NERINE reveals rare variant associations in gene networks across phenotypes and implicates an *SNCA-PRL-LRRK2* subnetwork in Parkinson's disease

Authors

Sumaiya Nazeen[1,2,3,4], Xinyuan Wang[3], Autumn Morrow[1,3], Ronya Strom[3], Elizabeth Ethier[3], Dylan Ritter[5], Alexander B. H. Henderson[6], Jalwa Afroz[5], Christopher S. Cassa[2,4], Nathan O. Stitziel[7,8], Rajat M. Gupta[2,9], Kelvin Luk[10], Lorenz Studer[5], Vikram Khurana**[1,4,11], Shamil R. Sunyaev**[1,2,4]

**Correspondence

1. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
2. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
3. Division of Movement Disorders, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
4. Broad Institute of MIT and Harvard, Cambridge, MA, USA
5. The Center for Stem Cell Biology, Sloan-Kettering Institute for Cancer Research, New York, NY, USA.
6. Department of Neurology, Sean M. Healey & AMG Center for ALS, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
7. Cardiovascular Division, John T. Milliken Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
8. Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
9. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
10. Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, PA, USA
11. Harvard Stem Cell Institute, Cambridge, MA, USA

Abstract

There are two primary approaches to study the genetic basis of human phenotypes. Experiments in model systems generate interpretable gene networks but, in isolation, do not establish relevance to the human condition. Statistical genetics identifies relevant association signals at the variant or gene level but lacks tools to test specific mechanistic models, as existing methods do not incorporate the topology of gene-gene interactions. We bridge these two strategies by introducing a method that competitively tests network hypotheses with rare variant associations. A hierarchical model-based association test NERINE for the first time incorporates gene network topology while remaining resilient to network inaccuracies. We demonstrate NERINE's ability to test network hypotheses derived from both canonical pathway databases and model system screens. Comprehensive database-wide search of pathway networks with NERINE uncovers compelling associations for breast cancer, cardiovascular diseases, and type II diabetes, which are undetected by single-gene tests. Testing bespoke networks from experimental screens targeting key PD pathologies: dopaminergic neuron survival and *α-synuclein* pathobiology, NERINE highlights rare variant burden in gene modules related to autophagy, vesicle trafficking, and protein homeostasis. Genome-scale CRISPRi-screening of *α-synuclein* toxicity modifiers in human neurons and NERINE converge on *PRL*, revealing an intraneuronal *α-synuclein*/*prolactin* stress response that may impact resilience to PD pathologies.

Introduction

Advancements in high-throughput screens, such as perturbation assays in model systems, have enabled the generation of vast amounts of data capturing crucial biological knowledge, often represented as gene and protein interaction networks[1-7]. Some networks represent well-characterized metabolic or signaling pathways, while others capture regulatory interactions, physical protein-protein interactions, or protein-DNA associations. More abstractly, genetic interactions, such as co-essentiality, reflect dependencies in gene contributions to cellular fitness and phenotypes. As a result, gene networks serve as fundamental frameworks for understanding polygenic trait variation, including complex disease mechanisms, enabling researchers to formulate testable biological hypotheses about human phenotypes[8-10]. However, establishing causal links between gene networks and human phenotypes requires direct evidence from human genetics.

Rare protein-coding variants offer a direct means of establishing such connections, as these mutations occur within gene bodies and can directly impact gene function[11,12]. When tested at the level of pathways and networks, they can reveal new insights about causal disease mechanisms[13-16]. Despite their potential, no existing rare variant association test leverages gene networks with defined topologies to assess their association with human phenotypes[14-25]. They treat pathways as either a single entity ("mega-gene") or a simple collection of genes ("bag of genes"), not accounting for the complex interactions between genes and the fact that not all genes in a pathway contribute equally to a phenotype.

To address this gap, we present NERINE (NEtwork-based Rare varIaNt Enrichment), a new statistical framework for rare variant association testing that integrates information on genes and their interactions into a hierarchical model, ensuring robustness against network inaccuracies. NERINE offers several key advantages. First, it allows for the competitive evaluation of network hypotheses, prioritizing network topologies and the experimental assays that define them by their relevance to human phenotypes. Second, it enhances statistical power by aggregating rare variants across functionally connected genes, leading to improved mechanistic insights into the directional effects of genes on a phenotype. Third, it moves beyond broad and overlapping gene classifications, such as gene ontology, by utilizing experimentally derived relationships with defined edge geometries and signed weights, thereby increasing the biological specificity of the burden test. Importantly, fourth, it allows for inaccuracies (e.g., false positive "nodes") in that network. This framework paves the way for integrative research strategies whereby genetic screens identify network-based hypotheses to test for association with human phenotypes. Networks demonstrating significant associations can subsequently be refined and validated through targeted

experimental approaches, bridging the gap between genetic discovery and functional understanding.

We demonstrate two ways of applying NERINE: first, using pre-existing pathway modules to analyze prevalent diseases—breast cancer (BRCA), type II diabetes (T2D), coronary artery disease (CAD), and myocardial infarction (MI) in UK and MGB biobanks; second, leveraging experimentally derived gene networks from model system screens to uncover rare variant associations in a complex neurodegenerative disease like Parkinson's disease (PD), where existing gene-level rare variant association tests have been thoroughly underpowered with existing cohorts[26,27].

Comprehensive testing of database pathway modules in large-scale biobanks with NERINE has enabled us to identify several novel associations, including in the estrogen receptor pathway in BRCA, an adipogenesis-related network in T2D, and non-lipid gene networks (e.g., complement system) in CAD and MI. In PD, NERINE tests network hypotheses generated from experimental screens targeting two core pathologies of PD: i) alpha-synuclein (αS) proteotoxicity and ii) human dopaminergic (DA) neuron survival. We uncover associations in gene modules linked to autophagy regulation (*HMGB1-OPTN-USP10* module) and vesicle trafficking and protein homeostasis (*LRRK2-PRL-SNCA* module)—findings that all have eluded single-gene tests.

One novel finding by NERINE—an association of rare damaging missense variants at the *PRL* locus encoding *prolactin* (*Prl*) with PD risk—converges with an independent genome-scale functional screen in a CRISPRi-induced synucleinopathy cortical neuron (CiS-CN) model designed for studying physical and genetic modifiers of *α-synuclein* (αS) toxicity in cortical neurons[28]. Subsequent functional validation using the CiS-CN model and a well-established mouse PFF model[29] uncovers a hitherto unknown intraneuronal role of *PRL* in *α-synuclein* stress response in PD. This underscores the transformative potential of experimental screens and human statistical genetics as mutually reinforcing methodologies to reveal novel disease mechanisms.

Results

Modeling rare variant burden in gene networks incorporating edge geometry

We present NERINE, a new statistical framework for assessing the cumulative effect of rare variants in a gene network on a dichotomous phenotype (**Figure 1**). NERINE uniquely incorporates information on network vertices (genes) and network edges (interactions) into a parametric model and is robust with respect to unimportant genes in the network (see **Supplementary Table T1** for distinction from other methods). Many data types can

represent interactions between genes and gene products (i.e., transcripts and proteins). Interactions may differ in terms of relationship types (from protein complexes to sets of co-expressed genes to protein-DNA or protein-RNA interactions) and scale (from extensive protein interaction networks to pathways of just a dozen genes). Regardless of the data source, we encode gene-gene relationships using a positive semidefinite matrix, Σ.

We model gene effect sizes, denoted by vector $\vec{\alpha}$, as draws from a multivariate skew-normal distribution $\vec{\alpha} \sim MSN(\vec{\alpha}\,|0, \theta \cdot \Sigma, \nu)$, with univariate skew-normal marginals. This hierarchical model captures the biological expectation that functionally related genes exhibit correlated effect sizes (when non-zero) and correlated probabilities of having no phenotypic effect. The parameter θ reflects the importance of the gene group and is the object of inference, while the skewness parameter, $\nu = f(N_{case}, N_{control})$ encodes the case–control imbalance (**Figure 1**, **Methods**). Here, N_{case} and $N_{control}$ represent the number of cases and controls, respectively. When case-control groups are perfectly balanced, this model reduces to a multivariate normal, $\vec{\alpha} \sim MVN(\vec{\alpha}\,|0, \theta \cdot \Sigma)$, with scalar normal marginals. This parameterization allows gene effect sizes to vary in both magnitude and direction while permitting some to be zero (in case of noisy networks).

NERINE infers θ using the maximum likelihood estimation (MLE) framework, where the likelihood is given by,

$$L(\theta|\boldsymbol{X}, \mathbf{Y}, \vec{\alpha}, \Sigma, N_{case}, N_{control}) = \int \left(\prod_i P(X_i|X_i + Y_i, \alpha_i) \right) P(\vec{\alpha}|\theta; \Sigma, \nu) d\vec{\alpha}$$

We approximate L as a weighted sum over multivariate quadrature points from the domain of integration achieved through a lookup table with pruning (**Methods, Supplementary Note**). To ensure computationally efficient evaluation of the quadrature points, we make several design choices. First, we model rare variant counts in member genes in cases ($\boldsymbol{X}$) and controls ($\boldsymbol{Y}$) as independent Poisson variables, with rates determined by population allele frequencies renormalized between cases and controls by $\vec{\alpha}$. This implies that the conditional probability of observed rare variant count in each gene in cases, given the total rare variant count in the cohort, follows a Binomial distribution (**Supplementary Note**). Next, we apply a custom transformation to map the unbounded univariate skew-normal marginals of $\vec{\alpha}$ to Beta distributions on [0, 1] (**Supplementary Note**), so that the transformed α_is can be used as Binomial success probabilities. We denote the transformed gene-effects as $\overrightarrow{\alpha'}$, which incorporates skewness (ν) by controlling the shape parameters of the Beta distribution (**Supplementary Note**). When the network has no effect on the trait, the expected case–control allele ratio reflects sample size proportions in each gene. When the

network influences the trait, allele count ratios deviate from this null expectation in relevant genes. The approximate likelihood is given by,

$$L\left(\theta|\boldsymbol{X},\mathbf{Y},\overrightarrow{\alpha'},\Sigma,N_{case},N_{control}\right)\approx\sum_{\overrightarrow{\alpha'}}\left(\prod_{i}\left(P_{binom}(X_i|X_i+Y_i,\alpha'_i)\right)P\left(\overrightarrow{\alpha'}|\theta;\Sigma,\nu\right)\right)$$

NERINE performs nested hypotheses testing, the test statistic being the log-likelihood ratio (LLR). We test the hypothesis $\theta = 0$ vs. $\theta > 0$ using the maximum likelihood estimate of θ (**Methods**). Since θ lies on the boundary of the parameter space, the test statistic, LLR, follows an asymptotic distribution arising from a weighted mixture of a point mass at zero and a chi-square distribution with one degree of freedom[30,31]. We confirmed this asymptotic behavior using extensive simulations with four well-studied canonical pathways (**Supplementary Figure S1** and **S2**), and artificially constructed topologies (**Supplementary Figure S3**) under the null model $(\theta = 0)$. NERINE provides asymptotic p-values for significant networks $(\theta > 0)$, and estimates the maximum likelihood effects of member genes under the inferred θ. However, gene-level p-values are not provided (**Methods**).

We benchmarked the performance of NERINE against existing rare variant tests in simulations under the alternative hypothesis $(\theta > 0)$ (**Figure 2**, **Supplementary Figures S4 and S5**, **Methods**). Compared with pathway-level adaptations of such tests, including CMC-Fisher, Fisher minimum p-value, Fisher combined test, SKAT-O, and pathway-based trend test, RVTT (only trait-increasing genes case), NERINE consistently showed higher empirical power, especially in noisy networks (**Figure 2, Supplementary Figures S4 and S5**). Positive control experiments using UK Biobank (UKBB) data on LDL and HDL cholesterol (LDL-C and HDL-C) confirmed NERINE's ability to detect significant rare variant burdens in key lipid-related pathways in European ancestry-specific (**Supplementary Figures S6, S7, and S8**, **Methods**) as well as pan-ancestry (**Supplementary Figure S9**, **Supplementary Table T2**, **Methods**) analyses. Compared to SKAT-O, NERINE provided lower p-values, demonstrating its ability to leverage network connectivity for improved power (**Supplementary Figure S6C**). NERINE's estimated effect sizes and directions for individual genes, correctly identified *PCSK9* and *APOB* as reducing LDL cholesterol and *LDLR* as increasing it, consistent with known biology[32] (**Supplementary Figure S6D**). Similarly, for HDL cholesterol, it linked *ABCA1*, *LCAT*, and *APOA1* to low levels and *CETP*, *LIPC*, *LIPG*, and *SCARB1* to high levels. (**Supplementary Figure S6E**).

Selecting the most informative network topology with NERINE

NERINE selects the most informative network topology when multiple alternatives exist for a given gene set, underscoring its ability to leverage the edge geometry of gene networks for

more powerful inference. We demonstrated this feature in simulations with (i) four well-characterized pathways of varying sizes (NOTCH, WNT, protein export, and EGFR signaling) for a simulated binary phenotype, and (ii) significant lipid-related pathways on LDL-C and HDL-C phenotypes in UKBB. For the former, NERINE consistently achieved a higher PPV at different p-value cutoffs with "true" networks compared to randomized counterparts, especially in noisy scenarios (**Supplementary Figure S10, Methods**). Similarly, in the latter analysis, NERINE achieved higher log-likelihood ratios and lower p-values with real network topologies compared to random ones (**Supplementary Figure S11, Methods**).

When gene-gene relationships in a network are available from multiple data sources, NERINE selects the optimal topology for estimating the burden of rare variants in a network, as showcased using lipid-related pathway gene networks for LDL-C and HDL-C in UKBB (**Supplementary Table T3**; Methods). For lipid-related pathway gene sets, we applied NERINE on network topologies constructed from three data sources --- (i) high confidence experimentally derived physical and genetic interactions of proteins from canonical databases, (ii) gene-gene co-expression in liver tissue in GTEx (v8), and (iii) co-essentiality of genes in liver cell lines in DepMap3 (v2023Q2) (**Methods**). Our results revealed that the data source providing optimal topology was highly dependent on the specific gene set under investigation (**Supplementary Table T3**). **Figure 2A** shows an example scenario where one data source provided more relevant topological information than the others. When comparing individuals with low HDL-C to those with high HDL-C in UKBB, liver tissue co-expression emerged as the most informative source for the core module of HDL-cholesterol-related genes (Bonf. p-value = 1.34e-73). Example scenarios where data sources, other than co-expression, captured the edge relationships of specific gene sets more accurately are shown in **Supplementary Figure S12**. For instance, physical and genetic interactions best described the VLDLR internalization and degradation pathway module associated with the LDL-cholesterol phenotype in UKBB (Bonf. p-value = 2.50e-31). Meanwhile, co-essentiality in liver cell lines provided the most relevant edge geometry for the pathway of LDL, HDL, and triglyceride metabolism (Bonf. p-value = 1.58e-46). These findings underscore a crucial insight: different data sources encode distinct and complementary aspects of gene relationships. NERINE could effectively exploit this phenomenon to select the most informative data source.

Applying NERINE across tissue-specific co-expression networks spanning all 52 tissue types from GTEx (v8) further reinforced its ability to identify the most informative biological context. Specifically, liver tissue emerged as the most relevant context for core HDL-cholesterol-related genes, where NERINE a high estimated network effect of $\theta = 0.8$ with the strongest statistical support (Bonf. p-value = 1.34e-73; **Figure 3B**, **Supplementary Table T4**). This

result highlights a key advantage of NERINE—its capacity to leverage the variability observed in edge relationships and their strengths across different tissue contexts for the same set of input genes, thereby selecting the optimal tissue context with biological interpretability.

Identifying rare variant associations in common diseases across canonical pathway networks

NERINE enables comprehensive rare variant burden testing in gene networks for common diseases using biobank-scale data from the UK Biobank (UKBB) and the MassGeneral Brigham Biobank (MGBBB) (**Figure 4A**, **Methods**). Applied to breast cancer (BRCA), type II diabetes (T2D), coronary artery disease (CAD), and early-onset myocardial infarction (MI)—which impact 3-13% of the population—NERINE identified significant rare loss-of-function (LoF) burdens in seven pathways for BRCA, one for T2D, two for CAD, and fourteen for MI (**Figure 4B**, **Supplementary Table T5, Supplementary Figures S13–S15**). Other than LoF variants, significant burdens were also detected in the damaging and damaging missense categories (**Supplementary Table T5**), with no enrichment of neutral or synonymous variants. Notably, networks with significant burden in the BRCA phenotype overlapped with biologically plausible pathways, such as cancer susceptibility (i.e., DNA repair, cell cycle regulation, apoptosis) and hormone signaling, while cardiovascular disease networks highlighted lipid metabolism, immune response, blood coagulation, and apoptosis (**Supplementary Figures S13–S15, Supplementary Tables T6 and T7**), indicating true-positive findings. NERINE-predicted directional effects of genes in significant pathways per disease phenotype are shown in **Supplementary Tables T8-T11.**

Unlike single-gene tests, which had previously identified only a handful of rare variant associations in these diseases (Genebass[33]— BRCA: 6, T2D: 3, CAD: 1, and MI: 1), NERINE captured aggregated effects of rare variants across gene networks, estimating directional contributions of member genes to putative disease pathways. For instance, in the T2D phenotype, we detected a significant burden of rare damaging variants (i.e., LoF and predicted damaging missense) in the adipogenesis pathway (i.e., BIOCARTA VOBESITY PATHWAY, Avg. $\hat{\theta}$ = 0.18, Bonf. p = 2.33E-03) (**Figure 4C**, **Supplementary Tables T5 and T9**), highlighting the potential roles of *RXRA*, *PPARG*, *ADIPOQ*, *TNF*, *NR3C1*, *LPL*, *RETN*, and *HSD11B1*—genes largely unrecognized in previous rare variant studies[19,33] (**Figure 4C**, **Supplementary Table T9**). Only *PPARG* and *LPL* were implicated in common variant GWAS studies[34-36]. Sensitivity analysis showed that while the inclusion of *LPL* was critical for the database-wide significant signal in this network, the inclusion of *PPARG* was not (**Supplementary Figure S16**). Five of the genes had been studied as drug targets before[37-43] (**Figure 4C**). Notably, NERINE's findings align with the fact that the deficiency of *adiponectin*—a protein encoded by *ADIPOQ*—increase insulin resistance[39] and the risk of

T2D, suggesting *ADIPOQ* as a potential candidate for therapeutic investigation. Additionally, NERINE's prediction of trait-decreasing roles of damaging variants in *RXRA* and *HSD11B1* supports their inhibition as possible therapeutic strategies for T2D[40-43], paralleling the example of *PCSK9* in cardiovascular disease where protective LoF and damaging missense variants guided effective therapy development[44,45]. However, further experiments are required to distinguish the roles of predicted damaging missenses from LoF variants contributing towards the observed trait-decreasing effect. While NERINE identified rare damaging variant burden in *TNF*, *NR3C1*, and *RETN* in the adipogenesis network, their mechanistic roles in T2D remain understudied, highlighting NERINE's ability to suggest potential gene-disease associations for further investigation.

Examples of new network-level associations in breast cancer and cardiovascular diseases

NERINE not only established rare variant connections between new gene networks and phenotypes but also highlighted how rare functional variants, aggregated at the gene level, could influence the phenotype within the network context. For instance, while the estrogen receptor pathway has been clinically studied for breast cancer, rare variant evidence for member genes beyond *BRCA1* has been limited[46,47]. NERINE identified a significant LoF burden at the network level, predicting how putative LoF variants in member genes might have exerted either trait-increasing or trait-decreasing effects on the phenotype. As expected, rare LoF variants in *BRCA1* were associated with an increased risk of BRCA (**Figure 5A: left**; Avg. $\hat{\theta}$ = 0.33, Bonf. p = 1.04e-04). However, after removing the effect of *BRCA1* LoF variants, the network still showed a nominally significant association with BRCA in UKBB, suggesting albeit small but non-zero contributions of other member genes (**Supplementary Figure S17**).

Among the estrogen receptor pathway members, 17 were classified as neither tumor suppressor genes (TSGs) nor oncogenes (**Figure 5A: right**). Within the context of this network, NERINE newly linked rare LoF variants in putative GWAS genes, *ESR1* and *CCND1*, to increased BRCA risk. Additionally, it shed light on the effects of loss-of-function of genes, which had unknown oncogenic roles and no genetic evidence. LoF variants in *TBP*, *GRIP1*, *GTF2A1*, *HDAC1*, *MED1*, *MEF2C*, and *PHB2* showed trait-increasing effects, while *HDAC5*, *NRIP1*, and *SRA1* loss-of-function had trait-decreasing effects (**Supplementary Table T8**). Indeed, *PHB2* (*prohibitin 2*) has recently emerged as a key biomarker and therapeutic target in BRCA because of its potential as a tumor suppressor through several possible mechanisms, including inhibition of estrogen receptor-alpha (*ERα*), interaction with *BIG3* (*brefeldin A-inhibited guanine nucleotide-exchange protein 3*), and promotion of *PIG3* (*p53-inducible gene 3*) transcription[48-50]. Furthermore, NERINE's prediction of *HDAC5* loss-of-

function variants decreasing the risk of BRCA aligns with the fact that the inhibition of *HDAC5* induces intrinsic apoptosis in human breast cancer cells, thereby exerting an anti-neoplastic effect[51,52], providing rationale for follow-up investigations for therapeutic potential.

Rare variant studies in cardiovascular diseases had been underpowered to uncover novel mechanisms beyond lipid-related pathways despite sequencing of over 100,000 subjects in multiple biobanks[33,53]. Here, we identified a significant enrichment for rare variants in several non-lipid pathway gene networks (**Figure 4B**, **Supplementary Figures S14 and S15**) whose member genes had not been identified in single-gene burden analyses[33]. One such pathway of note in MI is the intrinsic prothrombin activation and coagulation cascade (i.e., BIOCARTA AMI PATHWAY), with a significant LoF variant burden (Avg. $\hat{\theta}$ = 0.18, Bonf. p = 3.50e-05). This module consists of several genes associated with collagen and vascular basement membrane integrity (**Figure 5B: left, Supplementary Table T11**). Among the members, extracellular matrix proteins (i.e. *COL4A1, COL4A2, COL4A4*), and clotting factor (*PLG*) had previously been linked to myocardial infarction by common variant GWAS (**Figure 5B: right**) and in vivo studies in knockout mice[5,54-57]. NERINE provided more direct rare variant evidence for this gene module in MI.

Interestingly, NERINE identified an enrichment of rare LoF variants in the alternative complement pathway for both CAD (Avg. $\hat{\theta}$ = 0.03, Bonf. p = 1.94E-02) and MI (Avg. $\hat{\theta}$ = 0.30, Bonf. p = 1.76e-10), a connection not previously established through human genetics (**Figure 5C**). Functional studies in murine models linked complement components, *C3* and *C5*, to inflammation, endothelial dysfunction, and atherosclerosis[58-60]. In the network, we observed trait-increasing effects for *C3* LoF variants in both CAD and MI and *C5* LoFs in MI. Additionally, rare LoF variants in complement factors *C8A*, *CFB*, and *CFD* had trait-increasing effects on both conditions (**Supplementary Tables T10 and T11**). These directionalities can help reconcile the growing experimental evidence demonstrating that complement activation products in human atherosclerotic plaques can exhibit both protective and pro-atherogenic properties[60]. While the complement signal in cardiovascular diseases is intriguing, the observed differences between MGBBB and UKBB results might be because MGBBB participants had a higher frequency of cardiac events, with thrombosis and inflammation, compared to those in UKBB.

Bridging genetics and experimental biology with NERINE in Parkinson's disease

NERINE provides a robust framework for integrating insights from human genetics, cell systems, and model organisms simultaneously to facilitate the study of complex diseases. We illustrate this with Parkinson's disease (PD), a complex age-related neurodegenerative

disorder for which there are no disease-modifying therapies, but for which there is a growing number of genomic and biologic datasets[2,16,28,61-63] ideally suited for testing with NERINE. While GWAS have identified hundreds of PD-associated loci[64,65] and linkage analyses[66-68] have uncovered Mendelian rare variants in key genes—including *SNCA* (encoding αS itself), *GBA1*, and *LRRK2*—existing rare-variant association tests struggle to identify additional causal loci with replication across multiple independent cohorts[26,27]. For idiopathic PD, which accounts for ~80-85% of cases[69], the role of rare variants remained poorly understood. Traditional pathway-based analyses[70] have implicated broad PD mechanisms, including vesicle trafficking, G-protein–coupled receptor (GPCR) signaling, and immune pathways. However, their limited resolution and statistical power, due to aggregation of signal across many genes, prevent researchers from determining individual gene contributions or how genes interact to influence disease processes. Thus, such aggregated pathway- or ontology-level signal are less informative for generating experimentally testable hypothesis in specific cellular contexts.

It is possible to assemble the same canonical gene sets into molecular networks with well-defined edge geometries and apply NERINE to pinpoint specific genes and modules for targeted experimentation. We demonstrate this by applying NERINE to gene ontology (GO) biological process modules significantly enriched in PD GWAS-associated genes[64,65,71] (**Methods**, **Supplementary Figure S18A**, **Supplementary Table T12**). Using database edge-based topologies, NERINE identified a significant burden of rare LoF variants in the *peptidyl-threonine modification* co-essentiality module in two independent PD case-control cohorts from UKBB and AMP-PD (Avg. $\hat{\theta}$ = 0.9, Bonf. p-value = 4.34e-2; **Supplementary Figure 18B**, **Supplementary Table T13**). Within this module, NERINE suggested trait-increasing LoF burdens in *MCCC1* and *DYRK1A*, and trait-decreasing LoF effects in several other genes, including *FYN*, *USP8*, and *LRRK2* (**Supplementary Table T14**; see **Supplementary Note** for detailed discussion).

Experimentally derived networks to capture core PD pathologies

As a complementary approach, we turned to biologically informed networks constructed from large-scale genetic data in cellular and model-organism studies. These networks represent experimentally validated assemblies of genes with defined functional relationships, offering a degree of biological specificity not captured by canonical pathways or ontologies. NERINE can pinpoint gene- and module-level effect sizes within these networks, enabling targeted follow-up experiments. Specifically, we analyzed networks comprising genes that modulate *α-synuclein* toxicity and mistrafficking or support dopaminergic (DA) neuron survival, representing the two pathological hallmarks of Parkinson's disease (PD): *α-synuclein* aggregation in Lewy bodies and DA neuron loss[72,73]

(**Figure 6A**; **Methods**). We assessed these networks in independent PD case-control cohorts from UKBB and AMP-PD.

Rare variant burden in DA neuron essentiality genes involved in autophagy regulation

Using a genome-scale CRISPR screen of 19,993 genes (**Figure 6B; top**), we identified 693 essential genes for DA neuron survival with significant enrichment in ten GO modules, covering a wide range of biological processes, including autophagy regulation, mRNA processing, apoptosis, and cellular response to DNA damage (**Supplementary Table T15**, **Methods**). We generated networks from these ontological modules by grouping semantically similar GO biological process terms enriched in essentiality genes and then extracting the edge relationships of genes within each group from PPI, co-expression, and co-essentiality databases (**Methods**). NERINE identified a significant rare damaging variant burden the gene module linked to the *regulation of autophagy* (Avg. $\hat{\theta}$ = 0.03, Bonf. p-value: 1.86e-2; **Figure 6C**, **Supplementary Figure S19**, **Supplementary Tables T16 and T17**) in sporadic PD cases compared to controls in the AMP-PD sporadic and UKBB extreme (median age of controls at recruitment >= median age of cases) cohorts (**Methods**). NERINE predicted trait-increasing effects for damaging missense and damaging variants in key genes such as *HMGB1* and *OPTN* in PD (**Figure 6C**, **Supplementary Figure S19**, **Supplementary Tables T17**). The impairment of *HMGB1* function may inhibit autophagy and promote αS accumulation[74], while the disruption of *OPTN* function may result in the improper clearance of damaged mitochondria (or mitophagy), leading to neurodegeneration[75]. NERINE also identified a trait-increasing burden of rare damaging variants in *USP10* (**Supplementary Figure S19**, **Supplementary Table T17**) whose inactivation may disrupt αS-containing aggresome formation[76], exacerbating toxicity. These findings readily suggest a small number of potential candidate genes and testable hypotheses supported by orthogonal human genetic evidence, justifying in-depth future experimentation.

Rare variant burden in a LRRK2- and SNCA-containing αS proteotoxicity module related to vesicle trafficking and protein homeostasis

Next, we applied NERINE on an αS proteotoxicity network assembled through "yeast-to-neuron" discovery screens[2] targeting the "αS accumulation in Lewy bodies" feature of PD (**Figure 6B: bottom**). The αS network was generated using the TransposeNet methodology that connected hits from multiple genome-scale deletion and over-expression screens against αS proteotoxicity in yeast into an optimal topology by transposing yeast genes into the human interactome space and adding "predicted" genetic nodes as needed. TransposeNet introduced numerous human genes, without any known homologs in yeast, including *SNCA* (that encodes αS itself) and *LRRK2*, two genes that are among the most definitively connected directly to PD risk in GWAS and Mendelian linkage analyses[64-68].

Importantly, the network, validated in human iPSC cortical and DA neurons[2,16], converged with a proteome-scale proximity labeling screen for αS in neurons[63]. It spanned many relevant pathways, including vesicle trafficking, mRNA metabolism and translation, mitophagy, oxidative metabolism, calcium/NFAT signaling, Toll-like receptor signaling, and purine metabolism (**Supplementary Tables T18**).

Among the TransposeNet αS network branches (**Supplementary Figure S20**), we observed a screen-wide significant burden of rare damaging missense variants in the *LRRK2*- and *SNCA*-containing *vesicle trafficking and protein homeostasis*-related gene module associated with PD risk in AMP-PD and UKBB (Avg. $\hat{\theta}$ = 0.53, Bonf. p-value: 2.17e-2; **Figure 6D**, **Supplementary Figure S21**, **Supplementary Tables T18 and T19**). None of the member genes, except *LRRK2* showed an exome-wide significant rare-variant burden in previous single-gene analysis[26,33], and even this association could not be replicated in multiple cohorts. The network statistically supported by NERINE showed trait-increasing effects of rare damaging missenses in several *SNCA*-interactors in the stem, including *VDAC1*, *NEDD4*, and *TOR1A*, which were previously linked to PD pathology only in experimental models[61,62,77-81]. *PRL* also showed a strong trait-increasing effect, but it was not previously causally linked to PD. Beyond direct interactors of *SNCA* in the module, NERINE detected non-zero effects in genes such as *NDFIP1*, *PBX1*, and *RNF11* with PD risk, which had been linked to PD pathobiology only in functional studies[82-85] but not through rare variant studies.

Convergence of an unbiased functional genomics screen in CRISPRi-induced synucleinopathy model with NERINE on *PRL-SNCA* interaction

NERINE pinpointed genetic signals within the humanized TransposeNet αS proteinopathy network, which was originally derived from yeast-based screens[2] and therefore lacked the specific human cellular context for key network interactions. Brain-related disease like PD involves a plethora of cell types and peripheral systems that can influence brain function, offering many possibilities for follow-up experimentation. To focus on neuronal mechanisms, we sought convergence between NERINE's findings and an independent genome-scale functional screen in a CRISPR interference (CRISPRi) induced synucleinopathy cortical neuron (CiS-CN) model[28] designed for studying genetic determinants of αS toxicity (**Methods**). Briefly, this model was established in the WTC11 iPSC genetic background[28], and enabled—i) one-step trans-differentiation (AAVS1-Ngn2) into cortical glutamatergic neurons that are vulnerable in Lewy body diseases, including PD; ii) flexible gene knockdown via CRISPRi (dCas9-KRAB)[86]; iii) lentivirally-driven clonal αS expression at brain-like levels. The CiS-CN model, with a SNCA-overexpression clone (*SNCA*-high*)* and a control clone (*SNCA*-endo), thus provided a tractable but physiologically relevant system for studying αS aggregation and toxicity in cortical neurons[28,87].

CRISPRi was utilized to target physical and genetic interactors of αS in the CiS-CN model, including all genes from the TransposeNet αS proteinopathy network, using a custom sgRNA library (**Methods**). The full screen will be described in a forthcoming publication. Briefly, we assessed sgRNA representation at start, DIV28, and DIV42, and looked for the dropout of sgRNAs in three comparisons: (1) *SNCA*-high vs. *SNCA*-endo at DIV42, (2) DIV42 vs. DIV28 in *SNCA*-endo, and (3) DIV42 vs. DIV28 in *SNCA*-high (**Figures 7A,B**). Genes showing significant sgRNA dropout in MAGeCK-iNC analysis (FDR < 0.1) were classified as αS toxicity enhancers (**Methods**).

In these comparisons, *PRL* emerged as a particularly intriguing finding. NERINE had already pinpointed *PRL* within the LRRK2– and SNCA-containing branch of the same TransposeNet network (see above; **Figure 6D**), where damaging missense variants showed a strong trait-increasing effect. In the CiS-CN CRISPRi screen, sgRNAs targeting *PRL* significantly dropped out in *SNCA*-high versus *SNCA*-endo at DIV42 (Comparison 1) and from DIV28 to DIV42 in *SNCA*-high clones (Comparison 3), aligning with the NERINE result. Thus, in light of these convergent results, *PRL*, which previously lacked direct evidence of a causal role in PD, was selected as a key candidate for further experimental validation.

A novel intraneuronal *SNCA-PRL* stress response identified in human neurons

While *Prl* is predominantly expressed in neuroendocrine tissue, its unexpected identification in our neuronal functional genomics screen suggests a highly unexpected cell autonomous role in neurodegeneration. Previous studies, though conflicting, have suggested *PRL* expression in rodent neurons outside the pituitary under stress conditions[88], with speculation, albeit based on a few studies, that its expression may only be detected under conditions of neuronal stress. There is also evidence that its protein levels may be post-translationally regulated despite low mRNA expression[89]. Overall, the strong literature consensus is that any effects of *Prl* are likely to be through exogenous release of *Prl* from the pituitary gland.

To investigate the possibility that *Prl* may be part of an intraneuronal stress response directly related to αS, we performed immunostaining in neurons of different age. *Prl* expression was significantly increased in *SNCA-overexpression* neurons, with ~7-fold-higher expression in *SNCA-high*, and 2.5-fold-higher in *SNCA-intermediate*. than in *SNCA-endo* neurons at DIV7 (**Figure 7C**). This elevation was abrogated, as expected, by our sgRNAs directed to *PRL* (**Figure 7D**). Importantly, *Prl* levels were reduced over time, such that by DIV28, *Prl* levels in the *SNCA*-high lines, were only ~1.5-fold-higher than in *SNCA-endo* neurons (**Figure 7E**).

These data collectively suggest a potentially protective role of αS-induced *Prl* that diminishes over time.

We next tested whether inducing αS aggregation, thereby reducing soluble αS levels[87], in cellular and animal models via pre-formed αS fibrils (PFFs)[29,90] led to reduced *Prl* levels. Our CiS-CN neurons challenged with seven (7) days of exposure to PFFs showed a reduction of *Prl* levels as indicated by immunofluorescence (**Figure 7F**). We further considered whether this was transcriptional or post-transcriptional. mRNA expression levels of *PRL* were low in our models, possibly related to the immaturity of the iPSC-derived neurons. We thus turned to a well-established mouse PFF model[29]. We unilaterally injected PBS versus 5 μg of PFFs (2.5 μL volume) into the dorsal striatum of mice (**Figure 7G: left**, **Methods**). Mice were aged for 30 days post-injection. By this stage, as has been previously described[91], the αS aggregation pathology "spreads" distally and reaches the cortex and the amygdala, two brain regions highly susceptible to αS pathologies in later stage PD. We harvested mRNA from the amygdala and assessed gene expression with the Nanostring Neuropathology panel. In comparison to PBS-injected control animals which showed no pS129 positive inclusions across all brain regions (not shown here), PFF-injected mice showed a 16.6-fold downregulation of *Prl* mRNA expression in the amygdala region ipsilateral to the site of injection (n=5 per group; 3 males/2 females; **Figure 7G: right**), consistent and indeed stronger than our findings in the shorter-term human CiS neuronal model. At least in the intact rodent model, our data indicate that a component of this effect is transcriptional.

Our data indicate that in the context of aging and fibrillar αS pathologies, *Prl* levels drop intraneuronally, and potentially could sensitize the neurons to αS-induced cytotoxicity. To determine whether *Prl* directly protects against αS toxicity, we assessed its neuroprotective effects against oxidative stress in our CiS models. By DIV28, *SNCA-high* neurons exhibited heightened sensitivity to menadione-induced oxidative stress[92], measured with the fluorogenic probe CellRox (Thermo) (**Figure 7H**). In *SNCA-high* and *SNCA-endo* models, we pre-conditioned DIV6 CiS neurons with *prolactin* for 24hrs before treatment with menadione (**Figure 7I: top**). Importantly, exogenous *prolactin* treatment significantly decreased oxidative stress in CiS neurons (**Figure 7I: bottom**), strongly suggesting a neuroprotective effect against αS-induced toxicity. Our data collectively indicate that aged neurons in which αS aggregates leads to reduced Prl levels and increased vulnerability to oxidative stress. Taken together, the convergence of our forward genetics iPSC screen with NERINE uncovered an unexpected αS-*Prl* intraneuronal stress response, suggesting that *Prl* may play a crucial role in neuronal resilience to αS pathology in PD. These data underscore the utility of NERINE in uncovering novel biological insights into complex disease processes.

Discussion

NERINE is a powerful statistical test designed to bridge the gap between human genetics and experimental biology, adding a crucial mechanistic understanding to rare variant burden analysis. Unlike traditional rare variant tests, which focus on isolated genes or regions, NERINE contextualizes rare variant effects within biological networks, thereby enhancing our ability to interpret genetic associations. By incorporating network topologies, we provide a more structured framework for understanding how rare variants contribute to disease mechanisms.

One of NERINE's key advantages is its ability to select between competing network topologies for the same gene set—an ability no other method currently offers. By utilizing edge geometry, NERINE gains a statistical power boost, a feature demonstrated through extensive simulations and real-world data analyses, even in cases where network structures are poorly defined. Despite its strengths, NERINE has some limitations, particularly in cases when the networks are too large, or sample sizes are exceedingly small (see **Methods**). However, its ability to integrate human genetic data with functional insights from model system experiments makes it a valuable tool for hypothesis testing in disease genetics.

To showcase NERINE's utility, we applied it in two distinct contexts: first, to test network hypotheses from publicly available databases across four prevalent diseases—BRCA, T2D, CAD, and MI. NERINE provided novel rare variant signals for biologically plausible disease mechanisms, such as collagens and inflammatory response genes in cardiovascular diseases, adipogenesis pathway in T2D, and the network of estrogen receptor regulation in BRCA. The implication of collagens, key structural proteins in the extracellular matrix, in MI is consistent not only with their primary role in maintaining the stability of atherosclerotic plaques[93] but also with their indirect role in coagulation via platelet effects and interactions with clotting factors[94]. Furthermore, the significant rare variant burden in networks linking inflammatory response genes to the risk of both CAD and MI supports the potential of anti-inflammatory therapies as an alternative to LDL-lowering drugs for treating cardiovascular diseases[95]. Second, we applied NERINE to bespoke gene networks derived from experimental screens in Parkinson's disease (PD), a complex disorder where traditional rare-variant analyses have struggled due to limited sample sizes. Here, NERINE revealed a significant rare variant burden in gene modules related to autophagy regulation, vesicle trafficking, and protein homeostasis—signals largely undetected by single-gene tests.

Our findings further highlighted two major contributions: first, we provided rare variant evidence supporting hits from previous functional studies in model systems. For example, several member genes of the *LRRK2- and SNCA*-containing *vesicle trafficking and protein*

homeostasis subnetwork (e.g., *NEDD4*, *NDFIP1*, *VDAC1, PBX1, TOR1A*, and *RNF11*) which have been previously implicated in PD only in functional studies[61,62,77-85,96-98], are now corroborated with human genetics evidence by NERINE. Second, NERINE helped resolve conflicting hypotheses by providing genetic support for specific disease mechanisms. For instance, there is biological data suggesting haploinsufficiency of *DYRK1A*—a member of the PD GWAS gene module involved in *peptidyl-threonine modification* identified by NERINE—reduces its kinase activity and exacerbates DA neuron degeneration in mice[99]. There is also data arguing the opposite through phosphorylation of αS[100]. By providing genetic evidence in favor of the latter direction, we offered some resolution of the matter.

Although NERINE does not provide gene-level p-values within a network, it yields informative effect sizes and directions for member genes that often align with known biology, as demonstrated in our positive control experiments with lipid phenotypes in the UK Biobank. Importantly, NERINE captures both trait-increasing and trait-decreasing rare variant burdens at the gene level, derived from its joint modeling of network interactions and observed allele counts at each locus. While the latter may reflect gain-of-function effects requiring experimental validation, they can be of considerable therapeutic significance. A well-established precedent is *PCSK9*, where rare LoF variants are associated with reduced LDL cholesterol levels; this discovery directly enabled the development of PCSK9 inhibitors as an effective therapeutic strategy for hypercholesterolemia with significantly improved cardiac outcomes[44,45]. In this context, the trait-decreasing effects predicted by NERINE within disease-relevant networks may highlight genes of potential therapeutic interest, supporting future experimentation.

NERINE's integrative framework for unifying insights from model system screens with human genetics complements traditional pathway-based investigations in PD[70,101]. Traditional investigations rely on pathway-based polygenic risk scores from common variants or on rare variant aggregation tests, such as SKAT-O, applied at the pathway level[70,101] to implicate broad mechanisms, including endolysosomal trafficking, GPCR signaling, neuronal transmission, and immune response[70,101]. While informative, such approaches generally lack statistical power, and resolution to pinpoint contributions of individual genes and their specific interactions. A network-based approach like NERINE, which incorporates experimentally defined gene relationships, provides additional biological specificity, improves power to detect associations, and helps guide mechanistic hypothesis generation for follow-up studies.

Among NERINE's findings in PD, the most intriguing one is related to *PRL*, supported by converging evidence from NERINE and our functional CRISPRi cellular screens.

PRL encodes the protein *Prl*, a hormone secreted by the pituitary gland with no known connection to αS. The original TransposeNet network[2] unexpectedly introduced *Prl* to a small network of genes that also included key known PD genes SNCA (encoding αS itself) and *LRRK2* (encoding a protein with increased enzymatic kinase activity that has been implicated in PD). Mechanistically, a connection of *Prl* to αS had been elusive until this study. This is because, although not unprecedented, the presence of *Prl* in neurons outside the pituitary has been highly speculative[88,102]. While *Prl* showed some neuroprotective effects against oxidative stress[88,103-105], this effect had been attributed exclusively to an exogenous endocrine effect through pituitary secretion. It was thus remarkable for *Prl* loss-of-function to simultaneously be implicated by both NERINE from human genetic data and in a genome-scale screen for enhancers of αS proteotoxicity: predicted deleterious missense mutations were tied to PD risk by NERINE, and *Prl* knockdown enhanced αS toxicity within neurons. Our data suggest that the effect results from an early, possibly post-translational, intraneuronal mechanism in which *Prl* is specifically elevated in response to αS overexpression. This response diminishes with time and with αS aggregation, the latter effect being at least partly transcriptional in more chronic treatment of mice with fibrillar αS. This decrease is, in turn, associated with the sensitization to *PRL* knockdown and exogenous stress, demonstrated in this study with the oxidative stressor menadione. While we could not directly assess pituitary function in individuals carrying *PRL* variants in our study cohorts, literature evidence provides additional context. A recent proteomic study in the Harvard Biomarkers Study cohort found that *Prl* levels in the cerebrospinal fluid—likely reflecting non-pituitary sources—was the top feature in a classifier model distinguishing PD patients from controls, even after accounting for levodopa effects[106]. Thus, these converging findings from NERINE, our CRISPRi screen, mouse models, and literature strengthen the plausibility of a role for an unexpected role of *Prl* in PD that originates not in the pituitary but within neurons subjected to αS stress. Further work will now be necessary to clarify its mechanisms and to evaluate potential pleiotropic effects.

In sum, NERINE bridges the gap between human genetics and experimental biology, providing a robust framework for attributing putative disease-causing factors to experimentally derived molecular networks. Such a method is especially significant for complex diseases like PD, where the limited availability of well-phenotyped genomic data hinders traditional approaches. As large-scale molecular interaction networks become increasingly common, we anticipate that research programs where human genetics and experimental biology mutually reinforce each other—such as those enabled by NERINE—will play a growing role in identifying disease-relevant signals, resolving conflicting evidence, and uncovering druggable targets.

Methods

Overview of the NERINE methodology

NERINE models the "variant-gene-network" hierarchy as follows: it encodes gene-gene relationships in a network of *m* genes by a positive semidefinite matrix, Σ. We assume that phenotypic effects of genes within a network, represented as vector $\vec{\alpha}$, are drawn from a multivariate skew-normal distribution $\vec{\alpha} \sim MSN(0, \theta \cdot \Sigma, \nu)$. Here, θ is a parameter reflecting the cumulative effect of the gene network on a phenotype and is the object of inference (**Figure 1**). The skewness parameter $\nu = f(N_{case}, N_{control})$ adjusts for case-control imbalance. Here, $\mathrm{N_{case}}$ and $\mathrm{N_{control}}$ represent the case- and control-group sizes. For balanced cohorts (i.e., $\nu = 0$), the model reduces to $\vec{\alpha} \sim MVN(0, \theta \cdot \Sigma)$, where marginals are normally distributed with zero mean. Under this model, an edge between two genes in the network implies that they have either correlated non-zero effect sizes or correlated chances of having no phenotypic effects.

The network-effect, θ on a dichotomous phenotype is estimated using the maximum likelihood estimation (MLE) framework, where the likelihood is modeled as an integral over two components: (i) the product of conditional probabilities of observed case mutation counts given total mutation counts in network-genes, and (ii) the probability of gene effects given the network parameters:

$$L(\theta|\boldsymbol{X}, \mathbf{Y}, \vec{\alpha}, \Sigma, N_{case}, N_{control}) = \int \left(\prod_i P(X_i | X_i + Y_i, \alpha_i) \right) P(\vec{\alpha}|\theta; \Sigma, \nu) d\vec{\alpha}$$

We approximate this integral as a weighted sum over multivariate quadrature points from the domain of integration. The weight of each multivariate quadrature point is given by the product of the corresponding univariate weights. To make the computation of the integrand tractable, NERINE makes several design choices and employs a lookup table approach with pruning (**Supplementary Note**).

NERINE models rare variant counts in genes for cases and controls as two independent Poisson random variables as follows,

$$\text{Allele counts in cases in gene } i,\ X_i \sim Poisson(N_{case}, \lambda^i{}_{case})$$
$$\text{Allele counts in controls in gene } i,\ Y_i \sim Poisson(N_{control}, \lambda^i{}_{control})$$

Here, λ^i_{case} and $\lambda^i_{control}$ denote the rate parameters for case- and control- allele count distributions in gene i, and correspond to population allele frequencies renormalized between cases and controls by transformed network-gene effect, α_i. This implies that the conditional probability of observed rare variant count in each gene in cases, given the total rare variant count in the cohort, follows a Binomial distribution (**Supplementary Note**).

$$P(X_i = k|X_i + Y_i = n) = \binom{n}{k}\left(\frac{\lambda^i_{case}}{\lambda^i_{case} + \lambda^i_{control}}\right)^k \left(\frac{\lambda^i_{control}}{\lambda^i_{case} + \lambda^i_{control}}\right)^{n-k} \approx Binom(n, p)$$

, where $p = \phi(\alpha_i)$ (**Supplementary Note**).

Since each α_i represents unbounded univariate skew normal distributions, they cannot be directly used as Binomial success probabilities. Hence, we apply a custom variable transformation to map α_is to Beta distributions so that they fall on the interval [0, 1] (**Supplementary Note**). This transformation ensures that the mean gene effect is centered around $N_{case}/(N_{case} + N_{control})$, and the shape of the transformed distribution adapts with θ: small θs (close to 0) concentrate density near the mean, while large θs shift density toward the extremes. Case-control skew is incorporated using the shape parameters of the Beta distribution. The transformed gene-effects within the network are denoted $\overrightarrow{\alpha'}$ (**Supplementary Note**).

Thus, NERINE's approximate likelihood takes the form:

$$L(\theta|\boldsymbol{X}, \mathbf{Y}, \vec{\alpha}, \Sigma, N_{case}, N_{control}) \approx \sum_{\overrightarrow{\alpha'}} \left(\prod_{i=1}^{m} P(X_i|X_i + Y_i, \alpha'_i)\right) \cdot P\left(\overrightarrow{\alpha'}|\theta; \Sigma, \nu\right)$$

The conditional probability term is calculated using the probability density function of a standard Binomial distribution, which makes the likelihood computation fast and tractable. To calculate the probability of network-gene effects ($\overrightarrow{\alpha'}$) for a given network topology (Σ) and network effect (θ), we use a lookup table, where highly improbable $\overrightarrow{\alpha'}$s are pruned (**Supplementary Note**).

NERINE performs nested hypothesis testing; the null hypothesis being H_0: θ=0 (i.e., the gene network does not affect the trait and expected ratio of variant counts between cases and controls is proportional to the ratio of sample sizes in each gene) and the alternative set of hypotheses being H_z: $\theta = \theta_z$>0 (i.e., the gene network has an overall effect on the trait and the ratio of variant counts between cases and controls may deviate from the null expectation in relevant member genes). The test statistic of NERINE is the log-likelihood ratio (LLR):

$$LLR_z = 2 \times (log(L(H_z)) - log(L(H_0)))$$

We denote the maximum-likelihood estimate of network effect with $\hat{\theta} = \underset{\theta=\theta_z}{\operatorname{argmax}}(LLR_z)$. Since θ lies on the boundary of the parameter space, the test statistic asymptotically follows the distribution of a weighted mixture of a point mass at zero and a chi-square distribution with a degree of freedom of one ($dof = 1$)[30,31]. NERINE draws its asymptotic p-values from this distribution. For significant networks, NERINE calculates the maximum likelihood gene-specific effects ($\widehat{\overrightarrow{\alpha'}}$) under the estimated $\hat{\theta}$ as follows:

$$\widehat{\overrightarrow{\alpha'}} = \underset{\overrightarrow{\alpha'}}{\operatorname{argmax}}\, L_{\hat{\theta}} \approx \underset{\overrightarrow{\alpha'}}{\operatorname{argmax}} \left(\left(\prod_{i=1}^{m} P(X_i | X_i + Y_i, \alpha_i') \right) P\left(\overrightarrow{\alpha'} | \hat{\theta}; \Sigma, \nu\right) \right)$$

The search space for possible network-gene effects ($\overrightarrow{\alpha'}$) is determined by the estimated $\hat{\theta}$, the network structure (Σ), and the lookup table entries (**Supplementary Note**).

NERINE has several limitations. Currently the test is designed to accommodate dichotomous traits only. Continuous traits need to be dichotomized before NERINE can be run on them. Covariates correction is not directly included in the test. For continuous traits, it can be done as a preprocessing step using a regression-based approach and then the trait can be dichotomized. For binary traits, NERINE currently performs ancestry-stratified analysis and combines the p-values using Fisher's combined test post-hoc. For large networks with > 50 genes, NERINE's test statistic falls outside the asymptotic regime, hence we recommend using NERINE with networks up to 50 genes.

Simulations under the null model

To evaluate the performance of NERINE under the null model (θ=0), we performed extensive simulations with well-studied biological pathways, namely, NOTCH signaling (m=6), WNT signaling (m=24), protein export (m=24), and EGFR signaling (m=50), from the canonical pathways database (**Supplementary Figures S1 and S2**) and different simulated network architectures—(i) clique: complete graph with all nodes connected to each other; (ii) path: each node connected to two other nodes except the first and the last nodes; (iii) random: randomly generated scale-free graph of m nodes; and (iv) isolated genes: nodes not connected with each other (**Supplementary Figure S3**).

For the canonical pathways, gene lists were extracted from MSigDB (v7.3) and high confidence physical and genetic interactions from protein-protein interaction (PPI) databases were used as network edges between pathway genes. We simulated three different scenarios – (i) equal sized case/control groups ($N_{case} = 1{,}000;\ N_{control} = 1{,}000$) (ii) case group is larger ($N_{case} = 3{,}000;\ N_{control} = 1{,}000$), and (iii) control group is larger ($N_{case} = 1{,}000;\ N_{control} = 3{,}000$).

Under different network architectures and case-control skews, we simulated allele counts in cases and controls using independent Binomial distributions under the null model. For each gene, we assumed the presence of up to five qualifying loci each with minor allele frequency (MAF) of 0.001. The binomial probabilities for case- and control-groups are adjusted according to the group sizes and MAF of variants assuming no gene-specific effects

under the null model ($\theta = 0$). For each scenario, we performed 1,000 iterations to generate the QQ-plots.

We used the *pchibarsq* function from the *emdbook* (v1.3.13) in R (v4.3.2) to calculate the p-values from the mixture of chi-square distribution with one degree of freedom ($dof = 1$) and the delta function at zero (0). We calculated 95% bootstrap confidence intervals around NERINE's test-statistic for visualization.

For evaluating the null behavior of NERINE's test statistic in simulated networks of different sizes (m = 5, 10, and 25 genes) and different topological architectures (i.e., clique, path, random, and isolated nodes), we used equal sized case- and control-groups (**Supplementary Figure S3**). For each scenario, we performed 1,000 iterations to generate the QQ-plots. The allele counts in cases and controls were generated from independent Binomial distributions following the same procedure as above.

Performance benchmarking with simulated data

We evaluated the performance of NERINE under the alternative hypothesis (θ>0) in two sets of simulations—(i) when genes have only trait-increasing effects, and (ii) when genes have both trait-increasing and trait-decreasing effects using the same database pathway network topologies used for the null simulations (**Figure 2**, **Supplementary Figure S4**). For each scenario, we simulated different noise profiles, i.e., varying proportions of genes within the network with effects on the trait given the network topology. This mimics situations from having a very noisy network (i.e., ~10-30% genes with an effect on the trait) to a highly relevant network (i.e., ~70-90% genes with an effect on the trait). We simulated allele counts from cases and controls using independent Binomial distributions under the alternative model with $\theta = 0.2$. For each gene, we assumed the presence of up to five (5) qualifying loci per gene with minor allele frequency (MAF) of 0.001. The Binomial probabilities for case- and control-groups are adjusted according to the group sizes and MAF of variants assuming possible gene-specific effect configurations under the alternative model (i.e., $\vec{\alpha}$ given the network topology (Σ) and network effect, $\theta = 0.2$). Using this setup, we simulated a cohort of 2,000 cases and 2,000 controls.

Currently, there are no existing rare variant association tests that take gene network topology into account Thus, we compared the performance of NERINE with existing gene-level rare variant association tests adapted to the pathway level, namely, CMC-Fisher test[107], Fisher minimum p-value test[108], Fisher combined test[108], SKAT-O[18], and pathway-based rare variant trend test (RVTT)[15,16]. For each noise profile, we performed 250 iterations, resulting in 1,000 iterations per network. Empirical power of each method was measured as the positive

predictive value (PPV) across iterations using different p-value cutoffs (c): 1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5, and 1e-5. Here,

$$PPV = \frac{\text{\# Positive findings with p-value} < c}{\text{total \# of test cases}}$$

Additionally, we conducted performance benchmarking for NERINE on different simulated network architectures for 25 genes (**Supplementary Figure S5**). As before, we performed two sets of simulations—(i) when genes have only trait-increasing effects, and (ii) when genes have both trait-increasing and trait-decreasing effects for four network topologies—clique, path, random graph, and isolated nodes. We simulated allele counts in 1,000 cases and 1,000 controls using independent binomial distributions under the alternative model with a different network effect, $\theta = 0.5$. We simulated various network-noise profiles ranging from 0-90% following the same procedure as above. Empirical power of each method was calculated as PPV across 250 iterations per noise profile per network topology, resulting in 1,250 iterations per network. Since RVTT, by design, assumes that all qualifying rare variants in a pathway have the same direction of effects, we compared RVTT's performance with NERINE in simulations with genes having only trait-increasing effects. Moreover, RVTT computes a permutation-based p-value, and for 10,000 iterations its p-values cannot be less than 1e-4. Hence, we compared RVTT's performance only at cutoff values ≥ 1e-4.

Furthermore, we evaluated NERINE's performance with true versus randomized network architectures in simulations with the same four well-studied pathways (**Supplementary Figure S10**). We constructed the "true" network topologies for these pathways using the high-confidence protein–protein and genetic interactions from PPI databases as described above. For each gene set, we also generated 100 randomized networks by randomly introducing edges among the member genes. A binary trait was simulated in 2,000 cases and 2,000 controls with varying trait-increasing and trait-decreasing effects under the alternative model with a network effect, $\theta = 0.1$. We simulated different noise profiles (i.e., 10-30%, 30-50%, 50-70%, and 70-90%) and performed 250 iterations per noise profile. We measured NERINE's PPV at different p-value thresholds (i.e., 1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5, 1e-5).

Pathway database construction

We created a pathway database with all canonical pathways of five to fifty genes from the BIOCARTA database along with all lipid-, DNA replication-, DNA damage repair-, and cell cycle-related pathways from the REACTOME, KEGG, PID, and Wiki pathways databases. The lists of member genes for these pathways were extracted from the Molecular Signatures Database (MSigDB v7.3; https://www.gsea-msigdb.org/gsea/msigdb). The database

contained 306 pathways with a median pathway length of 25 genes (**Supplementary Table T20**). Due to the pleiotropy of genes, many biological pathways tend to overlap significantly with each other. Thus, we determined the effective number of independent hypotheses (t_{eff}) in our pathway database of t pathways by adapting Nyholt's approach[109].

First, we calculated the pairwise Jaccard similarity of pathways, and encode it in a matrix, $M_{t\times t}$, which is symmetric but not positive semi-definite. We then converted it to its nearest positive definite matrix using the *nearPD* function[110] from the *Matrix* package (v1.6-5) in R (v4.3.2). Next, we determined the eigenvalues ($\vec{\lambda}$) of this approximate covariance matrix. The effective number of hypotheses/pathways was computed using the following formula[109]:

$$t_{eff} = 1 + (t-1)\left(1 - \frac{Var(\vec{\lambda})}{t}\right)$$

For our pathway database, we found the effective number of independent hypotheses, t_{eff} to be 300, which was used for Bonferroni correction to determine database-wide significance.

Gene-network topology extraction

NERINE treats the gene-gene network as an input and can flexibly handle any symmetric pairwise relationship matrix that is positive-semidefinite. For a screen that provided the gene network topology, we used that network as is. For example, for the α-synuclein proteotoxicity screen in PD, we directly applied NERINE on the published TransposeNet humanized α-synuclein-modifier network[2] stems. In absence of the true network topology for a particular gene set, we adopted the following approach to construct one.

For canonical database pathway gene sets, we constructed network topologies by extracting physical and genetic interactions from several sources: (i) high confidence physical interactions (weight >= 0.7) from STRING v11.5[111], (ii) InWeb inBio Map database[112], (iii) HuRI[6], (iv) genetic interactions from the Megchelenbrink *et al.* study[113], and (v) humanized α-synuclein-, β-amyloid-, and TDP-43-modifier networks[2]. Using a heuristic approach, we first generated adjacency matrices for the genes, where edges were represented as binary values (1 = presence of an edge; 0 = absence of an edge), corresponding to the non-zero off-diagonal entries of the network. The diagonal was set to two (2) to indicate an equal prior on each gene. Because these modified adjacency matrices were not always positive semidefinite, we applied the *nearPD* function[110] from the **Matrix** package (v1.6-5) in R (v4.3.2) to compute the nearest positive-definite matrix. This matrix was input to NERINE as an approximate covariance matrix, Σ.

We also constructed co-expression networks in relevant tissue types for different phenotypes using the bulk tissue expression data from the Genotype Tissue Expression database[114] (GTEx v8). We computed a real-valued gene-gene co-expression networks in a specific tissue where the edges between two genes represented the Pearson correlation of their expression profiles in that tissue types. For lipid-related phenotypes, we constructed co-expression networks using the bulk expression data from liver tissue. For Parkinson's disease (PD), we used bulk expression data from the substantia nigra region of the mid-brain.

To construct co-essentiality networks in relevant cell lines for different phenotypes, we used the gene dependency data from CRISPR knockout screens from project Achilles, as well as genomic characterization data from the Cancer Cell Lines Encyclopedia (CCLE) project from the DepMap portal[115,116] (release: 2023Q2). We computed a gene-gene co-essentiality networks in relevant cell lines where the edges between two genes represent the Pearson correlation of their dependency profiles in those cell lines. While for lipid-related phenotypes, we constructed co-essentiality networks using data from liver cell lines, for PD, we used data from cell lines pertaining to the central nervous system (CNS). The cell lines used in this study are listed in the **Supplementary Table T21**.

Performance benchmark on lipid phenotypes in the UK biobank (UKBB)

Using jointly genotyped variant calls from the whole exomes of 469,589 individuals in the UK biobank (UKBB), we performed positive control experiments with NERINE in two lipid related phenotypes: direct LDL cholesterol (LDL-C; data field: 30780) and HDL cholesterol (HDL-C; data field: 30760) (**Supplementary Figures S6-S9**). We created two dichotomous phenotypes: (i) high LDL vs. low LDL and (ii) low HDL vs. high HDL, by selecting individuals belonging to the top and bottom quartiles of the distributions for LDL-C and HDL-C measurements. European ancestry groups had the largest sample sizes: LDL-C (30,007 cases and 28,673 controls), and HDL-C (26,800 cases and 27,178 controls). NERINE was competitively applied on these two phenotypes across networks from our canonical pathways database. We extracted the edge relationships of genes from high-confidence physical and genetic interactions from protein interaction databases for this exercise. This analysis served as an ideal positive control because the genetic determinants of these phenotypes are well-annotated.

Variants with a minor allele frequency (MAF) < 0.001 were considered rare for the test to not be influenced by artificial signal from common variant space propagated through linkage disequilibrium (LD). We stratified variants into six functional categories: LoF (i.e., frameshifts, insertions, deletions, and splice variants), damaging missense (i.e., missenses predicted to be damaging by in-silico tools), damaging (i.e., LoF and damaging missenses),

missense, neutral (i.e., missenses predicted to be bening by in-silico tools), and synonymous. We used neutral and synonymous variants in a pathway as control to safeguard against technical biases and LD leakage. Bonferroni correction, accommodating the presence of correlated hypotheses in the database[109], was used to control for Type I error.

To ensure our results for LDL-C phenotype were not driven by *LDLR* and *PCSK9*, which have large individual effect sizes on LDL-C levels, we performed sensitivity analysis by removing *LDLR* and *PCSK9* from the significant networks for the LDL phenotype. Even without *LDLR* and *PCSK9*, the module of core LDL-related genes, along with the LDL clearance and chylomicron clearance pathways remained significant for the LDL phenotype after Bonferroni correction (**Supplementary Figure S22**).

For many disease phenotypes, the large sample sizes available for the UKBB lipid phenotypes, are simply not attainable. Thus, we performed a down-sampling experiment to evaluate the consistency of NERINE's performance across different sample sizes and the robustness of NERINE's performance with small sample sizes. We downsampled the high LDL vs. low LDL cohort at different case-control ratios (1/3, 1/10, and 1/60) and competitively applied NERINE across the pathway database, demonstrating NERINE's effectiveness even in a cohort with as few as 500 cases and 500 controls (**Supplementary Figure S23**).

Since rare variants are more susceptible to subtle effects of population stratification than common variants, we performed stratified analysis of the LDL-C phenotype in all five major ancestry groups (i.e., European, American, African American, South Asian, and East Asian) and meta-analyzed the results using Fisher's combined test (**Supplementary Figure S9**, **Supplementary Table T2**). Bonferroni correction was applied on the meta p-values.

We further evaluated NERINE on real versus randomized network topologies for binarized LDL-C and HDL-C phenotypes (**Supplementary Figure S11**). For each phenotype, we selected database-wide significant pathways with rare damaging variant burden and used the most significant topologies for each pathway shown in **Supplementary Table T3** as "real" network topologies. We also generated 100 randomized networks per pathway by introducing random binary edges among member genes. We applied NERINE to both real and randomized networks and assessed its performance by comparing LLRs and p-values; higher LLRs and lower p-values indicated superior performance.

Study cohorts for common disease phenotypes

We used whole exome sequencing (WES) data from two population-scale biobanks, namely, the UK Biobank (UKBB) and the MassGeneral Brigham Biobank (MGBBB) and whole genome sequencing (WGS) data from the AMP-PD consortium for investigating network-level rare variant associations in complex diseases.

Our analysis focused on four representative, high-prevalence complex diseases of public health relevance: breast cancer (BRCA), type II diabetes mellitus (T2D), coronary artery disease (CAD), and early-onset myocardial infarction (MI). These phenotypes were chosen from the UK Biobank because large sample sizes were available, and replication cohorts existed in the MGB Biobank. They served to showcase NERINE's applicability in comprehensive database searches for rare variant network burden.

Cohorts for BRCA, T2D, CAD, and MI were selected primarily based on the summary diagnoses recorded in the data field 41270 as ICD-10 codes. For BRCA, the case group consisted of unrelated females of European ancestry with ICD-10 code C50 and the control group consisted of unrelated European females of age 60 or above with no history of neoplasms (ICD-10 codes: C00-C97 and D00-D48). The resulting cohort had 10,648 cases and 91,886 controls. For T2D, we included unrelated European individuals with ICD-10 code E11 in the case group and unrelated European individuals with no endocrine, nutritional and metabolic diseases (ICD10 codes: E00-E90) in the control group. The resulting cohort had 22,502 cases and 68,370 controls. We created an age-stratified case-control cohort for the CAD phenotype where cases consisted of unrelated European individuals of age <=65 with ICD-10 code I25 and controls consisted of unrelated European individuals of age > 65 with no diseases of the circulatory system (ICD-10 codes: I00-I99). This left us with 4,561 cases and 12,321 controls. Finally, for the MI phenotype, our cohort consisted of 2,521 cases and 5,012 controls. Cases included unrelated European individuals with ICD-10 code I 21. Only males with age <=55 and females with age <=65 were included in the case group. Whereas controls consisted of unrelated European individuals of age >= 69 who have no history of any disease of the circulatory system (ICD-10 codes: I00-I99).

Parkinson's disease (PD) was chosen as a case study to demonstrate NERINE's ability to bridge human genetics with model-system screens. We had access to two genome-scale screens, targeting the main pathological features of the disease and providing experimental gene networks for testing. We also had our in-house iPSC-derived cortical neuron model (CiS-CN) for functional validation. We focused on sporadic disease cases as they constitute ~85% of the PD population.

For sporadic PD phenotype, we created a discovery cohort from UKBB individuals by including unrelated European individuals with ICD-10 code G20 and no family history of PD as cases. The control group consisted of unrelated European individuals with no history of PD and other nervous system diseases (ICD-10 codes: G00-G99). In the original UKBB cohort, we observed notable differences in the incidence of type 2 diabetes (T2D) between PD cases (17.5%) and controls (7.1%), which could bias downstream analyses; therefore, we excluded all individuals with a history of T2D from both groups. We termed this cohort the "UKBB-Sporadic" cohort with 2,237 cases and 167,188 controls. Controls were on average younger than cases at recruitment in this cohort (**Supplementary Table T22**), which could introduce bias when investigating signatures of dopamine (DA) neuron degeneration. Thus, for the DA neuron essentiality screen analysis, we employed a super-control design by selecting older controls with a higher median age at recruitment than the cases (**Supplementary Table T22**), which are more likely to capture naturally occurring signatures of neurodegeneration and less likely to have sporadic PD later in life. We termed this cohort with super controls as UKBB-Extreme (Ncase = 2,237 and Ncontrol = 2,553).

For the database-wide investigations of BRCA, T2D, CAD, and MI, we created replication cohorts using the jointly genotyped WES data from 53,343 individuals in MGBBB, a biorepository of consented patient samples at Mass General Brigham (parent organization of Massachusetts General Hospital and Brigham and Women's Hospital). Same inclusion/exclusion criteria were used for each phenotype to ensure consistency between the biobanks. The resulting cohort sizes are as follows: BRCA (Ncase = 1,113; Ncontrol = 2,459), T2D (Ncase = 747; Ncontrol = 2,188), CAD (Ncase = 902; Ncontrol =1,488), and MI (Ncase = 326; Ncontrol = 2,068). For the BRCA cohort, we used an MAF cutoff of 0.001 to select rare variants. For the other three phenotypes, variants with MAF <0.03 were considered to be "rare". Since the cohort sizes in MGBBB were smaller compared to UKBB for T2D, CAD, and MI, we did this adjustment to the MAF cutoff and performed tests in the synonymous and neutral missense categories for each pathway to make sure that there was no LD leakage.

For PD, we created a replication cohort using the WGS data from 10,418 individuals from AMP-PD (Accelerating Medicines Partnership: Parkinson's Disease) v3 release (2022), which encompasses study participants at MassGeneral Brigham (i.e., Harvard Biomarkers Study 2.0). Any individual belonging to the genetic registry and genetic cohort group, as well as subjects without evidence of dopamine deficit (SWEDD) and subjects belonging to the prodromal categories and the AMP-LBD cohort were excluded from the analysis. As the AMP-PD cohort predominantly consists of individuals of European ancestry, we retained only unrelated individuals of the same ancestry group. We called the resulting cohort "AMP-PD-

sporadic," which consisted of 2,117 sporadic PD cases and 1,095 neurotypical controls. Both case and control groups had similar median ages in this cohort (**Supplementary Table 22**).

We performed both variant- and sample-level quality control (QC) steps on each dataset to ensure the study cohorts are free from technical biases as much as possible (see **Supplementary Note** for detailed steps). We retained only high-quality biallelic variants passing GATK best practices filters and having maximum 10% missingness for our analysis. We annotated variants with their functional consequences and gnomAD allele frequencies with VEP (v109) and dbNSFP (v4.3a) database. We used six masks to group variants into functional categories: (i) *Damaging missense*: missense variants predicted to be either "P'' or "D'' by PolyPhen2 or "deleterious'' by SIFT, (ii) *LoF*: variants labelled as splice donors, splice acceptors, splice region variants, stop-gained, stop-lost, start-lost, frameshifts, in-frame insertions, and in-frame deletions; (iii) *Damaging*: LoFs and damaging missenses, (iv) *Missense*, (v) *Neutral*: missense variants predicted to be either "B'' by PolyPhen2 or "tolerated'' by SIFT, and (vi) *Synonymous*. These masks were used in our analysis of binarized cholesterol phenotypes as well as complex diseases.

Notably, NERINE is agnostic to the variant annotation tool and can test any user-defined categories of rare variants. To assess robustness, we reclassified variants using REVEL[117] and AlphaMissense[118] scores in a positive control experiment with the LDL-C phenotype in UKBB. We defined three categories: (i) *pathogenic missense* (REVEL > 0.649 or AlphaMissense > 0.564), (ii) *pathogenic* (pathogenic missense plus LoF), and (iii) *benign missense* (REVEL < 0.29 and AlphaMissense < 0.34). Results were highly concordant with those obtained using PolyPhen2 and SIFT (**Supplementary Figure S24**).

We removed sample outliers based on Ts/Tv, Het/Hom ratios, and per-haploid SNV counts, where outliers were defined as samples that were three standard deviations away from the mean. We retained only unrelated European samples for our analyses because, at the time, individuals of European ancestry constituted the largest group in our study cohorts. Concentrated efforts in building large biobanks with diverse participants are already underway[119] and will enable NERINE to overcome this limitation and provide more insight into the contribution of rare variants to common disease etiology across populations.

Genome-wide CRISPR/Cas9 screen in midbrain DA neurons for essentiality genes

We performed genome-wide CRISPR/Cas9 screen in DA neurons differentiated from WA-09 (H9) embryonic stem cells. Guide RNAs (gRNAs) representing 19,993 genes were transduced into H9 hESC lines. At DIV14-16 of the differentiation, we induced iCas9

expression by doxycycline addition using the AAVS1 safe-harbor locus as previously described[120], while cells were neural progenitors. Then, we waited until DIV25 when they differentiated into DA neurons to take our initial sample to obtain gRNA representation through whole genome sequencing (DIV26). The remaining neurons were allowed to stay in the dish until our final collection time (DIV42) when neuronal cell death began. Stem cells were transduced with the Gattinara human CRISPR pooled knockout library[121] at an MOI of 0.3-0.5 and 1000x representation. Transduced stem cells were selected by puromycin and differentiated toward DA neurons until reaching the neural progenitor stage as described by Kim and colleagues[122]. Sample were processed for library preparation and sequenced and sequencing reads were aligned to the screened library and analyzed using *MAGeCK-MLE* from the *MAGeCKFlute*[123] (v2.6.0) package in R (v4.3.2). Essentiality genes were classified as having Wald test FDR-adjusted p-value < 0.05 and beta < -0.58. Broadly essential genes[123] that are not specific to DA neurons, were removed from the list of essentiality genes. We identified 693 essentiality genes. The complete screen will be described in a forthcoming publication.

Constructing ontology-based network topologies from gene sets

For PD GWAS genes as well essential genes for DA neuron survival, we first performed gene set enrichment analysis using the *enrichr* function from the *GSEApy* package (v 1.1.3) in python (v 3.12.4) and identified all GO biological processes with nominal significance (p-value < 0.05). We then grouped semantically similar GO terms using REVIGO (http://revigo.irb.hr/) to identify GO biological processes (BP) modules with minimal overlap. We only kept modules of 10 or more genes for our analysis. For PD GWAS genes, we identified six such modules (**Supplementary Table T12**) and for DA neuron essentiality genes, we identified 10 such modules (**Supplementary Table T15**). To impose network topology on these gene sets, we extracted edge relationships of genes in each group from three different data sources as described above: (i) high-confidence physical and genetics interactions from protein interaction databases, (ii) co-expression in the substantia nigra region of the mid-brain from GTEx v8, and (iii) co-essentiality in CNS cell types in DepMap v2023Q.

We also explored an alternative approach from a recent study[65] for constructing gene modules from PD GWAS loci, which involved running MAGMA[124] analysis on GWAS genes, followed by GO BP enrichment. This approach identified 21 significant conditionally independent GO terms[65]. We then imposed network topology on the overlapping genes with each GO term by extracting gene-gene relationships from the sources described above.

Human iPSCs culture and induced neuron differentiation

Human iPSCs were cultured in Stemflex (Gibco/Thermo Fisher Scientific; Cat. No. A33493) on 6-well plates coated with Matrigel Matrix (Corning; Cat. No. 356231) diluted 1:100 in Knockout DMEM (Gibco/Thermo Fisher Scientific; Cat. No. 10829-018). Briefly, Essential 8 Medium was replaced every day. When 80% confluent, cells were passaged with StemPro Accutase Cell Dissociation Reagent (Gibco/Thermo Fisher Scientific; Cat. No. A11105-01). Human iPSCs engineered to express *NGN2* under a doxycycline-inducible system in the AAVS1 safe harbor locus were differentiated following previously published protocol. Briefly, iPSCs were released as above, centrifuged, and resuspended in N2 Pre-Differentiation Medium containing the following: Knockout DMEM/F12 (Gibco/Thermo Fisher Scientific; Cat. No. 12660-012) as the base, 1X MEM Non-Essential Amino Acids (Gibco/Thermo Fisher Scientific; Cat. No. 11140-050), 1X N2 Supplement (Gibco/Thermo Fisher Scientific; Cat. No. 17502-048), 10ng/mL NT-3 (PeproTech; Cat. No. 450-03), 10ng/mL BDNF (PeproTech; Cat. No. 450-02), 1 µg/mL Mouse Laminin (Thermo Fisher Scientific; Cat. No. 23017-015), 10nM ROCK inhibitor, and 2µg/mL doxycycline hydrochloride (Sigma-Aldrich; Cat. No. D3447-500MG) to induce expression of mNGN2. iPSCs were counted and plated on Matrigel-coated plates in N2 Pre-Differentiation Medium for three days. After three days, hereafter Day 0, pre-differentiated cells were released and centrifuged as above, and pelleted cells were resuspended in Classic Neuronal Medium containing the following: half DMEM/F12 (Gibco/Thermo Fisher Scientific; Cat. No. 11320-033) and half Neurobasal-A (Gibco/Thermo Fisher Scientific; Cat. No. 10888-022) as the base, 1X MEM Non-Essential Amino Acids, 0.5X GlutaMAX Supplement (Gibco/Thermo Fisher Scientific; Cat. No. 35050-061), 0.5XN2 Supplment, 0.5XB27 Supplement (Gibco/Thermo Fisher Scientific; Cat. No. 17504-044), 10ng/mL NT-3, 10ng/mL BDNF, 1µg/mL Mouse Laminin, and 2µg/mL doxycycline hydrochloride. Pre-differentiated cells were subsequently counted and plated on BioCoat Poly-D-Lysine coated plates (Corning; Cat. No. 356470) in Classic Neuronal Medium. On DIV7 and each week after, medium change was performed without doxycycline added. In the PFF exposure experiment, a complete medium change to the medium containing 10 µg/mL synthetic PFF was performed on DIV21 neurons, and the neurons were fixed at DIV28 for immunostaining.

CRISPRi screen and analysis in the CiS-CN model

A customized CRISPRi library comprising 9,852 sgRNAs targeting 1,705 physical and genetic interactors of α-synuclein (αS)—including genes from the TransposeNet αS proteinopathy network—along with negative controls, was constructed. The library was packaged into lentivirus by the Virus Core at Boston Children's Hospital. Neurons were transduced with the viral library and cultured until the harvesting date (DIV 3, 28, and 42). Genomic DNA was extracted from harvested neurons, and the sgRNA-encoding regions were amplified for sequencing based on previously described protocols[86,125,126]. Sequencing and annotation

were conducted at Memorial Sloan Kettering Cancer Center. Data was analyzed using the *MAGeCK-iNC* pipeline[86]. Hits were classified as having an FDR-adjusted p-value < 0.1, and the gene product cutoff was selected by the *MAGeCK-iNC* pipeline dynamically for each comparison. Hits with significant sgRNA dropout were termed "toxicity enhancers." Details of the functional screen in its entirety will be described in a forthcoming publication from our group.

Immunostaining and microscopy imaging

Neurons are fixed with 4% PFA (EM Sciences; Cat. No. 15710) for 15 minutes at room temperature, and then permeabilized with 0.5% Triton X-100 and blocked with 0.05% Triton X-100 and 5% BSA in PBS for 1hr at room temperature. Samples were incubated with primary antibodies (PRL: Thermo Fisher; Cat. No. MA1-10597; 1:200 dilution) at 4°C overnight, followed by incubation with secondary antibody and Hoechst 33342 (Thermo Fisher; Cat. No. H3570; 1:2000 dilution) for 1hr at room temperature. Images were captured with identical settings for parallel cultures using Nikon Eclipse Ti microscope or Nikon TiE/C2 confocal microscope. Image analysis was performed with ImageJ (NIH). *Prl* level was determined by D (D = total *prolactin* intensity / total DAPI number in a given image). Immunostaining images were analyzed with ImageJ Macro Software (**Supplementary Note**).

Oxidative stress assay

DIV0 CiS neurons were seeded at a density of 40,000 cells/well of poly-L-ornithine-coated 96-well plate. At DIV6, the neuron media was fully changed for 100uL *Prolactin*-containing (Cat. No. 100-07-10UG media at a concentration of 0,1 or 10 nM. 24hrs post-treatment, the neuron media was fully changed for 100uL of Menadione-containing (Cat. No. ICN10225925) media at a concentration of 0 or 100uM and incubated for an hour at 37°C. After an hour, CELLROX™ Green Reagent (Cat. No. C10444) was added on top of the Menadione-treated media to a final concentration of 5 uM CELLROX™ for 30 minutes at 37°C. All media was then removed, and wells were washed 3 times with PBS. After the third wash, the PBS was replaced with neuron media. The plates were then taken to the Incucyte S3 live-cell analysis system (Sartorius) for imaging. Incucyte analysis was perform with S3 software, and CellRox = total integrated intensity/neuron was reported.

Mouse model

Wildtype B6C3F1 mice (Stock 100010; The Jackson Laboratories) were used for the stereotactic injection studies described. Animals were maintained on a 12-hour light/dark schedule and provided with food ad libitum. All housing, breeding, and procedures were performed according to the NIH Guide for the Care and Use of Experimental Animals and approved by the University of Pennsylvania Institutional Animal Care and Use Committee.

αS and PFF preparation

Full-length mouse αS was expressed in BL21 (DE3) RIL-competent E. coli cells (Agilent Technologies 230245) transformed with pRK172/mSyn containing αS cDNA. Protein purification was previously described (Luk et al., 2012; Volpicelli-Daley et al., 2014). Cultures expanded in Terrific Broth (12 g/L of Bacto-tryptone, 24 g/L of yeast extract 4% (vol/vol) glycerol, 17 mM KH2PO4 and 72 mM K2HPO4) containing ampicillin (Fisher Scientific) were harvested and sonicated in high salt buffer (750 mM NaCl in 10 mM Tris, pH 7.6,). After boiling for 15 mins, the supernatant was dialyzed against 10 mM Tris, pH 7.6, 50 mM NaCl, 1 mM EDTA overnight at 4ºC, filtered and concentrated using Amicon Ultra-15 centrifugal filter units (Millipore UFC901008). Gel filtration using a Superdex 200 column (Cytiva) was performed and fractions containing αS pooled and dialyzed in 10 mM Tris, pH 7.6, 50 mM NaCl, 1 mM EDTA overnight. The product was polished using a HiTrapQ HP column (Cytiva 645932) and eluted over an ionic gradient (25 to 1,000 mM NaCl). Fractions containing αS were combined and dialyzed into DPBS, sterile filtered and concentrated to 5 mg/mL and frozen at -80°C until used. PFFs were assembled by shaking monomer at 5 mg/mL using a Thermomixer C (Eppendorf) set at 1,000 rpm for 7 days at 37ºC. Fibril content was validated by sedimentation at 100,000 x g for 30 minutes and Thioflavin T fluorimetry.

Stereotaxic administration of PFFs

Prior to injection, PFFs were diluted to 2 mg/mL in Dulbecco's PBS and sonicated using a bath sonicator (Biorupter UCD-300, Diagenode) on high power for 10 cycles (30 sec on; 30 sec off) at 10ºC. Each mouse received a single unilateral injection of PFFs (5 µg of PFFs in 2.5 µL volume) into the dorsal striatum using a Hamilton syringe (33 gauge) using the following co-ordinates: AP +0.2 mm relative to Bregma; ML +2.0 mm; depth 2.6 mm beneath the dura. DPBS injected into the same region was used as a negative control. Mice were perfused transcardially with heparinized PBS at 30 d.p.i. and brains flash frozen at -80ºC until use.

RNA isolation and NanoString analysis

The amygdala region ipsilateral to PFF- or PBS injection was microdissected from each brain and homogenized in 1 ml of TRI Reagent (Sigma-Aldrich, T9424) using TissueRuptor II (Qiagen, 9002755) with a disposable probe (Qiagen, 990890). RNA was then isolated using Direct-zol RNA MiniPrep kit with in-column DNaseI treatment (Zymo, R2050). Samples were quantitated with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) and assayed for RNA integrity on a 4200 TapeStation (Agilent, G2991AA). NanoString hybridization of the resultant RNA was carried out for a constant 18 hours at 65° C on the *Mus musculus* Neuropathology panel (v1.0). Post-hybridization processing in the nCounter

Prep Station used the High Sensitivity settings. The cartridge scanning parameter was set at high (555 FOV). RNA isolation and NanoString studies were performed at the Wistar Genomics core facility. Genes with hybridization counts ≥20 were normalized to the geometric mean of panel positive controls as recommended by the manufacturer. Differential expression was analyzed using a generalized linear model in the nCounter module in Rosalind (Rosalind.bio). Adjusted p-values were calculated using the Benjamini-Hochberg method using treatment (i.e. PFF vs PBS).

Data and code availability

We downloaded the canonical pathway gene set collections under MSigDB (v7.3) from https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp. Physical interactions in humans in the STRING (v11.5) database were downloaded from https://string-db.org/cgi/download. The HuRI interaction database was downloaded from http://www.interactome-atlas.org/download. The inBio Map protein–protein interaction (PPI) network database can be obtained from[112] (https://www.intomics.com/inbio/map: last accessed on Jan 2022). Genetic interactions data from the Megchelenbrink *et al.* study was obtained from the supplement of the paper[113]. TransposeNet's humanized α-synuclein-, β-amyloid-, and TDP-43-modifier networks were obtained from our previously published study[2]. Bulk expression data in TPM format (GTEx v8) from different human tissue types were downloaded from the GTEx portal (https://www.gtexportal.org/). Data on gene dependencies in different cell lines from the DepMap project (release: 2023Q2) was downloaded from the DepMap portal (https://depmap.org/portal/data_page/?tab=allData).

UK Biobank 500K whole exome sequencing (WES) data was accessed through application 41250 and is available through https://ams.ukbiobank.ac.uk and was processed using the DNAnexus platform. MassGeneral Brigham (MGB) Biobank's 50k WES data was accessed through the https://biobankportal.partners.org/ (PI: Khurana). Whole genome sequencing data (v3 release, 2022) from the Accelerating Medicines Partnership Parkinson's disease (AMP-PD) was accessed through the AMP-PD Knowledge Platform (https://www.amp-pd.org).

The source code for NERINE is available on github (https://github.com/snz20/NERINE). RVTT was run adapting the code from https://github.com/snz20/RVTT. CMC-Fisher test, Fisher's combined test, SKAT-O, and MAGeCK-iNC analysis were performed using R (v4.3.2) packages, namely, stats (v4.3.2), SKAT (v2.2.5), and MAGeCKFlute (v2.6.0). Gene set enrichment of GO biological process terms was performed using the GSEApy (v1.1.3) package in python (v3.12.4).

Acknowledgement

We gratefully acknowledge Dr. Matthew Stevens, Dr. Richard Sherwood, Dr. Benjamin Neale, and Dr. Isabel Lam for their valuable insights. S.N. gratefully acknowledges the support from the National Institute of Health (NIH) grant R35GM127131, the Sudarsky Scholar Award from the BWH Movement Disorders Division, and the Australian Parkinson's Mission. S.S. is supported by the NIH grants U01HG012009, R35GM127131, and R01MH101244. V.K. is supported by the NIH grant R01NS109209. XW gratefully acknowledges support from the NIH grant T32AG000222 (PI: Yankner). Experiments in iPSC-derived CiS models were done with the support from the Aligning Science Across Parkinson's Initiative (ASAP) award ASAP-000472 (PI: Studer). We are indebted to the UK Biobank (www.ukbiobank.ac.uk), and its participants who provided biological samples and data for this study, performed under UK Biobank application 41250 (PI: Cassa) under the Mass General Brigham IRB protocol 2020P002093 supported by NIH R01HG010372 (PI: Sunyaev). UKBB is a major biomedical database that is globally accessible to approved researchers who are undertaking health-related research that's in the public interest and is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. Our study also included whole-exome and phenotypic data from ~50k participants from the Mass General Brigham Biobank, a biorepository of consented patient samples at Mass General Brigham (parent organization of Massachusetts General Hospital and Brigham and Women's Hospital). We further analyzed whole genome data from AMP-PD–a public-private partnership–managed by the FNIH and funded by Celgene, GSK, the Michael J. Fox Foundation for Parkinson's Research, the National Institute of Neurological Disorders and Stroke, Pfizer, and Verily. AMP-PD investigators have not participated in reviewing the data analysis or content of this manuscript. We thank all the participants and clinical and research teams who contributed to UKBB, MGBB, and AMP-PD.

Author contributions

S.N., V.K., and S.S. conceived of the project, interpreted the results; S.N., X.W., V.K., and S.S. wrote the manuscript; S.N. and S.S. developed the novel methodology NERINE; S.N. pre-processed and analyzed the study cohorts; A.M. assisted in the quality control and the pre-processing of the sequencing data; S.N., S.S., N.S. and R.G. interpreted the results in lipid and cardiac phenotypes; X.W. and V.K. designed the wet laboratory experiments to validate the NERINE's findings in PD in neuronal models; X.W., R.S., and E.E. generated the CiS neurons and performed the experiments; X.W. and S.N. analyzed the experimental data and interpreted the results with V.K.; D.R., A.H., J.A., and L.S. designed the genome-wide CRISPR/Cas9 experiments in mid-brain DA neurons, identified the essentiality genes, and curated the GO biological process modules; K.L. designed and performed the experiments

in the mouse PFF model, analyzed the data, and interpreted the results. All the authors read and helped edit the manuscript.

Declaration of conflict-of-interest

The authors have declared that no conflicts of interest exist.

Reference

1. Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol *4*, e1000117. 10.1371/journal.pcbi.1000117.
2. Khurana, V., Peng, J., Chung, C.Y., Auluck, P.K., Fanning, S., Tardiff, D.F., Bartels, T., Koeva, M., Eichhorn, S.W., Benyamini, H., et al. (2017). Genome-Scale Networks Link Neurodegenerative Disease Genes to alpha-Synuclein through Specific Molecular Pathways. Cell Syst *4*, 157-170 e114. 10.1016/j.cels.2016.12.011.
3. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576 e516. 10.1016/j.cell.2017.06.010.
4. Sledzieski, S., Singh, R., Cowen, L., and Berger, B. (2021). D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. Cell Syst *12*, 969-982 e966. 10.1016/j.cels.2021.08.010.
5. Schnitzler, G.R., Kang, H., Fang, S., Angom, R.S., Lee-Kim, V.S., Ma, X.R., Zhou, R., Zeng, T., Guo, K., Taylor, M.S., et al. (2024). Convergence of coronary artery disease genes onto endothelial cell programs. Nature *626*, 799-807. 10.1038/s41586-024-07022-x.
6. Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402-408. 10.1038/s41586-020-2188-x.
7. Yuan, Q., and Duren, Z. (2025). Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. Nat Biotechnol *43*, 247-257. 10.1038/s41587-024-02182-7.
8. Vidal, M., Cusick, M.E., and Barabasi, A.L. (2011). Interactome networks and human disease. Cell *144*, 986-998. 10.1016/j.cell.2011.02.016.
9. Barabasi, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat Rev Genet *12*, 56-68. 10.1038/nrg2918.
10. Ota, M., Spence, J.P., Zeng, T., Dann, E., Marson, A., and Pritchard, J.K. (2025). Causal modeling of gene effects from regulators to programs to traits: integration of genetic associations and Perturb-seq. bioRxiv. 10.1101/2025.01.22.634424.
11. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet *95*, 5-23. 10.1016/j.ajhg.2014.06.009.
12. Chen, W., Coombes, B.J., and Larson, N.B. (2022). Recent advances and challenges of rare variant association analysis in the biobank sequencing era. Front Genet *13*, 1014947. 10.3389/fgene.2022.1014947.
13. Richardson, T.G., Timpson, N.J., Campbell, C., and Gaunt, T.R. (2016). A pathway-centric approach to rare variant association analysis. Eur J Hum Genet *25*, 123-129. 10.1038/ejhg.2016.113.

14. Lee, S., Choi, S., Kim, Y.J., Kim, B.J., Consortium, T.d.-G., Hwang, H., and Park, T. (2016). Pathway-based approach using hierarchical components of collapsed rare variants. Bioinformatics *32*, i586-i594. 10.1093/bioinformatics/btw425.
15. Bendapudi, P.K., Nazeen, S., Ryu, J., Soylemez, O., Robbins, A., Rouaisnel, B., O'Neil, J.K., Pokhriyal, R., Yang, M., Colling, M., et al. (2024). Low-frequency inherited complement receptor variants are associated with purpura fulminans. Blood *143*, 1032-1044. 10.1182/blood.2023021231.
16. Hallacli, E., Kayatekin, C., Nazeen, S., Wang, X.H., Sheinkopf, Z., Sathyakumar, S., Sarkar, S., Jiang, X., Dong, X., Di Maio, R., et al. (2022). The Parkinson's disease protein alpha-synuclein is a modulator of processing bodies and mRNA stability. Cell *185*, 2035-2056 e2033. 10.1016/j.cell.2022.05.008.
17. Li, X., Quick, C., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Selvaraj, M.S., Sun, R., Dey, R., Arnett, D.K., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. Nat Genet *55*, 154-164. 10.1038/s41588-022-01225-6.
18. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Team, N.G.E.S.P.-E.L.P., Christiani, D.C., Wurfel, M.M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet *91*, 224-237. 10.1016/j.ajhg.2012.06.007.
19. Zhou, W., Bi, W., Zhao, Z., Dey, K.K., Jagadeesh, K.A., Karczewski, K.J., Daly, M.J., Neale, B.M., and Lee, S. (2022). SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. Nat Genet *54*, 1466-1469. 10.1038/s41588-022-01178-w.
20. Lee, S., Kim, Y., Choi, S., Hwang, H., and Park, T. (2018). Pathway-based approach using hierarchical components of rare variants to analyze multiple phenotypes. BMC Bioinformatics *19*, 79. 10.1186/s12859-018-2066-9.
21. Lee, S., Kim, S., Kim, Y., Oh, B., Hwang, H., and Park, T. (2019). Pathway analysis of rare variants for the clustered phenotypes by using hierarchical structured components analysis. BMC Med Genomics *12*, 100. 10.1186/s12920-019-0517-4.
22. Guo, M.H., Plummer, L., Chan, Y.M., Hirschhorn, J.N., and Lippincott, M.F. (2018). Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. Am J Hum Genet *103*, 522-534. 10.1016/j.ajhg.2018.08.016.
23. Zhao, J., Zhu, Y., Boerwinkle, E., and Xiong, M. (2015). Pathway analysis with next-generation sequencing data. Eur J Hum Genet *23*, 507-515. 10.1038/ejhg.2014.121.
24. Wu, G., and Zhi, D. (2013). Pathway-based approaches for sequencing-based genome-wide association studies. Genet Epidemiol *37*, 478-494. 10.1002/gepi.21728.
25. Pan, W., Kwak, I.Y., and Wei, P. (2015). A Powerful Pathway-Based Adaptive Test for Genetic Association with Common or Rare Variants. Am J Hum Genet *97*, 86-98. 10.1016/j.ajhg.2015.05.018.
26. Makarious, M.B., Lake, J., Pitz, V., Ye Fu, A., Guidubaldi, J.L., Solsberg, C.W., Bandres-Ciga, S., Leonard, H.L., Kim, J.J., Billingsley, K.J., et al. (2023). Large-scale rare variant

burden testing in Parkinson's disease. Brain *146*, 4622-4632. 10.1093/brain/awad214.

27. Pitz, V., Makarious, M.B., Bandres-Ciga, S., Iwaki, H., andMe Research, T., Singleton, A.B., Nalls, M., Heilbron, K., and Blauwendraat, C. (2024). Analysis of rare Parkinson's disease variants in millions of people. NPJ Parkinsons Dis *10*, 11. 10.1038/s41531-023-00608-8.
28. Nazeen, S., Wang, X., Zielinski, D., Lam, I., Hallacli, E., Xu, P., Ethier, E., Strom, R., Zanella, C.A., Nithianandam, V., et al. (2024). Deep sequencing of proteotoxicity modifier genes uncovers a Presenilin-2/beta-amyloid-actin genetic risk module shared among alpha-synucleinopathies. bioRxiv, 2024.2003.2003.583145. 10.1101/2024.03.03.583145.
29. Luk, K.C., Kehm, V., Carroll, J., Zhang, B., O'Brien, P., Trojanowski, J.Q., and Lee, V.M. (2012). Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. Science *338*, 949-953. 10.1126/science.1227157.
30. Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika *72*, 133-144. 10.1093/biomet/72.1.133.
31. Stoel, R.D., Garre, F.G., Dolan, C., and van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. Psychol Methods *11*, 439-455. 10.1037/1082-989X.11.4.439.
32. Paththinige, C.S., Sirisena, N.D., and Dissanayake, V. (2017). Genetic determinants of inherited susceptibility to hypercholesterolemia - a comprehensive literature review. Lipids Health Dis *16*, 103. 10.1186/s12944-017-0488-4.
33. Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G., Lu, W., Riley-Gillis, B.M., Tsai, E.A., Kim, H.I., Zheng, X., et al. (2022). Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. Cell Genom *2*, 100168. 10.1016/j.xgen.2022.100168.
34. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet *52*, 680-691. 10.1038/s41588-020-0637-y.
35. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet *40*, 638-645. 10.1038/ng.120.
36. Zhang, W., Zhang, L., Xiao, C., Wu, X., Cui, H., Yang, C., Yan, P., Tang, M., Wang, Y., Chen, L., et al. (2024). Bidirectional relationship between type 2 diabetes mellitus and coronary artery disease: Prospective cohort study and genetic analyses. Chin Med J (Engl) *137*, 577-587. 10.1097/CM9.0000000000002894.
37. Frkic, R.L., Richter, K., and Bruning, J.B. (2021). The therapeutic potential of inhibiting PPARgamma phosphorylation to treat type 2 diabetes. J Biol Chem *297*, 101030. 10.1016/j.jbc.2021.101030.

38. Geldenhuys, W.J., Lin, L., Darvesh, A.S., and Sadana, P. (2017). Emerging strategies of targeting lipoprotein lipase for metabolic and cardiovascular diseases. Drug Discov Today *22*, 352-365. 10.1016/j.drudis.2016.10.007.
39. Li, S., Shin, H.J., Ding, E.L., and van Dam, R.M. (2009). Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. JAMA *302*, 179-188. 10.1001/jama.2009.976.
40. Yamauchi, T., Waki, H., Kamon, J., Murakami, K., Motojima, K., Komeda, K., Miki, H., Kubota, N., Terauchi, Y., Tsuchida, A., et al. (2001). Inhibition of RXR and PPARgamma ameliorates diet-induced obesity and type 2 diabetes. J Clin Invest *108*, 1001-1013. 10.1172/JCI12864.
41. Miyazaki, S., Taniguchi, H., Moritoh, Y., Tashiro, F., Yamamoto, T., Yamato, E., Ikegami, H., Ozato, K., and Miyazaki, J. (2010). Nuclear hormone retinoid X receptor (RXR) negatively regulates the glucose-stimulated insulin secretion of pancreatic ss-cells. Diabetes *59*, 2854-2861. 10.2337/db09-1897.
42. Anderson, A., and Walker, B.R. (2013). 11beta-HSD1 inhibitors for the treatment of type 2 diabetes and cardiovascular disease. Drugs *73*, 1385-1393. 10.1007/s40265-013-0112-5.
43. Alberts, P., Nilsson, C., Selen, G., Engblom, L.O., Edling, N.H., Norling, S., Klingstrom, G., Larsson, C., Forsgren, M., Ashkzari, M., et al. (2003). Selective inhibition of 11 beta-hydroxysteroid dehydrogenase type 1 improves hepatic insulin sensitivity in hyperglycemic mice strains. Endocrinology *144*, 4755-4762. 10.1210/en.2003-0344.
44. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med *354*, 1264-1272. 10.1056/NEJMoa054013.
45. Sabatine, M.S., Giugliano, R.P., Keech, A.C., Honarpour, N., Wiviott, S.D., Murphy, S.A., Kuder, J.F., Wang, H., Liu, T., Wasserman, S.M., et al. (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med *376*, 1713-1722. 10.1056/NEJMoa1615664.
46. Clusan, L., Ferriere, F., Flouriot, G., and Pakdel, F. (2023). A Basic Review on Estrogen Receptor Signaling Pathways in Breast Cancer. Int J Mol Sci *24*. 10.3390/ijms24076834.
47. Wilcox, N., Dumont, M., Gonzalez-Neira, A., Carvalho, S., Joly Beauparlant, C., Crotti, M., Luccarini, C., Soucy, P., Dubois, S., Nunez-Torres, R., et al. (2023). Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. Nat Genet *55*, 1435-1439. 10.1038/s41588-023-01466-z.
48. Qi, A., Lamont, L., Liu, E., Murray, S.D., Meng, X., and Yang, S. (2023). Essential Protein PHB2 and Its Regulatory Mechanisms in Cancer. Cells *12*. 10.3390/cells12081211.
49. Yang, J., Li, B., and He, Q.Y. (2018). Significance of prohibitin domain family in tumorigenesis and its implication in cancer diagnosis and treatment. Cell Death Dis *9*, 580. 10.1038/s41419-018-0661-3.
50. Liu, M., Wang, Z., Li, S., Deng, Y., and He, N. (2022). Identification of PHB2 as a Potential Biomarker of Luminal A Breast Cancer Cells Using a Cell-Specific Aptamer. ACS Appl Mater Interfaces *14*, 51593-51601. 10.1021/acsami.2c12291.

51. Hsieh, T.H., Hsu, C.Y., Tsai, C.F., Long, C.Y., Wu, C.H., Wu, D.C., Lee, J.N., Chang, W.C., and Tsai, E.M. (2015). HDAC inhibitors target HDAC5, upregulate microRNA-125a-5p, and induce apoptosis in breast cancer cells. Mol Ther *23*, 656-666. 10.1038/mt.2014.247.
52. Oltra, S.S., Cejalvo, J.M., Tormo, E., Albanell, M., Ferrer, A., Nacher, M., Bermejo, B., Hernando, C., Chirivella, I., Alonso, E., et al. (2020). HDAC5 Inhibitors as a Potential Treatment in Breast Cancer Affecting Very Young Women. Cancers (Basel) *12*. 10.3390/cancers12020412.
53. Khera, A.V., and Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet *18*, 331-344. 10.1038/nrg.2016.160.
54. Aragam, K.G., Jiang, T., Goel, A., Kanoni, S., Wolford, B.N., Atri, D.S., Weeks, E.M., Wang, M., Hindy, G., Zhou, W., et al. (2022). Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nat Genet *54*, 1803-1815. 10.1038/s41588-022-01233-6.
55. Jeanne, M., Jorgensen, J., and Gould, D.B. (2015). Molecular and Genetic Analyses of Collagen Type IV Mutant Mouse Models of Spontaneous Intracerebral Hemorrhage Identify Mechanisms for Stroke Prevention. Circulation *131*, 1555-1565. 10.1161/CIRCULATIONAHA.114.013395.
56. Turner, A.W., Nikpay, M., Silva, A., Lau, P., Martinuk, A., Linseman, T.A., Soubeyrand, S., and McPherson, R. (2015). Functional interaction between COL4A1/COL4A2 and SMAD3 risk loci for coronary artery disease. Atherosclerosis *242*, 543-552. 10.1016/j.atherosclerosis.2015.08.008.
57. Zhang, L., Zhou, J., and Kong, W. (2025). Extracellular matrix in vascular homeostasis and disease. Nat Rev Cardiol. 10.1038/s41569-024-01103-0.
58. Hertle, E., Stehouwer, C.D., and van Greevenbroek, M.M. (2014). The complement system in human cardiometabolic disease. Mol Immunol *61*, 135-148. 10.1016/j.molimm.2014.06.031.
59. Onat, A., Can, G., Rezvani, R., and Cianflone, K. (2011). Complement C3 and cleavage products in cardiometabolic risk. Clin Chim Acta *412*, 1171-1179. 10.1016/j.cca.2011.03.005.
60. Kiss, M.G., and Binder, C.J. (2022). The multifaceted impact of complement on atherosclerosis. Atherosclerosis *351*, 29-40. 10.1016/j.atherosclerosis.2022.03.014.
61. Tardiff, D.F., Jui, N.T., Khurana, V., Tambe, M.A., Thompson, M.L., Chung, C.Y., Kamadurai, H.B., Kim, H.T., Lancaster, A.K., Caldwell, K.A., et al. (2013). Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. Science *342*, 979-983. 10.1126/science.1245321.
62. Chung, C.Y., Khurana, V., Auluck, P.K., Tardiff, D.F., Mazzulli, J.R., Soldner, F., Baru, V., Lou, Y., Freyzon, Y., Cho, S., et al. (2013). Identification and rescue of alpha-synuclein toxicity in Parkinson patient-derived neurons. Science *342*, 983-987. 10.1126/science.1245296.
63. Chung, C.Y., Khurana, V., Yi, S., Sahni, N., Loh, K.H., Auluck, P.K., Baru, V., Udeshi, N.D., Freyzon, Y., Carr, S.A., et al. (2017). In Situ Peroxidase Labeling and Mass-Spectrometry Connects Alpha-Synuclein Directly to Endocytic Trafficking and mRNA Metabolism in Neurons. Cell Syst *4*, 242-250 e244. 10.1016/j.cels.2017.01.002.

64. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol *18*, 1091-1102. 10.1016/S1474-4422(19)30320-5.
65. Kim, J.J., Vitale, D., Otani, D.V., Lian, M.M., Heilbron, K., andMe Research, T., Iwaki, H., Lake, J., Solsberg, C.W., Leonard, H., et al. (2024). Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. Nat Genet *56*, 27-36. 10.1038/s41588-023-01584-8.
66. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045-2047. 10.1126/science.276.5321.2045.
67. Nichols, W.C., Pankratz, N., Hernandez, D., Paisan-Ruiz, C., Jain, S., Halter, C.A., Michaels, V.E., Reed, T., Rudolph, A., Shults, C.W., et al. (2005). Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. Lancet *365*, 410-412. 10.1016/S0140-6736(05)17828-3.
68. Funayama, M., Nishioka, K., Li, Y., and Hattori, N. (2023). Molecular genetics of Parkinson's disease: Contributions and global trends. J Hum Genet *68*, 125-130. 10.1038/s10038-022-01058-5.
69. Towns, C., Richer, M., Jasaityte, S., Stafford, E.J., Joubert, J., Antar, T., Martinez-Carrasco, A., Makarious, M.B., Casey, B., Vitale, D., et al. (2023). Defining the causes of sporadic Parkinson's disease in the global Parkinson's genetics program (GP2). NPJ Parkinsons Dis *9*, 131. 10.1038/s41531-023-00533-w.
70. Bandres-Ciga, S., Saez-Atienzar, S., Kim, J.J., Makarious, M.B., Faghri, F., Diez-Fairen, M., Iwaki, H., Leonard, H., Botia, J., Ryten, M., et al. (2020). Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathol *140*, 341-358. 10.1007/s00401-020-02181-3.
71. Foo, J.N., Chew, E.G.Y., Chung, S.J., Peng, R., Blauwendraat, C., Nalls, M.A., Mok, K.Y., Satake, W., Toda, T., Chao, Y., et al. (2020). Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. JAMA Neurol *77*, 746-754. 10.1001/jamaneurol.2020.0428.
72. Maiti, P., Manna, J., and Dunbar, G.L. (2017). Current understanding of the molecular mechanisms in Parkinson's disease: Targets for potential treatments. Transl Neurodegener *6*, 28. 10.1186/s40035-017-0099-z.
73. Braak, H., Del Tredici, K., Rub, U., de Vos, R.A., Jansen Steur, E.N., and Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging *24*, 197-211. 10.1016/s0197-4580(02)00065-9.
74. Tang, D., Kang, R., Livesey, K.M., Cheh, C.W., Farkas, A., Loughran, P., Hoppe, G., Bianchi, M.E., Tracey, K.J., Zeh, H.J., 3rd, and Lotze, M.T. (2010). Endogenous HMGB1 regulates autophagy. J Cell Biol *190*, 881-892. 10.1083/jcb.200911078.

75. Wong, Y.C., and Holzbaur, E.L. (2015). Temporal dynamics of PARK2/parkin and OPTN/optineurin recruitment during the mitophagy of damaged mitochondria. Autophagy *11*, 422-424. 10.1080/15548627.2015.1009792.
76. Takahashi, M., Kitaura, H., Kakita, A., Kakihana, T., Katsuragi, Y., Nameta, M., Zhang, L., Iwakura, Y., Nawa, H., Higuchi, M., et al. (2018). USP10 Is a Driver of Ubiquitinated Protein Aggregation and Aggresome Formation to Inhibit Apoptosis. iScience *9*, 433-450. 10.1016/j.isci.2018.11.006.
77. Ham, S.J., Lee, D., Yoo, H., Jun, K., Shin, H., and Chung, J. (2020). Decision between mitophagy and apoptosis by Parkin via VDAC1 ubiquitination. Proc Natl Acad Sci U S A *117*, 4281-4291. 10.1073/pnas.1909814117.
78. Chu, Y., Goldman, J.G., Kelly, L., He, Y., Waliczek, T., and Kordower, J.H. (2014). Abnormal alpha-synuclein reduces nigral voltage-dependent anion channel 1 in sporadic and experimental Parkinson's disease. Neurobiol Dis *69*, 1-14. 10.1016/j.nbd.2014.05.003.
79. He, Y., Wang, W., Yang, T., Thomas, E.R., Dai, R., and Li, X. (2022). The Potential Role of Voltage-Dependent Anion Channel in the Treatment of Parkinson's Disease. Oxid Med Cell Longev *2022*, 4665530. 10.1155/2022/4665530.
80. Tofaris, G.K., Kim, H.T., Hourez, R., Jung, J.W., Kim, K.P., and Goldberg, A.L. (2011). Ubiquitin ligase Nedd4 promotes alpha-synuclein degradation by the endosomal-lysosomal pathway. Proc Natl Acad Sci U S A *108*, 17004-17009. 10.1073/pnas.1109356108.
81. McLean, P.J., Kawamata, H., Shariff, S., Hewett, J., Sharma, N., Ueda, K., Breakefield, X.O., and Hyman, B.T. (2002). TorsinA and heat shock proteins act as molecular chaperones: suppression of alpha-synuclein aggregation. J Neurochem *83*, 846-854. 10.1046/j.1471-4159.2002.01190.x.
82. Fu, X., Qu, L., Xu, H., and Xie, J. (2024). Ndfip1 protected dopaminergic neurons via regulating mitochondrial function and ferroptosis in Parkinson's disease. Exp Neurol *375*, 114724. 10.1016/j.expneurol.2024.114724.
83. Howitt, J., Gysbers, A.M., Ayton, S., Carew-Jones, F., Putz, U., Finkelstein, D.I., Halliday, G.M., and Tan, S.S. (2014). Increased Ndfip1 in the substantia nigra of Parkinsonian brains is associated with elevated iron levels. PLoS One *9*, e87119. 10.1371/journal.pone.0087119.
84. Villaescusa, J.C., Li, B., Toledo, E.M., Rivetti di Val Cervo, P., Yang, S., Stott, S.R., Kaiser, K., Islam, S., Gyllborg, D., Laguna-Goya, R., et al. (2016). A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. EMBO J *35*, 1963-1978. 10.15252/embj.201593725.
85. Pranski, E., Van Sanford, C.D., Dalal, N., Orr, A.L., Karmali, D., Cooper, D.S., Gearing, M., Lah, J.J., Levey, A.I., and Betarbet, R. (2013). NF-kappaB activity is inversely correlated to RNF11 expression in Parkinson's disease. Neurosci Lett *547*, 16-20. 10.1016/j.neulet.2013.04.056.
86. Tian, R., Gachechiladze, M.A., Ludwig, C.H., Laurie, M.T., Hong, J.Y., Nathaniel, D., Prabhu, A.V., Fernandopulle, M.S., Patel, R., Abshari, M., et al. (2019). CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. Neuron *104*, 239-255 e212. 10.1016/j.neuron.2019.07.014.

87. Lam, I., Ndayisaba, A., Lewis, A.J., Fu, Y., Sagredo, G.T., Zaccagnini, L., Sandoe, J., Sanz, R.L., Vahdatshoar, A., Martin, T.D., et al. (2022). Rapid iPSC inclusionopathy models shed light on formation, consequence and molecular subtype of α-synuclein inclusions. bioRxiv, 2022.2011.2008.515615. 10.1101/2022.11.08.515615.
88. Grattan, D.R. (2024). Does the brain make prolactin? J Neuroendocrinol *36*, e13432. 10.1111/jne.13432.
89. Featherstone, K., White, M.R., and Davis, J.R. (2012). The prolactin gene: a paradigm of tissue-specific gene regulation with complex temporal transcription dynamics. J Neuroendocrinol *24*, 977-990. 10.1111/j.1365-2826.2012.02310.x.
90. Lam, I., Ndayisaba, A., Lewis, A.J., Fu, Y., Sagredo, G.T., Kuzkina, A., Zaccagnini, L., Celikag, M., Sandoe, J., Sanz, R.L., et al. (2024). Rapid iPSC inclusionopathy models shed light on formation, consequence, and molecular subtype of alpha-synuclein inclusions. Neuron *112*, 2886-2909 e2816. 10.1016/j.neuron.2024.06.002.
91. Luk, K.C., Covell, D.J., Kehm, V.M., Zhang, B., Song, I.Y., Byrne, M.D., Pitkin, R.M., Decker, S.C., Trojanowski, J.Q., and Lee, V.M. (2016). Molecular and Biological Compatibility with Host Alpha-Synuclein Influences Fibril Pathogenicity. Cell Rep *16*, 3373-3387. 10.1016/j.celrep.2016.08.053.
92. Loor, G., Kondapalli, J., Schriewer, J.M., Chandel, N.S., Vanden Hoek, T.L., and Schumacker, P.T. (2010). Menadione triggers cell death through ROS-dependent mechanisms involving PARP activation without requiring apoptosis. Free Radic Biol Med *49*, 1925-1936. 10.1016/j.freeradbiomed.2010.09.021.
93. Di Nubila, A., Dilella, G., Simone, R., and Barbieri, S.S. (2024). Vascular Extracellular Matrix in Atherosclerosis. Int J Mol Sci *25*. 10.3390/ijms252212017.
94. Manon-Jensen, T., Kjeld, N.G., and Karsdal, M.A. (2016). Collagen-mediated hemostasis. J Thromb Haemost *14*, 438-448. 10.1111/jth.13249.
95. Libby, P. (2021). Inflammation during the life cycle of the atherosclerotic plaque. Cardiovasc Res *117*, 2525-2536. 10.1093/cvr/cvab303.
96. Sakata, T., Sakaguchi, H., Tsuda, L., Higashitani, A., Aigaki, T., Matsuno, K., and Hayashi, S. (2004). Drosophila Nedd4 regulates endocytosis of notch and suppresses its ligand-independent activation. Curr Biol *14*, 2228-2236. 10.1016/j.cub.2004.12.028.
97. Privman Champaloux, E., Donelson, N., Pyakurel, P., Wolin, D., Ostendorf, L., Denno, M., Borman, R., Burke, C., Short-Miller, J.C., Yoder, M.R., et al. (2021). Ring Finger Protein 11 (RNF11) Modulates Dopamine Release in Drosophila. Neuroscience *452*, 37-48. 10.1016/j.neuroscience.2020.10.021.
98. Wakabayashi-Ito, N., Doherty, O.M., Moriyama, H., Breakefield, X.O., Gusella, J.F., O'Donnell, J.M., and Ito, N. (2011). Dtorsin, the Drosophila ortholog of the early-onset dystonia TOR1A (DYT1), plays a novel role in dopamine metabolism. PLoS One *6*, e26183. 10.1371/journal.pone.0026183.
99. Barallobre, M.J., Perier, C., Bove, J., Laguna, A., Delabar, J.M., Vila, M., and Arbones, M.L. (2014). DYRK1A promotes dopaminergic neuron survival in the developing brain and in a mouse model of Parkinson's disease. Cell Death Dis *5*, e1289. 10.1038/cddis.2014.253.

100. Yong, Y., Wu, Q., Meng, X., Lu, R., Xia, H., Pei, F., and Yang, X. (2023). Dyrk1a Phosphorylation of alpha-Synuclein Mediating Apoptosis of Dopaminergic Neurons in Parkinson's Disease. Parkinsons Dis *2023*, 8848642. 10.1155/2023/8848642.
101. Ye, H., Robak, L.A., Yu, M., Cykowski, M., and Shulman, J.M. (2023). Genetics and Pathogenesis of Parkinson's Syndrome. Annu Rev Pathol *18*, 95-121. 10.1146/annurev-pathmechdis-031521-034145.
102. Costa-Brito, A.R., Goncalves, I., and Santos, C.R.A. (2022). The brain as a source and a target of prolactin in mammals. Neural Regen Res *17*, 1695-1702. 10.4103/1673-5374.332124.
103. Ramos-Martinez, E., Ramos-Martinez, I., Molina-Salinas, G., Zepeda-Ruiz, W.A., and Cerbon, M. (2021). The role of prolactin in central nervous system inflammation. Rev Neurosci *32*, 323-340. 10.1515/revneuro-2020-0082.
104. Molina-Salinas, G., Rodriguez-Chavez, V., Langley, E., and Cerbon, M. (2023). Prolactin-induced neuroprotection against excitotoxicity is mediated via PI3K/AKT and GSK3beta/NF-kappaB in primary cultures of hippocampal neurons. Peptides *166*, 171037. 10.1016/j.peptides.2023.171037.
105. Macias, F., Ulloa, M., Clapp, C., Martinez de la Escalera, G., and Arnold, E. (2024). Prolactin protects hippocampal neurons against H2O2-induced neurotoxicity by suppressing BAX and NOX4 via the NF-kappaB signaling pathway. PLoS One *19*, e0313328. 10.1371/journal.pone.0313328.
106. Karayel, O., Virreira Winter, S., Padmanabhan, S., Kuras, Y.I., Vu, D.T., Tuncali, I., Merchant, K., Wills, A.M., Scherzer, C.R., and Mann, M. (2022). Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. Cell Rep Med *3*, 100661. 10.1016/j.xcrm.2022.100661.
107. Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet *83*, 311-321. 10.1016/j.ajhg.2008.06.024.
108. Derkach, A., Lawless, J.F., and Sun, L. (2013). Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. Genet Epidemiol *37*, 110-121. 10.1002/gepi.21689.
109. Nyholt, D.R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet *74*, 765-769. 10.1086/383251.
110. Higham, N.J. (2002). Computing the nearest correlation matrix—a problem from finance. IMA journal of Numerical Analysis *22*, 329-343.
111. 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. 10.1093/nar/gkac1000.
112. Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Staerfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods *14*, 61-64. 10.1038/nmeth.4083.

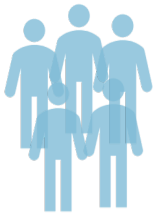
113. Megchelenbrink, W., Katzir, R., Lu, X., Ruppin, E., and Notebaart, R.A. (2015). Synthetic dosage lethality in the human metabolic network is highly predictive of tumor growth and cancer patient survival. Proc Natl Acad Sci U S A *112*, 12217-12222. 10.1073/pnas.1508573112.
114. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318-1330. 10.1126/science.aaz1776.
115. Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet *49*, 1779-1784. 10.1038/ng.3984.
116. Dempster, J.M., Rossen, J., Kazachkova, M., Pan, J., Kugener, G., Root, D.E., and Tsherniak, A. (2019). Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. bioRxiv, 720243. 10.1101/720243.
117. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet *99*, 877-885. 10.1016/j.ajhg.2016.08.016.
118. Cheng, J., Novati, G., Pan, J., Bycroft, C., Zemgulyte, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. 10.1126/science.adg7492.
119. All of Us Research Program Genomics, I. (2024). Genomic data in the All of Us Research Program. Nature *627*, 340-346. 10.1038/s41586-023-06957-x.
120. Gonzalez, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215-226. 10.1016/j.stem.2014.05.018.
121. DeWeirdt, P.C., Sangree, A.K., Hanna, R.E., Sanson, K.R., Hegde, M., Strand, C., Persky, N.S., and Doench, J.G. (2020). Genetic screens in isogenic mammalian cell lines without single cell cloning. Nat Commun *11*, 752. 10.1038/s41467-020-14620-6.
122. Kim, T.W., Piao, J., Koo, S.Y., Kriks, S., Chung, S.Y., Betel, D., Socci, N.D., Choi, S.J., Zabierowski, S., Dubose, B.N., et al. (2021). Biphasic Activation of WNT Signaling Facilitates the Derivation of Midbrain Dopamine Neurons from hESCs for Translational Use. Cell Stem Cell *28*, 343-355 e345. 10.1016/j.stem.2021.01.005.
123. Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., et al. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nat Protoc *14*, 756-780. 10.1038/s41596-018-0113-7.
124. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol *11*, e1004219. 10.1371/journal.pcbi.1004219.
125. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661. 10.1016/j.cell.2014.09.029.

126. Kampmann, M., Bassik, M.C., and Weissman, J.S. (2014). Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. Nat Protoc *9*, 1825-1847. 10.1038/nprot.2014.103.

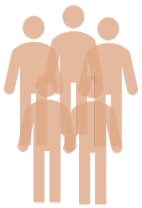
Figure 1

WGS/WES cohort

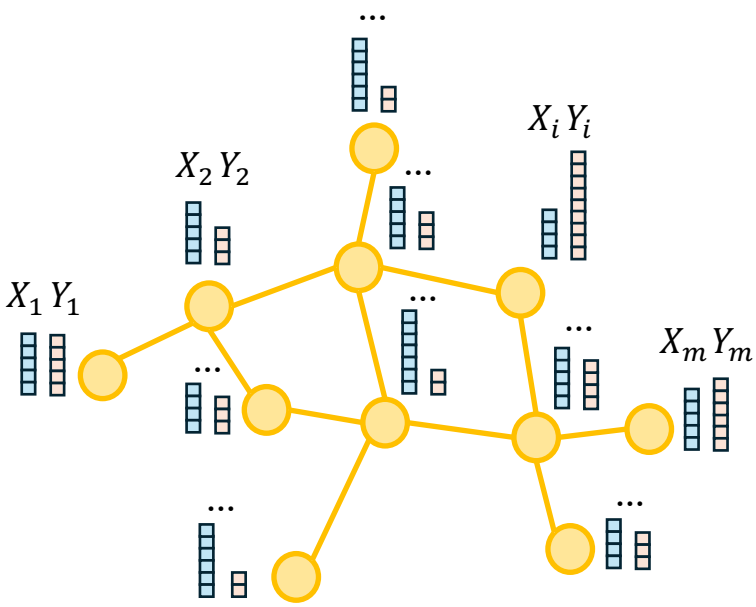
Gene Network: $\Sigma_{m\times m}$



$\boldsymbol{X}$: Case allele counts



$\boldsymbol{Y}$: Control allele counts



$$\vec{\alpha} \sim \mathrm{MSN}(0, \theta \cdot \Sigma, \nu = f(N_{case}, N_{control}))$$

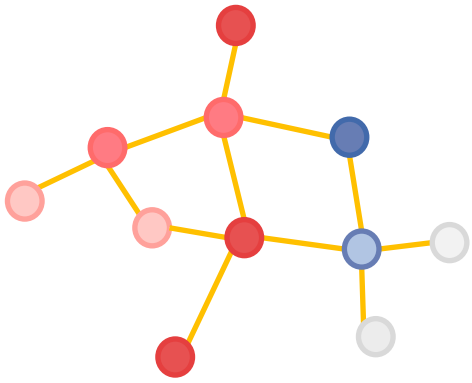
⇩

NERINE's MLE framework for inferring $\boldsymbol{\theta}$

$$\mathrm{L}(\theta|\boldsymbol{X}, \boldsymbol{Y}, \Sigma, N_{case}, N_{control}) = \int_{\vec{\alpha}} \left(\prod_{i=1}^{m} P(X_i|X_i + Y_i, \alpha_i) \right) \times P(\vec{\alpha}|\,\theta; \Sigma, \nu) d\vec{\alpha}$$

LLR ~ mixture of χ_1^2 and $\delta(x)$

⇩



Estimated network effect, $\hat{\theta}$

Network significance: *p*-value

MLE gene effects: $\hat{\vec{\alpha}}$

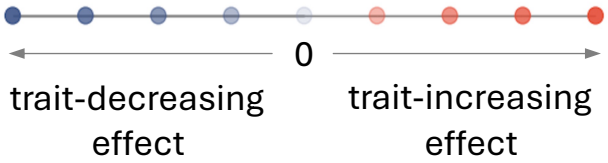


Figure 1. Overview of NERINE: rare variant test with unique ability to leverage the edge geometry of gene networks. Schematic diagram showing NERINE, a new statistical framework that infers the effect of rare variants aggregated across a gene network on a trait. Inputs include allele counts in cases ($\boldsymbol{X}$) and controls ($\boldsymbol{Y}$) from whole-genome or whole-exome sequencing datasets (WGS/WES) and a gene network encoded by a symmetric positive semidefinite matrix, $\Sigma_{\mathrm{m}\times\mathrm{m}}$. Network edges can represent diverse biological relationships (e.g., physical or genetic interactions, co-expression, co-essentiality, pathway membership). They can be extracted from either existing databases or novel experimental assays. NERINE models gene-effects, ($\vec{\alpha}$) using a multivariate skew-normal distribution parameterized by the input network structure (Σ), the network effect, θ, and the case-control skew $\nu = f(N_{case}, N_{control})$. This parameterization allows the network genes to have either zero effect or varying degrees of trait-increasing and decreasing effects. NERINE infers θ using a maximum likelihood estimation (MLE) framework where the likelihood function, L is an integral over the product of two terms—(i) the product of the per-gene conditional probability of observing allele count X_i in cases given the allele count in the overall cohort ($X_i + Y_i$) in the network, and (ii) the probability of observing a specific combination of gene effects determined by θ and Σ. NERINE performs nested hypothesis testing, with log-likelihood ratio (LLR) being the test statistic, and provides an asymptotic p-value from the mixture of the delta function, $\delta(x)$, and a chi-square distribution with one degree of freedom ($\chi^2_{1_}$). It also estimates the most likely combination of gene effects, $\hat{\vec{\alpha}}$ with the estimated network effect, $\hat{\theta}$ (see Methods).

Figure 2

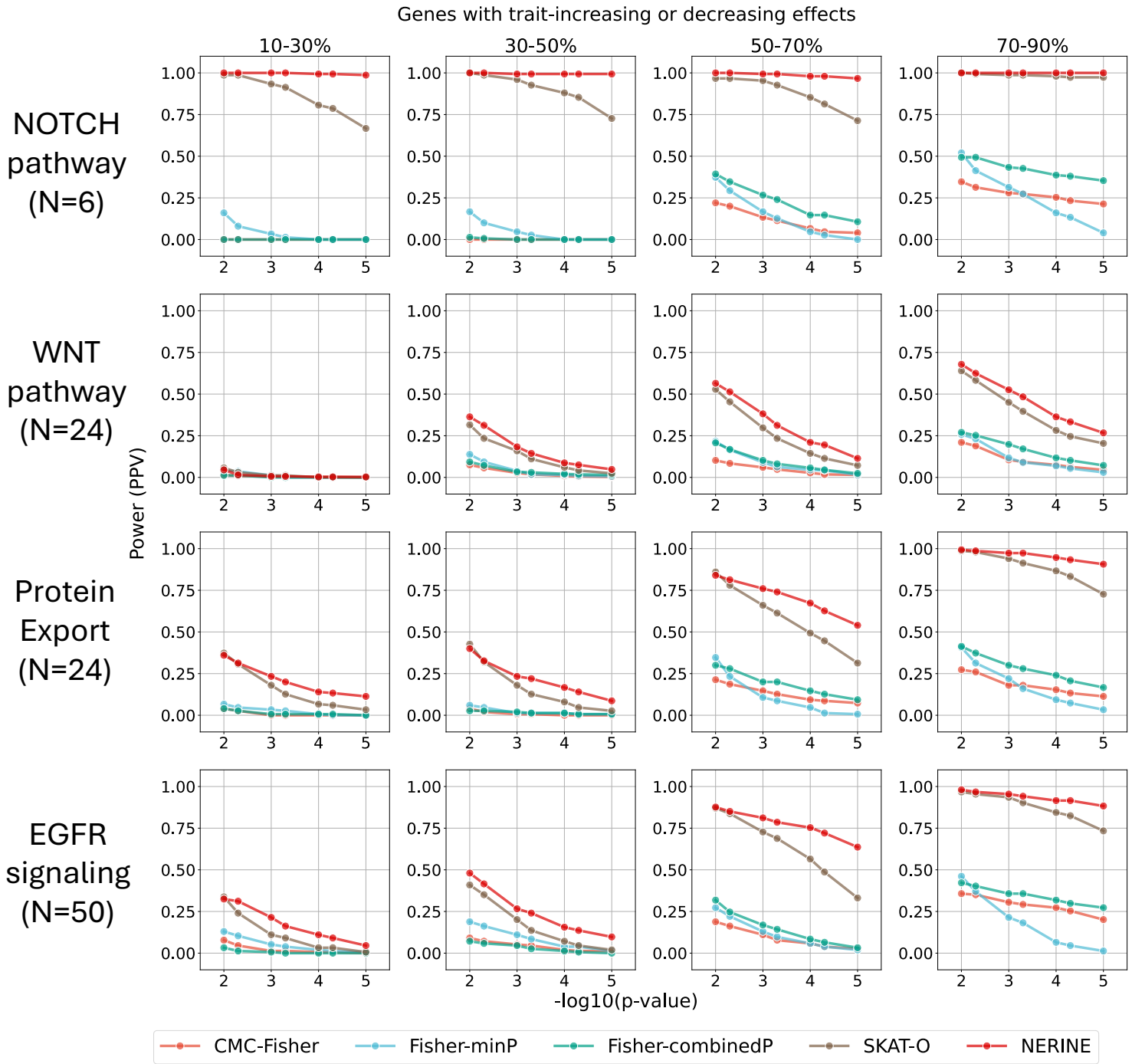

Genes with trait-increasing or decreasing effects
10-30%
30-50%
50-70%
70-90%
NOTCH pathway (N=6)
WNT pathway (N=24)
Protein Export (N=24)
EGFR signaling (N=50)
Power (PPV)
1.00
0.75
0.50
0.25
0.00
2
3
4
5
-log10(p-value)
CMC-Fisher
Fisher-minP
Fisher-combinedP
SKAT-O
NERINE

Figure 2. NERINE outperforms existing rare variant association tests in power simulations. The empirical power of each method was evaluated using a simulated binary disease trait in a cohort of 2,000 cases and 2,000 controls under a non-zero network effect ($\theta = 0.2$). Four well-characterized biological pathways—NOTCH signaling, WNT signaling, protein export, and EGFR signaling—were used as representative networks. Within each network, genes were assigned both trait-increasing and trait-decreasing effects. Power was assessed across scenarios in which varying proportions of genes (10–90%) exerted non-zero effects on the trait, representing conditions from highly noisy (only 10–30% genes with effect) to highly informative (70–90% genes with effect) networks. For each noise profile, 250 iterations were performed per network, and power was quantified as the positive predictive value (PPV) at different significance thresholds ($p = 1 \times 10^{-2}$, 5×10^{-3}, 1×10^{-3}, 5×10^{-4}, 1×10^{-4}, 5×10^{-5}, and 1×10^{-5}). Across all conditions, NERINE consistently outperformed existing rare-variant association tests, with the greatest gains observed in noisy network settings.

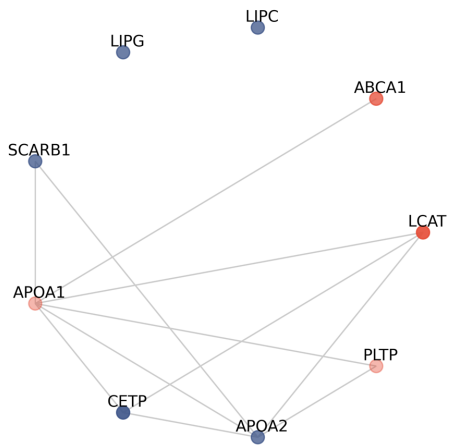
Figure 3

A.

Phenotype: Low HDL vs High HDL

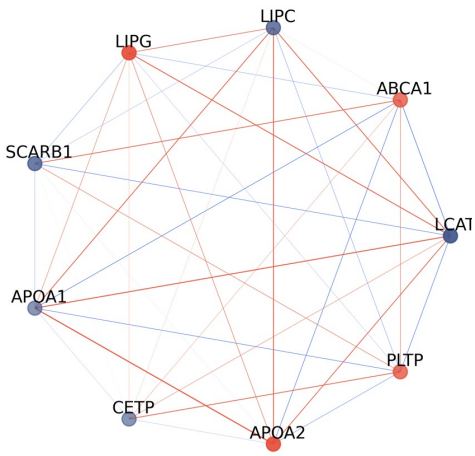
Variant class: Damaging, MAF cutoff: 0.001

Physical & Genetic interactions



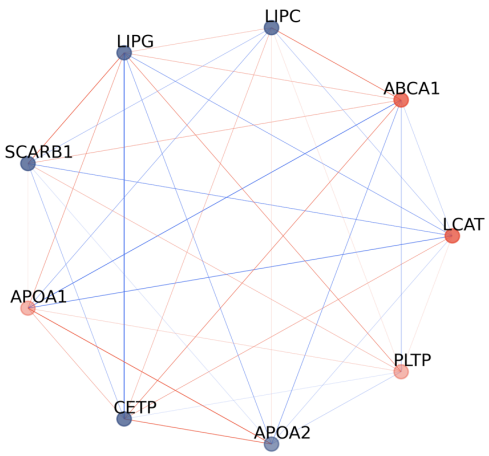
Bonf. p-value: 1.8360e-71
Effect size, $\hat{\theta} = 0.5$

Coexpression in Liver (GTEx v8)

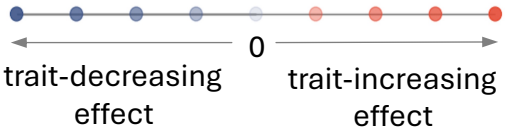


Bonf. p-value: **1.3435e-73**
Effect size, $\hat{\theta} = \mathbf{0.8}$

Coessentiality in Liver cell lines (DepMap)



Bonf. p-value: 2.0260e-66
Effect size, $\hat{\theta} = 0.2$



B.

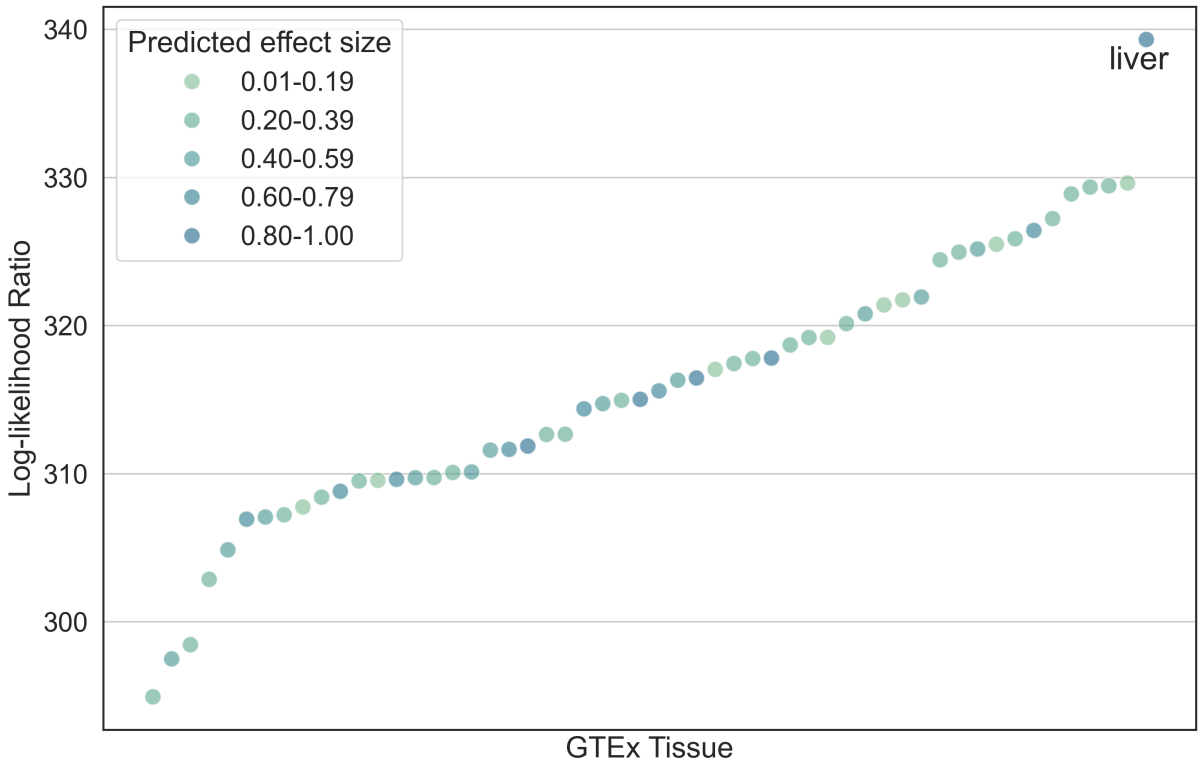
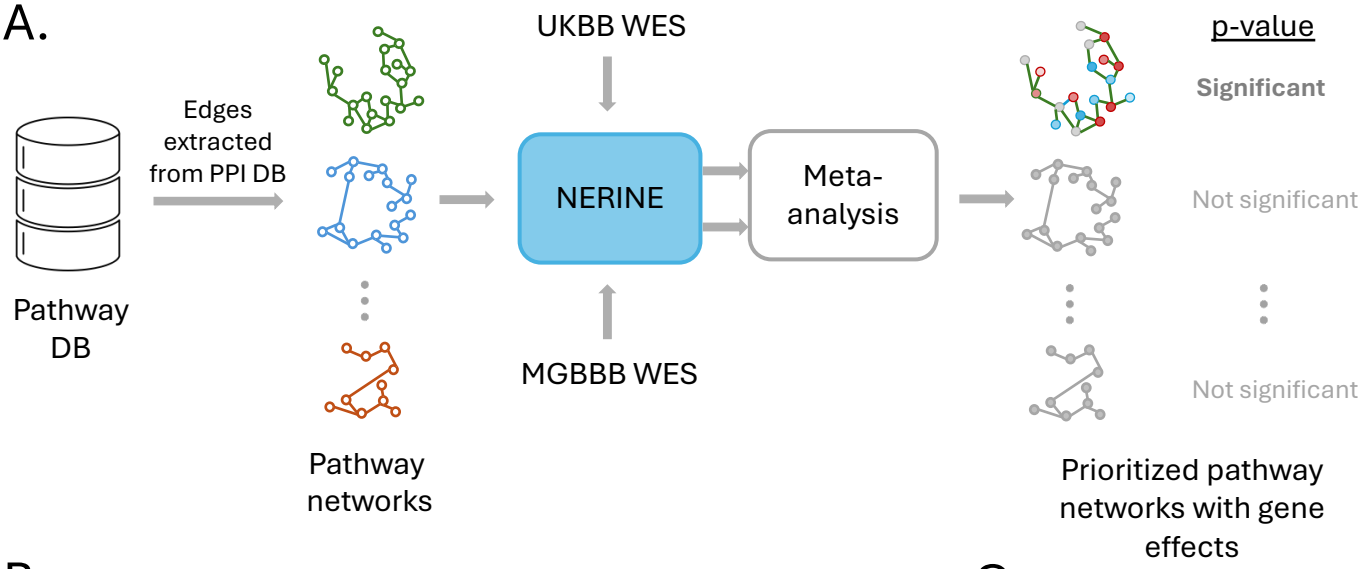


Figure 3. NERINE helps select the most informative network topology

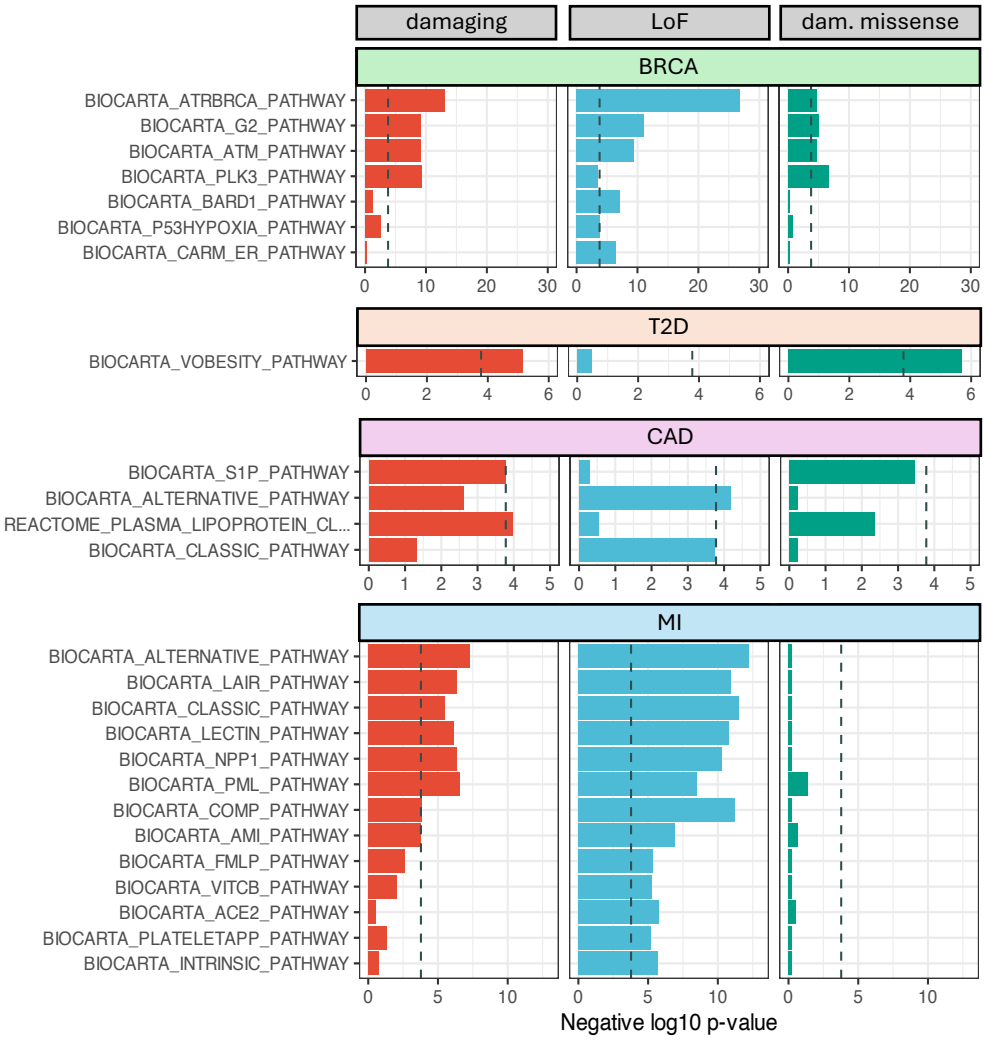
A. NERINE selects the most informative edge geometry describing the relationships of a given gene set. For core HDL-cholesterol-related genes, NERINE competitively selects co-expression as the most informative data source among the three data sources used to generate network topologies—(i) high-confidence physical or genetic interactions from protein-protein interaction (PPI) databases, (ii) co-expression in liver from GTEx (v8), and (iii) co-essentiality in liver cell lines from DepMap (2023Q2) (Methods). For each topology, NERINE assessed the enrichment of rare damaging variants in core HDL-cholesterol-related genes in individuals with low HDL (bottom 10%) as compared to individuals with high HDL (top 10%) in the UK biobank (UKBB). This finding demonstrates NERINE's unique ability to leverage edge geometry unlike existing rare variant tests. Here, trait-increasing effects are represented by shades of orange, and trait-decreasing effects are represented by shades of purple, as shown on the scale. A darker color represents a more pronounced effect in each direction. For co-expression and co-essentiality networks, a positive correlation between two genes is indicated by red edges and a negative correlation is indicated by blue edges. The thickness of the edge corresponds to the strength of the correlation. The edges in the physical and genetic interactions network do not represent correlations but binary relationships and are therefore colored in gray. **B.** NERINE identifies liver as the most informative tissue context for the core HDL-cholesterol-related gene module among all tissue types in GTEx (v8). We extracted the co-expression relationships for all 52 tissue types in GTEx (Supplementary Table T4) for the same core module of HDL-related genes. We tested each network topology for rare damaging variant enrichment in individuals with low HDL vs. high HDL cholesterol in UKBB with NERINE. Our method achieved the most significant log-likelihood ratio (LLR) and p-value with the co-expression network derived from liver tissue.

Figure 4

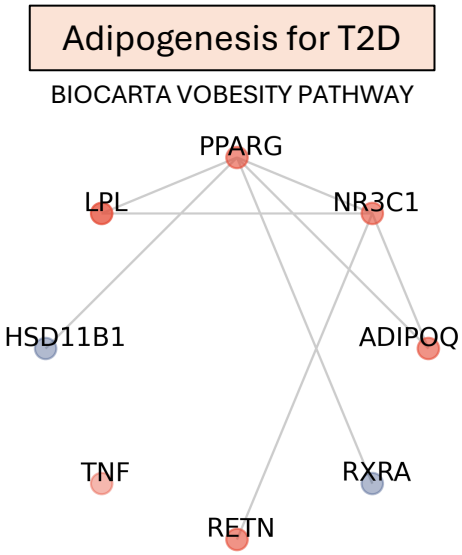
A.



B.



C.



Bonf. p-value: **2.33e-3**

Effect size, $\hat{\theta} = \mathbf{0.18}$

Gene symbol	GWAS hit	Known rare variant hit	Potential drug target
ADIPOQ	X	X	✓
HSD11B1	X	X	✓
LPL	✓	X	✓
NR3C1	X	X	?
PPARG	✓	X	✓
RETN	X	X	?
RXRA	X	X	✓
TNF	X	X	?

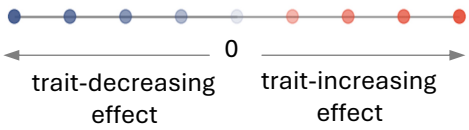
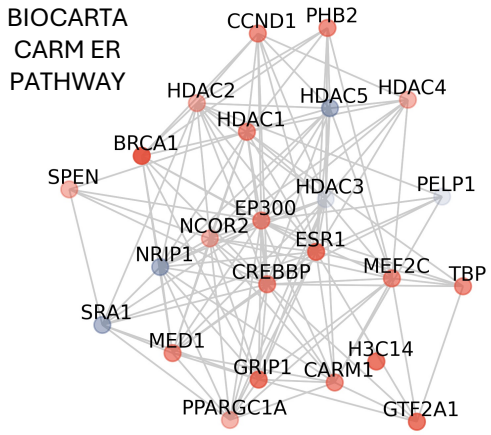


Figure 4. NERINE identifies significant rare variant burden in canonical pathway networks in common diseases in the UK and MGB biobanks.

A. Schematic diagram showing NERINE's approach to testing rare variant burden across our canonical pathway database networks. Network topologies for all 306 database pathways were extracted from high-quality physical and genetics interactions databases (Methods). NERINE tests each network for four phenotypes, namely, breast cancer (BRCA), type II diabetes (T2D), coronary artery disease (CAD), and early-onset myocardial infarction (MI) in UKBB and MGBBB for six variant categories —(i) LoF, (ii) damaging missense, (iii) damaging, (iv) missense, (v) synonymous, and (vi) neutral missense. NERINE's p-values from the two biobanks were meta-analyzed for each network using Fisher's combined test. Bonferroni correction with Nyholt's adjustment was applied to the meta p-values for multiple hypotheses correction (Methods), nominating database-wide significant pathway networks for each phenotype. **B.** Bonferroni-significant findings for different disease phenotypes across the pathway database gene networks in three functional categories of rare variants (i.e., LoF, damaging missense, and damaging) are shown. Negative log-transformed Fisher's combined p-values are reported for significant pathways. The dashed gray line represents the Bonferroni threshold of 0.05. **C.** The adipogenesis pathway (i.e., BIOCARTA VOBESITY pathway) gene module, which showed significant rare damaging variant-burden in T2D, is shown as an example of how NERINE enables novel findings. Individual gene effects predicted by NERINE averaged across the biobanks, are shown by colored nodes in the depicted topology. Trait-increasing effects are represented by shades of orange, and trait-decreasing effects are represented by shades of blue, as shown on the scale. A darker color represents a more pronounced effect. Member genes include *RXRA*, *PPARG*, *ADIPOQ*, *TNF*, *NR3C1*, *LPL*, *RETN*, and *HSD11B1*—none of which were previously implicated in T2D through rare variant association studies (bottom). Only *PPARG* and *LPL* were previously identified in GWAS studies (bottom). The information on the gene-phenotype association in germline population genetics was derived from the GWAS catalog (https://www.ebi.ac.uk/gwas/), Genebass[33], SAIGE-GENE+[19], and disease-specific literature[34-36]. Five out of the eight members of the network are currently under investigation as potential drug targets for T2D[37-43], which further highlights NERINE's ability to uncover therapeutically targetable pathway networks for disease phenotypes.

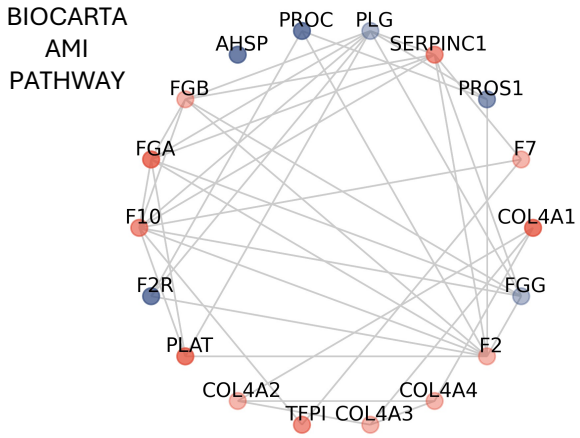
Figure 5

A. Regulation of estrogen receptor for BRCA

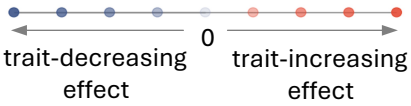


Gene symbol	GWAS hit	Known rare variant hit	Somatic hit	Role in cancer
BRCA1	X	✓	✓	TSG
CARM1	X	X	X	unknown
CCND1	✓	X	✓	oncogene, fusion
CREBBP	X	X	✓	oncogene, TSG, fusion
EP300	X	X	✓	TSG, fusion
ESR1	✓	X	✓	oncogene, TSG, fusion
GRIP1	X	X	X	unknown
GTF2A1	X	X	X	unknown
H3C14	X	X	X	unknown
HDAC1	X	X	X	unknown
HDAC2	X	X	X	unknown
HDAC3	X	X	X	unknown
HDAC4	X	X	X	unknown
HDAC5	X	X	X	unknown
MED1	X	X	X	unknown
MEF2C	X	X	X	unknown
NCOR2	X	X	✓	TSG
NRIP1	✓	X	X	unknown
PELP1	X	X	X	unknown
PHB2	X	X	X	unknown
PPARGC1A	X	X	X	unknown
SPEN	X	X	✓	TSG
SRA1	X	X	X	unknown
TBP	X	X	X	unknown

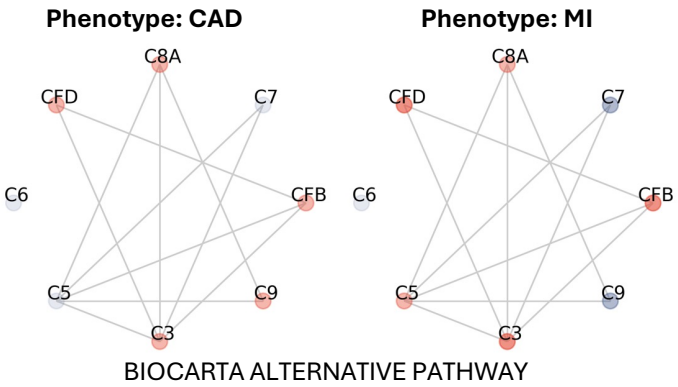
B. prothrombin activation and coagulation for MI



Gene symbol	GWAS hit	Known rare variant hit	Role relevant to early MI
AHSP	X	X	hemoglobin stabilization
COL4A1	✓	X	ECM/vascular homeostasis
COL4A2	✓	X	ECM/vascular homeostasis
COL4A3	X	X	ECM/vascular homeostasis
COL4A4	✓	X	ECM/vascular homeostasis
F10	X	X	blood coagulation
F2	X	X	blood coagulation
F2R	X	X	platelet aggregation and thrombus formation
F7	X	X	blood coagulation
FGA	X	X	blood coagulation
FGB	X	X	blood coagulation
FGG	X	X	blood coagulation
PLAT	X	X	breakdown of blood clots
PLG	✓	X	breakdown of blood clots
PROC	X	X	anticoagulant
PROS1	X	X	anticoagulant
SERPINC1	X	X	regulate blood coagulation
TFPI	X	X	regulate blood coagulation



C. Alternative complement pathway for CAD & MI



Gene symbol	GWAS hit in CAD or MI	Known rare variant hit	Role relevant to cardiovascular diseases
C3	X	X	atherosclerotic plaque formation
C5	X	X	mediating pro-inflammatory effects
C6	X	X	formation of the membrane attack complex
C7	X	X	mediating immune cell activation and endothelial injury
C8A	X	X	formation of the membrane attack complex
C9	X	X	formation of the membrane attack complex
CFB	X	X	atherosclerotic plaque formation
CFD	X	X	promoting complement activation

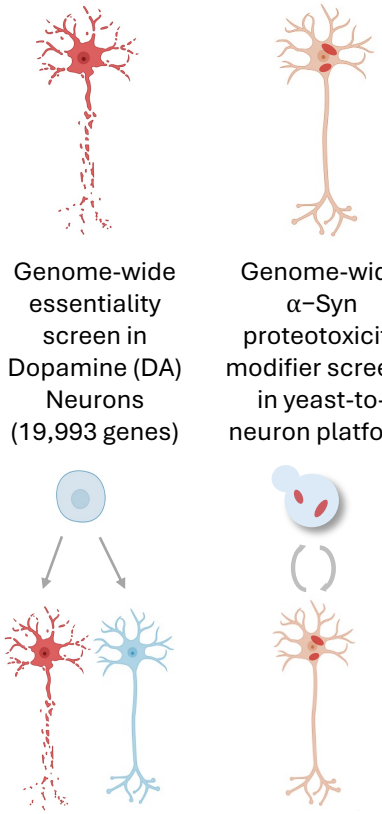
Figure 5. Examples of NERINE's novel findings in breast cancer and cardiovascular diseases. A. (Top) Individual gene effects predicted by NERINE, averaged across the biobanks, are shown by colored nodes in the regulation of estrogen receptor pathway network (i.e., BIOCARTA CARM ER PATHWAY) in breast cancer (BRCA). (Bottom) Known information about member genes based on somatic and germline genetics, along with their functional annotations. **B.** (Top) Individual gene effects predicted by NERINE, averaged across the biobanks, are shown by colored nodes in the depicted network for the intrinsic prothrombin activation and coagulation pathway (i.e., BIOCARTA AMI PATHWAY) in early-onset myocardial infarction (MI). (Bottom) Known information about member genes based on germline genetics along with functional annotations. **C.** Individual gene effects predicted by NERINE averaged across the biobanks are shown by colored nodes in the alternative complement pathway (i.e., BIOCARTA ALTERNATIVE PATHWAY) in (Left) coronary artery disease (CAD) and (Middle) early-onset myocardial infarction (MI). (Right) Known information about member genes based on germline genetics along with functional annotations. The complement signal in cardiovascular diseases showed some disparities between UKBB and MGBBB, which might stem from the fact that MGBBB participants had a higher frequency of cardiac events, with thrombosis and inflammation, compared to those in UKBB. In each depiction of gene networks, trait-increasing effects of gene nodes are represented by shades of orange, and trait-decreasing effects are represented by shades of purple, as shown on the scale. A darker color means a more pronounced effect. The information on whether a member gene was a GWAS hit was derived from the GWAS catalog (https://www.ebi.ac.uk/gwas/). For rare variant hits, we consulted the Genebass[33] and the SAIGE-GENE+[19] analyses and disease-specific literature[47,53,54]. Functional annotations for estrogen receptor pathway members were extracted from the COSMIC database (https://cancer.sanger.ac.uk/cosmic). Annotations indicating known roles of genes relevant to MI and CAD were extracted from the SynGO portal (https://www.syngoportal.org/).

Figure 6

A. Pathological features of PD

(i) DA neuron degeneration

(ii) Lewy body formation

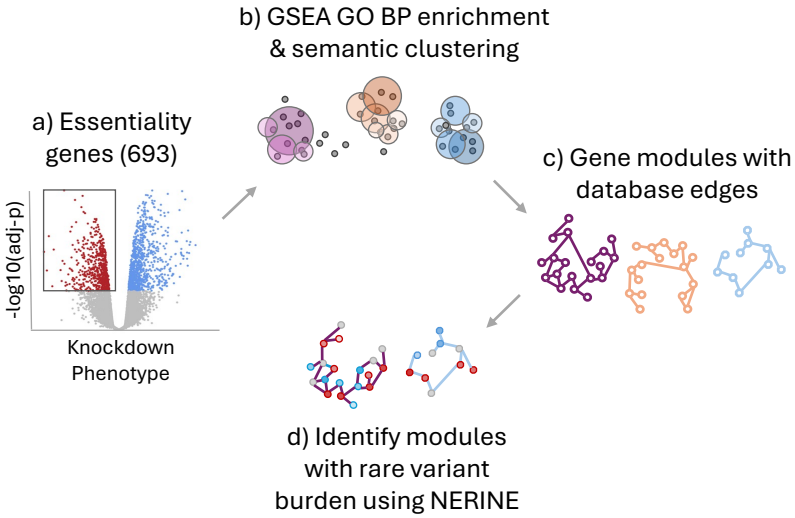


Genome-wide essentiality screen in Dopamine (DA) Neurons (19,993 genes)

Genome-wide α−Syn proteotoxicity modifier screens in yeast-to-neuron platform

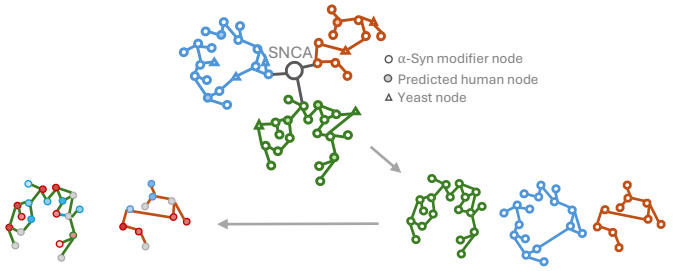
B. NERINE applied on screen-based gene modules

(i) DA neuron essentiality screen



(ii) α-Syn proteotoxicity modifier screens

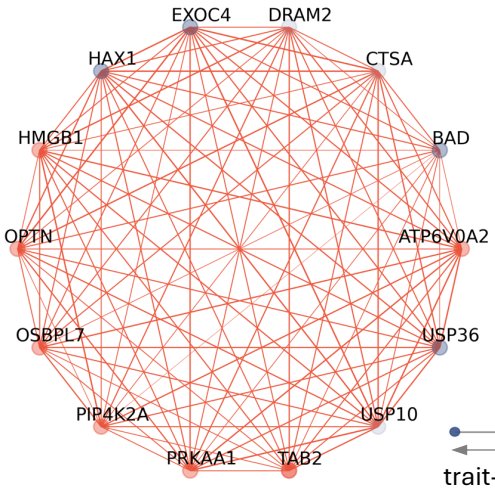
a) TransposeNet: humanized αS network



C. DA neuron essentiality modules

Regulation of autophagy

Category	Avg. $\hat{\theta}$	Fisher combo-p	Screen-wide Bonf. p
Dam. Missense	**0.03**	**1.86E-03**	**0.0186**
Neutral	0	5.97E-01	1
Synonymous	0	5.97E-01	1



D. α-Syn proteotoxicity modules

LRRK2-SNCA vesicle trafficking stem

Category	Avg. $\hat{\theta}$	Fisher combo-p	Screen-wide Bonf. p
Dam. Missense	**0.53**	**1.28E-03**	**0.0217**
Neutral	0	5.97E-01	1
Synonymous	0	5.97E-01	1

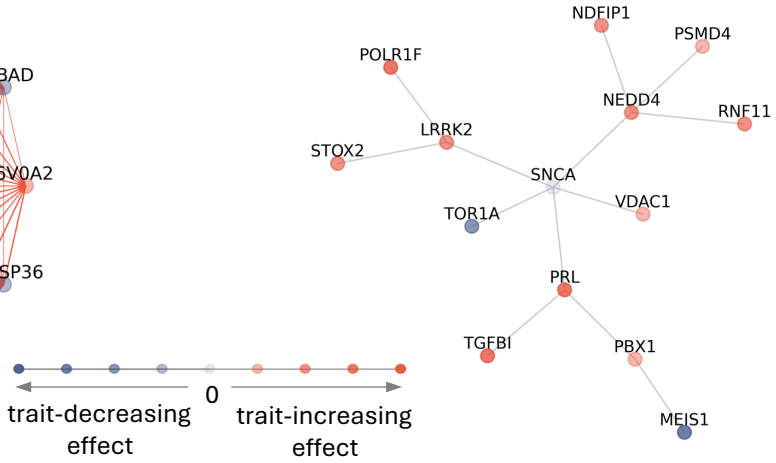


Figure 6. NERINE identifies rare variant burden in bespoke networks from model system screens targeting key pathological features of Parkinson's disease (PD). A. Key pathological features of PD, namely, Lewy body formation and the loss of dopaminergic (DA) neurons, targeted by two screens: (i) DA neuron essentiality screen, and (ii) α-Synuclein (αS) proteotoxicity screen. **B.** Schematic diagram showing the process of network hypotheses generation for interrogating PD pathophysiology using NERINE. Network hypotheses were generated based on (i) essentiality genes for DA neurons, and (ii) α-Synuclein (αS) toxicity modifier genes. From DA essentiality genes, we identified 10 GO biological process (BP) modules, and from the TransposeNet humanized αS modifier gene network, we tested 17 subnetworks. For the DA essentiality gene modules, we tested the edge relationships based on (i) high-confidence physical or genetic interactions from PPI databases, (ii) co-expression of genes in the substantia nigra region of the mid-brain in GTEx v8, and (iii) co-essentiality of genes in cell lines from central nervous system (CNS) in DepMap v2023Q2 (Methods). For αS modifier genes, we directly used the TransposeNet network topology[2]. **C.** NERINE identifies a screen-wide Bonferroni significant burden of rare damaging missense variants across AMP-PD and UKBB cohorts in the GO biological process module related to the *regulation of autophagy* containing key genes, such as, *HMGB1* and *OPTN*. Co-expression in the midbrain *substantia nigra* region provided the most informative edge relationships for this module. **D.** A *LRRK2-SNCA*-containing *vesicle trafficking and protein homeostasis*-related subnetwork of αS-modifier genes show screen-wide significant burden of rare damaging missense variants in sporadic PD cases compared to controls in both AMP-PD and UKBB datasets. For **C** and **D**, screen-wide significance was determined by applying Bonferroni correction over Fisher combined p-values with number of gene modules being tested in each screen. The absence of enrichment of neutral missense and synonymous variants in the networks served as an internal control. In each scenario, trait-increasing effects are represented by shades of orange and trait-decreasing effects are represented by shades of purple, as shown on the scale. A darker color means a more pronounced effect in each direction. For the co-expression network in **C**, color of the edge indicates the sign of the correlation (red: positive; blue: negative) and the thickness of the edge corresponds to the strength of the correlation. The edges in the TransposeNet module in **D** do not represent correlations, rather binary relationships and are therefore colored in grey.

Figure 7

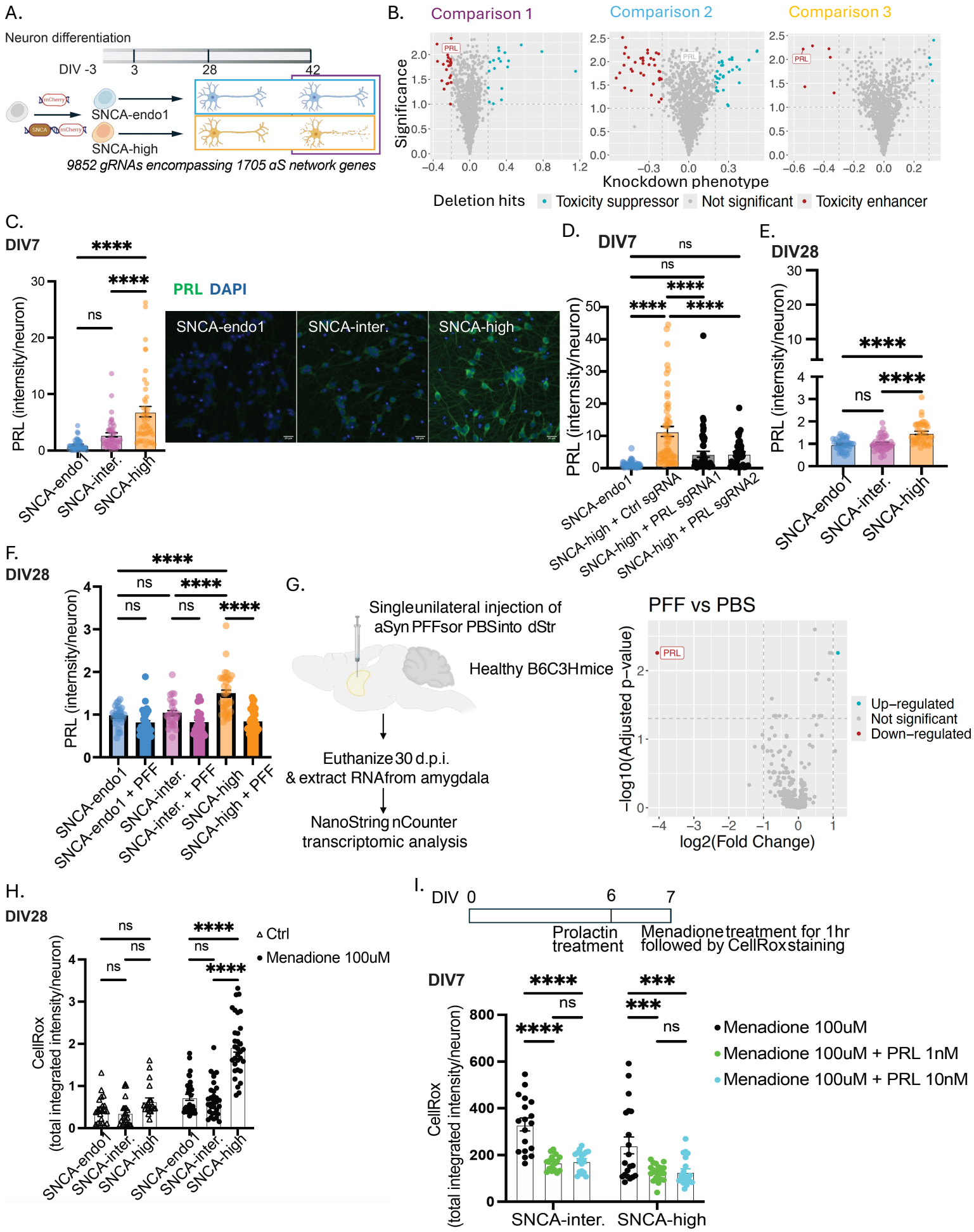

A.
Neuron differentiation
DIV -3
3
28
42
mCherry
SNCA
SNCA-endo1
SNCA-high
9852 gRNAs encompassing 1705 αS network genes
B.
Comparison 1
Comparison 2
Comparison 3
Significance
PRL
Knockdown phenotype
Deletion hits
Toxicity suppressor
Not significant
Toxicity enhancer
C.
DIV7
PRL (internsity/neuron)
ns
****
PRL DAPI
SNCA-endo1
SNCA-inter.
SNCA-high
D.
DIV7
SNCA-high + Ctrl sgRNA
SNCA-high + PRL sgRNA1
SNCA-high + PRL sgRNA2
E.
DIV28
F.
DIV28
SNCA-endo1 + PFF
SNCA-inter. + PFF
SNCA-high + PFF
G.
Single unilateral injection of
aSyn PFF or PBS into dStr
Healthy B6C3H mice
Euthanize 30 d.p.i.
& extract RNA from amygdala
NanoString nCounter
transcriptomic analysis
PFF vs PBS
−log10(Adjusted p−value)
log2(Fold Change)
Up−regulated
Not significant
Down−regulated
H.
DIV28
CellRox
(total integrated intensity/neuron)
Ctrl
Menadione 100uM
I.
DIV 0
6
7
Prolactin treatment
Menadione treatment for 1hr
followed by CellRox staining
DIV7
***
Menadione 100uM
Menadione 100uM + PRL 1nM
Menadione 100uM + PRL 10nM

Figure 7. Unbiased functional genomics screens converge on an intraneuronal *SNCA-PRL* stress response. A. Top: Timeline of CiS-CN differentiation. Bottom: iPSCs were transduced to overexpress either mCherry (control) or SNCA-mCherry, generating the *SNCA-endo1* and *SNCA*-high lines, respectively. At DIV0, neurons were transduced with the sgRNA library. Neuronal samples were harvested and sequenced at DIV3, DIV28, and DIV42 using next-generation sequencing. *Comparison 1* shows the comparison of the sgRNA frequencies in DIV42 *SNCA*-high neuron vs DIV42 *SNCA-endo1* neurons. *Comparison 2* compares the sgRNA frequencies of *SNCA-endo1* neurons in DIV42 vs DIV28. *Comparison 3* compares the sgRNA frequencies of *SNCA*-high neurons in DIV42 to DIV28. **B.** Left: *PRL* sgRNA containing neurons dropped out in Comparison 1, indicating PRL knockdown was toxic to *SNCA*-high neurons compared to *SNCA-endo1* neurons. Middle: No significant drop-out in *PRL* sgRNA containing neurons was observed in Comparison 2, indicating PRL knock-down was non-toxic to DIV42 *SNCA*-endo1 neurons compared to DIV28 *SNCA-endo1* neurons. Right: *PRL* sgRNA containing neurons dropped out in Comparison 3, indicating *PRL* knockdown was toxic to DIV42 *SNCA*-high neurons compared to DIV28 *SNCA*-high neurons. **C.** Left: Immunostaining data shows prolactin was upregulated in DIV7 *SNCA*-high neurons (n=3, one-way ANOVA). Right: Representative images acquired with Nikon Eclipse Ti microscope, with solar power at 10%, 50ms exposure time for *prolactin*. Scale bar: 20um. **D.** *PRL* knockdown was validated with two different *PRL* sgRNAs from the screen library (n=3, one-way ANOVA). DIV0 *SNCA*-high neurons were transduced with either control sgRNA, or PRL sgRNA at MOI=5. At DIV7, neurons were stained with *prolactin* antibody and Hoechst. Images captured with Nikon Eclipse Ti microscope. *Prolactin* intensity per neuron is reported here. **E.** Immunostaining data shows *prolactin*-upregulation was diminished in DIV28 *SNCA*-high neurons (n=3, one-way ANOVA). Images were captured with Nikon Eclipse Ti microscope, with solar power at 30% and 30ms exposure time for *prolactin*. **F.** CellRox assay shows that, with menadione treatment, oxidative stress in DIV28 SNCA-high neurons was significantly higher than in others of the same age (n=3, two-way ANOVA). **G.** Immunostaining data shows *prolactin*-upregulation was diminished in DIV28 *SNCA*-high neurons treated with PFF (n=3, one-way ANOVA). Images were captured with Nikon Eclipse Ti microscope, with solar power at 30% and 30ms exposure time for *prolactin*. **H.** Left: Illustration of PFF-induced mouse model. Amygdala tissue, micro-dissected from mice and injected with αS PFFs, was subjected to transcriptomic analysis using the NanoString Neuropathology panel (see Methods). Right: In comparison to PBS-injected control animals which showed no pS129 positive inclusions across all brain regions, PFF-injected mice showed a 16.6-fold downregulation of *Prl* mRNA expression in the amygdala region ipsilateral to the site of injection (n=5 per group; 3 males/2 females). **I.** Top: Timeline of exogenous *prolactin* assay. Bottom: CellRox assay shows that exogenous *prolactin* significantly decreased menadione-triggered oxidative stress in DIV7 *SNCA*-OE neurons (n=3, two-way ANOVA).