NERINE reveals rare variant associations in gene networks across multiple 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 Henderson[6], Jalwa Afroz[5], 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. Massachusetts General Hospital, 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

Gene networks encapsulate biological knowledge, often linked to polygenic diseases. While model system experiments generate many plausible gene networks, validating their role in human phenotypes requires evidence from human genetics. Rare variants provide the most straightforward path for such validation. While single-gene analyses often lack power due to rare variant sparsity, expanding the unit of association to networks offers a powerful alternative, provided it integrates network connections. Here, we introduce NERINE, a hierarchical model-based association test that integrates gene interactions that integrates gene interactions while remaining robust to network inaccuracies. Applied to biobanks, NERINE uncovers compelling network associations for breast cancer, cardiovascular diseases, and type II diabetes, undetected by single-gene tests. For Parkinson's disease (PD), NERINE newly substantiates several GWAS candidate loci with rare variant signal and synergizes human genetics with experimental screens targeting cardinal PD pathologies: dopaminergic neuron survival and alpha-synuclein pathobiology. CRISPRi-screening in human neurons and NERINE converge on *PRL*, revealing an intraneuronal α-synuclein/prolactin stress response that may impact resilience to PD pathologies.

Introduction

Gene networks encapsulate a significant portion of the accumulated knowledge in biology for a good reason: genes and their products do not act in isolation but in close coordination. Some of these networks represent classic metabolic or signaling pathways, while others describe sets of interconnected regulatory interactions, physical interactions between proteins, or between proteins and DNA. Genetic interactions, such as coessentiality, capture more abstract dependencies of a gene's contributions to fitness or phenotypes. It is common practice to ascribe genetic phenomena—such as variations in polygenic phenotypes or the development of complex diseases—using the concepts of pathways and networks.

Genetic screens in model organisms and cellular systems, increasingly assisted by advanced genomic perturbation technologies, aim at identifying pathways, networks, or gene modules that drive specific phenotypes or diseases. Existing comprehensive pathway databases and individual experiments targeting specific phenotypes provide a source of hypotheses about disease mechanisms. However, the actual relevance of these network hypotheses to human biology is a matter of debate. A reasonable way to settle this debate is to directly test the association of genetic variation in a network with the relevant phenotype in population-level data.

Aggregate coding rare variant burden (or overdispersion) tests provide the most interpretable way to connect genes to common diseases. Existing methods collapse rare variants within a gene into a single statistical test[1,2]. Applications of such tests to biobank scale datasets have resulted in numerous interpretable discoveries[2-5]. Still, the tests suffer from low power primarily due to the sparsity of rare variants in a single gene[6]. Expanding the unit of association from single genes to networks or pathways presents a natural solution to this limitation, as demonstrated by recent gene set-based analyses[7-12]. Such approaches can enhance power while prioritizing biologically relevant pathways, yet no robust method currently exists that fully integrates network connections and accounts for potential inaccuracies in network inference.

Testing rare variant associations at the level of networks offers three apparent advantages. First, it enables the competitive evaluation of network hypotheses, prioritizing networks—and the experimental assays that define them—by their relevance to human phenotypes. Second, it boosts statistical power by aggregating rare variants across functionally connected genes and offers an improved mechanistic understanding of the phenotype. Third, it minimizes reliance on broad or overlapping gene classifications, such as gene ontology, opting instead for experimentally derived relationships. These relationships are

characterized by defined edge geometries and weights, which enhance the biological specificity of the burden test. This perspective opens new opportunities for integrative research programs where broad genetic screens can identify network hypotheses, which are then tested for associations with human phenotypes. Networks showing significant associations can be refined and validated through focused experimental workflows, combining human statistical genetics with insights from model organisms and cell systems. Such an approach promises to bridge the gap between genetic discovery and mechanistic understanding.

Here, we present NERINE (NEtwork-based Rare varIaNt Enrichment), a new statistical framework for rare variant association testing that directly incorporates information on both network vertices (genes) and network edges (interactions) into a hierarchical model. Our method is designed to be robust against inaccuracies in network specification. NERINE accommodates networks with genes that exert both trait-increasing and trait-decreasing effects, each with varying magnitudes. NERINE performs nested hypothesis testing in a maximum likelihood framework, providing an asymptotic p-value. It has well-controlled Type I error and substantial power, as evidenced by simulations and sub-sampling experiments on dichotomized lipid phenotypes in UK Biobank (UKBB) treated as positive control. NERINE enables the selection of the most informative assays, data sources, and tissue types.

Using NERINE, we conducted a comprehensive scan of canonical pathways for associations with breast cancer, type II diabetes (T2D), coronary artery disease (CAD), and early onset myocardial infarction (MI). We uncovered several compelling biological associations that were not detectable by single-gene tests. Notable findings include rare variant signals in non-lipid gene networks linked to cardiovascular diseases, associations in genes outside traditional tumor suppressors or oncogenes for breast cancer, and an adipogenesis-related gene network implicated in T2D.

Finally, to showcase the potential of an integrated experimental and statistical genetics approach, we applied NERINE to Parkinson's disease (PD), a complex neurodegenerative disease for which rare-variant analyses are thoroughly underpowered and consequently uninformative. We tested network hypotheses derived from common variant GWAS genes as well as genome-scale screens for genetic modifiers of the two core pathologies of PD: ii) alpha-synuclein (αS) proteotoxicity and iii) human dopaminergic (DA) neuron survival. NERINE reinforced specific GWAS candidate loci (*DYRK1A*, *FYN*, *MCCC1*, *USP8*) with rare variant signal and uncovered rare variant burden in experimentally-derived gene modules linked to autophagy regulation (*HMGB1-USP10* module) and vesicle trafficking (*LRRK2-PRL-SNCA* module)—signals that eluded single-gene tests. While several specific genes,

including *NEDD4*, *NDFIP1*, *PBX1*, and *RNF11*, had previously been tied to PD through cell biology, the rare variant risk signal associated with these genes was new. We validated one novel finding by NERINE—an association of rare damaging missense variants at the *PRL* locus encoding prolactin with PD risk—through network-scale CRISPRi screening and targeted analysis in a transgenic induced pluripotent stem cell (iPSC)-based αS toxicity model. This work underscores the transformative potential of experimental screens and human statistical genetics as mutually reinforcing methodologies to reveal novel disease mechanisms.

Results

Overview of NERINE and evaluation of its performance

We present NERINE, a new statistical framework for assessing the cumulative effect of rare variants in a gene network on a dichotomous phenotype. NERINE directly 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. Many data types can represent interactions between genes and gene products (i.e., transcripts and proteins). Interactions may differ in types of relationships (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). Irrespective of the data source, we encode gene relationships using a positive semidefinite matrix, Σ. We assume that the phenotypic effects of genes, represented as vector $\vec{\alpha}$, are drawn from a multivariate normal distribution $\vec{\alpha} \sim MVN(0, \theta \cdot \Sigma)$. Here, θ is a parameter reflecting the importance of the group of related genes and is the object of inference (**Figure 1A**, Methods). This hierarchical model captures the biological expectation that functionally related genes have correlated effect sizes (if the effects are non-zero) and correlated chances that they do not have phenotypic effects.

We model rare variant counts in genes for cases and controls as samples from different Poisson distributions with the rate parameters corresponding to 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 (**Figure 1A**, Methods).

NERINE performs nested hypotheses testing, the test statistic being the log-likelihood ratio (LLR). The likelihood is in the form $L(\theta|\boldsymbol{X}, \mathbf{Y}, \vec{\alpha}, \Sigma) = \int \prod P(X_i|X_i + Y_i, \alpha_i) \cdot P(\vec{\alpha}|\theta; \Sigma) \cdot d\vec{\alpha}$. Here, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are allele counts in cases and controls. We test the hypothesis $\theta = 0$ vs. $\theta > 0$ using the maximum likelihood estimate of θ. Since θ lies on the boundary of the parameter space, the test statistic 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[13,14]. NERINE draws p-values from this distribution. For significant networks with $\theta > 0$, our procedure estimates the maximum likelihood gene-specific effects under the estimated θ (Methods).

We evaluated the performance of NERINE under the null model ($\theta = 0$) in extensive simulations with well-studied biological pathways from the canonical pathway database (**Figure 1B**). Since the performance of existing rare variant methods tends to suffer in the presence of case-control imbalance, we simulated three different scenarios – (i) equal-sized case/control groups, (ii) the case group is larger, and (iii) the control group is larger (Methods). Canonical pathway gene lists were extracted from MSigDB v7.3, and corresponding network edges were extracted from the high confidence physical and genetic interactions in protein-protein interaction (PPI) databases (Methods). In all scenarios, NERINE's test statistic follows the asymptotic distribution (**Figure 1B**). Confidence bands in the QQ-plots represent 95% bootstrap confidence intervals around NERINE's test-statistic. We also tested the null behavior of NERINE's test statistic in simulated networks of different sizes and different topological architectures and demonstrated that it asymptotically follows the theoretical distribution (Supplementary Figure S1).

Next, we evaluated the performance of NERINE under the alternative hypothesis ($\theta > 0$) 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 different well-studied biological pathways from the canonical pathway database (Figure 2, Methods). Under each scenario, we simulated networks with different proportions of genes affecting the trait, mimicking situations from having a very noisy network to a highly relevant one. Currently, no existing rare variant association tests take gene network topology into account. Thus, we compared the performance of NERINE with gene-level rare variant association tests adapted to the pathway level, namely, CMC-Fisher test[15], Fisher minimum p-value test[16], Fisher combined test[16], SKAT-O[17], and pathway-based rare variant trend test (RVTT)[7,8]. The empirical power of each method was measured as the positive predictive value (PPV) at different p-value cutoffs: 1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5, and 1e-5 (Methods). In both scenarios, NERINE outperforms other rare variant tests (**Figure 2**), especially in noisy networks. We also demonstrated that NERINE outperforms other rare variant association tests on different simulated network architectures (Supplementary Figure S2).

NERINE recaptures known biology in lipid-related traits

We analyzed two lipid-related phenotypes in the UK Biobank (UKBB), direct LDL cholesterol (data field: 30780) and HDL cholesterol (data field: 30760), using publicly available exome sequencing data. We created two dichotomous phenotypes---(i) high LDL vs. low LDL

(30,007 cases and 28,673 controls), and (ii) low HDL vs. high HDL (26,800 cases and 27,178 controls), by selecting individuals of European ancestry belonging to the top and bottom quartiles of the distributions for LDL and HDL cholesterol measurements (Methods). We competitively applied NERINE on these two phenotypes across networks from a pathway database consisting of 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 Wikipathways databases. We extracted the edge relationships of genes from high-confidence physical and genetic interactions from protein interaction databases (Methods). 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[18], was used to control for Type I error (Methods).

NERINE identified a significant burden of rare LoF, damaging missense, damaging, and missense variants in key LDL-cholesterol related pathways across the pathway database in both LDL and HDL cholesterol phenotype in UK biobank (UKBB) after multiple hypotheses correction (**Figures 3A** and **3B**, Supplementary Figures S3 and S4). No significant burden of neutral missense or synonymous variants in these pathways was observed for either phenotype, serving as an internal control against technical biases. For the LDL phenotype, the core module of LDL-related genes was the most significant network (Bonferroni p-value = 6.72e-51; damaging variants). In contrast, the core module of HDL-related genes was the most significant for the HDL phenotype (Bonferroni p-value = 6.18e-68; damaging variants). For both phenotypes, the metabolic pathway of LDL, HDL and triglycerides, composition of lipid particles pathway, plasma lipoprotein clearance pathway, cholesterol metabolism pathway, and statin pathway were among the significant hits. No non-lipid-related pathways were significant for either LDL- or HDL-cholesterol phenotypes. For most of the significant pathways in both phenotypes, NERINE provided a lower p-value than SKAT-O, which was applied at the pathway level aggregating the allele counts from member genes (**Figure 3C**). This result demonstrates that NERINE efficiently incorporates network connectivity information to gain additional power.

For the significant pathways, NERINE provided maximum likelihood estimates of signed effect sizes for individual genes, although these estimates are not equipped by p-values. For example, NERINE estimated *PCSK9* and *APOB* to have trait-decreasing effects and *LDLR* to have a trait-increasing effect on LDL cholesterol which conformed with known biology[19] (**Figure 3D**). Similarly for HDL cholesterol, NERINE estimated *ABCA1*, *LCAT*, and *APOA1* to be associated with low HDL cholesterol levels, and *CETP*, *LIPC*, *LIPG*, and *SCARB1* with high HDL cholesterol levels (**Figure 3D**). To ensure our results were not driven by *LDLR* and *PCSK9*, which have large individual effect sizes, 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 S5).

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 (Methods). Even with a cohort consisting of one-tenth of the original cohort size, NERINE recovered a significant rare damaging variant burden in most of the lipid-related pathways that were significant in the original analysis. For a cohort with as few as 500 cases and 500 controls, most top pathways showed nominal significance in the functional categories (LoF, damaging, and damaging missense) without inflation in the neutral missense and synonymous categories (Supplementary Figure S6).

Since rare variants are more susceptible to subtle effects of population stratification than common variants, we performed stratified analysis of LDL phenotype in European, African American, American, South Asian, and East Asian ancestries individually and meta-analyzed the results to identify gene networks with significant rare variant burden for LDL phenotype across different ancestries (Methods). Most lipid-related pathways remained significant networks after Bonferroni correction across ancestries; the most significant results were identified in the European individuals, constituting the largest ancestry group in our dataset (Supplementary Figure S7 and Supplementary Table T1).

Elucidating the rare variant genetic architecture of common diseases with NERINE

We applied NERINE to identify pathway gene networks with significant rare variant burden for four common disease phenotypes in the UK Biobank (UKBB) and MGB Biobank (MGBBB),

namely, breast cancer (BRCA), type II diabetes mellitus (T2D), coronary artery disease (CAD), and early onset myocardial infarction (MI) (**Figure 4**; Methods). These diseases are prevalent in 3-13% of the population.

Previously, in the Genebass study[20] on UKBB exome sequencing data, only a few genes were found to have exome-wide significant rare variant burden for phenotypes like BRCA (6), T2D (3), CAD (1), and MI (1). There might be other genes that collectively play a causal role but evade single-gene analyses due to small individual effect sizes. NERINE can detect such aggregated effects of gene modules. We applied NERINE competitively across the pathway database by first stratifying rare variants into six categories—LoF, damaging missense, damaging, missense, neutral, and synonymous for each phenotype in UKBB and MGBBB. We then meta-analyzed the results for the two biobanks using Fisher's combined test. Significant pathways were selected after Bonferroni correction, adjusting for overlap between the member genes (Methods). **Figure 4** shows the database-wide Bonferroni-significant pathways identified by NERINE in the four disease phenotypes and representative network topologies for each phenotype with predicted gene effects.

For different cancer phenotypes, tumor suppressor genes (TSG) and oncogenes are the primary drug targets, although they explain only a small proportion of the heritability due to common and rare variant burden[21]. For BRCA, NERINE identified seven (7) database-wide Bonferroni-significant pathway gene modules with rare variant burden in at least one of the LoF, damaging, and damaging missense categories (**Figure 4A** and Supplementary Table T2). No significant rare variant burden was observed in synonymous or neutral missense categories (Supplementary Figure S8). These pathways are highly plausible and are involved in cancer susceptibility, ataxia telangiectasia-mutated gene (ATM) signaling, the G2/M DNA damage checkpoint, regulation of cell cycle progression by *PLK3*, hypoxic stress-induced *P53* accumulation and *P53*-dependent apoptosis, *BRCA1*-dependent ubiquitin ligase activity, and regulation of the estrogen receptor (**Figure 4A**, Supplementary Figure S8, and Supplementary Table T2). These pathways are significantly overlapping except the regulation of the estrogen receptor pathway (Supplementary Table T3). Thus, we visualized the cancer susceptibility gene module containing *ATR*, *BRCA1*, and *BRCA2*, and the regulation of estrogen receptor pathway as representative examples in **Figure 4B**. The cancer susceptibility network (BIOCARTA ATRBRCA PATHWAY; Fisher's combined p-value: 1.55e-27 for LoF variants) contained several tumor suppressor genes (TSG), oncogenes, or fusion oncogenes (i.e., genes that undergo fusion with other genes to act as oncogenes). Many of these genes, such as, *BRCA1*, *BRCA2*, *ATM*, *ATR*, *CHEK2*, *FANCD2*, *FANCG*, *RAD50*, and *TP53* were previously implicated in Mendelian genetics[22] and GWAS for breast cancer[23].

NERINE-predicted gene effects in each significant pathway per variant category are provided in the Supplementary Table T4.

Interestingly, none of the member genes of the regulation of estrogen receptor pathway (BIOCARTA CARM ER PATHWAY; Fisher's combined p-value: 3.47e-7 for LoF variants) except *BRCA1* were previously identified to have significant rare variant burden in single-gene analyses; 17 out of 24 member genes were neither TSGs nor oncogenes (Supplementary Table T4). Among the GWAS hits in this pathway—*ESR1*, *CCND1*, and *NRIP1*—only *ESR1* was predicted to have a rare LoF variant burden towards an increased risk of BRCA by NERINE hinting at its causal role in the disease. Notably, *ESR1* is a well-known risk factor for resistance to hormonal therapies but has not been implicated in BRCA risk to date. Among the genes that were neither cancer drivers nor GWAS hits, rare LoF variants in *CARM1*, *TBP*, *GTF2A1*, *HEC14*, *MEF2C*, and *PELP1* were predicted to increase the risk of BRCA by NERINE (**Figure 4B** and Supplementary Table T4); *TBP* already has an approved anti-neoplastic drug, Etoposide, which is used for BRCA treatment. In contrast, NERINE identified a trait-decreasing effect of rare LoF variants in *MRE11*, a member of the MRN complex (*MRE11-RAD50-NBS1*), which detects DNA double-strand breaks (DSBs) and initiates repair, supporting the inhibition of *MRE11* as a therapeutic strategy for BRCA[24]. To test how sensitive our finding is to the presence of *BRCA1*, we ran NERINE on the estrogen receptor network by removing—(i) the observed mutation counts in *BRCA1* but keeping the gene and its connections in the network, and (ii) the *BRCA1* gene and its edges from the network. The network retained significance in the UKBB cohort, indicating the importance of other genes in the network concerning the BRCA phenotype (Supplementary Figure S9).

Exome-wide rare variant association studies for type II diabetes mellitus (T2D) to date identified only a handful of significant genes[20,25], such as *GCK*, *GIGYF1*, and *HNF1A*, primarily involved in insulin secretion, insulin growth factor signaling, and pancreatic beta cell differentiation pathways. We applied NERINE across the pathway database on the T2D cohorts from UKBB and MGBBB (Methods) and detected significant cumulative burden of rare damaging variants in the adipogenesis pathway (BIOCARTA VOBESITY PATHWAY) module (**Figures 4A** and **4C**; Supplementary Tables T2 and T5). This module consisted of *RXRA*, *PPARG*, *ADIPOQ*, *TNF*, *NR3C1*, *LPL*, *RETN*, and *HSD11B1*, none of which were previously implicated in T2D in rare variant association studies (Supplementary Table T5). Only *PPARG* and *LPL* were implicated in GWAS studies[26-28]. Notably, low levels of the protein adiponectin, encoded by *ADIPOQ*, is known to be associated with insulin resistance[29], increasing the risk of T2D. Our finding provided human genetics corroboration for *ADIPOQ* as a potential therapeutic target for T2D. The roles of *TNF*, *NR3C1*, and *RETN* in increasing the risk of T2D, as observed in NERINE's output, are understudied and warrant further

investigation. NERINE's prediction of trait-decreasing roles of *RXRA* and *HSD11B1* aligned with the fact that inhibitors of these genes are currently being considered as potential therapeutic agents for T2D[30-33].

The primary gene network associated with the risk of coronary artery disease (CAD) and myocardial infarction (MI) consists of lipid-related genes, including *LDLR, PCSK9, LPL, APOA5, APOC3, ANGPTL4,* and *LPA*, where a convergence of common and rare-variant signals was previously observed[34-36]. Although several non-lipid pathways, including neovascularization angiogenesis, vascular remodeling, thrombosis, immune response and inflammation, proliferation and transcriptional regulation, have also been hypothesized to be also involved in these diseases, significant rare variant burden in their member genes have not yet been identified in exome-wide analysis[36,37]. We applied NERINE across all gene modules from the pathway database on cardiovascular phenotypes, CAD and MI, in both UKBB and MGBBB (Methods).

For the CAD phenotype, we recaptured significant rare damaging variant burden in lipid-related pathways—plasma lipoprotein clearance and *SREBP* control of lipid synthesis (**Figures 4A** and **4D**, Supplementary Figure S10, and Supplementary Tables T2 and T6). Among the member genes, only *LDLR* was previously identified in both GWAS and exome-wide rare variant analysis of CAD and is a known therapeutic target[38]. NERINE predicted rare damaging variants in *APOA1, CUBN, SCARB1, CES3, SOAT2, AMN, LIPC, LDLRAP1, NCEH1, HMGCS1, SREBF1*, and *SREBF2* to increase the risk of CAD (**Figure 4D**, Supplementary Figure S10, and Supplementary Table T6). We also identified a significant burden of rare LoF variants in the classic and alternative complement system pathways (**Figures 4A** and **4D**; Supplementary Figure S10, and Supplementary Table T2). NERINE predicted several complement genes, including *C3, C6, C7, C8A, C9, C1QC, CFB*, and *CFD*, involved in the inflammatory response, to have trait-increasing effects on CAD (**Figure 4D** and Supplementary Table T6).

When applied to the MI phenotype in UKBB and MGBB, NERINE identified 13 gene modules with significant rare LoF variant burden across the pathway database (**Figures 4A** and Supplementary Figure S11). Since there were significant overlaps among the member genes (Supplementary Table T7), we grouped them into four broad classes – (i) collagens and coagulation, (ii) inflammatory response, (iii) regulation of transcriptional activity, and (iv) MAPK signaling cascade (Supplementary Figure S11). **Figure 4E** shows one representative pathway per group. Among the genes in these pathways, structural components of the basement membrane—collagen type IV alpha 1,2, and 4 (*COL4A1, COL4A2*, and *COL4A4*), plasminogen (*PLG*), complement component II (*C2*), and interleukin-8 (*CXCL8*)—an

activator of the MAPK signaling cascade, were previously implicated in common variant GWAS for the MI phenotype (Supplementary Table T8). The finding around the MAPK signaling pathway might also be an artifact due to the CHIP (Clonal Hematopoiesis of Indeterminate Potential) effect because the original biobank samples were primarily from blood. Notably, NERINE implicated rare LoF variants in several non-lipid-related genes including, basement membrane gene, *COL4A3*, genes involved in blood clotting (*FGA, FGB, FGG, F7, F12, PLAT, KLKB1*, and *TFPI*), apoptosis (*CREBBP* and *SIRT1*), inflammatory response (*TNF, TNFRSF1A, TNFRSF1B, C3, C4B, C5, C8A, CFB, CFD, CXCL8*, and *IL6*), vascular calcification (*ALPL* and *ENPP1*), calcium regulation (*PLCB1, PPP3CA, PPP3CB, PPP3CC, CAMK1*, and *CAMK1G*) and retinoic acid signaling (*RARA*) to be associated with an increased risk of MI (**Figure 4E**, Supplementary Figure S10, and Supplementary Table T8).

Selecting the most informative assay and tissue context with NERINE

NERINE directly incorporates gene-gene relationships from diverse data sources and competitively selects the most appropriate network for a given analysis when multiple competing network topologies exist. We demonstrated this functionality of NERINE using the LDL-direct and HDL-direct phenotypes in UKBB (Methods, Supplementary Table T9). For the lipid-related pathway gene sets in our pathway database, we created network topologies considering 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 DepMap (release 2023Q2) (Methods).

Our results indicated that different data sources described the functional relationships better depending on the gene set under question. **Figure 5A** shows an example scenario where one data source provides more relevant information than the others. When comparing individuals with low HDL cholesterol with the ones with high HDL cholesterol in UKBB, co-expression in liver tissue served as the best source of information for the core module of HDL-cholesterol-related genes (p-value: 4.48e-76). Example scenarios where data sources other than co-expression better described the edge relationships of a specific gene set are shown in Supplementary Figure S12. When comparing individuals with high LDL cholesterol to those with low LDL cholesterol, the VLDLR internalization and degradation pathway genes were best described using physical and genetic interactions (NERINE p-value: 8.33e-34). In contrast, we achieved a p-value of 5.25e-49 for the genes in the metabolic pathway of LDL, HDL, and triglycerides (TG) with the co-essentiality network constructed from the DepMap liver cell lines data. These observations implied that different data sources provided different levels of information about the edge relationships for the same set of input genes,

and NERINE could effectively exploit this phenomenon to select the most informative data source.

To demonstrate the potential for helping select the most informative tissue or cellular context, we applied NERINE to different co-expression networks representing the core module of HDL-related genes in all 52 tissue types available in GTEx. The co-expression network in liver tissue achieved the most significant p-value (**Figure 5B**), which indicated that liver tissue provides the most relevant context for the core HDL-related genes: *LCAT, ABCA1, LIPC, LIPG, SCARB1, APOA1, CETP, APOA2*, and *PLTP*. This result was driven by the fact that the edge relationships varied in different tissue contexts for the same set of input genes, and NERINE could effectively exploit this phenomenon.

NERINE reveals novel gene modules with rare variant burden in PD

Synergizing insights from human genetics, cell systems, and model organism experiments provides a powerful framework for unraveling the complexities of human diseases. This approach is exemplified in our investigation of a complex neurodegenerative disease, PD. PD is the most common neurodegenerative movement disorder defined neuropathologically by two central features: the degeneration of dopaminergic (DA) neurons in the substantia nigra (SN) region of the midbrain and the accumulation of intraneuronal αS in so-call pathological "inclusions" called Lewy bodies[39,40]. Genome-wide association studies (GWAS) previously identified hundreds of loci mapped to approximately 100 genes implicated in PD[41,42]. However, the largest population-based genetic study of PD[43] to date identified only a handful of replicable rare variants in *PRKN*, *GBA1*, and *LRRK2*—genes already known to play roles in Mendelian forms of PD. At the gene-level, only *GBA1* was found to have an exome-wide significant rare variant burden replicated in multiple independent cohorts[43,44]. For sporadic PD, which constitutes ~80-85% of the patient population[45], the role of rare variants remains poorly understood. To address this gap, we employed a three-pronged strategy with NERINE. First, we leveraged NERINE to analyze gene modules constructed with GWAS-identified genes. Second and third, we used key biological networks comprised of genes that modulate the two central phenotypes of PD, dopaminergic survival and αS proteinopathy, respectively. We assessed these in two independent PD case-control cohorts from the UKBB (Ncase = 2,237 and Ncontrol = 167,188) and AMP-PD (Ncase = 2,117 and Ncontrol = 1,095) (**Figure 6A**, Methods).

Rare variant burden in a gene module of PD GWAS hits reveals potential therapeutic targets

We applied NERINE to six GO biological process modules for which top PD GWAS-associated genes showed enrichment (**Figure 6A**: left, Supplementary Table T10). We generated networks from these process modules by grouping semantically similar GO

biological process terms enriched in PD GWAS hits and then extracting the edge relationships of genes in each group from—(i) high-confidence physical and genetics interactions from protein interaction databases, (ii) co-expression in the substantia nigra region of mid-brain, and (iii) co-essentiality in CNS cell types (Methods). By testing these network topologies competitively with NERINE, we identified a significant rare LoF variant burden in the co-essentiality module related to *Peptidyl-threonine modification* in both AMP-PD and UKBB (Fisher's combined p-value = 7.23e-3; **Figure 6B**, Supplementary Tables T11 and T12). This finding is consistent with kinase-phosphatase dysregulation being heavily implicated in PD and synucleinopathy[46,47]. We did not observe an enrichment of neutral missense and synonymous variants in this module.

Notably, rare LoF variants in *LRRK2* showed an overall trait-decreasing effect, aligning with well-known biological understanding that the inhibition of *LRRK2* kinase activity is protective for PD[48,49]. NERINE newly predicted rare LoF variants in *MCCC1* and *DYRK1A* genes to have a trait-increasing effect on PD (Supplementary Table T12), providing confirmatory support of these genes being the true genes-of-origin for the original GWAS signal, but also that loss-of-function is likely to be the mechanism through which they confer PD risk. LoF variants in *MCCC1* and *DYRK1A* might lead to mitochondrial dysfunction and exacerbate the degeneration of DA neurons[50] , respectively, thereby increasing PD risk. In contrast, rare LoF variants in *FYN* and *USP8*, are predicted by NERINE to have a protective role in PD (Supplementary Table T12), highlighting their inhibition as a potential therapeutic strategy. Inhibition of *FYN* was previously shown to protect DA neurons and prevent locomotor deficits[51-53], and *USP8* inhibition was known to reduce αS accumulation in experimental models[54,55]. Thus, NERINE's findings provide much-needed human genetic corroboration (and resolution of controversies) for experimental manipulation of these genes in model systems, and suggest inhibition of *FYN* and *USP8* may be worthy of therapeutic consideration as targets in PD.

A dopamine neuron essentiality gene module with rare variant burden in PD implicating HMGB1 and USP10

Our second line of investigation focused on an unbiased whole-genome CRISPR screen to identify genes essential for DA neuron survival, with the aim of pinpointing those relevant to PD risk (**Figure 6A**: middle). Guide RNAs (gRNAs) representing 19,993 genes were transduced into H9 hESC lines that had been knocked in for a doxycycline-inducible Cas9 at the *AAVS1* safe-harbor locus. In the DA neural progenitor stage, Cas9 was induced, and a gRNA representation was obtained through next-generation sequencing. A parallel culture was aged to DIV42, and gRNA representation was assessed to look for dropout. If gRNAs representing a specific gene consistently dropped out, that gene was considered essential

(Methods). We identified 693 essentiality genes that converged onto ten (10) GO biological process modules (Supplementary Table T13). We generated networks from these process modules by grouping semantically similar GO biological process terms enriched in essentiality genes and then extracting the edge relationships of genes in each group from PPI, co-expression, and co-essentiality databases as described before (Methods). We tested these modules competitively with NERINE on sporadic PD cases compared to controls in AMP-PD and UKBB with extreme controls, i.e., controls with age >= 85 (Methods).

NERINE uncovered a screen-wide significant burden of rare damaging variants in the gene module linked to the *regulation of autophagy* (Fisher's combined p-value: 6.69e-4; Figure 6C, Supplementary Table T14), with no inflation observed in neutral missense and synonymous variants. Co-essentiality in CNS cell types provided the most informative gene-gene relationship for this set of genes. Key genes with trait-increasing effect in PD (Supplementary Table T15) included *HMGB1* and *USP10*, both highly intolerant to loss-of-function variants. While damaging variants in *HMGB1* may lead to autophagy inhibition and αS accumulation[56], such variants in *USP10* may lead to αS toxicity in PD by disrupting αS-containing aggresome formation[57]—a finding now corroborated through human genetic analysis by NERINE.

Rare variant burden in PD-associated submodule in an αS proteinopathy network pinpoints highly plausible PD risk genes

Finally, we focused on a network of α-synuclein modifier genes assembled through "yeast-to-neuron" discovery screens[58] (**Figure 6A**: right). 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[41,42,59]. The network, validated in human iPSC cortical and DA neurons[58], converged with a proteome-scale proximity labeling screen for αS in neurons[60].

The TransposeNet αS network consisted of 17 branches representing different biological processes (Supplementary Figure S13) which provided the hypotheses to be tested with NERINE. Among these branches, we observed a screen-wide significant burden of rare damaging missense variants with NERINE in the *LRRK2* and *SNCA*-containing *vesicle trafficking and protein homeostasis*-related gene subnetwork associated with PD risk in both AMP-PD and UKBB datasets (Fisher's combined p-value: 1.27e-3; Figure 6D and Supplementary Tables T16 and T17). Neutral missense and synonymous variants in this

network showed no significant enrichment. None of the genes in the subnetwork was previously shown to have a replicable exome-wide significant rare-variant burden in single-gene analysis[20,44].

Intriguingly, the subnetwork highlighted by NERINE contained *SNCA* and its direct interactor *LRRK2,* which had previously been implicated in PD by GWAS and Mendelian genetics[41,59]. Several other direct interactors of *SNCA* in this subnetwork were previously implicated to PD-relevant pathology and pathobiology in different contexts: *VDAC1*[61-63], *NEDD4*[64-66], and *TOR1A*[67]. For the remaining direct interactor of *SNCA*, *PRL*, there was limited direct evidence of a causal role in PD. Beyond the direct interactors of *SNCA*, NERINE also predicted a trait-increasing role for several other genes that were linked to PD in functional studies, including *NDFIP1*[68,69], *PBX1*[70], and *RNF11*[71].

Convergence of unbiased functional genomics screen and NERINE on a *PRL-SNCA* interaction

Since conventional iPSC models for neurodegenerative proteinopathies suffer from poor reproducibility and tractability, and levels of endogenous α-synuclein far lower than found in the brain, recently we[72,73] have established a suite of tractable iPSC models to study different aspects of neurodegenerative disease biology. Extra copy numbers of wild-type *SNCA* through gene multiplication at the *SNCA* locus can lead to early-onset PD and dementia with pathology in DA and cortical glutamatergic neurons, similar to sporadic PD with dementia. To create a better surrogate for this pathobiology, we engineered a CRISPRi-induced synucleinopathy model in cortical neurons (CiS-CN) in the widely used WTC11 genetic background[72]. This model—i) allows for cortical glutamatergic neurons generation through one-step trans-differentiation with *Ngn2* (an integrated dox-inducible transgene at the *AAVS1* locus), ii) allows for flexible target gene knock-down with CRISPRi (dCas9-KRAB) knocked in to the *CLYBL* locus[74]; iii) have lentivirally transduced αS to express the protein at different levels: three hiPSC clones in CiS model were utilized in this study, namely two *SNCA*-overexpression clones (*SNCA-high, SNCA-intermediate)*, and a control clone (*SNCA-endo1*). These lines express αS at brain-like levels with a higher propensity for αS aggregation and toxicity than conventional PD iPSC models[72,73].

In this model, we first performed a surrogate human functional genomics screen "in the dish" to complement NERINE and look for points of convergence between the methods in an unbiased way. Specifically, we transduced into our CiS models a CRISPRi-optimized gRNA library comprising 9,852 gRNAs targeting 1,705 individual genes in our αS gene-interaction networks, including our TransposeNet αS proteinopathy network (**Figure 7A**). At DIV0 (3 days post-Ngn2 induction), CiS neurons were transduced with the sgRNA library. We

cultured transduced neurons in parallel until the harvesting day. Neurons were collected at DIV3, 28, and 42. The proportion of cells expressing each sgRNA at each time point was determined by next-generation sequencing of the sgRNA-encoding locus. The *MAGeCK-iNC* pipeline was used to analyze the screen results (Methods). Based on the depletion or enrichment of sgRNAs targeting specific genes at different time points, we identified hit genes for which knockdown was αS toxicity enhancer or suppressor. We made three comparisons – sgRNAs that dropped out at DIV42 in *SNCA-high* versus *SNCA-endo*, sgRNAs that selectively dropped out at DIV42 versus DIV28 in *SNCA-endo* and sgRNAs that selectively dropped out at DIV42 versus DIV28 in *SNCA-high*. We refer to these comparisons as Comparisons 1, 2, and 3, respectively, in **Figures 7A** and **7B**. We defined genes passing a false-discovery rate (FDR) < 0.1 as hit genes. For each comparison, the gene product cutoff for hit genes was selected dynamically by the *MAGeCK-iNC* pipeline.

Comparisons 1 and 2 provided the key αS toxicity-specific enhancers from these functional genomics screens in human neurons. Intriguingly, *PRL* knockdown was close to the top hit in both comparisons. Thus, our screen indicated that *PRL* knockdown within neurons is an enhancer of αS toxicity. This result converged with NERINE's PD genetics finding (**Figure 6D**), in which the TransposeNet subnetwork with rare variant burden pinpointed damaging missense variants in *PRL* as conferring PD risk. This finding was even more remarkable because, like *SNCA* and *LRRK2, PRL* was a genetic "node" predicted by the TransposeNet algorithm in our original αS network[58].

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

The convergence of signals from the CRISPRi screen and NERINE at the *PRL* locus, which encodes the hormone *prolactin*, highlighted it as a compelling candidate for further experimental validation. *Prolactin* is expressed predominantly in neuroendocrine tissue, specifically the anterior pituitary gland. While prolactin has been extensively studied in the context of neuroprotection[75] and has even been linked to PD[76-79], it is generally considered to be of exogenous pituitary origin. It was thus most surprising that it appeared within a neuronal functional genomics screen. Notably, there is a literature, albeit controversial and conflicting, documenting *PRL* expression in different brain regions of the rodent[80], with speculation that its expression may only be detected under conditions of neuronal stress. There is also evidence that, despite mRNA expression being very low, its expression can be post-translationally regulated because expression of the protein is often disconnected from transcript levels[81].

Our data raised the possibility that *prolactin* may be part of an intraneuronal stress response directly related to αS. To further investigate, we performed immunostaining in neurons of

different age. *Prolactin* 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 αS-induced *prolactin* that diminishes over time.

We next tested whether inducing αS aggregation (that in turn reduces soluble αS levels)[73] could also reduce prolactin levels. αS aggregation can be induced by exposing cellular and animal models to pre-formed αS amyloid fibrils (PFFs)[82,83]. We challenged our neurons with seven (7) days of exposure to pre-formed αS fibrils (PFFs). This challenge led to a reduction of prolactin levels as indicated by immunofluorescence (**Figure 7F**). In our models, mRNA expression levels of *PRL* were low. We assumed this might relate to the immaturity of the iPSC-derived neurons. We thus turned to a well-established mouse PFF model[82]. We unilaterally injected PBS versus 5 µg of PFFs (2.5 µL volume) into the dorsal striatum of mice (**Figure 7G**: left). Mice were aged for 30 days post-injection. By this stage, as has been previously described[84], 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. Thus, in the context of aging and fibrillar αS pathologies, *prolactin* levels drop intraneuronally, and neurons become highly sensitized to its removal.

To investigate whether prolactin can directly protect from αS toxicity, we assessed neuroprotection against oxidative stress sensitization phenotype in our CiS models. In aged neurons by DIV28, *SNCA-high* neurons became sensitized to the oxidative stressor, menadione (**Figure 7H**). Menadione generates reactive oxygen species (O2-) through redox cycling[85]. These species can be identified with the fluorogenic probe CellRox (Thermo). In *SNCA-high* and *SNCA-endo* models, we pre-conditioned DIV6 CiS neurons with prolactin for 24hrs before treatment with menadione (**Figure 7I**: top). As anticipated, exogenous prolactin treatment significantly decreased oxidative stress in CiS neurons (**Figure 7I**: bottom). Taken together, a notable convergence of a forward genetics iPSC screen with NERINE, identified

an unexpected αS-prolactin intraneuronal stress response that may have important implications for resilience to αS pathology in PD.

Discussion

There are two tracks in human disease genetics: one centers on the development of experimental techniques in easy-to-manipulate systems. It is propelled by advances in gene editing technology coupled with massively parallel phenotyping at the cellular level. These experimental studies generate mechanistic biological hypotheses, but by themselves do not establish relevance to human phenotypes. The second track focuses on finding associations in rapidly growing human genomic datasets using the increasingly sophisticated computational and statistical methods. This line of research identifies signals highly relevant to human conditions but is unable to provide mechanistic knowledge. The promising path towards building mechanistic hypotheses is through combining statistical human genetics and direct experimentation. We developed NERINE to directly integrate results of experimental screens, competitively test hypotheses about gene relationships, and guide follow-up focused experiments. NERINE shows improved performance over existing rare variant tests, albeit with a few limitations (see Methods).

We demonstrated two avenues of applications of NERINE. First, we used pre-existing biological knowledge in the form of gene networks systematically assembled by the community with no regard to a specific human phenotype. We tested these networks for associations with four highly prevalent human diseases---breast cancer (BRCA), type II diabetes (T2D), coronary artery disease (CAD), and early onset myocardial infarction (MI). Second, we demonstrated an integrative approach by testing bespoke gene networks originated from genome-wide experimental screens in model systems targeting two main pathophysiological features of Parkinson's disease—the degeneration of DA neurons and α-synuclein proteotoxicity and helping prioritize targets for downstream experiment.

Our investigation of common diseases in UKBB and MGBB with NERINE provided new insights into the biology of BRCA, T2D, CAD, and MI by uncovering a significant rare variant burden in non-lipid-related pathways in cardiovascular diseases, adipogenesis in T2D, and the estrogen receptor pathway in BRCA, elucidating their causal role. 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[86] but also with their indirect role in coagulation via platelet effects and interactions with clotting factors[87]. 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[88].

Our findings further highlighted two key insights. First, we provided human genetics corroboration for genes identified only in functional studies in model systems. For example, several member genes of the *LRRK2-SNCA*-containing *vesicle trafficking and protein homeostasis* subnetwork of the αS proteotoxicity network, including *NEDD4*, *NDFIP1*, *VDAC1, PBX1, TOR1A*, and *RNF11* have been previously implicated in PD in model systems experiment[61-71,89-91]—now corroborated with human genetics evidence by NERINE. Second, our study provides additional clarity on the directionality of gene effects in networks where conflicting evidence exists from functional studies. For example, there is biological data suggesting haploinsufficiency of *DYRK1A*—a member of the *peptidyl-threonine modification* gene module implicated in PD by NERINE—reduces its kinase activity and exacerbates DA neuron degeneration in mice[50]. There is also data arguing the opposite through phosphorylation of αS[92]. By providing genetic evidence in favor of the latter direction, NERINE provides some resolution of the matter.

We do recognize the fact that currently individuals of European ancestry constitute the largest group in our study cohorts. Thus, real data applications of NERINE in this study primarily focused on individuals of European ancestry. Concentrated efforts in building large biobanks with diverse participants are already underway[93] and will enable NERINE to overcome this limitation and provide more insight into the contribution of rare variants to common disease etiology across populations. Additionally, the cohort sizes in MGBBB are smaller compared to UKBB for several phenotypes requiring us to adjust the MAF cutoff accordingly for MGBBB in some comparisons. For cardiovascular phenotypes, our results included several pathways involved in the complement system cascade, especially in the MGBBB cohort which might either be a true signal for cardiac phenotypes or a technical artifact because bio-samples were likely collected from patients at a time close to the occurrence of a cardiac event. Similarly, the significant rare variant burden near the MAPK signaling pathway in the MI phenotype might also be due to the CHIP effect since the patient samples primarily came from blood.

Prolactin, a pleiotropic hormone, is encoded by *PRL*—a gene where NERINE's findings from human genetics converged with experimental evidence from forward genetic screens in our tractable synucleinopathy hiPSC and mouse models. This was a notable finding, especially given the abundant controversies about whether *PRL* is expressed outside the pituitary gland. Interestingly, there has been considerable speculation that prolactin may be responding to intracellular stress, explaining some of the controversy[80], and this is indeed what our data point to. Numerous studies have linked prolactin to neuroprotection against a variety of toxic insults through reduction in oxidative stress, cytotoxicity, and

inflammation[75,94,95], but with minimal evidence from genetics. Our own data indicate an early and likely post-translational intraneuronal response in which prolactin is elevated in specific response to αS overexpression. This response diminishes with time, and with αS aggregation. This decrease in turn associated with a sensitization to *PRL* knockdown and to exogenous stress. This reduction in *PRL* expression also occurs with a more chronic treatment of mice with fibrillar αS. Interestingly, even beyond the known effect of dopamine on prolactin reduction, *prolactin* levels in CSF emerged as the top feature of a classifier model for classifying PD patients versus controls[78]. Together with human genetic evidence from NERINE, our cellular and mouse data suggests a causal connection between the *PRL-SNCA* stress loop in PD risk and resilience. Thus, NERINE systematically brings experimental biology and human genetics together to enable mechanistic discoveries in human phenotypes. This may be particularly powerful in complex diseases like PD for which population sizes are limited and existing rare-variant methodology have been hitherto unrevealing.

Methods

Estimating rare-variant burden in a gene network with NERINE

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 normal distribution $\vec{\alpha} \sim MVN(0, \theta \cdot \Sigma)$. Here, θ is a parameter reflecting the cumulative effect of the gene network on a phenotype and is the object of inference (Figure 1**A**). 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. Since the marginal distributions of the multivariate normal distributions are univariate normal distributions which are unbounded, we leveraged variable transformation to map the marginals to beta distributions bounded between 0 and 1 (Supplementary Note). The transformed gene-effects within the network are represented by a vector, $\overrightarrow{\alpha'}$. We account for the skew in the sizes of case- and control-groups within the cohort by controlling the shape parameters of the $\overrightarrow{\alpha'}$ distributions (Supplementary Note).

We model rare variant counts in genes for cases and controls as two independent Poisson distributions 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, N_{case} and $N_{control}$ represent the case- and control-cohort sizes and λ^i_{case} and $\lambda^i_{control}$ denote the rate parameters for case- and control- allele count distributions in gene i. The rate parameters 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).

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

NERINE infers the network-effect, θ on a dichotomous phenotype using the maximum likelihood estimation (MLE) framework, where the likelihood is given by,

$$L(\theta|\boldsymbol{X}, \mathbf{Y}, \vec{\alpha}, \Sigma) = \int \left(\prod_{i=1}^{m} P(X_i|X_i + Y_i, \alpha_i)\right) \cdot P(\vec{\alpha}|\theta; \Sigma) \cdot d\vec{\alpha}$$

We approximate the integral leveraging multivariate Gaussian-Hermite quadrature with pruning (Supplementary Note):

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

We calculate the conditional probability of allele counts in cases and counts in each gene using the probability density function of a standard binomial distribution as described above which makes the 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 approach (Supplementary Note).

NERINE performs nested hypothesis testing; the null hypothesis being H_0: θ=0 and the alternative set of hypotheses being H_z: $\theta = \theta_z$>0. 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 $\hat{\theta} = \underset{\theta=\theta_z}{\text{argmax}}(LLR_z)$. 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 ($dof = 1$)[13,14]. NERINE draws its asymptotic p-values from this distribution. For significant networks with θ>0, NERINE calculates the maximum likelihood gene-specific effects ($\widehat{\overrightarrow{\alpha'}}$) under the estimated $\hat{\theta}$ as follows:

$$\widehat{\overrightarrow{\alpha'}} = \underset{\overrightarrow{\alpha'}}{\text{argmax}}\, L\left(\overrightarrow{\alpha'}|\boldsymbol{X}, \mathbf{Y}, \hat{\theta}, \Sigma\right) \approx \underset{\overrightarrow{\alpha'}}{\text{argmax}} \left(\prod_{i=1}^{m} P(X_i|X_i + Y_i, \alpha'_i)\right) \cdot P\left(\overrightarrow{\alpha'}|\hat{\theta}; \Sigma\right)$$

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 (see 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 pathway database (**Figure 1B**) 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 S1). 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.

We also tested 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) for equal sized case- and control-groups (Supplementary Figure S1). 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 four database pathway network topologies used for the null simulations. For each scenario, we simulated different noise profiles i.e., different 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 1-5 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[15], Fisher minimum p-value test[16], Fisher combined test[16], SKAT-O[17], and pathway-based rare variant trend test (RVTT)[7,8]. For each noise profile, we performed 250 iterations. 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 benchmarked NERINE's performance against other rare variant association tests on different simulated network architectures for 25 genes (Supplementary Figure S2). 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 network effect, $\theta = 0.5$. We simulated different 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. 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 < 1e-4. Hence, we compared RVTT's performance only at cutoff values ≥ 1e-4.

Study cohorts

We used whole exome sequencing (WES) data from two population-scale biobanks, namely, the UK biobank (UKBB) and the Mass General Brigham biobank (MGBBB) and whole genome sequencing (WGS) data from the AMP-PD consortium.

Utilizing jointly genotyped variant calls from 469,589 individuals in the UKBB, we first 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). 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 and HDL cholesterol 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). We performed stratified analysis of 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. Bonferroni correction was applied on the meta p-values. For the down-sampling experiment under the LDL-C phenotype in the European ancestry group, we created three additional cohorts by randomly sampling individuals without replacement from cases and controls: (i) 10,000 cases vs 10,000 controls, (ii) 3,000 cases vs 3,000 controls, and (iii) 500 cases vs 500 controls.

We performed database-wide investigations for four disease phenotypes in the UKBB with NERINE, namely, breast cancer (BRCA), type II diabetes mellitus (T2D), coronary artery disease (CAD), and early onset myocardial infarction (MI). The cohorts 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).

For the three-pronged analysis of the Parkinson's disease (PD) phenotype, we created a discovery cohort from the UKBB individuals by including 2,237 unrelated European individuals of age 40 or above with ICD-10 code = G20, no history of T2D, and no family history of PD as cases. Control group consisted of 167,188 unrelated European individuals of age >= 60 who had no history of T2D, PD, and other diseases of the nervous system (ICD-10 codes: G00-G99). We termed this cohort as "UKBB-Sporadic". For analyzing the dopamine (DA) neuron essentiality screen gene modules, we used a subset of the controls consisting of individuals who had an age of 85 or above as super controls. The case group was the same as before. We termed this cohort with super controls as UKBB-Extreme (Ncase = 2,237 and Ncontrol = 2,553). For the UKBB-sporadic cohort, we used MAF < 0.001 as the cutoff to select rare variants. For the UKBB-extreme cohort, the cutoff was fixed at 0.01.

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" and tests for synonymous and neutral missense categories were performed for each pathway to make sure that there is no LD leakage.

For the three-pronged analysis of 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). 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 AMP-PD cohort predominantly consists of individuals of European ancestry, we included only unrelated individuals of the same ancestry group in our analysis. We called the resulting cohort as "AMP-PD-sporadic" which consisted of 2,117 sporadic PD cases and 1,095 neurotypical controls. We used an MAF cutoff of 0.001 to select rare variants for this cohort.

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*. After removing sample outliers based on Ts/Tv, Het/Hom ratios, and per-haploid SNV counts, we retained only unrelated European samples for our analyses.

Pathway database construction

For this study, 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 S18). 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 (m_{eff}) in our pathway database of m pathways by adapting Nyholt's approach[18]. First, we calculated the pathway-by-pathway correlation matrix, *M*, using Jaccard similarity and converted it to its nearest positive definite matrix using the *nearPD* function from the *Matrix* package (v1.6-5) in R (v4.3.2). Then we determined the eigenvalues ($\vec{\lambda}$) of the positive-definite correlation matrix. The effective number of hypotheses/pathways was computed using the following formula[18]:

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

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

Gene-network topology extraction

NERINE's methodology treated the gene-gene network as an input. 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 used the published TransposeNet humanized α-synuclein-modifier network stems with our method. In absence of the true network topology for a particular gene set, we adopted the following approach to construct one.

For the pathway gene sets in the database, we constructed physical/genetic network topologies by extracting the binary edge relationships of genes from the following sources—(i) high confidence physical interactions (weight >= 0.7) from STRING v11.5[96], (ii) InWeb inBio Map database[97], (iii) HuRI[98], (iv) genetic interactions from Megchelenbrink *et al.* study[99], and (v) humanized α-synuclein-, β-amyloid-, and TDP-43-modifier networks[58]. For a specific pathway gene set, NERINE represented these edge relationships by a gene-by-gene variance-covariance matrix, Σ, where off-diagonal elements contained information about the edges (i.e., covariances) and the diagonal contains priors on genes (i.e., variances). We considered all genes to have the same variance for this study; the magnitude of the variance being set at two (2). The binary edge relationships were represented by either one (1: presence of an edge) and zero (0: absence of an edge). Finally, for this network matrix, Σ, we computed the nearest positive definite matrix using the *nearPD* function from the *Matrix* package (v1.6-5) in R (v4.3.2).

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[100] (GTEx v8). We computed a real-valued gene-gene co-expression networks in a

specific tissue where the edges between two genes represent 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, 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[101,102] (release: 23Q2). 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. For lipid-related phenotypes, we constructed co-expression networks using the bulk expression data from liver cell lines. For Parkinson's disease, we used bulk expression data from cell lines pertaining to the central nervous system (CNS). The cell lines used in this study are listed in Supplementary Table S19.

Genome-wide CRISPR/Cas9 screen in midbrain DA neurons

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[103], 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[104] 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[105]. 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*[106] (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[106] that are not specific to DA neurons, were removed from the list of essentiality genes. We identified 693 essentiality genes.

Constructing GO biological process modules

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 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 T10) and for DA neuron essentiality genes, we identified 10 such modules (Supplementary Table T13). 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, and (iii) co-essentiality in CNS cell types.

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

A customized CRISPRi library containing 9,852 sgRNAs targeting 1,705 α-synuclein network genes and negative controls were selected from the CRISPRi v2 H1 library[107]. The library was packaged into lentivirus by the Virus Core at Boston Children's Hospital. A small-scale preliminary experiment was conducted to determine the functional MOI. In this experiment, D0 neurons were transduced with the viral library. At DIV3, neurons were released, and a flow-based assay was performed to assess BFP++ population. The titer was calculated to achieve a functional MOI (= the fraction of BFP-positive cells) of 0.43. Based on these preliminary data, screening experiments were conducted using 56 million D0 neurons per CiS line with a functional MOI of 0.43, corresponding to a library representation of ~400 cells per library element. Briefly, 6 BioCoat Poly-D-Lysine 10-cm dish (Corning; Cat. No. 356469) were seeded with D0 neurons/each CiS line in 10mL of the virus-containing differentiation medium supplemented with Dox and ROCK inhibitor and left in the tissue culture hood for 15 minutes to allow even distribution and attachment before moving to the incubator. At DIV3, 2 dishes/each CiS line were pelleted and stored at -80C. Full medium change was performed for other dishes. Weekly medium change schedule was followed starting at DIV7. 2 dishes/each CiS line was pelleted DIV28 and DIV42 and stored at -80C. Genomic DNA was extracted from all neuron pellets in parallel with NucleoBond Xtra Maxi EF or Midi EF (Macherey-Nagel; Cat. No. 740424.10 or 740420.50, respectively). Samples were prepared, and sgRNA-encoding region were amplified for sequencing based on previously described protocols[74,108,109]. Sequencing and annotation were conducted at Memorial Sloan Kettering Cancer Center. Data was analyzed using the *MAGeCK-iNC* pipeline[74]. Hits were classified as having FDR-adjusted p-value < 0.1 and gene product cutoff was selected by the *MAGeCK-iNC* pipeline dynamically for each comparison.

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 ImagJ (NIH). *Prolactin* 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 $\geq$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. HuRI ppi was downloaded from http://www.interactome-atlas.org/download. The inBio Map protein–protein interaction (PPI) network database can be obtained from[97] (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[99]. TransposeNet's humanized α-synuclein-, β-amyloid-, and TDP-43-modifier networks were obtained from our previously published study[58]. 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. MGB Biobank 50k WES data was accessed through the https://biobankportal.partners.org/. Whole genome sequencing data (v3 release, 2022) from the Accelerating Medicines Partnership Parkinson's disease (AMP-PD) is available on 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-MLE analysis were performed using R (v4.3.2) packages 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. Christopher Cassa, Dr. Richard Sherwood, and Dr. Benjamin Neale for their valuable insights and resource sharing. 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). Part of this research has been conducted using the whole exome sequencing data and phenotypic information from ~500k voluntary participants from the UK Biobank (www.ukbiobank.ac.uk), a major biomedical database that is globally accessible to approved researchers who are undertaking health-related research that's in the public interest. UK Biobank 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 Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014). https://doi.org/10.1016/j.ajhg.2014.06.009

2 Chen, W., Coombes, B. J. & Larson, N. B. Recent advances and challenges of rare variant association analysis in the biobank sequencing era. *Front Genet* **13**, 1014947 (2022). https://doi.org/10.3389/fgene.2022.1014947

3 Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021). https://doi.org/10.1038/s41588-021-00870-7

4 Zhou, W. *et al.* SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet* **54**, 1466-1469 (2022). https://doi.org/10.1038/s41588-022-01178-w

5 Li, X. *et al.* Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nat Genet* **55**, 154-164 (2023). https://doi.org/10.1038/s41588-022-01225-6

6 Kryukov, G. V., Shpunt, A., Stamatoyannopoulos, J. A. & Sunyaev, S. R. Power of deep, all-exon resequencing for discovery of human trait genes. *Proc Natl Acad Sci U S A* **106**, 3871-3876 (2009). https://doi.org/10.1073/pnas.0812824106

7 Hallacli, E. *et al.* The Parkinson's disease protein alpha-synuclein is a modulator of processing bodies and mRNA stability. *Cell* **185**, 2035-2056 e2033 (2022). https://doi.org/10.1016/j.cell.2022.05.008

8 Bendapudi, P. K. *et al.* Low-frequency inherited complement receptor variants are associated with purpura fulminans. *Blood* **143**, 1032-1044 (2024). https://doi.org/10.1182/blood.2023021231

9 Lee, S. *et al.* Pathway-based approach using hierarchical components of collapsed rare variants. *Bioinformatics* **32**, i586-i594 (2016). https://doi.org/10.1093/bioinformatics/btw425

10 Lee, S., Kim, Y., Choi, S., Hwang, H. & Park, T. Pathway-based approach using hierarchical components of rare variants to analyze multiple phenotypes. *BMC Bioinformatics* **19**, 79 (2018). https://doi.org/10.1186/s12859-018-2066-9

11 Lee, S. *et al.* Pathway analysis of rare variants for the clustered phenotypes by using hierarchical structured components analysis. *BMC Med Genomics* **12**, 100 (2019). https://doi.org/10.1186/s12920-019-0517-4

12 Richardson, T. G., Timpson, N. J., Campbell, C. & Gaunt, T. R. A pathway-centric approach to rare variant association analysis. *Eur J Hum Genet* **25**, 123-129 (2016). https://doi.org/10.1038/ejhg.2016.113

13 Stoel, R. D., Garre, F. G., Dolan, C. & van den Wittenboer, G. On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychol Methods* **11**, 439-455 (2006). https://doi.org/10.1037/1082-989X.11.4.439

14 SHAPIRO, A. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72**, 133-144 (1985). https://doi.org/10.1093/biomet/72.1.133

15 Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-321 (2008). https://doi.org/10.1016/j.ajhg.2008.06.024
16 Derkach, A., Lawless, J. F. & Sun, L. 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 (2013). https://doi.org/10.1002/gepi.21689
17 Lee, S. *et al.* 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 (2012). https://doi.org/10.1016/j.ajhg.2012.06.007
18 Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**, 765-769 (2004). https://doi.org/10.1086/383251
19 Paththinige, C. S., Sirisena, N. D. & Dissanayake, V. Genetic determinants of inherited susceptibility to hypercholesterolemia - a comprehensive literature review. *Lipids Health Dis* **16**, 103 (2017). https://doi.org/10.1186/s12944-017-0488-4
20 Karczewski, K. J. *et al.* Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* **2**, 100168 (2022). https://doi.org/10.1016/j.xgen.2022.100168
21 Weiner, D. J. *et al.* Polygenic architecture of rare coding variation across 394,783 exomes. *Nature* **614**, 492-499 (2023). https://doi.org/10.1038/s41586-022-05684-z
22 Hu, C. *et al.* The Contribution of Germline Predisposition Gene Mutations to Clinical Subtypes of Invasive Breast Cancer From a Clinical Genetic Testing Cohort. *J Natl Cancer Inst* **112**, 1231-1241 (2020). https://doi.org/10.1093/jnci/djaa023
23 Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017). https://doi.org/10.1038/nature24284
24 Alblihy, A. *et al.* Untangling the clinicopathological significance of MRE11-RAD50-NBS1 complex in sporadic breast cancers. *NPJ Breast Cancer* **7**, 143 (2021). https://doi.org/10.1038/s41523-021-00350-5
25 Huerta-Chagoya, A. *et al.* Rare variant analyses in 51,256 type 2 diabetes cases and 370,487 controls reveal the pathogenicity spectrum of monogenic diabetes genes. *Nat Genet* **56**, 2370-2379 (2024). https://doi.org/10.1038/s41588-024-01947-9
26 Vujkovic, M. *et al.* 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 (2020). https://doi.org/10.1038/s41588-020-0637-y
27 Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* **40**, 638-645 (2008). https://doi.org/10.1038/ng.120
28 Zhang, W. *et al.* Bidirectional relationship between type 2 diabetes mellitus and coronary artery disease: Prospective cohort study and genetic analyses. *Chin Med J (Engl)* **137**, 577-587 (2024). https://doi.org/10.1097/CM9.0000000000002894
29 Li, S., Shin, H. J., Ding, E. L. & van Dam, R. M. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA* **302**, 179-188 (2009). https://doi.org/10.1001/jama.2009.976

30 Yamauchi, T. *et al.* Inhibition of RXR and PPARgamma ameliorates diet-induced obesity and type 2 diabetes. *J Clin Invest* **108**, 1001-1013 (2001). https://doi.org/10.1172/JCI12864
31 Miyazaki, S. *et al.* Nuclear hormone retinoid X receptor (RXR) negatively regulates the glucose-stimulated insulin secretion of pancreatic ss-cells. *Diabetes* **59**, 2854-2861 (2010). https://doi.org/10.2337/db09-1897
32 Anderson, A. & Walker, B. R. 11beta-HSD1 inhibitors for the treatment of type 2 diabetes and cardiovascular disease. *Drugs* **73**, 1385-1393 (2013). https://doi.org/10.1007/s40265-013-0112-5
33 Alberts, P. *et al.* Selective inhibition of 11 beta-hydroxysteroid dehydrogenase type 1 improves hepatic insulin sensitivity in hyperglycemic mice strains. *Endocrinology* **144**, 4755-4762 (2003). https://doi.org/10.1210/en.2003-0344
34 Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease. *Cell* **177**, 132-145 (2019). https://doi.org/10.1016/j.cell.2019.02.015
35 Li, W. *et al.* Rare and common coding variants in lipid metabolism-related genes and their association with coronary artery disease. *BMC Cardiovasc Disord* **24**, 97 (2024). https://doi.org/10.1186/s12872-024-03759-5
36 Rocheleau, G. *et al.* Rare variant contribution to the heritability of coronary artery disease. *Nat Commun* **15**, 8741 (2024). https://doi.org/10.1038/s41467-024-52939-6
37 Schnitzler, G. R. *et al.* Convergence of coronary artery disease genes onto endothelial cell programs. *Nature* **626**, 799-807 (2024). https://doi.org/10.1038/s41586-024-07022-x
38 Cadby, G. *et al.* Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. *Nat Commun* **13**, 3124 (2022). https://doi.org/10.1038/s41467-022-30875-7
39 Maiti, P., Manna, J. & Dunbar, G. L. Current understanding of the molecular mechanisms in Parkinson's disease: Targets for potential treatments. *Transl Neurodegener* **6**, 28 (2017). https://doi.org/10.1186/s40035-017-0099-z
40 Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* **24**, 197-211 (2003). https://doi.org/10.1016/s0197-4580(02)00065-9
41 Nalls, M. A. *et al.* 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 (2019). https://doi.org/10.1016/S1474-4422(19)30320-5
42 Kim, J. J. *et al.* Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. *Nat Genet* **56**, 27-36 (2024). https://doi.org/10.1038/s41588-023-01584-8
43 Pitz, V. *et al.* Analysis of rare Parkinson's disease variants in millions of people. *NPJ Parkinsons Dis* **10**, 11 (2024). https://doi.org/10.1038/s41531-023-00608-8
44 Makarious, M. B. *et al.* Large-scale rare variant burden testing in Parkinson's disease. *Brain* **146**, 4622-4632 (2023). https://doi.org/10.1093/brain/awad214

45 Towns, C. *et al.* Defining the causes of sporadic Parkinson's disease in the global Parkinson's genetics program (GP2). *NPJ Parkinsons Dis* **9**, 131 (2023). https://doi.org/10.1038/s41531-023-00533-w
46 Gitler, A. D. *et al.* Alpha-synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. *Nat Genet* **41**, 308-315 (2009). https://doi.org/10.1038/ng.300
47 Dzamko, N., Zhou, J., Huang, Y. & Halliday, G. M. Parkinson's disease-implicated kinases in the brain; insights into disease pathogenesis. *Front Mol Neurosci* **7**, 57 (2014). https://doi.org/10.3389/fnmol.2014.00057
48 Lee, B. D. *et al.* Inhibitors of leucine-rich repeat kinase-2 protect against models of Parkinson's disease. *Nat Med* **16**, 998-1000 (2010). https://doi.org/10.1038/nm.2199
49 Taymans, J. M. *et al.* Perspective on the current state of the LRRK2 field. *NPJ Parkinsons Dis* **9**, 104 (2023). https://doi.org/10.1038/s41531-023-00544-7
50 Barallobre, M. J. *et al.* DYRK1A promotes dopaminergic neuron survival in the developing brain and in a mouse model of Parkinson's disease. *Cell Death Dis* **5**, e1289 (2014). https://doi.org/10.1038/cddis.2014.253
51 Saminathan, H. *et al.* Fyn Kinase-Mediated PKCdelta Y311 Phosphorylation Induces Dopaminergic Degeneration in Cell Culture and Animal Models: Implications for the Identification of a New Pharmacological Target for Parkinson's Disease. *Front Pharmacol* **12**, 631375 (2021). https://doi.org/10.3389/fphar.2021.631375
52 Panicker, N. *et al.* Fyn kinase regulates misfolded alpha-synuclein uptake and NLRP3 inflammasome activation in microglia. *J Exp Med* **216**, 1411-1430 (2019). https://doi.org/10.1084/jem.20182191
53 Guglietti, B. *et al.* Fyn kinase inhibition using AZD0530 improves recognition memory and reduces depressive-like behaviour in an experimental model of Parkinson's disease. *bioRxiv*, 2021.2006.2016.448746 (2021). https://doi.org/10.1101/2021.06.16.448746
54 Alexopoulou, Z. *et al.* Deubiquitinase Usp8 regulates alpha-synuclein clearance and modifies its toxicity in Lewy body disease. *Proc Natl Acad Sci U S A* **113**, E4688-4697 (2016). https://doi.org/10.1073/pnas.1523597113
55 Mauri, S. *et al.* USP8 Down-Regulation Promotes Parkin-Independent Mitophagy in the Drosophila Brain and in Human Neurons. *Cells* **12** (2023). https://doi.org/10.3390/cells12081143
56 Tang, D. *et al.* Endogenous HMGB1 regulates autophagy. *J Cell Biol* **190**, 881-892 (2010). https://doi.org/10.1083/jcb.200911078
57 Takahashi, M. *et al.* USP10 Is a Driver of Ubiquitinated Protein Aggregation and Aggresome Formation to Inhibit Apoptosis. *iScience* **9**, 433-450 (2018). https://doi.org/10.1016/j.isci.2018.11.006
58 Khurana, V. *et al.* Genome-Scale Networks Link Neurodegenerative Disease Genes to alpha-Synuclein through Specific Molecular Pathways. *Cell Syst* **4**, 157-170 e114 (2017). https://doi.org/10.1016/j.cels.2016.12.011
59 Funayama, M., Nishioka, K., Li, Y. & Hattori, N. Molecular genetics of Parkinson's disease: Contributions and global trends. *J Hum Genet* **68**, 125-130 (2023). https://doi.org/10.1038/s10038-022-01058-5

60 Chung, C. Y. *et al.* 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 (2017). https://doi.org/10.1016/j.cels.2017.01.002
61 Ham, S. J. *et al.* Decision between mitophagy and apoptosis by Parkin via VDAC1 ubiquitination. *Proc Natl Acad Sci U S A* **117**, 4281-4291 (2020). https://doi.org/10.1073/pnas.1909814117
62 Chu, Y. *et al.* Abnormal alpha-synuclein reduces nigral voltage-dependent anion channel 1 in sporadic and experimental Parkinson's disease. *Neurobiol Dis* **69**, 1-14 (2014). https://doi.org/10.1016/j.nbd.2014.05.003
63 He, Y. *et al.* The Potential Role of Voltage-Dependent Anion Channel in the Treatment of Parkinson's Disease. *Oxid Med Cell Longev* **2022**, 4665530 (2022). https://doi.org/10.1155/2022/4665530
64 Chung, C. Y. *et al.* Identification and rescue of alpha-synuclein toxicity in Parkinson patient-derived neurons. *Science* **342**, 983-987 (2013). https://doi.org/10.1126/science.1245296
65 Tofaris, G. K. *et al.* Ubiquitin ligase Nedd4 promotes alpha-synuclein degradation by the endosomal-lysosomal pathway. *Proc Natl Acad Sci U S A* **108**, 17004-17009 (2011). https://doi.org/10.1073/pnas.1109356108
66 Tardiff, D. F. *et al.* Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. *Science* **342**, 979-983 (2013). https://doi.org/10.1126/science.1245321
67 McLean, P. J. *et al.* TorsinA and heat shock proteins act as molecular chaperones: suppression of alpha-synuclein aggregation. *J Neurochem* **83**, 846-854 (2002). https://doi.org/10.1046/j.1471-4159.2002.01190.x
68 Fu, X., Qu, L., Xu, H. & Xie, J. Ndfip1 protected dopaminergic neurons via regulating mitochondrial function and ferroptosis in Parkinson's disease. *Exp Neurol* **375**, 114724 (2024). https://doi.org/10.1016/j.expneurol.2024.114724
69 Howitt, J. *et al.* Increased Ndfip1 in the substantia nigra of Parkinsonian brains is associated with elevated iron levels. *PLoS One* **9**, e87119 (2014). https://doi.org/10.1371/journal.pone.0087119
70 Villaescusa, J. C. *et al.* A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. *EMBO J* **35**, 1963-1978 (2016). https://doi.org/10.15252/embj.201593725
71 Pranski, E. *et al.* NF-kappaB activity is inversely correlated to RNF11 expression in Parkinson's disease. *Neurosci Lett* **547**, 16-20 (2013). https://doi.org/10.1016/j.neulet.2013.04.056
72 Nazeen, S. *et al.* 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 (2024). https://doi.org/10.1101/2024.03.03.583145
73 Lam, I. *et al.* Rapid iPSC inclusionopathy models shed light on formation, consequence and molecular subtype of α-synuclein inclusions. *bioRxiv*, 2022.2011.2008.515615 (2022). https://doi.org/10.1101/2022.11.08.515615

74 Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255 e212 (2019). https://doi.org/10.1016/j.neuron.2019.07.014
75 Molina-Salinas, G., Rodriguez-Chavez, V., Langley, E. & Cerbon, M. Prolactin-induced neuroprotection against excitotoxicity is mediated via PI3K/AKT and GSK3beta/NF-kappaB in primary cultures of hippocampal neurons. *Peptides* **166**, 171037 (2023). https://doi.org/10.1016/j.peptides.2023.171037
76 Duc Nguyen, H. *et al.* Prolactin and Its Altered Action in Alzheimer's Disease and Parkinson's Disease. *Neuroendocrinology* **112**, 427-445 (2022). https://doi.org/10.1159/000517798
77 Duc Nguyen, H. *et al.* Association between Serum Prolactin Levels and Neurodegenerative Diseases: Systematic Review and Meta-Analysis. *Neuroimmunomodulation* **29**, 85-96 (2022). https://doi.org/10.1159/000519552
78 Karayel, O. *et al.* Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Rep Med* **3**, 100661 (2022). https://doi.org/10.1016/j.xcrm.2022.100661
79 Al-Kuraishy, H. M., Jabir, M. S., Al-Gareeb, A. I. & Albuhadily, A. K. The conceivable role of prolactin hormone in Parkinson disease: The same goal but with different ways. *Ageing Res Rev* **91**, 102075 (2023). https://doi.org/10.1016/j.arr.2023.102075
80 Grattan, D. R. Does the brain make prolactin? *J Neuroendocrinol* **36**, e13432 (2024). https://doi.org/10.1111/jne.13432
81 Featherstone, K., White, M. R. & Davis, J. R. The prolactin gene: a paradigm of tissue-specific gene regulation with complex temporal transcription dynamics. *J Neuroendocrinol* **24**, 977-990 (2012). https://doi.org/10.1111/j.1365-2826.2012.02310.x
82 Luk, K. C. *et al.* Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949-953 (2012). https://doi.org/10.1126/science.1227157
83 Lam, I. *et al.* Rapid iPSC inclusionopathy models shed light on formation, consequence, and molecular subtype of alpha-synuclein inclusions. *Neuron* **112**, 2886-2909 e2816 (2024). https://doi.org/10.1016/j.neuron.2024.06.002
84 Luk, K. C. *et al.* Molecular and Biological Compatibility with Host Alpha-Synuclein Influences Fibril Pathogenicity. *Cell Rep* **16**, 3373-3387 (2016). https://doi.org/10.1016/j.celrep.2016.08.053
85 Loor, G. *et al.* Menadione triggers cell death through ROS-dependent mechanisms involving PARP activation without requiring apoptosis. *Free Radic Biol Med* **49**, 1925-1936 (2010). https://doi.org/10.1016/j.freeradbiomed.2010.09.021
86 Di Nubila, A., Dilella, G., Simone, R. & Barbieri, S. S. Vascular Extracellular Matrix in Atherosclerosis. *Int J Mol Sci* **25** (2024). https://doi.org/10.3390/ijms252212017
87 Manon-Jensen, T., Kjeld, N. G. & Karsdal, M. A. Collagen-mediated hemostasis. *J Thromb Haemost* **14**, 438-448 (2016). https://doi.org/10.1111/jth.13249
88 Libby, P. Inflammation during the life cycle of the atherosclerotic plaque. *Cardiovasc Res* **117**, 2525-2536 (2021). https://doi.org/10.1093/cvr/cvab303

89 Sakata, T. *et al.* Drosophila Nedd4 regulates endocytosis of notch and suppresses its ligand-independent activation. *Curr Biol* **14**, 2228-2236 (2004). https://doi.org/10.1016/j.cub.2004.12.028
90 Privman Champaloux, E. *et al.* Ring Finger Protein 11 (RNF11) Modulates Dopamine Release in Drosophila. *Neuroscience* **452**, 37-48 (2021). https://doi.org/10.1016/j.neuroscience.2020.10.021
91 Wakabayashi-Ito, N. *et al.* Dtorsin, the Drosophila ortholog of the early-onset dystonia TOR1A (DYT1), plays a novel role in dopamine metabolism. *PLoS One* **6**, e26183 (2011). https://doi.org/10.1371/journal.pone.0026183
92 Yong, Y. *et al.* Dyrk1a Phosphorylation of alpha-Synuclein Mediating Apoptosis of Dopaminergic Neurons in Parkinson's Disease. *Parkinsons Dis* **2023**, 8848642 (2023). https://doi.org/10.1155/2023/8848642
93 All of Us Research Program Genomics, I. Genomic data in the All of Us Research Program. *Nature* **627**, 340-346 (2024). https://doi.org/10.1038/s41586-023-06957-x
94 Ramos-Martinez, E., Ramos-Martinez, I., Molina-Salinas, G., Zepeda-Ruiz, W. A. & Cerbon, M. The role of prolactin in central nervous system inflammation. *Rev Neurosci* **32**, 323-340 (2021). https://doi.org/10.1515/revneuro-2020-0082
95 Macias, F., Ulloa, M., Clapp, C., Martinez de la Escalera, G. & Arnold, E. Prolactin protects hippocampal neurons against H2O2-induced neurotoxicity by suppressing BAX and NOX4 via the NF-kappaB signaling pathway. *PLoS One* **19**, e0313328 (2024). https://doi.org/10.1371/journal.pone.0313328
96 Szklarczyk, D. *et al.* 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 (2023). https://doi.org/10.1093/nar/gkac1000
97 Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* **14**, 61-64 (2017). https://doi.org/10.1038/nmeth.4083
98 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402-408 (2020). https://doi.org/10.1038/s41586-020-2188-x
99 Megchelenbrink, W., Katzir, R., Lu, X., Ruppin, E. & Notebaart, R. A. 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 (2015). https://doi.org/10.1073/pnas.1508573112
100 Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020). https://doi.org/10.1126/science.aaz1776
101 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784 (2017). https://doi.org/10.1038/ng.3984
102 Dempster, J. M. *et al.* Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv*, 720243 (2019). https://doi.org/10.1101/720243

103 Gonzalez, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215-226 (2014). https://doi.org/10.1016/j.stem.2014.05.018

104 DeWeirdt, P. C. *et al.* Genetic screens in isogenic mammalian cell lines without single cell cloning. *Nat Commun* **11**, 752 (2020). https://doi.org/10.1038/s41467-020-14620-6

105 Kim, T. W. *et al.* Biphasic Activation of WNT Signaling Facilitates the Derivation of Midbrain Dopamine Neurons from hESCs for Translational Use. *Cell Stem Cell* **28**, 343-355 e345 (2021). https://doi.org/10.1016/j.stem.2021.01.005

106 Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756-780 (2019). https://doi.org/10.1038/s41596-018-0113-7

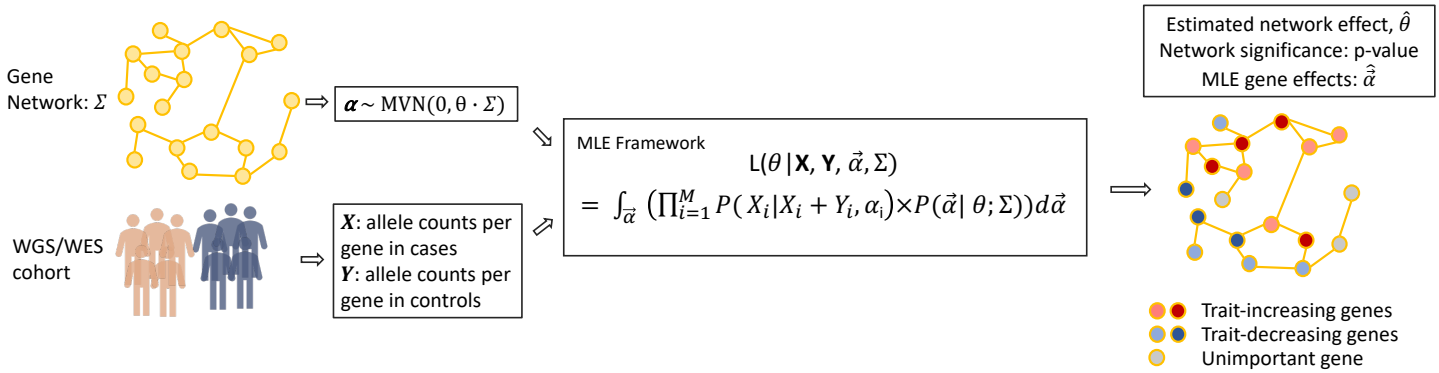
107 Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5** (2016). https://doi.org/10.7554/eLife.19760

108 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014). https://doi.org/10.1016/j.cell.2014.09.029

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

Figure 1.

A.


Gene Network: Σ
$\boldsymbol{\alpha} \sim \mathrm{MVN}(0, \theta \cdot \Sigma)$
WGS/WES cohort
$\boldsymbol{X}$: allele counts per gene in cases
$\boldsymbol{Y}$: allele counts per gene in controls
MLE Framework
$L(\theta \mid \mathbf{X}, \mathbf{Y}, \vec{\alpha}, \Sigma)$
$= \int_{\vec{\alpha}} \left(\prod_{i=1}^{M} P(X_i \mid X_i + Y_i, \alpha_i) \times P(\vec{\alpha} \mid \theta; \Sigma)\right) d\vec{\alpha}$
Estimated network effect, $\hat{\theta}$
Network significance: p-value
MLE gene effects: $\hat{\vec{\alpha}}$
Trait-increasing genes
Trait-decreasing genes
Unimportant gene

B.

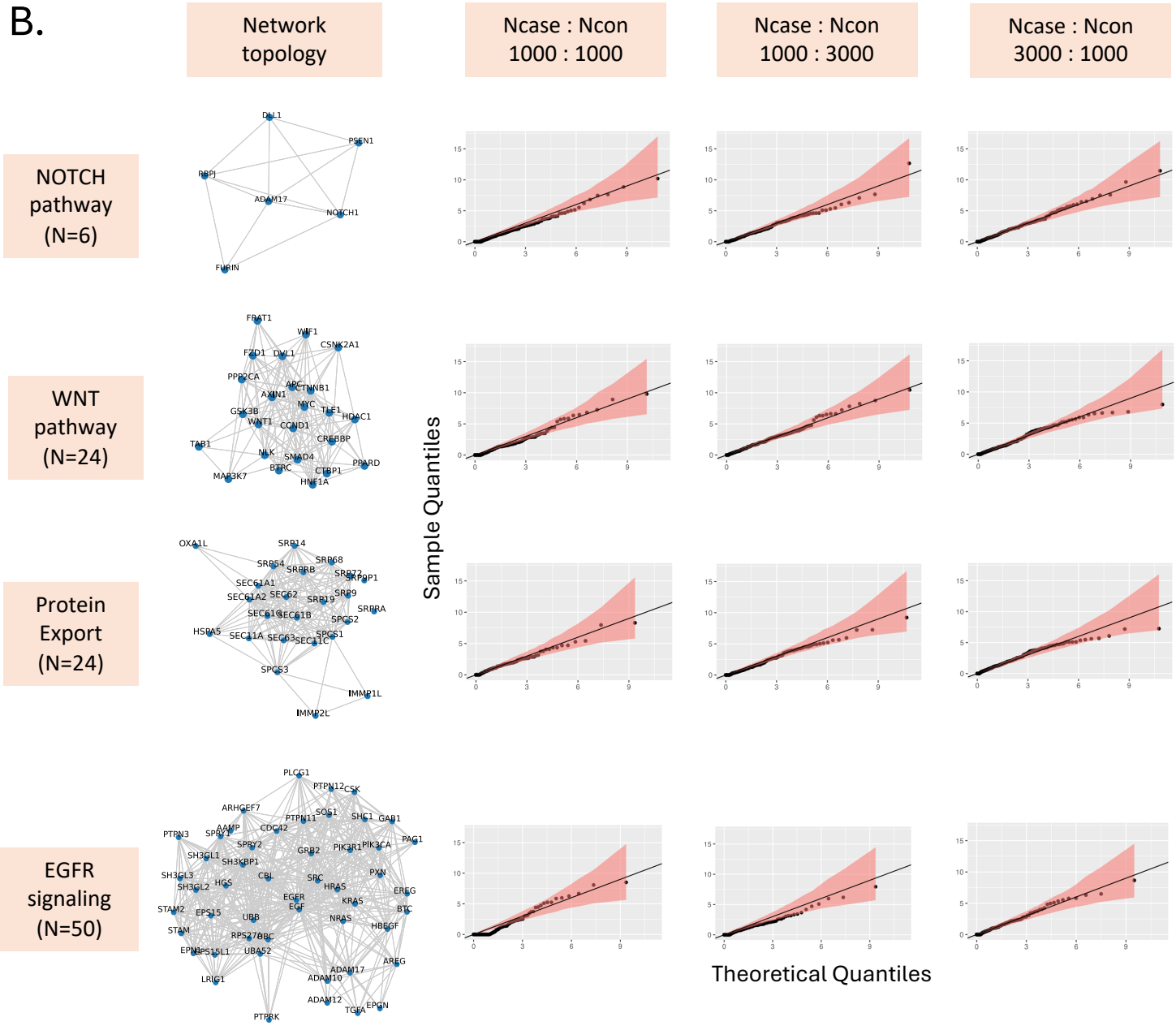
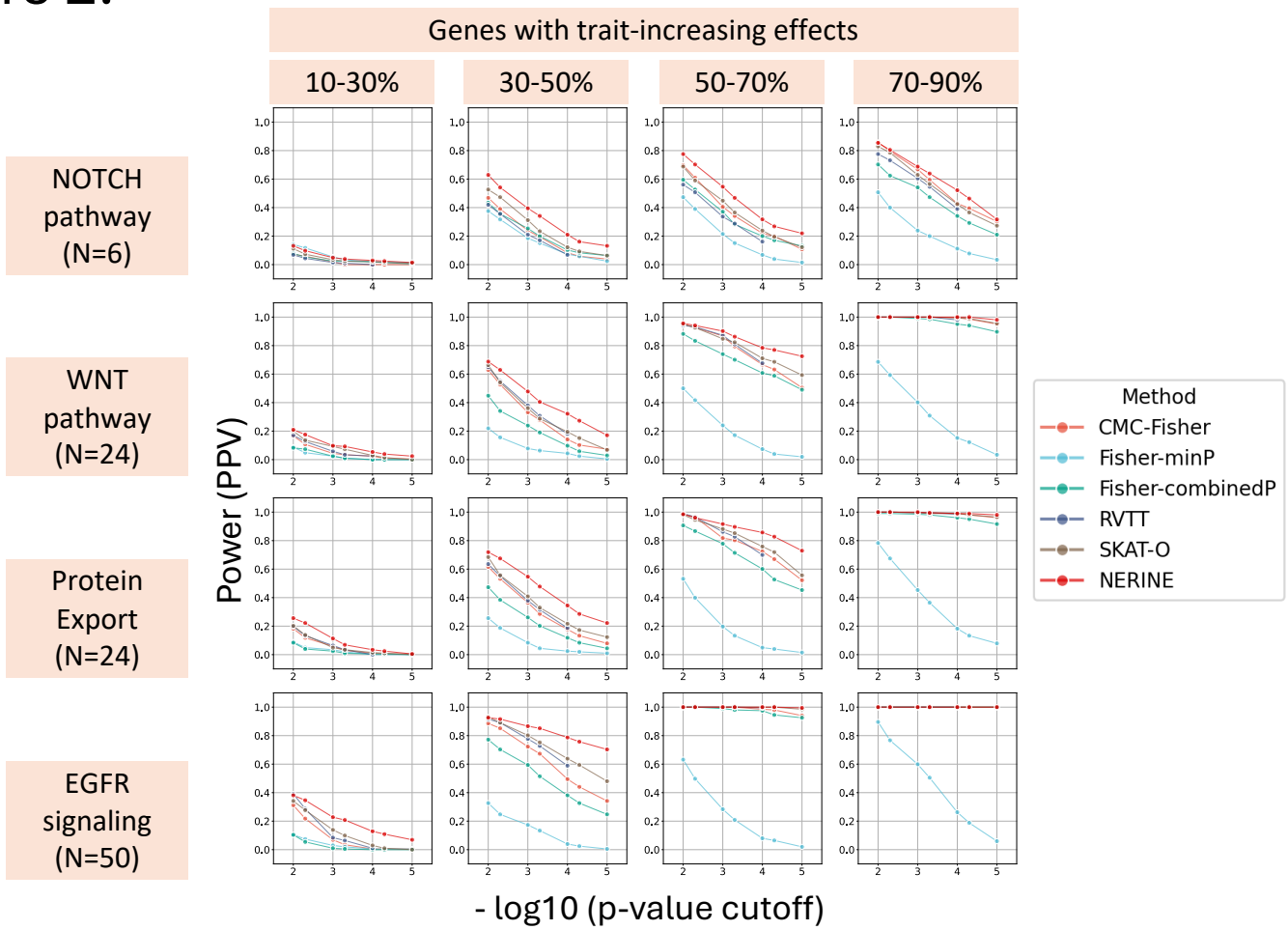

Network topology
Ncase : Ncon 1000 : 1000
Ncase : Ncon 1000 : 3000
Ncase : Ncon 3000 : 1000
NOTCH pathway (N=6)
WNT pathway (N=24)
Protein Export (N=24)
EGFR signaling (N=50)
Sample Quantiles
Theoretical Quantiles

Figure 1. Overview of NERINE and its performance in null simulations.
A. Schematic diagram of NERINE, a new statistical framework that infers the effect of rare variants aggregated across a gene network on a trait. Input to NERINE includes observed allele counts per gene in cases (denoted by X) and controls (denoted by Y) in whole-genome or whole-exome sequencing (WGS or WES) datasets and a gene network represented as a symmetric positive-definite matrix, Σ. Network edges between genes can encode a wide range of relationships including physical and genetic interactions, co-expression, co-essentiality, pathway membership, etc. and can be extracted either from existing databases, literature, or novel experimental assays. NERINE models individual gene-effects, $\vec{\alpha}$, using a multivariate normal distribution whose variance-covariance parameter is controlled by the input network structure, Σ and the overall network effect, θ (object of inference). This 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 in cases given the allele count in the overall cohort in the network, and (ii) the probability of observing a particular combination of gene-effects given the network structure and network effect size. NERINE tests for network significance by nested hypothesis testing using the log-likelihood ratio (LLR) as the test-statistic and provides an asymptotic p-value from the mixture of delta function and a chi-square distribution with a degree-of-freedom of one. It also estimates the most likely combination of gene effects, $\hat{\vec{\alpha}}$ with the estimated network effect, $\hat{\theta}$. **B.** Simulations at null showing that NERINE's test-statistic asymptotically follows the theoretical distribution of a mixture of the delta function and a chi-square distribution with a degree-of-freedom of one. Simulations were performed with different network architectures for well-studied pathways of different sizes namely, NOTCH pathway, WNT pathway, protein export pathway, and EGFR signaling pathway. Pathway 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 (see Methods). Simulations were performed in cohorts with different case-control skews. Confidence bands in the QQ-plots represent 95% bootstrap confidence intervals around NERINE's test-statistic.

Figure 2.

A.


Genes with trait-increasing 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)
- log10 (p-value cutoff)
Method
CMC-Fisher
Fisher-minP
Fisher-combinedP
RVTT
SKAT-O
NERINE

B.

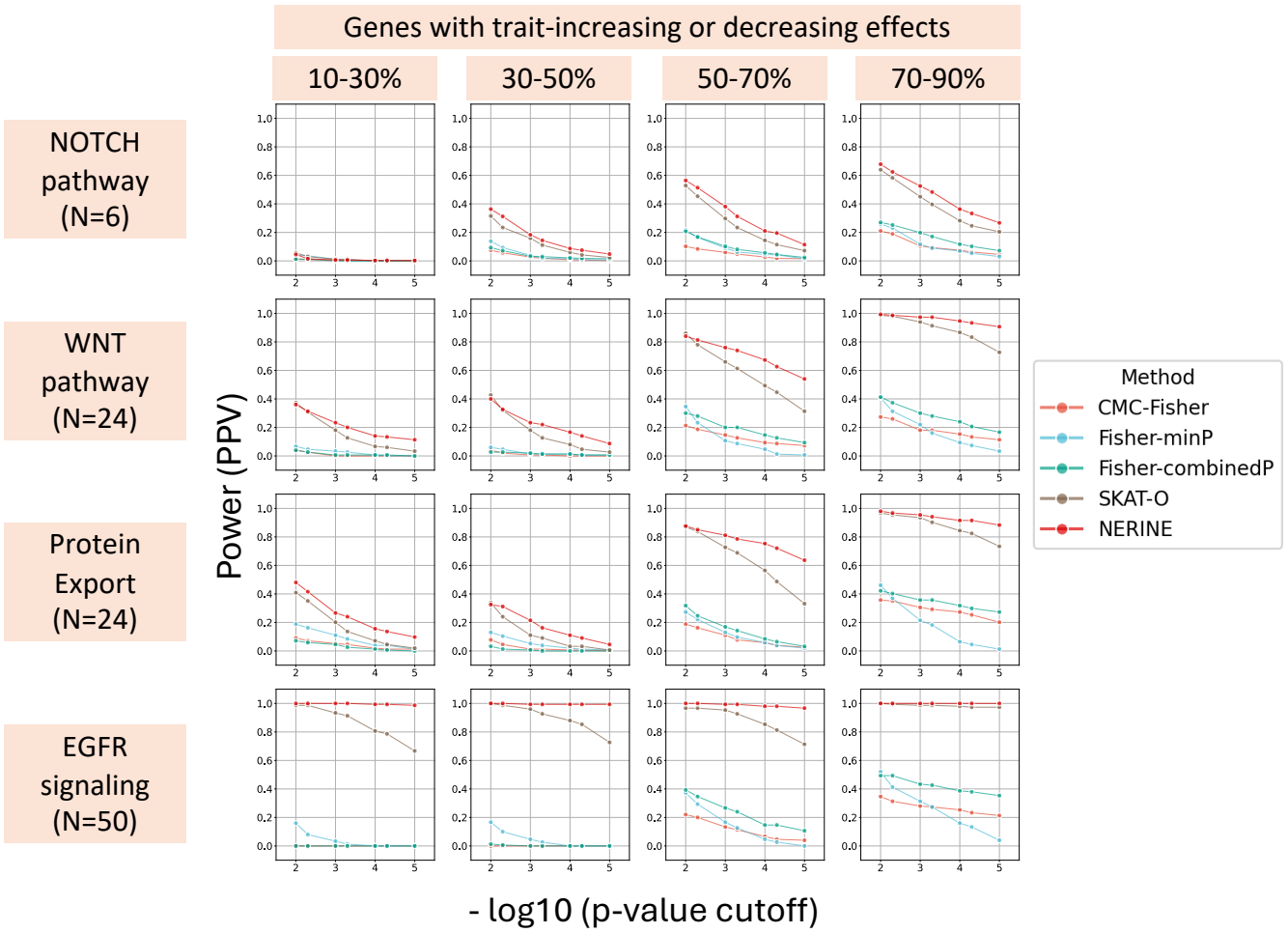
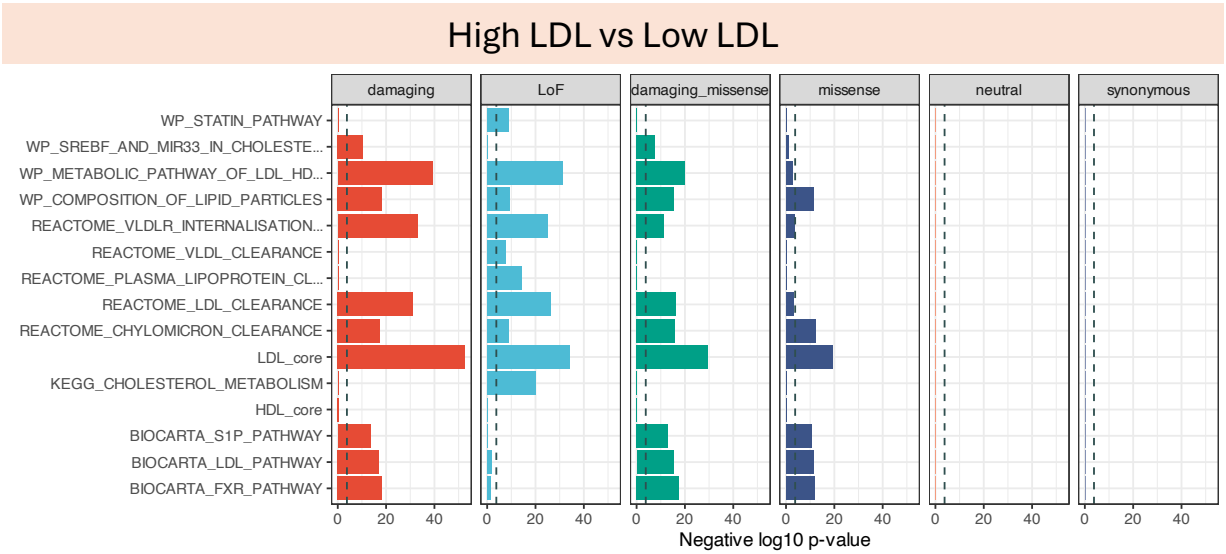

Genes with trait-increasing or decreasing effects
10-30%
30-50%
50-70%
70-90%
NOTCH pathway (N=6)
WNT pathway (N=24)
Protein Export (N=24)
EGFR signaling (N=50)
Power (PPV)
- log10 (p-value cutoff)
Method
CMC-Fisher
Fisher-minP
Fisher-combinedP
SKAT-O
NERINE

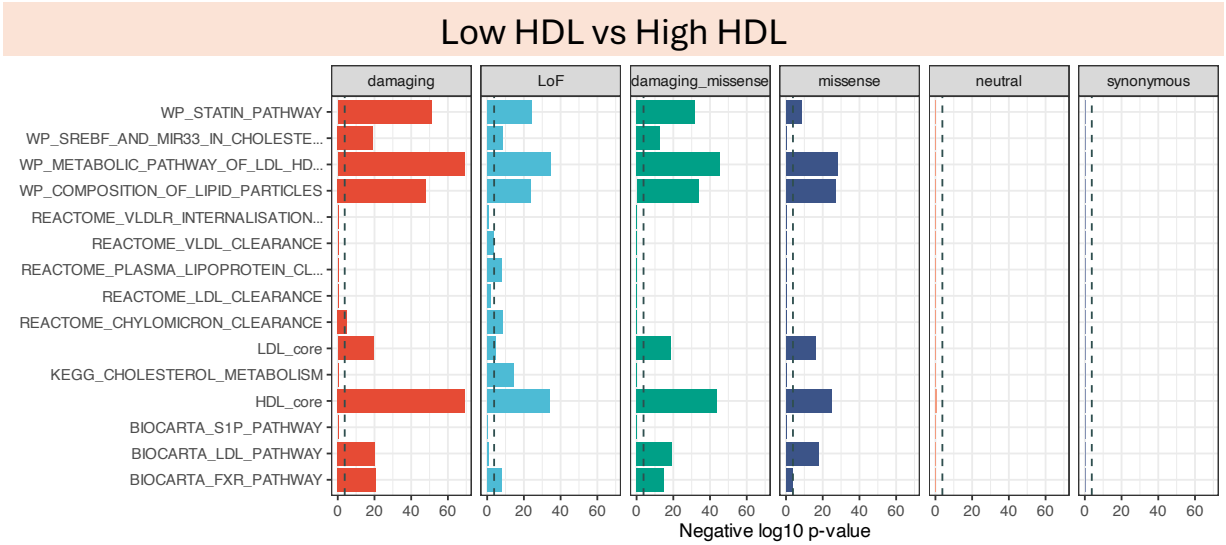
Figure 2. Comparing the performance of NERINE with existing rare variant association tests in simulations. Empirical power of the methods was measured for a simulated binary disease trait in a cohort of 2,000 cases and 2,000 controls with network effect, $\theta = 0.2$, in four well-studied database pathways in two scenarios: (i) genes in the network having only trait-increasing effects, and (ii) genes in the network having both trait-increasing and trait-decreasing effects. From left to right, the power plots show settings in which different proportions of genes within the network have effects on the trait, mimicking scenarios from having a very noisy network to a highly relevant network. Power was calculated as the positive predictive value at different p-value cutoffs: 1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5, and 1e-5. In all scenarios, NERINE outperforms existing rare variant tests. Note that, RVTT p-values were calculated from 10,000 permutations, hence we don't report RVTT's power for the cutoffs below 1e-4. Also, RVTT is a test to look for monotonic trends in rare variant occurrences within a pathway. Hence, it was excluded from the comparison in scenario (ii).

Figure 3

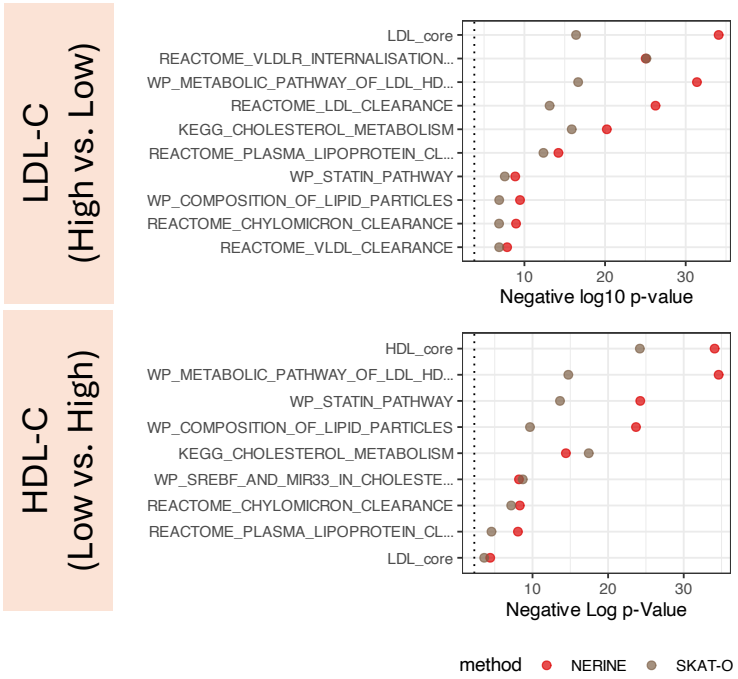
A.



B.



C.



D.

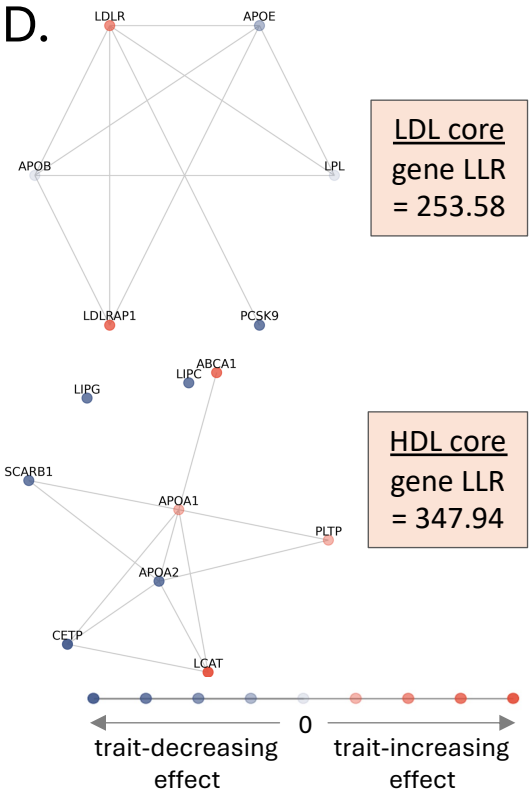
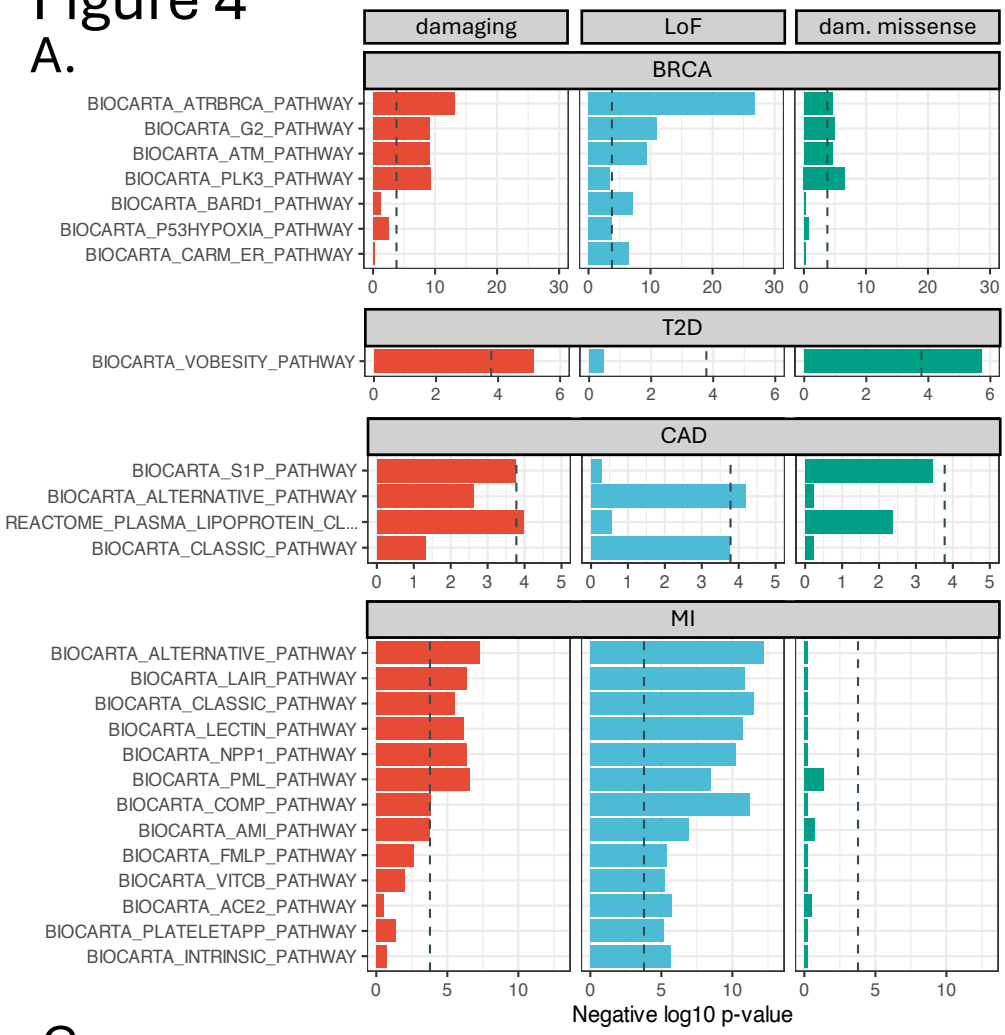
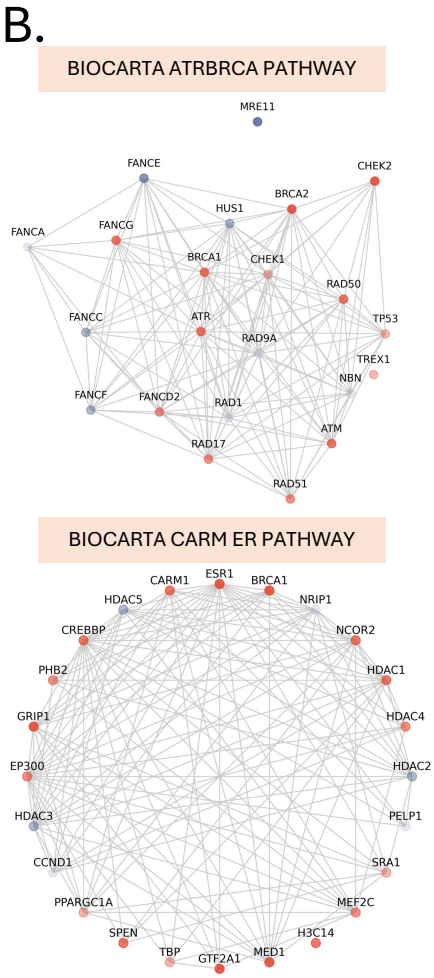
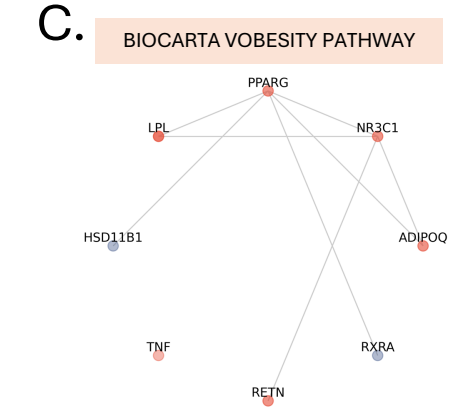


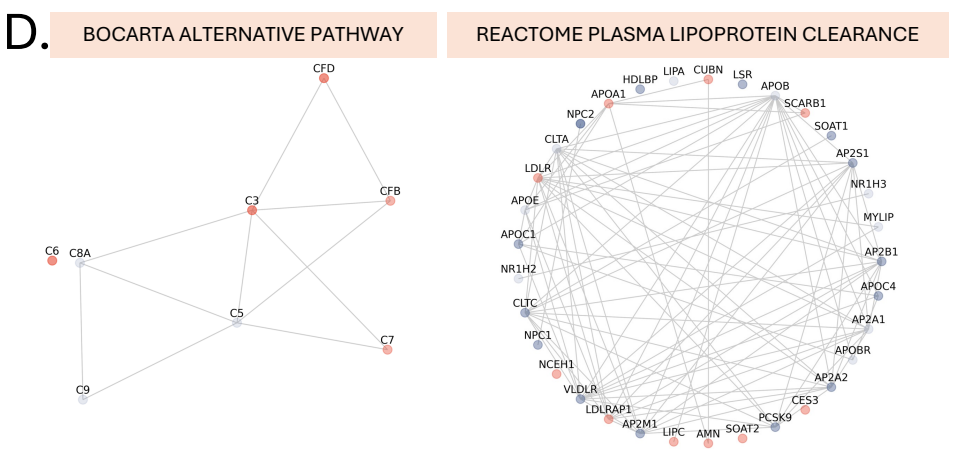
Figure 3. Performance of NERINE while comparing high LDL-direct vs low LDL-direct individuals and low HDL-direct vs high HDL-direct individuals in the UK biobank.
A. While comparing individuals with high LDL cholesterol with the ones with low LDL cholesterol in UKBB, NERINE identifies significant cumulative effect of rare (MAF < 0.001) variants in LoF (i.e., frameshifts, insertions, deletions, and splice), damaging missense (i.e., missenses predicted to be damaging by in-silico tools), damaging (i.e., damaging missense and LoF), and missense categories in key lipid-related pathways. No significant burden of neutral missense and synonymous variants was observed. The tests were performed across our canonical pathway database of 306 pathways. Pathway 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 (see Methods). The dashed grey line represents the Bonferroni-corrected p-value threshold of 0.05. The core module of LDL genes, which contains *LDLR* and *PCSK9*, was identified as the most significant hit which serves as a positive control. **B.** While comparing individuals with low HDL cholesterol with the ones with high HDL cholesterol in UKBB, NERINE identifies significant cumulative effect of rare (MAF < 0.001) variants in LoF, damaging missense, damaging, and missense categories in key lipid-related pathways. Notably, the module of core HDL genes, which contains *ABCA1*, *CETP*, *LIPC*, and *LIPG*, was the most significant hit, serving as a positive control. No significant burden of neutral missense and synonymous variants was observed. The tests were performed across the same canonical pathway database. The dashed grey line represents the Bonferroni-corrected p-value threshold of 0.05. **C.** Comparison of SKAT-O p-values against NERINE p-values for rare LoF variant burden across the significant pathways in both the LDL (high vs low) and HDL (low vs high) phenotypes. For most of the pathways in both phenotypes, NERINE provides a lower p-value than SKAT-O. SKAT-O was applied at the pathway level aggregating the allele counts from member genes. **D.** NERINE's estimates of gene effects in the most significant pathways with rare damaging variant burden in the LDL (high vs low) and HDL (low vs high) phenotypes. Trait-increasing effects are represented by different shades of orange and trait-decreasing effects are represented by different shades of blue as shown on the scale. Darker color represents more pronounced effect. For example, *PCSK9* and *APOB* shows trait-decreasing effects and *LDLR* shows a trait-increasing effect on LDL cholesterol (high vs low) phenotype. Similarly, *ABCA1* and *LCAT* show trait-increasing effects, while *LIPC* and *LIPG* show trait-decreasing effects for the HDL (low vs high) phenotypes. These findings agree with known lipid biology. Note, NERINE's predicted gene effects represent the "most likely scenario" with the observed allele counts per gene and the gene-gene network topology under the estimated network effect. NERINE does not provide p-values on per-gene predictions.

Figure 4


A.
damaging
LoF
dam. missense
BRCA
BIOCARTA_ATRBRCA_PATHWAY
BIOCARTA_G2_PATHWAY
BIOCARTA_ATM_PATHWAY
BIOCARTA_PLK3_PATHWAY
BIOCARTA_BARD1_PATHWAY
BIOCARTA_P53HYPOXIA_PATHWAY
BIOCARTA_CARM_ER_PATHWAY
T2D
BIOCARTA_VOBESITY_PATHWAY
CAD
BIOCARTA_S1P_PATHWAY
BIOCARTA_ALTERNATIVE_PATHWAY
REACTOME_PLASMA_LIPOPROTEIN_CL...
BIOCARTA_CLASSIC_PATHWAY
MI
BIOCARTA_ALTERNATIVE_PATHWAY
BIOCARTA_LAIR_PATHWAY
BIOCARTA_CLASSIC_PATHWAY
BIOCARTA_LECTIN_PATHWAY
BIOCARTA_NPP1_PATHWAY
BIOCARTA_PML_PATHWAY
BIOCARTA_COMP_PATHWAY
BIOCARTA_AMI_PATHWAY
BIOCARTA_FMLP_PATHWAY
BIOCARTA_VITCB_PATHWAY
BIOCARTA_ACE2_PATHWAY
BIOCARTA_PLATELETAPP_PATHWAY
BIOCARTA_INTRINSIC_PATHWAY
Negative log10 p-value


B.
BIOCARTA ATRBRCA PATHWAY
MRE11
FANCE
CHEK2
BRCA2
HUS1
FANCA
FANCG
BRCA1
CHEK1
RAD50
FANCC
ATR
TP53
RAD9A
TREX1
NBN
FANCF
FANCD2
RAD1
ATM
RAD17
RAD51
BIOCARTA CARM ER PATHWAY
CARM1
ESR1
BRCA1
HDAC5
NRIP1
CREBBP
NCOR2
PHB2
HDAC1
GRIP1
HDAC4
EP300
HDAC2
HDAC3
PELP1
CCND1
SRA1
PPARGC1A
MEF2C
SPEN
H3C14
TBP
GTF2A1
MED1


C.
BIOCARTA VOBESITY PATHWAY
PPARG
LPL
NR3C1
HSD11B1
ADIPOQ
TNF
RXRA
RETN


D.
BOCARTA ALTERNATIVE PATHWAY
CFD
C3
CFB
C6
C8A
C5
C7
C9
REACTOME PLASMA LIPOPROTEIN CLEARANCE
HDLBP
LIPA
CUBN
LSR
APOB
APOA1
SCARB1
NPC2
SOAT1
CLTA
AP2S1
LDLR
NR1H3
APOE
MYLIP
APOC1
AP2B1
NR1H2
APOC4
CLTC
AP2A1
NPC1
APOBR
NCEH1
AP2A2
VLDLR
CES3
LDLRAP1
PCSK9
AP2M1
LIPC
AMN
SOAT2

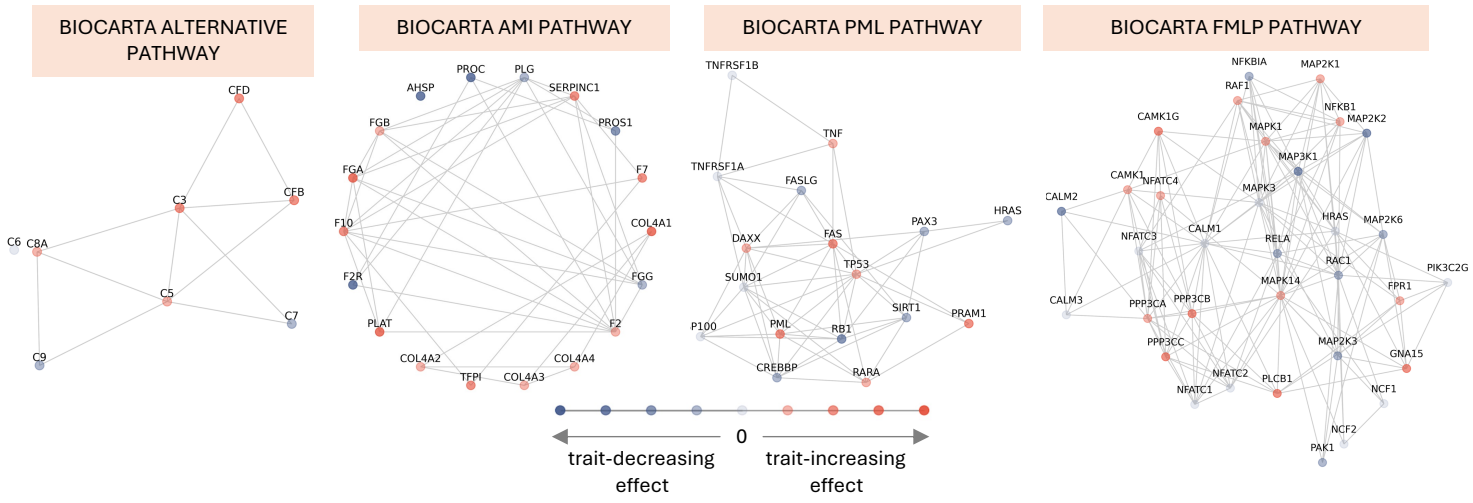

E.
BIOCARTA ALTERNATIVE PATHWAY
CFD
C3
CFB
C6
C8A
C5
C7
C9
BIOCARTA AMI PATHWAY
PROC
PLG
AHSP
SERPINC1
FGB
PROS1
FGA
F7
F10
COL4A1
F2R
FGG
PLAT
F2
COL4A2
COL4A4
TFPI
COL4A3
BIOCARTA PML PATHWAY
TNFRSF1B
TNF
TNFRSF1A
FASLG
HRAS
PAX3
DAXX
FAS
TP53
SUMO1
SIRT1
PRAM1
P100
PML
RB1
CREBBP
RARA
BIOCARTA FMLP PATHWAY
NFKBIA
MAP2K1
RAF1
NFKB1
CAMK1G
MAP2K2
MAPK1
MAP3K1
CAMK1
NFATC4
MAPK3
CALM2
HRAS
MAP2K6
NFATC3
CALM1
RELA
RAC1
PIK3C2G
MAPK14
FPR1
CALM3
PPP3CA
PPP3CB
PPP3CC
MAP2K3
GNA15
NFATC2
PLCB1
NFATC1
NCF1
NCF2
PAK1
0
trait-decreasing effect
trait-increasing effect

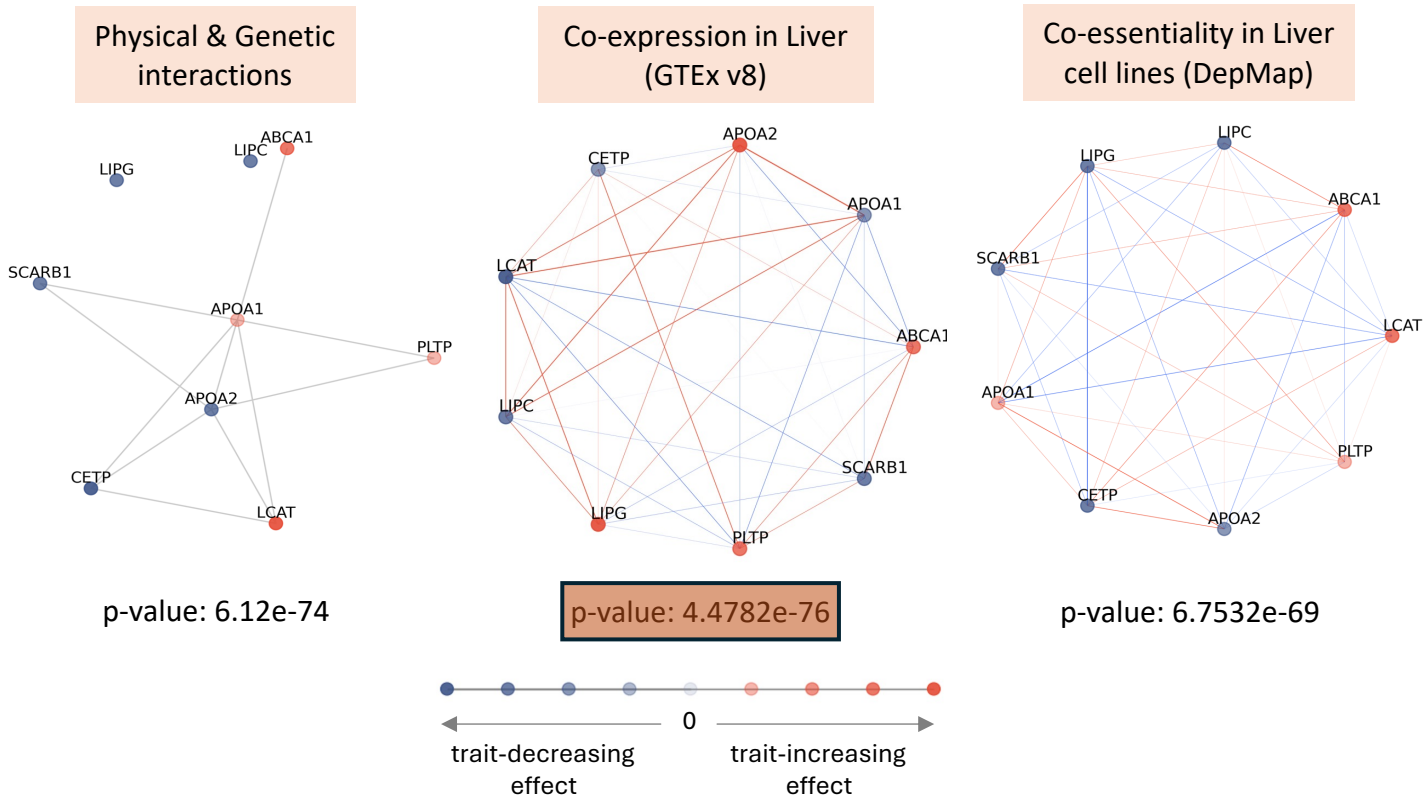
Figure 4. NERINE identifies significant rare variant burden in pathway gene modules in breast cancer (BRCA), type II diabetes (T2D), coronary artery disease (CAD), and early onset myocardial infarction (MI) in the UK and MGB biobanks.
A. Bonferroni-significant findings for different disease phenotypes across a database of 306 pathways in three functional categories of rare variants—(i) LoF (i.e., frameshifts, insertions, deletions, and splice), (ii) damaging missense (i.e., missense variants predicted to be damaging by in-silico tools), and (iii) damaging (i.e., LoF and damaging missense variants) are shown. For each variant category, pathways were tested individually in UKBB and MGBBB cohorts and p-values were meta-analyzed with Fisher's combined test. Figure shows negative log-transformed Fisher's combined p-values for significant pathways. The dashed gray line represents the Bonferroni threshold of 0.05. **B.** Representative pathway gene modules with significant rare LoF variant-burden in BRCA are shown with individual gene effects. Since there were significant overlaps between BIOCARTA ATRBRCA PATHWAY (cancer susceptibility pathway) with the other significant pathways except BIOCARTA CARM ER PATHWAY (estrogen receptor pathway), we showed two pathways with minimal overlap (i.e., one gene, *BRCA1*) as representatives. NERINE predicted a trait-increasing role for known tumor suppressor genes (TSG), oncogenes, or fusion oncogenes such as, *BRCA1*, *BRCA2*, *ATM*, *ATR*, *CHEK2*, *FANCD2*, *FANCG*, *RAD50*, and *TP53* in the cancer susceptibility pathway. It also indicated a trait-increasing role for several genes in the estrogen receptor pathway which were neither cancer drivers nor GWAS hits for BRCA, such as, *CARM1*, *TBP*, *GTF2A1*, *HEC14*, *MEF2C*, and *PELP1*. **C.** Individual gene effects predicted by NERINE in the adipogenesis pathway (i.e., BIOCARTA VOBESITY PATHWAY) are shown for the T2D phenotype. Member genes include *RXRA*, *PPARG*, *ADIPOQ*, *TNF*, *NR3C1*, *LPL*, *RETN*, and *HSD11B1*—none of which were previously implicated in T2D in rare variant association studies. **D.** NERINE's estimates of gene effects in the two database-wide significant pathways—plasma lipoprotein clearance and alternative complement cascade—in CAD are shown. NERINE predicts rare damaging variants in *APOA1, CUBN, SCARB1, CES3, SOAT2, AMN, LIPC, LDLRAP1, NCEH1, HMGCS1, SREBF1*, and *SREBF2,* and rare LoF variants in complement genes, including *C3, C6, C7, C8A, C9, C1QC, CFB*, and *CFD*, to have trait-increasing effects on CAD. **E.** NERINE identified 13 gene modules with significant rare LoF variant burden in MI, broadly grouped into four classes – inflammatory response, extracellular matrix proteins and coagulation, regulation of transcriptional activity, and the MAPK signaling cascade. One representative pathway per group is shown. NERINE newly implicated rare LoF variants in several genes involved in blood clotting, myocardial remodeling, apoptosis, inflammatory response, vascular calcification, calcium regulation, and retinoic acid signaling. Trait-increasing effects are represented by different shades of orange and trait-decreasing effects are represented by different shades of blue as shown on the scale. Darker color represents more pronounced effect.

Figure 5.

A.

Phenotype: Low HDL vs High HDL

Variant class: Damaging, MAF cutoff: 0.001



B.

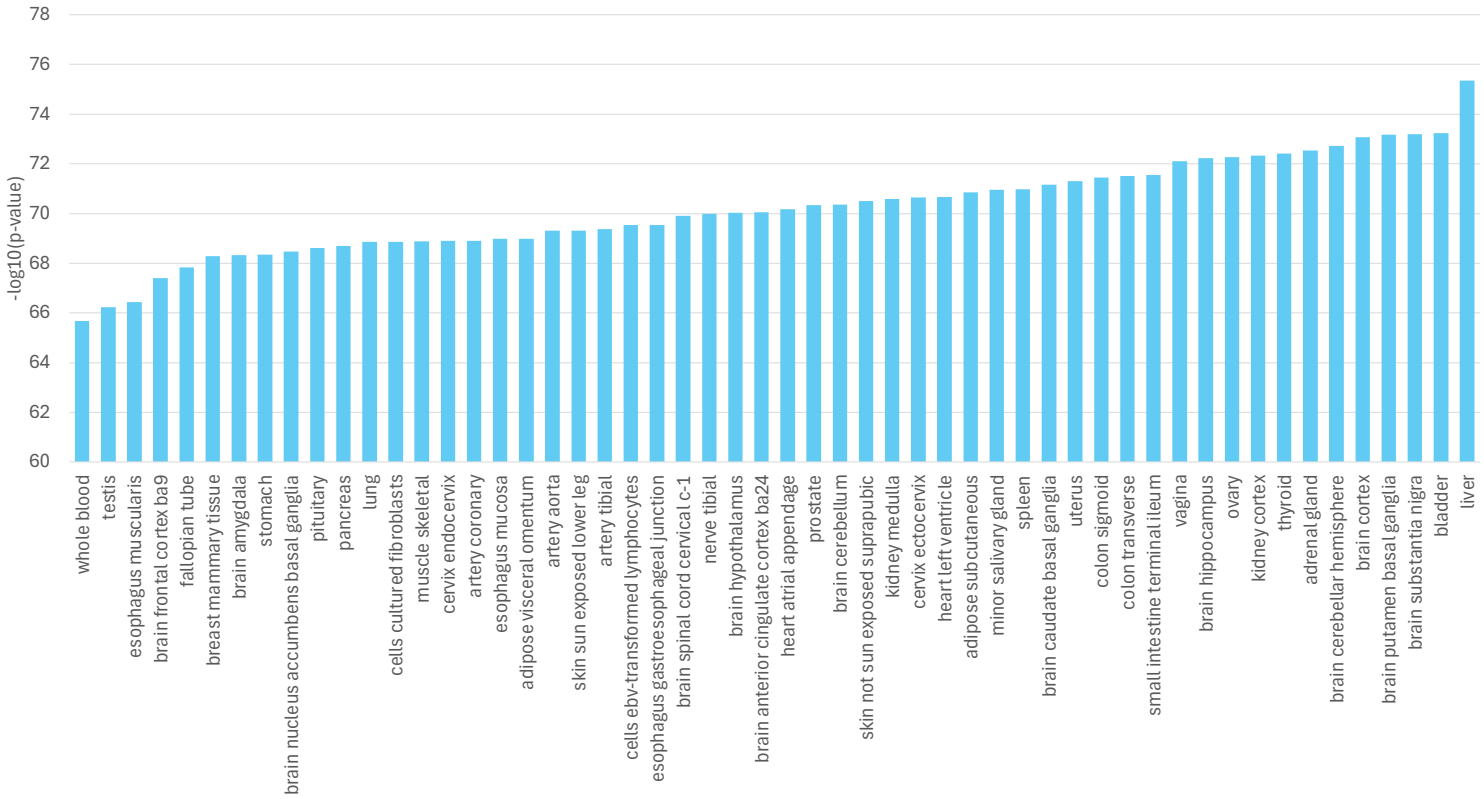


Figure 5. NERINE selects the most informative data source and context describing gene-gene relationships. A. For core HDL-cholesterol-related genes, NERINE competitively selects co-expression in the liver tissue as the most informative source among three data sources describing the edge relationship of genes—(i) high-confidence physical or genetic interactions from protein-protein interaction (PPI) databases, (ii) co-expression in a specific tissue from GTEx v8, and (iii) co-essentiality in liver cell lines from DepMap. For each data source, NERINE assessed the enrichment of rare damaging variants in core HDL-related genes in individuals with low HDL (bottom 10%) as compared to individuals with high HDL (top 10%) in the UK biobank (UKBB). We achieved the most-significant p-value with the co-expression network suggesting that co-expression might provide the most accurate representation of the relationship of these genes. Trait-increasing effects are represented by different shades of orange and trait-decreasing effects are represented by different shades of purple as shown on the scale. Darker color represents more pronounced effect in each direction. For co-expression and co-essentiality networks in **A**, positive correlation between two genes is indicated by red edges and 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-related gene module among all tissue types in GTEx. For all 52 tissue types in GTEx, we extracted the co-expression network relationships for the core module of HDL-related genes. We tested each network topology for the enrichment of rare damaging variants in low HDL vs high HDL individuals in UKBB with NERINE. Our method achieved the most significant p-value with the co-expression network in liver tissue.

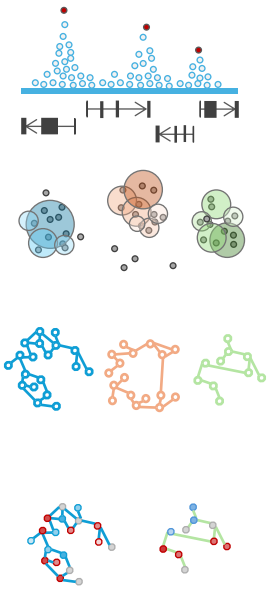
Figure 6.

A.

Parkinson's disease

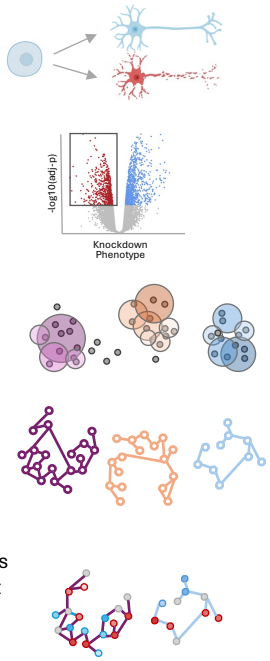
(i) GWAS genes

GWAS loci
↓
Nearby genes
↓
GSEA GO BP enrichment & semantic clustering
↓
Gene modules with database edges
↓
Identify modules with rare variant burden in case-control cohorts with NERINE



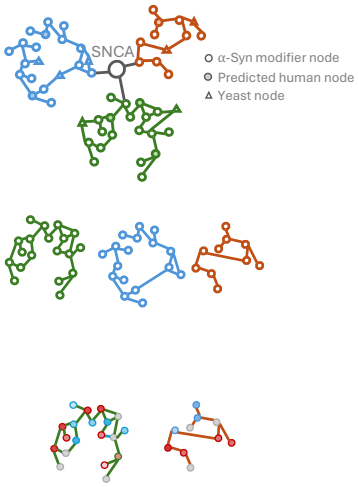
(ii) DA neuron essentiality genes

Genome-wide essentiality screen in DA neurons (19,993 genes)
↓
Essentiality genes (693)
↓
GSEA GO BP enrichment & semantic clustering
↓
Gene modules with database edges
↓
Identify modules with rare variant burden in case-control cohorts with NERINE



(iii) α-Syn proteinopathy network

Genome-wide screen in yeast-to-neuron platform
↓
TransposeNet: humanized αS network
↓
Gene modules: TransposeNet network stems
↓
Identify modules with rare variant burden in case-control cohorts with NERINE



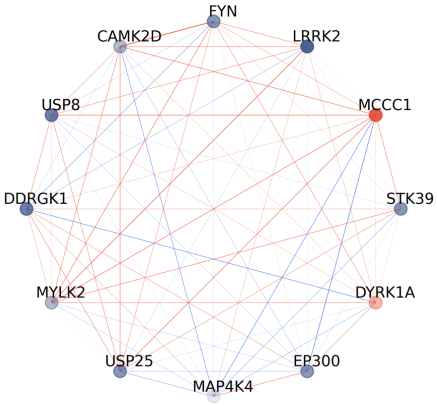
B.

GWAS loci-enriched modules

Peptidyl-threonine modification

STK39, MCCC1, LRRK2, FYN, CAMK2D, USP8, DDRGK1, MYLK2, USP25, MAP4K4, EP300, DYRK1A

Category	UKBB sporadic	AMP-PD	Fisher combo-p
LoF	4.93E-02	1.83E-02	**7.23E-03**
Neutral	0.5	0.5	5.97E-01
Synonymous	0.5	0.5	5.97E-01



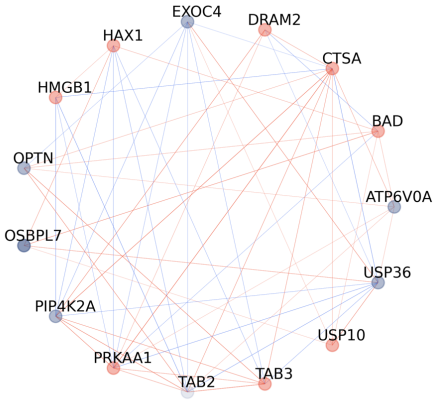
C.

DA neuron essentiality modules

Regulation of autophagy

ATP6V0A2, BAD, CTSA, DRAM2, EXOC4, HAX1, HMGB1, OPTN, OSBPL7, PIP4K2A, PRKAA1, TAB2, TAB3, USP10, USP36

Category	UKBB extreme	AMP-PD	Fisher combo-p
Damaging	1.75E-03	3.58E-02	**6.69E-04**
Neutral	0.5	0.5	5.97E-01
Synonymous	0.5	0.5	5.97E-01



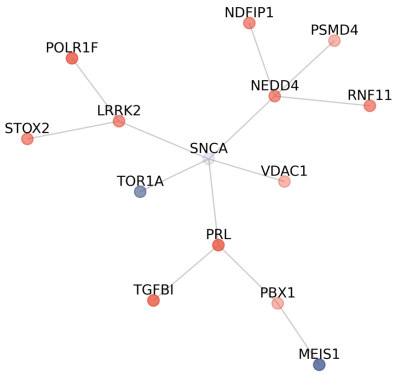
D.

α-Syn proteotoxicity modules

LRRK2-SNCA vesicle trafficking stem

PSMD4, NEDD4, LRRK2, SNCA, MEIS1, RNF11, POLR1F, STOX2, NDFIP1, PRL, TOR1A, PBX1, TGFBI, VDAC1

Category	UKBB sporadic	AMP-PD	Fisher combo-p
Dam. Missense	3.01E-03	4.25E-02	**1.28E-03**
Neutral	0.5	0.5	5.97E-01
Synonymous	0.5	0.5	5.97E-01



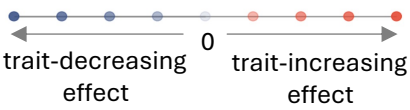


Figure 6. NERINE identifies networks enriched in rare qualifying variants in Parkinson's disease (PD) patients. A. Schematic diagram indicates a three-pronged approach for interrogating PD pathophysiology using NERINE. Network hypotheses were generated based on (i) known GWAS genes, (ii) essentiality genes for dopaminergic (DA) neurons, and (iii) α-Synuclein (αS) toxicity modifier genes. Among GWAS genes, we identified six GO biological process (BP) modules; from essentiality genes, we identified 10 GO BP modules, and from the TransposeNet humanized α-Synuclein modifier gene network, we tested 17 subnetworks. For GO biological process modules identified from GWAS and DA essentiality genes, 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, and iii) co-essentiality of genes in cell lines from central nervous system (CNS) in the DepMap database (see Methods). For αS modifier genes, we directly used the TransposeNet network topology[57]. **B.** GWAS genes enriched in the GO biological process module related to *Peptidyl-threonine modification* show significant rare LoF variant (i.e., frameshifts, insertions, deletions, and splice variants) burden in sporadic PD cases compared to controls in two independent datasets: AMP-PD and UK Biobank (UKBB). Co-essentiality in CNS cell-types provided the most informative edge relationships for this module. Notably, rare LoF variants in *LRRK2* shows an overall trait-decreasing effect which aligns known biology (whereby LRRK2 gain-of-function mutations are considered detrimental), serving as a positive control. **C.** In sporadic PD cases compared to extreme controls (controls with age >= 85) in both the AMP-PD and UKBB, there is screen-wide significant burden of rare damaging variants (i.e., predicted damaging missense variants, frameshifts, insertions, deletions, and splice variants) in the GO biological process module related to the *regulation of autophagy*. Co-essentiality in CNS cell-types provided the most informative edge relationships for this module. **D.** A *LRRK2-SNCA*-containing *protein trafficking and 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 **B**, **C**, and **D**, screen-wide significance was determined by applying Bonferroni correction over the Fisher combined p-values. 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 different shades of orange and trait-decreasing effects are represented by different shades of purple as shown on the scale. Darker color represents more pronounced effect in each direction. For co-essentiality networks in **B** and **C**, positive correlation between two genes is indicated by red edges and negative correlation is indicated by blue edges. The thickness of the edge corresponds to the strength of the correlation. The edges in the TransposeNet module in **D** do not represent correlations but binary relationships and are therefore colored in gray.

Figure 7

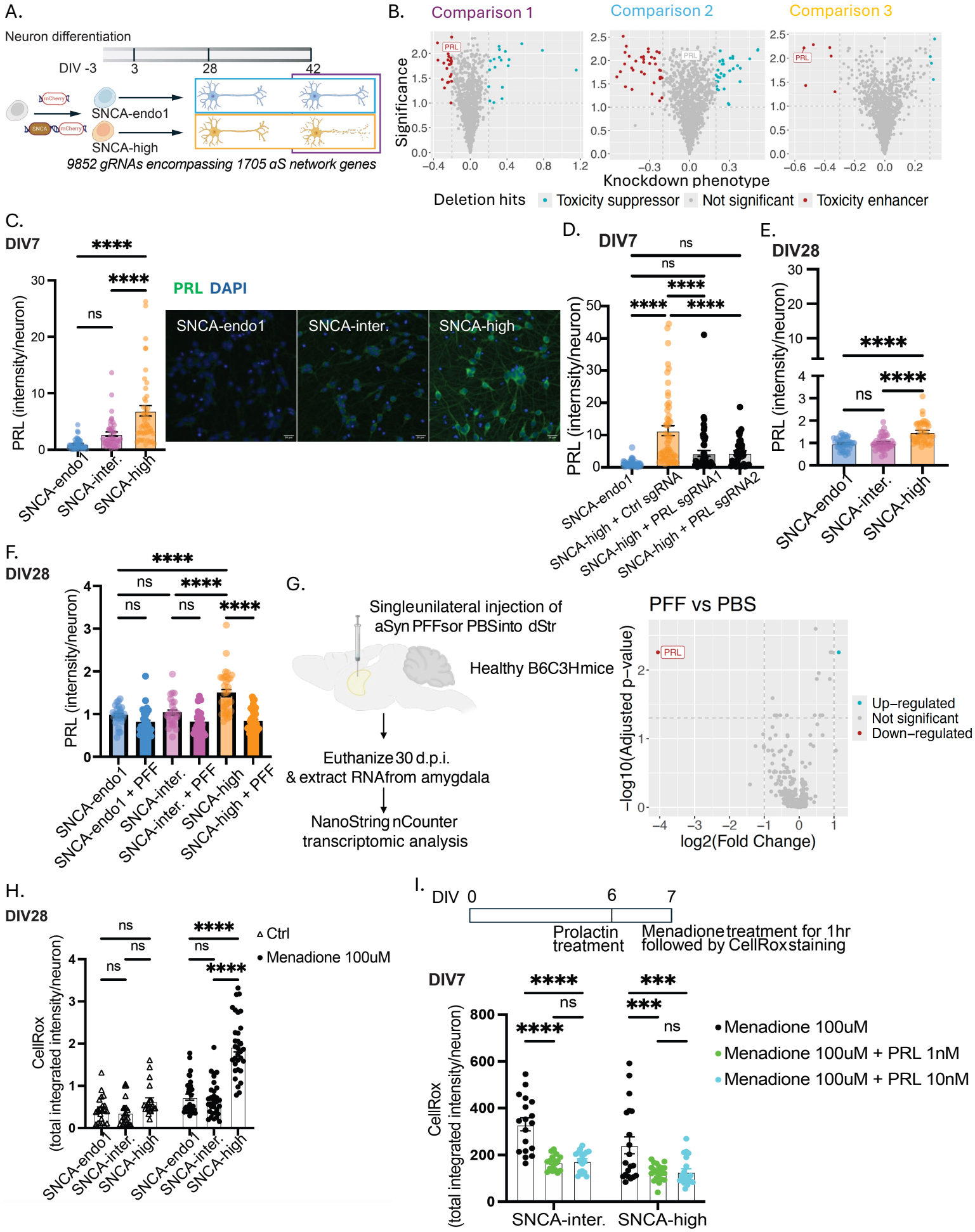

A.
Neuron differentiation
DIV -3
3
28
42
SNCA-endo1
SNCA-high
9852 gRNAs encompassing 1705 αS network genes
B.
Comparison 1
Comparison 2
Comparison 3
Significance
Knockdown phenotype
Deletion hits
Toxicity suppressor
Not significant
Toxicity enhancer
C.
DIV7
PRL (internsity/neuron)
SNCA-endo1
SNCA-inter.
SNCA-high
PRL DAPI
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-endo neurons. Comparison 2 compares the sgRNA frequencies of SNCA-endo 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-endo neurons compared to DIV28 SNCA-endo 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.** 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*. **G.** 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). **H.** 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). **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).