=true |
| Human HipSci iPSC line: WTSii048-A | EBiSC & ECACC | https://www.ebi.ac.uk/ena/browser/view/ERZ368931?show=analyses |

Software and algorithms

| | | |
|----------------------|--------------------------|------------------|
| MSConvert | Command line software | Version 3 |
| DIA-NN | Github | Version 1.9 |
| ImageJ | National Inst. Of Health | Version 1.54i |
| Cytoscape | Open source | Version 3.10.0 |
| AnnotationDbi | R package | Version 1.66.0 |
| biomaRt | R package | Version 2.60.1 |
| BioNERO | R package | Version 1.12.0 |
| clusterProfiler | R package | Version 4.12.6 |
| dynamicTreeCut | R package | Version 1.63-1 |
| edgeR | R package | Version 4.2.1 |
| ggraph | R package | Version 2.2.1 |
| igraph | R package | Version 2.0.3 |
| impute | R package | Version 5.1-3 |
| limma | R package | Version 3.60.4 |
| org.Hs.e.g...db | R package | Version 3.19.1 |
| pheatmap | R package | Version 1.0.12 |
| ruv | R package | Version 0.9.7.1 |
| stats | R package | Version 4.4.1 |
| SummarizedExperiment | R package | Version 1.34.0 |
| tidygraph | R package | Version 1.3.1 |
| WGCNA | R package | Version 1.72-5 |
| Proteome Discoverer | Thermo Fisher Scientific | Version 2.2.0388 |

Other

| | | |
|--|---|---|
| LSRFortessa Cell Analyzer | BD Bioscience | https://www.bdbiosciences.com/en-us/products/instruments/flow-cytometers/research-cell-analyzers/bd-lsrfortessa |
| nano-LC chromatography system | Thermo Fisher Scientific (UltiMate 3000 RSLCnano, Dionex) | https://www.thermofisher.com/order/catalog/product/ULTIM3000RSLCNANO |
| Orbitrap Eclipse Tribrid Mass Spectrometer | Thermo Fisher Scientific | https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-systems/orbitrap-lc-ms/orbitrap-tribrid-mass-spectrometers/orbitrap-eclipse-tribrid-mass-spectrometer.html |

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human induced pluripotent stem cell (iPSC) lines

Three human iPSC lines were used in this study. All lines were derived from skin fibroblasts of healthy female donors with no prior history of cardiac disease.

- (1) Individual 1 (UCSD131i-77-1): An iPSC line derived from a 21-year-old Asian-Chinese female donor that is part of the iPSCORE collection at the University of California San Diego. The cell line was generated by Dr. Kelly A. Frazer as part of the National Heart, Lung and Blood Institute Next Generation Sequencing Consortium⁸⁵ and is available through the biorepository at WiCell Research Institute (Madison, WI, USA), or through contacting Dr. Kelly A. Frazer at the University of California, San Diego.

- (2) Individual 2 (UCSD143i-87-1): An iPSC line derived from a 23-year-old Asian-Chinese female donor that is part of the iPSCORE collection at the University of California San Diego. The cell line was generated by Dr. Kelly A. Frazer as part of the National Heart, Lung and Blood Institute Next Generation Sequencing Consortium⁸⁵ and is available through the biorepository at WiCell Research Institute (Madison, WI, USA), or through contacting Dr. Kelly A. Frazer at the University of California, San Diego.
- (3) Individual 3 (WTSli048-A): An iPSC line derived from a 72-year-old White British female donor that is part of the HipSci project, funded by the Wellcome Trust and Medical Research Council.⁸⁶ The cell line is available from the European Bank of induced pluripotent Stem Cells (EBiSC) and the European Collection of Authenticated Cultures (ECACC).

All iPSC lines were maintained in feeder-free culture conditions with mTeSR1 medium on Matrigel-coated plates. Cells were cultured at 37°C in 5% CO₂ and 95% humidity and routinely passaged using EDTA-based dissociation. All iPSC lines were tested and confirmed negative for mycoplasma contamination prior to experimentation.

Ethical approvals and consent

- (1) The iPSCORE lines (UCSD131i-77-1 and UCSD143i-87-1) were generated with informed written consent from donors and approval by the Institutional Review Boards of the University of California San Diego and The Salk Institute (IRB protocol: 110776ZF).
- (2) The HipSci line (WTSli048-A) was generated with donor consent and approved by the East of England - Cambridge Central Research Ethics Committee (REC 15/EE/0049).

Sex and gender consideration

This study used iPSC lines derived from female donors. As such, the influence of biological sex on gene expression and phenotype could not be evaluated across sexes. This represents a limitation in terms of generalizability of our findings to male-derived cardiomyocytes. Future studies including male donor-derived iPSC lines are needed to address potential sex-specific effects in the observed phenotypes.

Sample size and allocation to experimental groups

This study included three human iPSC lines from three donors. The sample size was determined based on proteomics benchmarks and established molecular responses to DOX. Cardiomyocytes from each individual were treated with DOX or Vehicle. The treatment was replicated three times for Individual 3, allowing for modeling of technical variation in the experimental process. 10 samples in total were generated. Cardiomyocytes from each individual were randomly assigned to the DOX or Vehicle treatment groups. Protein was extracted from all samples in one batch.

METHOD DETAILS

iPSC culture

Feeder-independent iPSCs were cultured in mTESR1 (85850, Stem Cell Technology, Vancouver, BC, Canada) media with 1% Penicillin/Streptomycin (30-002-CI, Corning, Bedford, MA, USA) on hESC-qualified Matrigel Matrix (354277, Corning, Bedford, MA, USA) at a dilution of 1:100. iPSCs were passaged with dissociation reagent (0.5 mM EDTA, 300 mM NaCl in PBS) when they attained 70–80% confluency, approximately every 3–5 days.

Cardiomyocyte differentiation

Cardiomyocyte differentiations were performed as previously described.²⁶ Briefly, iPSC lines were seeded in Matrigel-coated culture dishes (Days –6/-5) and cultured until 85–90% confluent. Differentiations were initiated (Day 0) by adding 12 μM of the GSK3 inhibitor, CHIR99021 trihydrochloride (4953, Tocris Bioscience, Bristol, UK) in Cardiomyocyte Differentiation Media (CDM) [500 mL RPMI 1640 (15-040-CM, Corning), 10 mL B-27 minus insulin (A1895601, ThermoFisher Scientific, Waltham, MA, USA), 5 mL GlutaMAX (35050-061, ThermoFisher Scientific), and 5 mL of Penicillin/Streptomycin (100X) (30-002-CI, Corning)]. After 24 h (Day 1), the media was replaced with fresh CDM. On Day 3 (after 48 h) media was replaced with CDM containing 2 μM of the Wnt signaling inhibitor Wnt-C59 (5148, Tocris Bioscience) in CDM. CDM was replaced on Day 5, 7, 10 and 12. We observed spontaneously beating cells between Day 7–10. iPSC-CMs were purified by metabolic selection with glucose-free, lactate-containing media [500 mL RPMI without glucose (11879, ThermoFisher Scientific), 106.5 mg L-Ascorbic acid 2-phosphate sesquimagnesium salt (sc228390, Santa Cruz Biotechnology, Santa Cruz, CA, USA), 3.33 mL 75 mg/mL Human Recombinant Albumin (A0237, Sigma-Aldrich, St Louis, MO, USA), 2.5 mL 1 M lactate in 1 M HEPES (L(+)-Lactic acid sodium (L7022, Sigma-Aldrich)), and 5 mL Penicillin/Streptomycin] added on Day 14, 16 and 18. On Day 20, iPSC-CMs were detached with 0.05% Trypsin-EDTA solution (25-053 CI, Corning), and a single cell suspension was generated by straining. iPSC-CMs were counted with a Countess 2 machine. 1.5 million iPSC-CMs were plated per well of a 0.1% Gelatin-coated 6-well plate in 3 mL Cardiomyocyte Maintenance Media (CMM) [500 mL DMEM without glucose (A1443001, ThermoFisher Scientific), 50 mL FBS (S1200-500, Genemate), 990 mg Galactose (G5388, Sigma-Aldrich), 5 mL 100 mM

sodium pyruvate (11360–070, ThermoFisher Scientific), 2.5 mL 1 M HEPES (SH3023701, ThermoFisher Scientific), 5 mL Glutamax (35050–061, ThermoFisher Scientific), 5 mL Penicillin/Streptomycin]. The iPSC-CMs were matured in culture for 10 days, with CMM replaced on Day 23, 25, 27, 28, and 30.

iPSC-CM purity determination using flow cytometry

After differentiation of iPSCs from Individuals 2 & 3, the iPSC-CM purity was determined using flow cytometry. Between Day 25–27 of differentiation the iPSC-CMs were dissociated with 0.05% Trypsin-EDTA solution and strained to generate a single cell suspension. One million cells were stained with Zombie Violet Fixable Viability Kit (423113, BioLegend) for 30 min at 4°C prior to fixation and permeabilization (FOXP3/Transcription Factor Staining Buffer Set, 00–5523, ThermoFisher Scientific) for 30 min at 4°C. Cells were stained with 5 mL PE Mouse Anti-Cardiac Troponin T antibody (564767, clone 13–11, BD Biosciences, San Jose, CA, USA) for 45 min at 4°C. Cells were washed three times in permeabilization buffer and re-suspended in autoMACS Running Buffer (130-091-221, Miltenyi Biotec, Bergisch Gladbach, Germany). We used several negative controls in each flow cytometry experiment: 1) iPSCs, which should not express TNNT2, 2) an iPSC-CM sample that has not been labeled with viability stain or TNNT2 antibody, 3) an iPSC-CM sample that is only labeled with the viability stain and 4) an iPSC-CM sample that is only labeled with TNNT2. 10,000 cells were captured and profiled on the BD LSRFortessa Cell Analyzer. Multiple gating steps were performed to determine the proportion of TNNT2-positive cells: 1) Cellular debris was removed by gating out cells with low granularity on FSC versus SSC density plots, 2) From this population, live cells were identified as the violet laser-excitable, Pacific Blue dye-negative population, 3) TNNT2-positive cells were identified within the set of live cells and any cells that overlap the profiles of the negative control samples were excluded. iPSC-CM purity is reported as the proportion of TNNT2-positive live cells.

Drug treatment of iPSC-CMs

On Day 29, iPSC-CMs were treated with 0.5 μM of Doxorubicin (D1515, Sigma-Aldrich) or vehicle (Molecular Biology grade water) in fresh CMM media for 24 h. The treatment for Individual 3 was replicated two additional times yielding 10 samples in total across three individuals. Post-treatment, cells were washed twice and scraped in ice-cold PBS. iPSC-CMs were flash-frozen and stored at –80°C prior to further processing.

γH2AX immunofluorescence staining and quantification of DNA double-strand breaks

300,000 iPSC-CMs were seeded per well of a 24-well plate in CMM media. Cells were treated with 0.5 μM DOX or vehicle (DMSO). The treated cells were fixed in 4% paraformaldehyde for 15 min and permeabilized with 0.25% DPBS-T for 10 min at room temperature. Cells were incubated with 5% BSA:DPBS-T for 30 min at room temperature, then incubated with a 1:500 dilution of γ-H2AX primary antibody in BSA:DPBS-T overnight at 4°C (Phospho-Histone H2A.X (Ser139) Rabbit mAb; NC1602516; Fisher Scientific). Cells were incubated with Fluorochrome-conjugated secondary antibody (Donkey anti-Rabbit Alexa Fluor 594, A-21207, Invitrogen) for 1 h at room temperature at a 1:1000 dilution in DPBS-T. Cell nuclei were counterstained with Hoechst 33342 nucleic acid stain (PI62249, Thermo Scientific) for 10 min in the dark. Stained cells were subjected to fluorescence microscopy. The total number of nuclei and γH2AX-positive nuclei were quantified using the cell counter plugin of ImageJ software.⁸⁷ The number of γH2AX-positive nuclei were divided by the total number of nuclei to determine the percentage of γH2AX-positive nuclei in DOX- and vehicle-treated iPSC-CMs for three different individuals. The percentage of γH2AX-positive cells between vehicle- and DOX-treated iPSC-CM samples was compared by t test.

Protein isolation and quantification

Protein was isolated from iPSC-CMs by lysing the cells with RIPA buffer [1.5 mL 5 M NaCl, 1 mL Triton X-100, 1 g Na deoxycholate, 1 mL 10% SDS, 1 mL 1 M Tris pH 7.4 and 45 mL dH₂O] with protease inhibitor for 1 h at 4°C. Isolated proteins were quantified by using the BCA Protein Assay kit (23227, Thermo Scientific) according to the manufacturer's instructions.

Protein digestion

The samples were prepared similarly as previously described.⁸⁸ Briefly, 15 μg of protein were solubilized with 60 μL of 50 mM Tris-hydroxymethyl ammonium bicarbonate (TEAB) pH 7.55. The proteins were then reduced with 10 mM Tris(2-carboxyethyl) phosphine (TCEP) (77720, Thermo) and incubated at 65°C for 10 min. The sample was then cooled to room temperature and 1 μL of 500 mM iodoacetamide acid was added and allowed to react for 30 min in the dark. Then, 3.3 μL of 12% phosphoric acid was added to the protein solution followed by 200 μL of binding buffer (90% Methanol, 100mM TEAB pH 8.5). The resulting solution was added to S-Trap spin column (protifi.com) and passed through the column using a bench top centrifuge (60 s spin at 1,000 g). The spin column is washed with 150 μL of binding buffer and centrifuged. This is repeated two times. 30 μL of 20 ng/μL Trypsin is added to the protein mixture in 50 mM TEAB pH 8.5, and incubated at 37°C overnight. Peptides were eluted twice with 75 μL of 50% acetonitrile, 0.1% formic acid. Aliquots of 20 μL of eluted peptides were quantified using the Quantitative Fluorometric Peptide Assay (Pierce, Thermo Fisher Scientific). Eluted volume of peptides corresponding to 5.5 μg of peptides are dried in a speed vac and resuspended in 27.5 μL 1.67% acetonitrile, 0.08% formic acid, 0.83% acetic acid, 97.42% water and placed in an autosampler vial.

Nanoflow liquid chromatography mass spectrometry

Peptide mixtures were analyzed by nanoflow liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) using a nano-LC chromatography system (UltiMate 3000 RSLCnano, Dionex), coupled on-line to a Thermo Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a nanospray ion source. Instrument performance was verified by analyzing a standard six protein mix digest before the sample set run, between each experimental block and at the end of the experiment. The six protein mix data files were analyzed to confirm that instrument performance remained consistent throughout the experiment. A direct injection method using 3 μL of digest onto an analytical column was used; Aurora (75 $\mu\text{m} \times 25 \text{ cm}$, 1.6 μm) from (IonOpticks). After equilibrating the column in 98% solvent A (0.1% formic acid in water) and 2% solvent B (0.1% formic acid in acetonitrile (ACN)), the samples (2 μL in solvent A) were injected (300 nL/min) by gradient elution onto the C18 column as follows: isocratic at 2% B, 0–5 min; 2%–6%, 5–5.1 min; 6%–25% 5.1–105 min, 25%–50% B, 105–120 min; 50%–90% B, 120–122 min; isocratic at 90% B, 122–124 min; 90%–5%, 124–125 min; isocratic at 5% B, 125–126 min; 5%–90% 126–128 min; isocratic for 1 min; 90%–2%, 129–130 min; and isocratic at 2% B, till 150 min.

NanoLC MS/MS analysis for DDA

All data were acquired using an Orbitrap Eclipse in positive ion mode using a top speed data-dependent acquisition (DDA) method with a 3 s cycle time and a spray voltage of 1600 V. The survey scans (m/z 375–2000) were acquired in the Orbitrap at 120,000 resolution (at $m/z = 400$) in profile mode, with a maximum injection time of 100 ms and an AGC target of 1,000,000 ions. The S-lens RF level was set to 30. Isolation was performed in the quadrupole with a 1.6 Da isolation window, and HCD MS/MS acquisition was performed in profile mode using the orbitrap at a resolution of 15,000 using the following settings: parent threshold = 5,000; collision energy = 30%; AGC target at 125,000 using the default settings. Monoisotopic precursor selection (MIPS) and charge state filtering were on, with charge states 2–10 included. Dynamic exclusion was used to remove selected precursor ions, with a ± 10 ppm mass tolerance, for 30 s after acquisition of one MS/MS spectrum.

NanoLC MS/MS analysis for DIA

All LC-MS/MS data were acquired using an Orbitrap Eclipse in positive ion mode using a data-independent acquisition (DIA) method with 16 Da windows from 400 to 1000 and a loop time of 3 s. The survey scans (m/z 350–1500) were acquired in the Orbitrap at 60,000 resolution (at $m/z = 400$) in centroid mode, with a maximum injection time of 118 ms and an AGC target of 100,000 ions. The S-lens RF level was set to 60. Isolation was performed in the quadrupole, and HCD MS/MS acquisition was performed in profile mode using the orbitrap at a resolution of 30000 using the following settings: collision energy = 33%, IT 54 ms, AGC target = 50,000. A pooled sample was used to create spectral libraries that we search the individual samples against by injecting 5 times using narrow (4 Da), staggered windows over 100 m/z ranges from 400 to 900 m/z in a technique called gas phase fractionation as described in Searle et al.⁸⁹

Database searching for DDA proteins

Tandem mass spectra were extracted and charge state deconvoluted using Proteome Discoverer (Thermo Fisher, version 2.2.0388). Deisotoping was not performed. All MS/MS spectra were searched against a Uniprot Human database using Sequest and the Minora node used to perform Label-Free Quan (LFQ) using the MS peak areas for each of the peptide-spectral matches (PSMs). Searches were performed with a parent ion tolerance of 5 ppm and a fragment ion tolerance of 0.02 Da. Trypsin was specified as the enzyme, allowing for two missed cleavages. Fixed modification of carbamidomethyl (C) and variable modifications of oxidation (M) and deamidation were specified in Sequest. Protein identities reported at 1% false discovery rate were considered for filtering and downstream analysis.

Database searching for DIA proteins

The raw data was demultiplexed to mzML with 10 ppm accuracy after peak picking in MSConvert.⁹⁰ The resulting mzML files were searched in MSFragger⁹¹ and quantified via DIA-NN⁹² using the following settings: peptide length range 7–50, protease set to Trypsin, 2 missed cleavages, 3 variable modifications, clip N-term M on, fixed C carbamidomethylation, variable modifications of methionine oxidation and n-terminal acetylation, MS1 and MS2 accuracy set to 20 ppm, 1% FDR, and DIANN quantification strategy set to Robust LC (high accuracy). The files were searched against a database of human acquired from Uniprot (18th December, 2023). The gas-phase fractions were used only to generate the spectral library, which was used for analysis of the individual samples.

Abundance matrix filtering and imputation

We removed four non-human or uncharacterized proteins (UniProt: Q6ZSR9, P15252, P25691, P00761). Proteins that were missing across 50% or more of the samples were removed. For the remaining 246 proteins, missing values were imputed using k-Nearest Neighbors from the impute package knn. with, parameters k (number of neighbors to be used in the imputation) and rowmax (the maximum percent missing data allowed in any row).³⁰ Using k = 10 in the impute.knn function leverages all available samples to impute missing values, given the small sample size of our dataset. The rowmax = 0.4 parameter allows for imputing rows with up to 40% missing values, balancing the need to retain as much data as possible while maintaining the quality of the imputations. These steps led to a total of 4,178 analyzable proteins across all samples.

Removal of unwanted technical variation and normalization

To eliminate unwanted technical variation from the \log_2 -transformed imputed abundance matrix, we adjusted the 10 sample data matrix using both negative controls and the replicate data with the RUV-III function in the *ruv* R package.⁹³ Negative control proteins were defined as the 5% least variable proteins across all samples. The RUV-corrected abundance matrix and RUV factors were used in downstream analysis.

Comparison of the iPSC-CM proteome with the proteome across human tissues

Protein abundance values across 26 tissues was obtained from Jiang et al.²⁰ We identified the set of proteins expressed across the 26 tissues and our iPSC-CMs. We calculated the median abundance across individuals for the set of 26 tissues, as well as for our iPSC-CMs. The median abundance between our iPSC-CMs and each of the 26 tissues was correlated using Pearson correlation.

WGCNA network construction

We adopted the Weighted Gene Co-expression Network Analysis (WGCNA) methodology to investigate correlations between protein abundances in our RUV-corrected abundance matrix of DOX- and VEH-treated iPSC-CMs.³² The WGCNA framework, primarily utilizing wrapper functions from BioNERO, was implemented for this analysis.⁹⁴ We first established a scale-free network topology, achieved by determining the appropriate soft threshold power using the *SFT_fit* function. After iterating different soft power thresholds (β), the linear regression of $\log_{10}(k)$ versus $\log_{10}(p(k))$ indicates that by setting $\beta = 20$, the network is close to a scale-free network, where k is the whole network connectivity and $p(k)$ is the corresponding frequency distribution. This setting resulted in our network attaining a scale-free fit index of 0.71 (quantification of how well the network approximates a scale-free topology), with a mean node connectivity of 65.4 and median connectivity of 53.6. We identified modules of co-expressed proteins using a modified workflow as previously described.^{32,94} The workflow was encapsulated in the *exp2gcn2* function from the BioNERO package. We selected a signed network with a soft power threshold of 20, merging threshold of 0.85, and the Pearson correlation method. We computed the Topological Overlap Matrix (TOM), a measure of network connectivity that emphasizes the shared neighbors between protein pairs to enhance the robustness and reliability of the calculated adjacency network. Using hierarchical clustering on the dissimilarity TOM (dissTOM), we identified initial modules of co-expressed proteins. The dynamic tree cut method using the *cutreeDynamicTree* function with *maxTreeHeight* of 3, minimum module sizes of 40 and no deep splitting was applied to the protein dendrogram to define these modules by modifying the *exp2gcn2* function. The module eigenproteins (MEs), representing the first principal component for the module, were then calculated across modules with *moduleEigengenes*. These eigenproteins served as characteristic expression profiles of proteins within a module and were used to assess the interrelation between modules. Similar modules were merged based on the eigenprotein dissimilarity, ensuring that highly correlated modules were combined. Modules whose eigenproteins had a correlation of 0.85 or greater were merged.

We used three types of connectivity metrics to describe how each node/protein in the network related to other nodes/proteins. Total connectivity was calculated by summing the weighted correlations between each protein and all other proteins in the network (*kTotal*). Intra-modular connectivity (*kIN*) was calculated by summing the weighted correlations between each protein and all other proteins in the assigned module. Extra-modular connectivity (*kOut*) was calculated by summing the weighted correlations between each protein and all other proteins outside the assigned module.

Identification of hub proteins

We identified hub proteins that might play central roles in the biological processes represented by each module. We used the *get_hubs_gcn* function to identify hub proteins within our protein abundance correlation network. Hub proteins within modules are defined as the proteins with the highest intra-modular connectivity (*kIN*) score (top 10%) and the highest Pearson correlation value with the module eigenprotein (>0.8).

Module correlation to DOX treatment and individual

To assess the relationship between each module's eigenprotein and the defined trait (DOX treatment or individual), we performed a Pearson correlation analysis using the *cor.test* function from the *stats* package in R.⁹⁵ This analysis provided a correlation coefficient for each module, indicating the strength and direction of the association between the module's expression profile and the experimental condition. The statistical significance of these correlations was determined by extracting *p* values for each module-trait correlation using the *cor.test* function, where *p* values less than 0.01 were determined to be significant.

Hub protein correlation network visualization in DOX-correlated modules

The correlation network for hub proteins in DOX-correlated modules was visualized using the *igraph* package in R.⁹⁶ Abundance correlations between proteins were obtained using the *get_edge_list* function from BioNERO.⁹⁴ Protein pairs with positive Pearson correlations of 0.9 or more were selected for visualization.

Linear modeling to identify differentially abundant proteins

We utilized the *limma* package to fit protein abundances to a linear model across conditions.^{97,98} We randomly selected one technical replicate per treatment group from Individual 3 so as not to confound technical and biological variation in the linear modeling process.

The RUV-III corrected abundance matrix from six samples was quantile normalized using the `normalizeBetweenArrays` function. Drug treatment (DOX or VEH) was modeled as a fixed effect, whereas individual (IND) was treated as a random effect estimated using the `duplicateCorrelation` function. The linear model fitting was done using `lmFit`, which incorporated the block effect from the individuals and the design matrix. This model was then passed through the empirical Bayes moderation in the `eBayes` function to obtain moderated t-statistics. We defined contrasts in the linear model to compare the differential expression between DOX and VEH conditions such that positive \log_2 -fold change values correspond to increased abundance in the DOX-treated group, and negative values correspond to decreased abundance in the DOX-treated group. The model was refitted with these contrasts, and empirical Bayes moderation was applied again to adjust the statistics. We performed multiple-testing correction with the Benjamini-Hochberg method and denoted those proteins that meet an adjusted p -value threshold of 0.05 as differentially abundant proteins (DAPs).

Comparison of the DIA and DDA datasets

We detected 4,501 proteins by DDA and removed all proteins containing at least one missing protein abundance value, resulting in 3,384 measured proteins. We detected 4,261 proteins by DIA and removed all proteins containing at least one missing protein abundance value, resulting in 3,934 measured proteins. We selected the 3,027 proteins shared between these two sets and calculated the mean \log_2 abundance for each protein across all samples. DDA and DIA protein \log_2 abundance were fit to a linear model using the `lmFit` function to determine the strength and significance of the correlation between datasets. We identified differentially abundant proteins in the DDA dataset as described above for DIA, using the same preprocessing and modeling parameters. We then compared the response effect sizes (\log_2 fold change) between proteins shared between the two datasets using the `lmFit` function to determine the strength and significance of the correlation between datasets.

Comparison of proteins elevated in heart tissue across modules

We sourced data on proteins whose expression is elevated in heart tissue compared to other tissue types from the Human Protein Atlas.³³ Elevated proteins correspond to those with at least a 4-fold higher mRNA level in a particular tissue compared to any other tissue. Proteins identified as elevated in heart tissue were categorized into their respective modules within our correlation network. We then calculated their percentages relative to the total protein count within each respective module.

Comparison of heart ventricle tissue specificity scores across modules

Tissue specificity values for 'Heart.Ventricle' were obtained for our 4,178 proteins.²⁰ We tested for differences in tissue specificity across modules using a Wilcoxon rank-sum test for each module. We compared scores between all proteins in a given module, and all proteins not contained within that module, where significant differences were denoted when $p < 0.05$.

Cellular localization of module proteins

We retrieved lists of proteins classified as signal peptides, voltage-gated channels, secreted, intracellular, membrane-bound, or plasma proteins from the Human Protein Atlas database.³³ We utilized the UniProt database³⁴ to identify proteins experimentally shown to be located in subcellular structures including autophagosomes, membranes, cytoplasm, cytoskeleton, endoplasmic reticulum, Golgi apparatus, lysosomes, mitochondria, nucleus, and sarcomere. We classified each protein in the correlation network according to the above annotations, and performed a Fisher's exact test to determine which modules were enriched for proteins from a particular classification. Fisher's exact test $p < 0.05$ was considered significant.

Biological process enrichment across modules

To interpret the biological significance of the WGCNA-derived modules correlated with the DOX treatment, enrichment analysis was performed based on annotated Gene Ontologies (GO).⁹⁹ The `enrichGO` function from the `clusterProfiler` R package was used to test the enrichment of terms associated with biological processes against the background of all network proteins.¹⁰⁰ Enriched terms were those with a Benjamini-Hochberg adjusted $p < 0.05$.

Functional categorization of module proteins

We retrieved lists of proteins classified as human transcription factors,³⁵ RNA binding proteins (RBPs),³⁷ enzymes and transporter proteins,³³ and mammalian stress granule (MSG) proteins.³⁶ Enrichment of different protein categories within each module was determined by Fisher's exact test. Enriched terms were those with a Benjamini-Hochberg adjusted $p < 0.05$.

Protein family enrichment amongst module proteins

We queried UniProt for protein family names for the set of network proteins.³⁴ We assigned each protein to its set of families and performed a Fisher's exact test to determine which modules were enriched for proteins from a particular family. Those with a Benjamini-Hochberg adjusted $p < 0.05$ were considered significant.

pQTL protein enrichment

We obtained plasma pQTL data from the UK Biobank.¹⁷ We selected pQTLs that were identified independent of genetic ancestry. pQTLs were further classified as *cis*-pQTLs or *trans*-pQTLs. For pQTL protein enrichment analysis, we collated the set of unique

pQTL proteins. We tested for enrichment of pQTL proteins among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins, compared to the set of proteins not contained within each of those sets using the Fisher's exact test. A Fisher's exact $p < 0.05$ was considered statistically significant.

Comparison of pQTL SNP effect sizes

From the aforementioned set of *cis*-pQTLs and *trans*-pQTLs, we selected the SNP with the highest effect size for each protein. We tested for differences in effect sizes between DOX-correlated and non-DOX-correlated proteins, hub proteins and non-hub proteins, and DOX-correlated hub proteins and non-DOX-correlated hub proteins using a Wilcoxon rank-sum test, where a $p < 0.05$ was considered statistically significant.

Proportion of pQTLs across network protein connectivity deciles

For each protein in the network, we calculated the normalized intra-modular connectivity by taking the *k*_{IN} of each protein and dividing it by one less than the number of module connections. Proteins were assigned to one of two groups based on their DOX-correlation status, and stratified into deciles based on their normalized intra-modular connectivity. We then calculated the proportion of proteins in each decile that were pQTLs (either *cis*- or *trans*-pQTLs).

pLI, pHaplo and pTripto comparisons across modules

We obtained probability of loss-of-function intolerance (pLI) scores¹⁰¹ and probability of Haploinsufficiency (pHaplo) and Triplosensitivity (pTripto) scores.⁴⁴ We tested whether the set of scores for each metric was different for each module compared to all proteins in the network outside the module using the Wilcoxon rank-sum test. $p < 0.05$ was considered to be significant.

Comparisons between pLI scores and network connectivity

Proteins with a pLI ≥ 0.9 are considered intolerant to mutation, while proteins with a pLI ≤ 0.1 are considered tolerant to mutation. We compared the normalized *k*_{IN} values of mutation-intolerant proteins to mutation-tolerant proteins using the Wilcoxon rank-sum test. $p < 0.05$ was considered to be significant.

We used three types of connectivity metrics to describe how each protein in the network related to pLI scores across connectivity deciles. Total connectivity was calculated by summing the weighted correlations between each protein and all other proteins in the network (*k*_{Total}). Normalized *k*_{IN} was used as described above. Extra-modular connectivity (*k*_{Out}), the correlation between proteins within each module to all proteins outside the module, was normalized by dividing *k*_{Out} by the number proteins outside the target protein's module. Proteins were assigned into two groups based on their DOX-correlation status, and deciles generated for each aforementioned type of connectivity.

To ascertain enrichment for mutation-tolerant or intolerant proteins among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins, the Fisher's exact test was conducted. A Fisher's exact $p < 0.05$ was considered statistically significant.

CVD GWAS enrichment testing

We obtained GWAS summary statistics from the GWAS catalog.⁵⁰ We downloaded summary statistics for "heart disease" (EFO_0003777) to obtain risk proteins for "atrial fibrillation" and "heart failure". As per the catalog's metadata, each trait is associated with a set of SNPs and each SNP is associated with a 'mapped gene'. If the SNP is located in a chromosomal region that falls within a gene, then that gene is considered the mapped gene. For intergenic SNPs, SNPs are mapped to the nearest gene upstream or downstream of its chromosomal location. These 333 mapped AF genes and 115 mapped HF genes were intersected with our set of iPSC-CM expressed proteins yielding 70 AF and 20 HF proteins to interrogate for enrichment across modules. Module-wise enrichment testing was then performed by comparing to the set of all proteins not contained within the module of interest. This analysis involved constructing contingency tables for each module-gene set pair and conducting Fisher's exact tests to determine the overlap.

CVD risk protein interactor enrichment testing

Using the set of 'heart disease' CVD risk proteins as mentioned above, protein UniProt IDs were uploaded to STRINGdb and PPI networks were generated.¹⁰² We focused on proteins with experimental interaction evidence (confidence score ≥ 0.9). The protein interactors for CVD risk proteins were denoted "CVD protein interactors". To ascertain enrichment for CVD proteins and CVD protein interactors among hub proteins, DOX-correlated proteins, and DOX-correlated hub proteins, the Fisher's exact test was conducted. A Fisher's exact $p < 0.05$ was considered statistically significant.

pLI comparison between CVD proteins and CVD protein interactors

Using the set of CVD proteins and CVD protein interactors as mentioned above, we assigned each protein their pLI score as previously described. We compared the pLI values between the two groups using the Wilcoxon rank-sum test. $p < 0.05$ was considered to be significant.

We also generated a degree-randomized PPI network with resampling to ensure robustness of results. We utilized the set of CVD proteins and their physical protein-protein interaction partners, as determined by STRINGdb with confidence scores above 0.9. The degree of each GWAS protein within the physical PPI network was calculated, resulting in a degree distribution for the CVD proteins.

This distribution highlighted the percentage of proteins with varying degrees of interaction. CVD proteins were assigned group names based on their degree, and we subsequently calculated the degree for every protein in the entire network. Proteins in the network with degrees similar to those of the CVD proteins were assigned to the same group, while those with degrees not captured by the CVD-proteins were assigned to the group with the nearest degree. Each protein in the network was thus assigned to a group reflective of the degree distribution of the GWAS proteins, ensuring comprehensive sampling without excluding proteins based on their degree. We resampled while maintaining a degree-based proportion comparable to that of the original CVD proteins. We simulated 10,000 subnetworks among sampled proteins and their interactors, comparing pLI scores between CVD proteins and their physical interactors using the Wilcoxon rank-sum test.

PPI network construction

UniProt IDs for all proteins in the iPSC-CM network were imported into Cytoscape¹⁰³ to generate a PPI network. PPIs were determined using the STRINGdb application in Cytoscape. We selected the ‘physical subnetwork’ of PPIs with confidence scores >0.4, prior to later sub-setting PPIs with confidence scores >0.9 in R. The combination of both moderate and stringent cutoffs was used for sensitivity analysis of our results. We then annotated proteins in the network by whether they were DOX-correlated, hub proteins or CVD proteins, as well as by which module each protein belonged to. Edges in the network represent the WGCNA-derived weighted correlation between proteins. We next generated a subnetwork from the total network by selecting PPIs where at least one protein in the PPI was a CVD protein to center our analyses on the differences between CVD proteins and their direct physical protein interactors. We then used the tidygraph¹⁰⁴ and ggraph¹⁰⁵ packages in the R programming environment to visualize a network of CVD proteins and their physical protein interactors.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R version 4.4.1 unless otherwise noted. Analyses were performed using R packages including: WGCNA, limma, edgeR, clusterProfiler, ggplot2, igraph, AnnotationDbi, biomaRt, pheatmap, BioNERO, dynamicTreeCut, ggraph, ruv, impute, org.Hs.e.g.,.db, stats, SummarizedExperiment, and tidygraph (see [key resources table](#)). Statistical details—including test types, *n* values, *p* values and adjusted *p* values are provided in the figure legends, [results](#) and [STAR Methods](#) sections, and Supplemental Tables.

For γ H2AX immunofluorescence analysis ([Figure 1C](#)), *n* represents three biological replicates derived from three independent iPSC-CM lines (one per individual). For proteomics-based comparisons, *n* = 10 total samples, derived from iPSC-CMs generated from three individuals. Individuals 1 and 2 each contributed one sample per treatment group (DOX and VEH), while Individual 3 contributed three technical replicates per treatment group. For network analyses and enrichment testing, *n* represents the number of proteins in each module or test set, as detailed in [Table S2](#).

A two-tailed paired Student’s *t* test was used to compare γ H2AX-positive nuclei percentages between DOX- and VEH-treated cardiomyocytes ($p < 0.05$). Pearson correlation (via `cor.test`, R stats package) was used to assess relationships between module eigenproteins and traits (e.g., DOX treatment, individual), with $p < 0.01$ considered significant. Differential abundance analysis of proteins was performed using a linear model, and multiple testing correction was applied using the Benjamini-Hochberg method (adjusted $p < 0.05$). The Wilcoxon rank-sum test was used to compare distributions of tissue specificity scores across modules, SNP effect sizes in pQTL proteins, mutation tolerance metrics (pLI, pHaplo, pTripto) across modules, and connectivity distributions of mutation-tolerant vs. intolerant proteins. Fisher’s exact test was applied to assess enrichment of Gene Ontology (GO) terms and biological processes, protein family domains and subcellular localization categories, pQTL overlap and mutation intolerance across hub proteins, GWAS gene/protein enrichment in modules or networks, and drug target enrichment in DOX-correlated modules.

Unless otherwise specified, $p < 0.05$ was considered statistically significant. Multiple testing was performed using Benjamini-Hochberg false discovery rate (FDR) control where applicable. Values are expressed as mean unless otherwise indicated. Error bars represent mean \pm standard deviation (SD) as noted in [Figure 1C](#).

ADDITIONAL RESOURCES

This study does not involve a clinical trial and is not registered in a clinical trial database.