**Article**

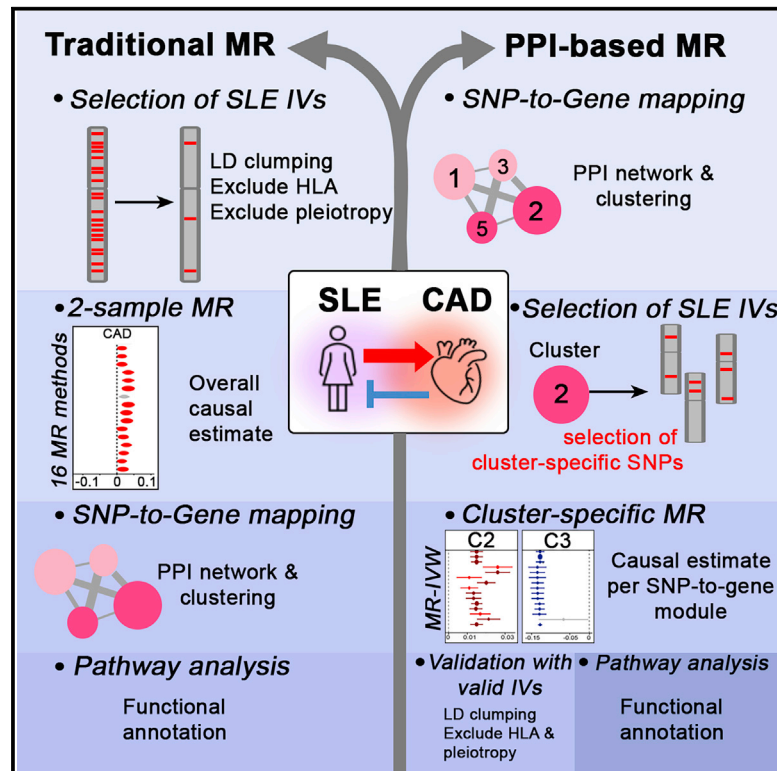# Mendelian randomization and pathway analysis demonstrate shared genetic associations between lupus and coronary artery disease

## Graphical abstract

## Authors

Jessica Kain, Katherine A. Owen, Miranda C. Marion, Carl D. Langefeld, Amrie C. Grammer, Peter E. Lipsky

## Correspondence

kate.owen@ampelbiosolutions.com

## In brief

Kain et al. report a positive genetic relationship between SLE and CAD using traditional and comprehensive Mendelian randomization (MR) approaches. MR analysis coupled with network modeling highlights the shared genetic risk underlying the two diseases and identifies putative molecular pathways involved in the development of CAD in SLE.

## Highlights

- Genetic predisposition to SLE confers risk of CAD

- Increased CVD risk in SLE involves atherosclerosis rather than cardiac dysfunction

- PPI-based MR identifies pathways with causal implications between SLE and CAD

- Pathway analysis informs therapeutic selection for managing CAD in SLE

CellPress

## Article

# Mendelian randomization and pathway analysis demonstrate shared genetic associations between lupus and coronary artery disease

Jessica Kain,[1,2] Katherine A. Owen,[1,2,4,*] Miranda C. Marion,[3] Carl D. Langefeld,[3] Amrie C. Grammer,[1,2] and Peter E. Lipsky[1,2]

[1]AMPEL BioSolutions, LLC, Charlottesville, VA, USA
[2]The RILITE Research Institute, Charlottesville, VA, USA
[3]Department of Biostatistics and Data Science, and Center for Precision Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
[4]Lead contact
*Correspondence: kate.owen@ampelbiosolutions.com
https://doi.org/10.1016/j.xcrm.2022.100805

## SUMMARY

Coronary artery disease (CAD) is a leading cause of death in patients with systemic lupus erythematosus (SLE). Despite clinical evidence supporting an association between SLE and CAD, pleiotropy-adjusted genetic association studies are limited and focus on only a few common risk loci. Here, we identify a net positive causal estimate of SLE-associated non-HLA SNPs on CAD by traditional Mendelian randomization (MR) approaches. Pathway analysis using SNP-to-gene mapping followed by unsupervised clustering based on protein-protein interactions (PPIs) identifies biological networks composed of positive and negative causal sets of genes. In addition, we confirm the casual effects of specific SNP-to-gene modules on CAD using only SNP mapping to each PPI-defined functional gene set as instrumental variables. This PPI-based MR approach elucidates various molecular pathways with causal implications between SLE and CAD and identifies biological pathways likely causative of both pathologies, revealing known and novel therapeutic interventions for managing CAD in SLE.

## INTRODUCTION

Systemic lupus erythematosus (SLE) is a female-predominant autoimmune disease characterized by immune dysregulation and multi-organ inflammation that is frequently associated with the development of cardiovascular disease (CVD).[1,2] SLE exhibits hyperactivity of the innate and adaptive immune systems, increased production of numerous autoantibodies, and disturbed cytokine balance.[3] Although CVD is not a diagnostic criterion of SLE and was not included in the original descriptions of the disease, it is currently the main cause of death in SLE,[4–6] with coronary artery disease (CAD) directly responsible for one-third to one-half of all CVD cases and 30% of deaths.[7–9] Notably, whereas mortality from infections and active disease has decreased in SLE patients, CVD-related death rates have not improved,[10] and the standardized mortality ratio related to CVD has actually increased.[11] Women with SLE have a significantly increased risk of stroke and myocardial infarction along with elevated incidence of asymptomatic atherosclerosis compared with the general population.[12,13] Furthermore, traditional CVD risk factors, such as cholesterol, blood pressure, and smoking status, fail to fully account for the overall higher risk of acute CVD events in SLE, although the underlying mechanisms remain unknown.[14–17] This lack of an understanding for the increased risk of CVD in SLE has resulted in limited treatment options and the puzzling juxtaposition that despite the efficacy of statins and angiotensin-converting enzyme inhibitors/angiotensin receptor blockers in treating the general population, they appear to have little effect on CVD outcomes in SLE patients.[5,18] As a result, even though SLE has a prevalence of only about 70 per 100,000, it ranks among the leading causes of death in young women,[1] despite the omission of lupus diagnoses in almost half of SLE patients' death certificates.[19,20]

Genetic predisposition imposes important risk factors for both SLE and CVD.[21–23] To date, genetic association studies of SLE patients with and without CVD have been limited in size and have detected only a few common genetic risk loci, including *IRF8*, *STAT4*, *IL19*, and *SRP54-AS1*.[22,24,25] Mendelian randomization (MR) is a causal inference method using genotypes as "treatments" when randomized controlled trials are not feasible. By measuring and correlating the effect sizes of exposure-associated genetic variants in large-scale genetic association studies on traits of interest, a causal effect of the exposure on the outcome can be estimated. Here, we report the application of multiple, complementary MR methods to identify causal paths from SLE-associated variants to CAD using summary statistics from genetic association studies. Using multiple MR algorithms, we have identified large sets of SLE causal variants that also impart genetic risk

for CAD as well as those that appear to diminish the risk of CAD. Using innovative approaches to build molecular pathways from genetic risk factors,[26] we have developed a map of SLE-derived biological processes with causal implications for CAD that may account for the genetic basis of the association between these two apparently dissimilar clinical entities and may also provide insights into the shared mechanisms underlying each. Understanding the pathogenesis of genetic variants underlying the increased CAD risk in SLE can ultimately provide insight into the immune and inflammatory components of atherosclerosis, as well as revealing opportunities for targeted therapeutics.

## RESULTS

### Pathway analysis reveals gene networks implicated by genetic variants associated with both SLE and CAD

To explore the shared genetic predispositions for SLE and CAD, we first identified single nucleotide polymorphisms (SNPs) associated with each trait in five SLE and one CAD multi-ancestral genetic association study.[27–32] In total, 96 SNPs were associated with both conditions (Figure S1A). Notably, the majority of the overlapping SNPs mapped to the human leukocyte antigen (HLA) region of chromosome 6. To identify putative gene(s) influenced by each of the 96 SNPs associated with both SLE and CAD, we mapped causal SNPs to genes,[26] identifying 189 unique genes encoding 135 proteins in STRINGdb (Figure S1B). Stratified linkage disequilibrium score regression (S-LDSC) was then used to validate the biological relevance of SNP-predicted gene and protein sets by assessing whether they captured more disease heritability than expected by chance with respect to all genes and STRINGdb proteins, respectively.[33] Application of S-LDSC using genome-wide association study (GWAS) summary statistics for SLE (GCST003155[27]), CVD (GCST004280[34]), and two CAD datasets (CAD-I, GCST000998[35] and CAD-II, GCST001479[36]) determined that the 189 genes predicted by the 96 overlapping SNPs were significantly ($p < 0.05$) enriched for genomic regions capturing the genetic heritability of CAD and CVD (Figure S1C). Nearly identical results were also obtained using the smaller subset of 135 protein-coding genes (Figure S1C). However, application of standard LDSC, which was restricted to the use of the relatively small, European-only SLE GWAS,[27] did not reveal a significant level of genetic correlation between the diseases (Figure S1D).

To assess molecular networks encoded by the set of 135 protein-encoding genes predicted from the overlapping SLE/CAD SNPs, a protein-protein interaction (PPI) network was generated, and unsupervised clustering revealed 12 distinct gene clusters that were functionally enriched in a diverse range of immunological and cellular categories (Figure S1E), many with relevance to SLE and CAD/CVD, including cluster 1 characterized by canonical pathways for "Antigen presentation pathway" and "B cell development," along with "Sudden cardiac death," and cluster 3 enriched in "Atherosclerosis signaling" and "Lupus erythematosus, systemic," among others (Table S1). Although the molecular pathways associated with SNP-predicted genes suggested a convergence of biological processes underlying SLE and CAD, it remained uncertain whether the finding of overlapping SNPs implied shared genetic causation. The subsequent studies presented here explore this in detail.

### MR estimates a positive correlation between effects of SLE-associated non-HLA variants on SLE and CAD

MR methods were employed to estimate the association between effect sizes of relevant variants on SLE and CAD. We first applied six MR methods using various sets of SLE-associated instrumental variables (IVs) to determine whether they tend to confer shared (positive) effects on SLE and CAD, noting that this initial approach did not satisfy all assumptions for IV validity or IV independence and therefore could only provide an estimated association (Figure 1A). Initial exploratory analyses employing IVs derived from the Immunochip and GWASs suggested a net positive association for non-HLA SLE-associated SNPs on CAD (Figures 1B and 1C). Even when using SLE IVs determined with the more stringent significance level and removal of known pleiotropic associations to CVD or confounders such as cholesterol, obesity, blood pressure, insulin resistance, and smoking, the indication of a positive causal relationship between SLE and CAD remained (Figure 1C, bottom row).

To validate the robustness of our estimated associations by satisfying stringent requirement for IV selection, we carried out two-sample MR analyses using multi-ancestral, non-HLA SNPs strongly associated ($p < 5 \times 10^{-8}$) with SLE, excluding SNPs weakly associated ($p < 10^{-5}$) with CVD or confounders (Table S2), followed by stringent linkage disequilibrium (LD) clumping to ensure IV independence[37] ($R^2 = 0.001$, 100 kb window, 1000 Genomes EA reference population) (Figure 2A). SLE GWAS summary statistics were used for exposure,[27] and multiple CAD GWASs were used for outcome (GCST005194,[32] CAD-a and GCST005195,[32] CAD-b). Since CAD is causative of myocardial infarction (MI) and atherosclerosis is common to CAD and ischemic stroke (IS), MR was carried out using summary statistics for two additional MI GWASs (GCST003117,[38] MI-a and GCST011365,[39] MI-b) and IS (GCST006906[40]). Summary statistics for cardiomyopathy from the FinnGen biobank analysis (finn-b-I9_CARDMYO, CM) and atrial fibrillation (GCST006414,[41] AFib), which are not associated with atherosclerosis or CAD, were also included for comparison. After LD clumping, 60 independent SNPs were included in the SLE GWAS and then harmonized with each outcome-GWAS pair before use as IVs for SLE exposure (Table S9). Between 43 and 56 harmonized IVs for SLE exposure were then tested using 16 MR methods, some of which account for additional IV invalidity, pleiotropy, or heterogeneity, to estimate causal relationships with the various atherosclerotic and cardiac conditions. The majority of MR methods resulted in significant ($p < 0.05$) positive causal estimates of SLE-associated variants on CAD-a and both MI GWASs, but not for cardiomyopathy or AFib (Figures 2B and S2A; Table S3). Directional pleiotropy was only detected between the SLE and CAD-b GWASs by the MR-Egger intercept test (Figure S2A), indicating potential bias in the causal estimates based on effect sizes using these summary statistics. Overall, inverse variance weighting (IVW), weighted median, penalized weighted median, maximum likelihood, robust adjusted profile score (RAPS), and pleiotropy residual sum and outlier (PRESSO) were significant in four out of five outcome GWASs (Figure 2B). These results establish a positive causal effect of SLE on CAD and suggest that the increased CVD risk associated
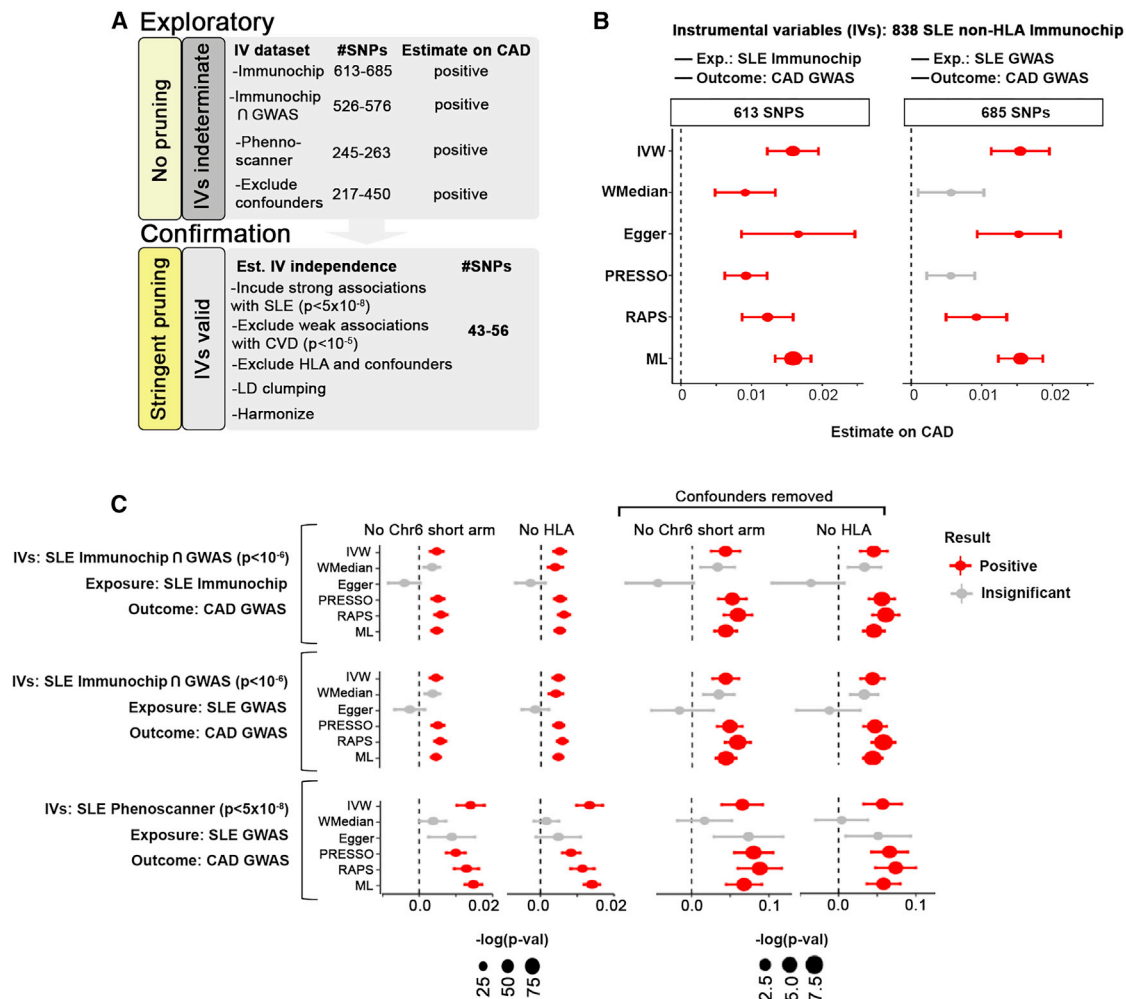
**Figure 1. MR demonstrates a positive association of effect sizes of SLE-associated non-HLA SNPs on SLE and CAD**

(A) Graphical depiction of the two-stage approach for an initial exploratory analysis using expanded groups of SNPs as IVs followed by a confirmatory analysis using highly curated IVs.

(B and C) Forest plots of six MR causal estimates ($\beta \pm$ standard error). For results, gray indicates insignificant (p > 0.05) and red indicates positive causal estimates determined by each MR method. (B) Immunochip-derived SLE-associated non-HLA SNPs were used as IVs for SLE; summary statistics from both the SLE Immunochip study (left) and SLE GWAS (right) were used for the exposure; summary statistics from the CAD GWAS were used for the outcome. (C) Additional MR analyses using SNPs associated (p < $10^{-6}$) with SLE in the Immunochip and GWAS study (rows 1 and 2) or Phenoscanner reaching genome-wide significance (p < 5 × $10^{-8}$, row 3) were used as IVs; summary statistics for the exposure and outcome are indicated. MR analyses excluding the entire short arm of chromosome 6 and excluding only the extended HLA region (chr6: 27–34 Mb, left columns). Right columns show MR analyses using the same sets of SNPs excluding pleiotropic SNPs associated (p < $10^{-5}$) with either CAD directly or CVD-related confounders, included on the Phenoscanner platform.

with SLE is likely to involve atherosclerosis rather than other aspects of cardiac pathology.

To eliminate the possibility that the positive causal estimate of SLE on CAD is bidirectional and therefore unlikely to represent a true causal relationship, MR was also carried out in the reverse direction, with CAD or MI as exposure and SLE as the outcome. Importantly, none of the 14 methods yielded a significant positive causal estimate of CAD or MI on SLE. Of interest, however, significant (p < 0.05) negative causal estimates of CAD and MI on SLE were observed for approximately half of the 14 MR methods tested (Figure S2B and Table S3).

To understand the pathways underlying the positive causal estimates of SLE on CAD in greater detail, all SLE-associated SNPs included as putative IVs before harmonization with each GWAS were mapped to genes. Consistent with satisfying the exclusion restriction criteria and independence assumption with respect to traits imposing significant CVD risk, S-LDSC results demonstrated that the 284 genes and 160 predicted proteins captured a significant portion of SLE heritability (p = 3.46 × $10^{-5}$ and 3 × $10^{-5}$, respectively), but not that of CVD or CAD (Figure 2C).

Proteins predicted from the SLE IVs were then integrated into connectivity networks in STRINGdb (Figure 2D). Cluster annotations were dominated by processes commonly dysregulated in SLE as expected, including canonical pathways for "Systemic lupus erythematosus in B cell signaling," "Th1 pathway," and
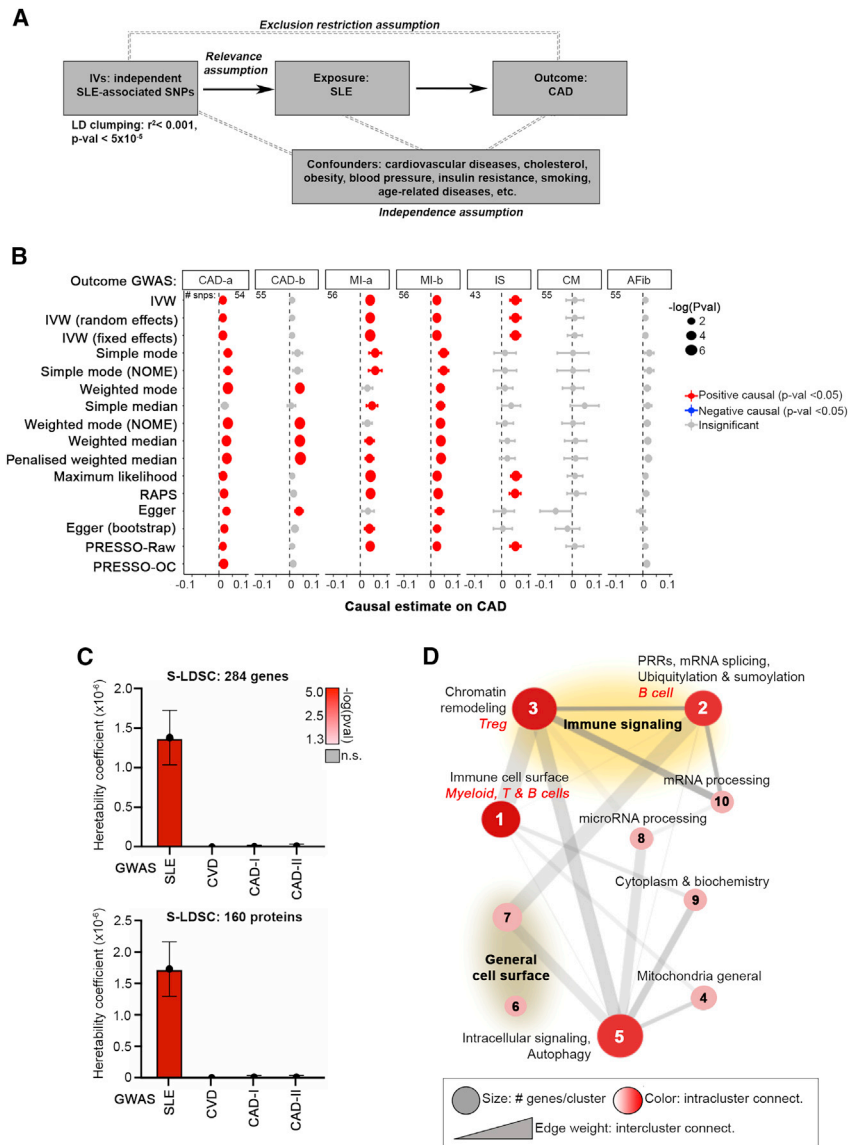
**Figure 2. MR demonstrates a net positive causal effect of SLE-associated non-HLA SNPs on CAD**

(A) MR diagram for testing the causal effects of SLE on CAD with respect to instrument relevance to the exposure, exclusion from the outcomes (i.e., CAD, MI, IS), and independence from confounding factors. LD clumping ($R^2 < 0.001$) was used to obtain independent IVs.

(B) Forest plots of MR causal estimates ($\beta \pm$ standard error) for SLE on CAD (CAD-a, CAD-b), MI (MI-a, MI-b), IS, cardiomyopathy (CM), and atrial fibrillation (AFib) GWAS using 16 MR methods. Missing PRESSO-OC (outlier correction) estimates indicate insignificant global tests for horizontal pleiotropy. For results, gray indicates insignificant ($p > 0.05$), red indicates positive ($p < 0.05$), and blue indicates negative ($p < 0.05$) causal estimates determined by each MR method. Numbers within forest plots indicate the SNPs used as IVs after harmonization. NOME, no measurement error.

(C) Application of S-LDSC using summary statistics for SLE, CVD, and CAD GWAS to estimate the heritability (coefficient $\pm$ standard error) of the 284 SNP-predicted genes (top) and 160 SNP-predicted proteins from STRINGdb (bottom). Bar color indicates coefficient significance.

(D) Cluster metastructures for the 160 putative protein-coding genes are based on PPI networks. Functional and cell-type enrichments for each cluster were determined using BIG-C (black labels) and I-scope (red labels), respectively. Black labels over colored shadings represent shared functional annotations for the clusters they surround.

the HLA region, reveal that the majority of negative causal SNPs are located on the short arm of chromosome 6; all but one were tightly packed around the HLA region, spanning chr6:28,014,374–33,683,352 (Figure S3). When excluding the short arm of chromosome 6 and SNPs associated with CVD or confounders, SSMR identified 80 and 96 SLE-associated variants with significant ($p < 0.05$) positive (Figure 3A, top 25) and negative (Figure 3B, top 25) causal estimates on CAD, respectively (Figure S4 and Table S3). The majority of positive causal SNPs were distributed on chromosomes 1, 2, and 4, whereas over 50% of negative causal SNPs were on chromosomes 7, 11, and 17 (Figures 3C and 3E).

Non-HLA SLE variants with either significant positive or negative causal estimates on CAD were separately mapped to 236 (Figure 3D) and 244 (Figure 3F) predicted genes, respectively, for clustering and pathway analysis. Positive SNP-predicted gene clusters were enriched in the canonical pathway for "Antigen presentation" and functional categories for major histocompatibility complex (MHC) class I, as well as epigenetic processes, transcription, and endocytosis (Figure 3E and Table S5), whereas negative SNP-predicted gene clusters were dominated by processes related to cell death ("Pyroptosis" [GO:0070269], cluster 2; "Regulation of oxidative stress-induced cell death" [GO:1903201], cluster 6) and

"Th2 pathway," as well as gene ontology (GO) terms for "Regulation of immune response" (GO:0050776) and "Negative regulation of B cell activation" (GO:0050869) (Table S4). Interestingly, disease associations were enriched in various autoimmune diseases ("Lupus erythematosus, systemic," "Aicardi-Goutières syndrome," and "Hashimoto's disease"), along with cardiovascular dysfunction, such as "Arterial embolism and thrombosis," "Hypertension," "Plaque, atherosclerotic," and "Ischemic heart disease" (Table S4).

## Single-SNP MR identifies gene networks implicated by SLE-associated variants with positive and negative causal estimates on CAD

We next employed single-SNP MR (SSMR) to identify specific SLE-associated variants with positive or negative estimates on CAD. SSMR applied to SLE-associated SNPs, including those in
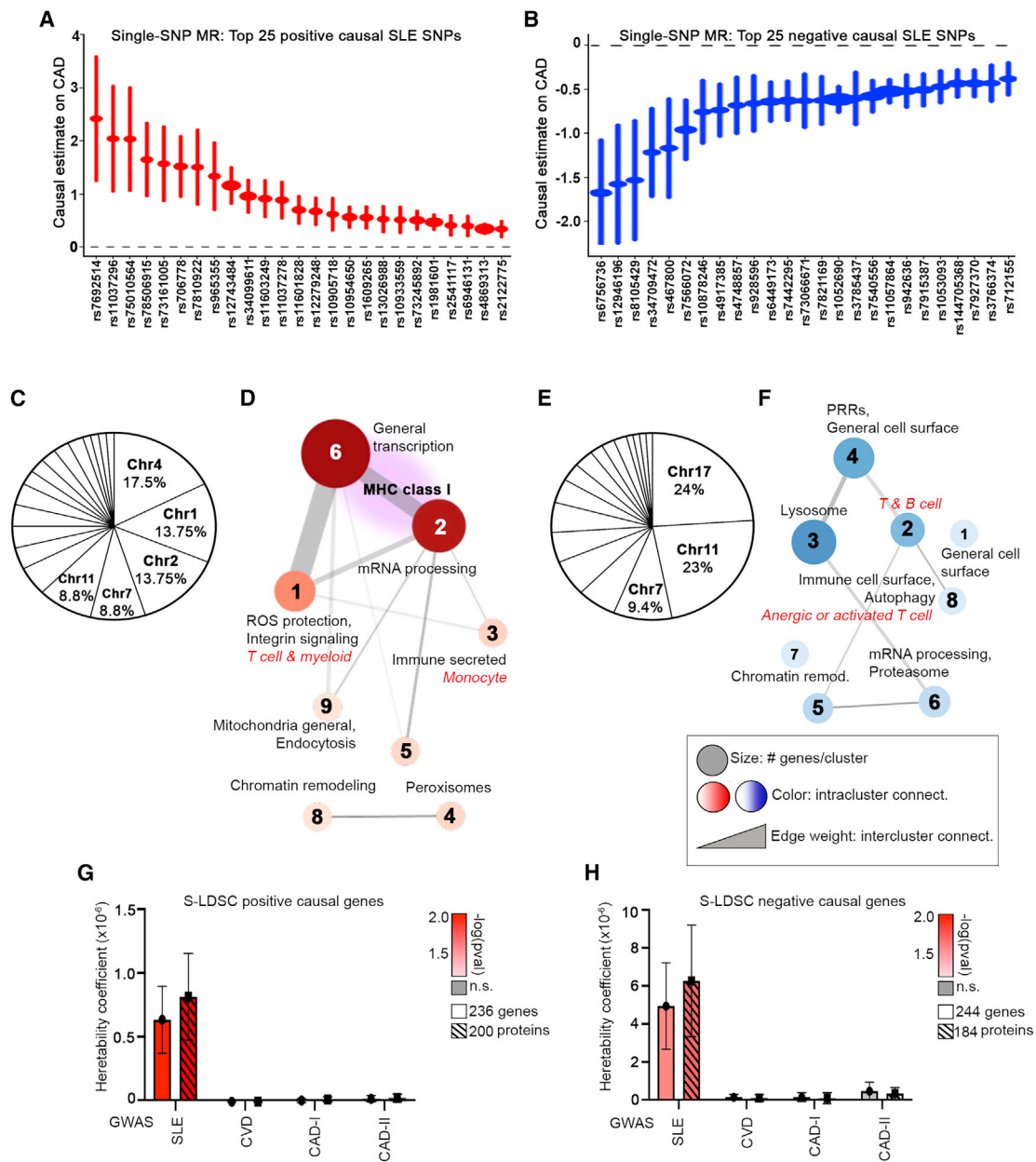
**Figure 3. Analysis of SLE-associated SNP-predicted genes with causal effects on CAD by single-SNP MR**

(A and B) Forest plots (β ± standard error) of the top 25 (by absolute value of causal estimates) positive (A) and negative (B) causal non-HLA SNPs identified by single-SNP MR (SSMR) using the Wald ratio method.

(C and E) Pie charts illustrating the chromosomal distribution of 80 positive (C) and 96 negative (E) causal SLE SNPs on CAD.

(D and F) Cluster metastructures for the 200 (D) and 184 (F) predicted genes from positive and negative causal SNPs identified by single-SNP MR. Functional and cell-type enrichments for each cluster were determined using BIG-C (black labels) and I-scope (red labels), respectively. Bold black labels over colored shadings represent shared functional annotations for the clusters they surround.

(G and H) S-LDSC using summary statistics for SLE, CVD, and CAD GWAS to estimate the heritability (coefficient ± standard error) of genes (open bars) and SNP-predicted proteins (hashed bars) predicted by positive (G) and negative (H) causal SNPs determined by SSMR. Bar color indicates coefficient significance.

protein degradation ("Proteasome," cluster 6 and "Autophagy," cluster 8; Figure 3F and Table S5). Finally, gene sets predicted by both positive and negative causal SNPs captured a significant portion of SLE heritability but not that of CAD or CVD, consistent with their selection as IVs for SLE (Figures 3G and 3H).

## Pathway analysis of HLA-region variants associated with SLE risk and protective of CAD

Risk haplotypes in the HLA region heavily contribute to susceptibility for SLE[42] and CAD.[43] However, accurate genotyping of HLA alleles and corresponding GWAS effect size estimates

are notoriously unreliable.[44] Additionally, the complex genetic architecture of this region makes mapping HLA variants to genes especially challenging given the extensive LD and high density of genes in this region. Nonetheless, an examination of the HLA area (chr6:28.5–33.5 Mb) revealed 30 SNPs significantly ($p < 10^{-6}$) associated with both SLE and CAD in their respective GWAS. While these SNPs are not independently associated variants, all 30 SNPs had positive effect sizes for SLE but were negative for CAD (Figure S5A), possibly reflecting the extensive LD in this region. Connectivity mapping and clustering of the 69 protein-encoding genes predicted from these 30 SNPs revealed six distinct clusters dominated by processes dysregulated in SLE, including the functional categories for MHC class I and MHC class II in clusters 1 and 6, along with canonical pathways for "Th1 and Th2 activation," "B cell development," and "Notch signaling", as well as GO term for "Interferon-gamma mediated signaling pathway" (GO:0060334) (Figures S5B and S5C). Other pathways of interest involving "Complement system abnormalities," "LXR/RXR activation," and "21-hydroxylase deficiency" were predicted by cluster 3 (Figure S5C).

## PPI-based MR predicts specific sets of SLE-associated variants and gene pathways causal of CAD

To obtain a more comprehensive view of the possible impact of SLE-derived molecular pathways on atherosclerosis, we mapped SLE-associated, non-HLA Immunochip SNPs with net positive causal estimates on CAD by MR to genes and pathways regardless of their associations with CVD-related traits. In total, 838 SNPs predicted 2,336 putative genes and 1,501 proteins that collectively captured a significant amount of SLE, but not CAD or CVD, heritability (Figure 4A); these 1,501 proteins clustered into 46 distinct clusters based on PPI connectivity (Figure 4B). We then grouped SLE-associated SNP mapping to genes in each of the 46 PPI-based clusters for use as SLE IVs to estimate cluster-specific associations with atherosclerotic traits. Initial application of MR-IVW to these 46 subsets of SLE SNP-derived IVs yielded 16 and 9 significant ($p < 0.05$) positive and negative causal estimates, respectively (Figures 4B, 4C, and S6A) for CAD. Additional MR methods, including mode- and median-based methods, MR-Egger, MR-RAPS, MR-PRESSO, and maximum likelihood, were carried out for further validation of the PPI-based MR-IVW causal estimates (Table S3). Clusters were grouped into tiers with respect to consistency across the various MR methods, with tier 1 clusters yielding significant positive or negative causal estimates by almost all (at least 14/16) MR methods and tier 2 clusters yielding significant positive or negative causal estimates by MR-IVW or at least seven MR methods (Figures 4B and 4C). Finally, when examined individually, 20 of the 46 clusters specifically captured SLE heritability by PPI-based S-LDSC, many with significant causal estimates on CAD by PPI-based MR (Figure 4D and Table S6).

In an effort to support these results by expanding the size of the network, we added 914 multi-ancestral, non-HLA SNPs associated with SLE on the Phenoscanner database to the analysis. Overall, 1,708 unique SNPs predicted 3,272 putative genes and 1,972 proteins that collectively captured a signifi-

cant amount of SLE heritability, but not that of CAD or CVD (Figure 5A) and clustered into 67 distinct sets of protein-coding genes (Figure 5B). PPI-based MR-IVW using these 67 clusters of SLE SNPs as IVs yielded 24 and 11 significant ($p < 0.05$) positive and negative causal estimates on CAD, respectively (Figures 5C, 5D, and S6B), many of which captured SLE heritability, but not that of CAD or CVD, by PPI-based S-LDSC (Figure 5E and Table S6).

To ensure that the majority of predicted causal clusters are not a result of random chance or multiple-hypothesis testing, simulations were carried out to estimate the false discovery rate. Results from these simulations, which account for both LD and pleiotropy, indicate that SLE-derived and PPI-clustered modules, as opposed to randomly generated SNP-to-gene modules, demonstrate a higher rate of significant causal estimates on CAD (Figure S7). Furthermore, to assess the reproducibility of the cluster-specific causal estimates, PPI-based MR was repeated using CVD-related GWAS datasets on the MR-Base platform.[45] The PPI-based MR-IVW causal estimates were highly consistent using summary statistics from two CAD and two MI GWAS on MR-Base, but not cardiomyopathy or AFib (Table S7), suggesting that the stratified causal estimates on CAD are associated with the atherosclerotic component of CVD. Together, these results support the conclusion that the PPI-based MR results are atherosclerosis specific and unlikely to be trivial results of random chance or multiple-hypothesis testing.

SLE-derived clusters in all positive and negative causal tiers were annotated using multiple functional and cellular composition tools (Figure 5B and Table S8). These results show that a wide range of SNP-predicted biological functions known to be involved in SLE pathogenesis have causal implications for CAD by MR, such as "Neutrophil degranulation" (clusters 2 and 43), "Th1 and Th2 activation" and/or "Th17 activation" (clusters 3, 5, 8, and 9), "Interferon signaling" (cluster 8), "Leukocyte extravasation signaling" (cluster 12), "Leukocyte transendothelial migration" (cluster 28), and "Leukocyte adhesion to endothelial cells" (cluster 2). In addition to immune-related pathways, many of these positive causal clusters were enriched in disease phenotypes associated with CVD, including "Th1 cell activation and proliferation in atherosclerosis" (cluster 9) and "Lipid and atherosclerosis" (cluster 12). Interestingly, several clusters were enriched in autonomic nervous system control related to cardiac function ("Cardiac muscle contraction," clusters 13 and 33) or "Neuroinflammation" (cluster 60) (Figure 5B and Table S8).

In contrast, SLE-derived clusters with negative causal estimates on CAD were enriched for oxidative stress (cluster 10), nitric oxide (clusters 24, 40, and 64), and high-density lipoprotein cholesterol (clusters 24 and 50) (Figure 5B and Table S8). Pathway enrichment was further reflected in assigned functional categories, with reactive oxygen species protection (clusters 24 and 45), nuclear receptor transcription (cluster 40), and ubiquitylation and SUMOylation (cluster 64) dominating clusters with protective estimates on CAD (Figure 5B). These results are highly consistent with the enrichments associated with negative causal variants in our single-SNP MR and HLA-specific pathway analyses.
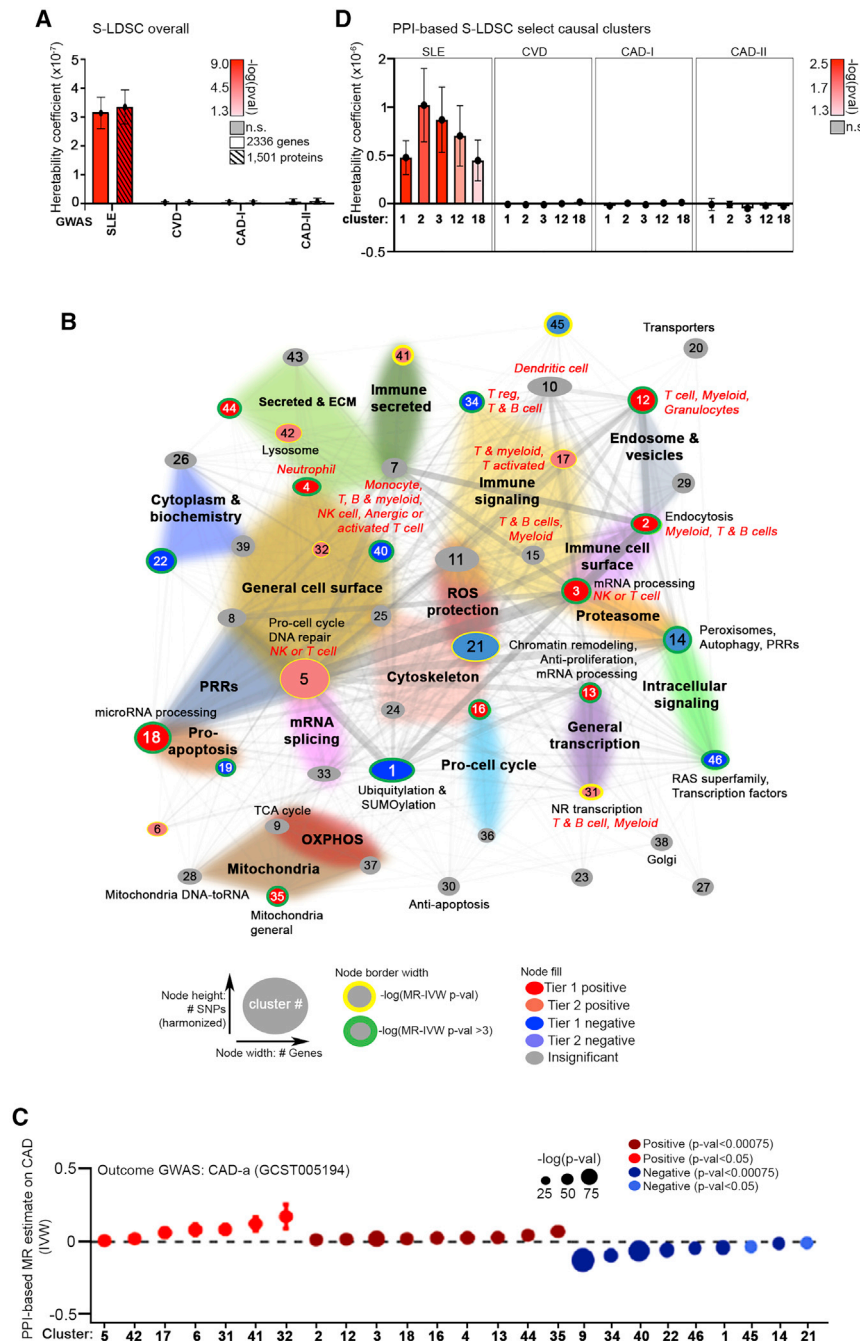
**Figure 4. SLE-derived gene network with causal implications for CAD and PPI-based MR**

(A) S-LDSC using summary statistics for SLE, CVD, and CAD GWAS to estimate the heritability (coefficient ± standard error) of the 2,336 genes (open bars) and 1,501 proteins (hashed bars) predicted by 838 Immunochip SNPs associated with SLE. Bar color indicates coefficient significance.

(B) Functional and cell-type enrichments for cluster metastructures were determined using BIG-C (black labels) and I-scope (red labels), respectively. Bold black labels over colored shadings represent shared BIG-C functional annotations for the clusters they surround. Node size is proportional to the number of SNPs (height) mapping to the genes in each cluster (width). For node color, red and blue indicate significant positive or negative estimates, respectively for 14/16 MR methods used (tier 1); light red and light blue indicate significant positive or negative estimates by MR-IVW or at least 7/16 MR methods (tier 2); gray indicates insignificant. Thickness of the yellow border is roughly proportional to the negative log of the MR-IVW p value. Green border indicates clusters with −log(MR-IVW p value) > 3.

(C) Forest plots from PPI-based MR showing estimates (β ± standard error) calculated by MR-IVW for select positive and negative clusters.

(D) PPI-based S-LDSC (coefficient ± standard error) using GWAS summary statistics for SLE, CVD, and CAD. Bar color indicates coefficient significance.

positive, significant, and consistent than those based on all SNPs (Figures 6B and 6C; Table S3). Similarly, negative causal estimates for SLE on CAD were obtained using IVs meeting the negative tier 1 and MR-IVW p value <0.00075 thresholds from the 67-cluster network. In contrast, IVs derived from clusters with insignificant causal estimates generally failed to reach significance in either direction. While these trends were observed using summary statistics from both the SLE GWAS and SLE Immunochip, causal estimates were more significant using the SLE Immunochip, consistent with its larger sample size. Importantly, these results demonstrate that PPI-based MR can be used to identify independent IVs satisfying MR assumptions that underlie both positive and negative causal effects of SLE on CAD.

## PPI-based MR stratifies SNPs, genes, and networks underlying the positive and negative causal effects of SLE on CAD

To further validate the causal effects of the 67 SNP-to-gene modules identified by PPI-based MR (Figure 6A), we carried out additional MR analyses with respect to PPI-based MR cluster groupings after accounting for pleiotropy and LD. Causal estimates of SLE on CAD with IVs derived from clusters meeting the tier 1 or the tier 1 and 2 criteria, as well as those that surpassed the MR-IVW p value <0.00075 threshold, were universally more

## Pathway analysis facilitates drug prediction

Pathways associated with positive causal clusters were used to facilitate identification of new therapeutic interventions for managing the unique inflammatory environment contributing to CAD in SLE (Figure 7A). Canonical pathways related to immune function in clusters 2, 3, 5, and 8 predicted drugs
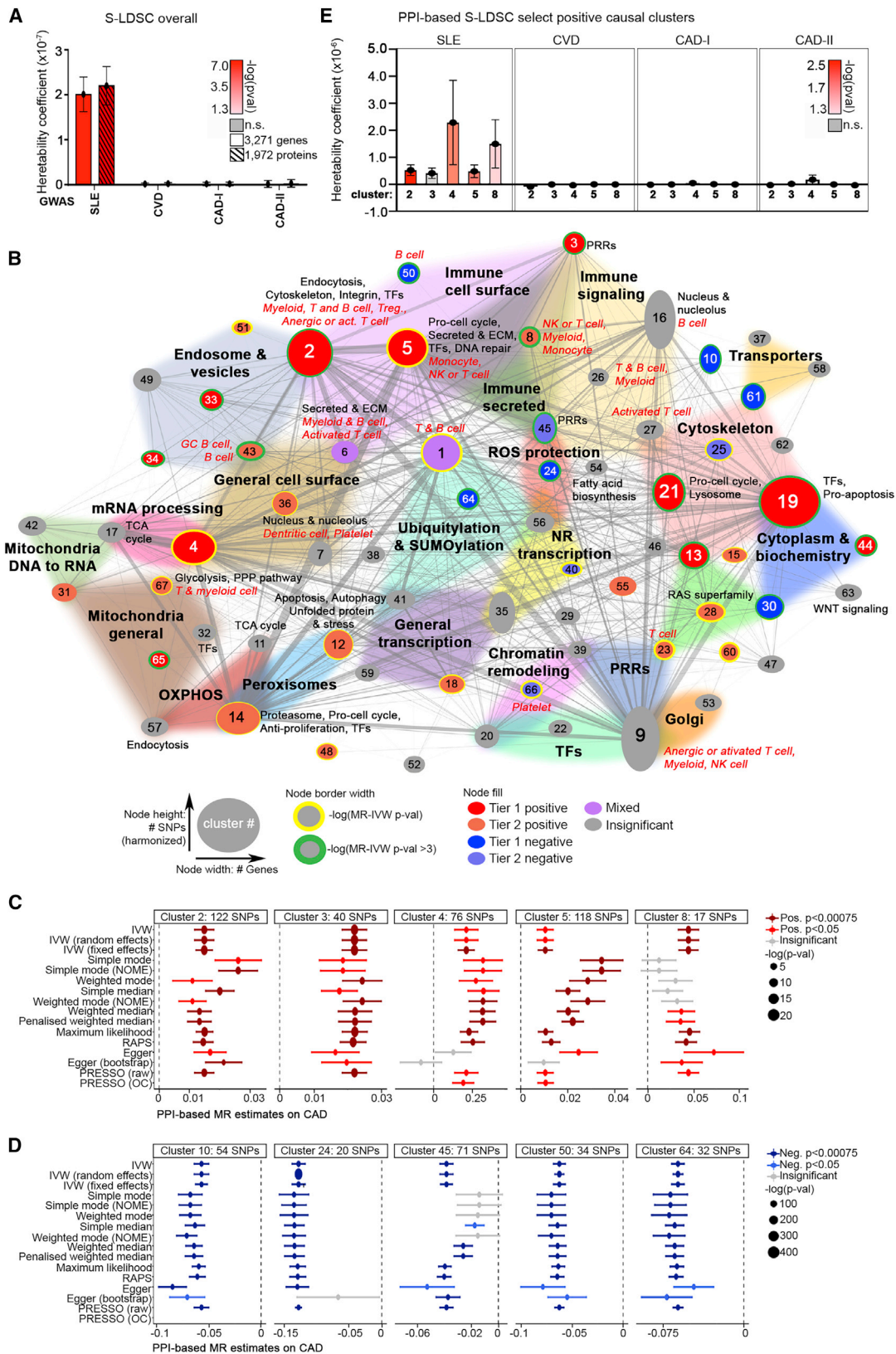
(legend on next page)

targeting T and B cells and inflammatory cytokines, including daratumumab (CD38), belimumab (TNFSF13), elotuzumab (SLAMF7), abatacept (CD80/86), iberdimide (IKZF1/IKZF3), and sarliumab (IL6R). Broader analysis of pathway categories also suggested the utility of targeting interferon signaling with anifrolumab (cluster 8), as well as antiplatelet/coagulant therapy to combat dyslipidemia (cluster 5).[46] Additional noteworthy targets include PCSK9 (cluster 5), a protease involved in the degradation and recycling of the low-density lipoprotein (LDL) receptor targeted by alirocumab and evolocumab, and oxidated LDL molecules (cluster 5) targeted by orticumab (Figure 7B).

## DISCUSSION

Although genetic association studies have been successful in mapping disease loci in both immune and cardiovascular diseases, the genetic and molecular basis for the increased CAD predisposition in SLE patients has remained largely unexplained. Considering the limited data on CAD in SLE, we developed an approach that utilized GWAS summary statistics for both diseases to identify and interpret various sets of SLE-associated variants with causal implications for CAD. New findings suggest that the causal relationship with SLE appears to be focused on the atherosclerotic process, evidenced by positive estimates with CAD, MI, and IS, but not other cardiac conditions such as cardiomyopathy or AFib. Furthermore, we developed and carried out PPI-based MR approach to identify specific sets of SLE variants mapping to biologically relevant gene sets with causal implications for CAD. By coupling various MR methods with network modeling and variant interpretation, we not only provided substantial evidence of shared genetic risk but also identified the putative molecular pathways involved in the development of CAD in SLE. Moreover, a number of the immune and inflammatory pathways identified in these analyses could well contribute to the pathogenesis of CAD even in the absence of SLE or other recognized autoimmune conditions. This points to the larger implication that CAD itself is a heterogeneous condition and subpopulations, such as those driven by SLE-associated processes, might require potentially distinct treatment strategies, at least partially motivated by unique genetic predispositions.

Causal inference using traditional MR methods rely on strict assumptions for independent IVs; however, given the extensive pleiotropy underlying complex traits such as SLE and CVD, ef-

forts to satisfy these assumptions can result in biasing the analyses by excluding previously established associations. Furthermore, the exclusion of SNPs associated with CVD-related traits results in the loss of relevant molecular information. While the use of SLE IVs that are also associated with CVD or confounders in traditional MR disqualifies the causal estimates from representing an effect on CAD directly through SLE, these SNPs can be just as important with respect to understanding the relevant biological pathways underlying CAD in SLE. Similarly, stringent LD clumping to obtain an independent set of IVs not only reduces the statistical power of MR [47] but also can omit additional SNPs, genes, and pathways underlying CAD in SLE. Due to our rigorous efforts to satisfy the assumptions and account for LD in the traditional MR analyses while also employing numerous MR methods that account for IV invalidity, pleiotropy, or heterogeneity, these results may give overly conservative estimates of the causal effects and underlying mechanisms as a result of overpruning.

To overcome these limitations of traditional MR, we developed and employed a PPI-based MR approach using networks comprehensively derived from large sets of SLE-associated SNPs, regardless of their associations with CVD-related traits. By generating cluster-specific associations between effect sizes on SLE and CAD, biologically relevant SNP-to-gene modules can be categorized as having shared (positive estimates) or opposing (negative estimates) effects on SLE and CAD. Traditional MR using independent, SLE-specific IVs mapping to positive and negative clusters separately, confirmed that the groups of causal clusters are representative of positive and negative causal effects on CAD through SLE, respectively. We believe that our PPI-based MR approach is particularly beneficial in cases when the exposure is complex and heterogeneous, such as SLE which embodies a diverse range of molecular and pathophysiological mechanisms that we expect to impose unique casual effects on CAD.

Genetic variants are typically mapped to genes with respect to genomic location, identifying genes containing and/or nearby the SNPs of interest. Additionally, more recent advances have given rise to identification of *trans*-acting genomic regions that can epigenetically and/or transcriptionally influence genes at distant locations. This is especially important for complex, polygenic traits, such as SLE and CAD, of which most associated variants are non-coding. Here, we link SNPs to genes via amino acid changes in encoded proteins, proximity, expression quantitative trait loci predictions, and regulatory elements in an effort to

**Figure 5. Comprehensive PPI-based MR predicts sets of SLE-associated variants and pathways causal of CAD**
(A) S-LDSC using GWAS summary statistics for SLE, CVD, and CAD to estimate the heritability (coefficient ± standard error) of the 3,272 genes (open bars) and 1,972 protein-coding genes (hashed bars) predicted by 1,708 combined Immunochip- and Phenoscanner-derived SNPs. Bar color indicates coefficient significance.
(B) Functional and cell-type enrichments for cluster metastructures were determined using BIG-C (black labels) and I-scope (red labels), respectively. Bold black labels over colored shadings represent shared BIG-C functional annotations for the clusters they surround. Node size is proportional to the number of SNPs (height) mapping to the genes in each cluster (width). For node color, red and blue indicate significant positive or negative estimates, respectively for 14/16 MR methods used (tier 1); light red and light blue indicate significant positive or negative estimates by MR-IVW or at least 7/16 MR methods (tier 2); purple indicates mixed estimates; gray indicates insignificant. Thickness of the yellow border is roughly proportional to the negative log of the MR-IVW p value. Green border indicates clusters with a negative log of the MR-IVW p value >3.
(C and D) Forest plots from PPI-based MR showing estimates (β ± standard error) calculated by 16 MR methods for select positive (C) and negative (D) clusters. The number of SNPs used as IVs for each cluster are indicated in the plots.
(E) PPI-based S-LDSC (coefficient ± standard error) using GWAS summary statistics for SLE, CVD, and CAD. Bar color indicates coefficient significance.
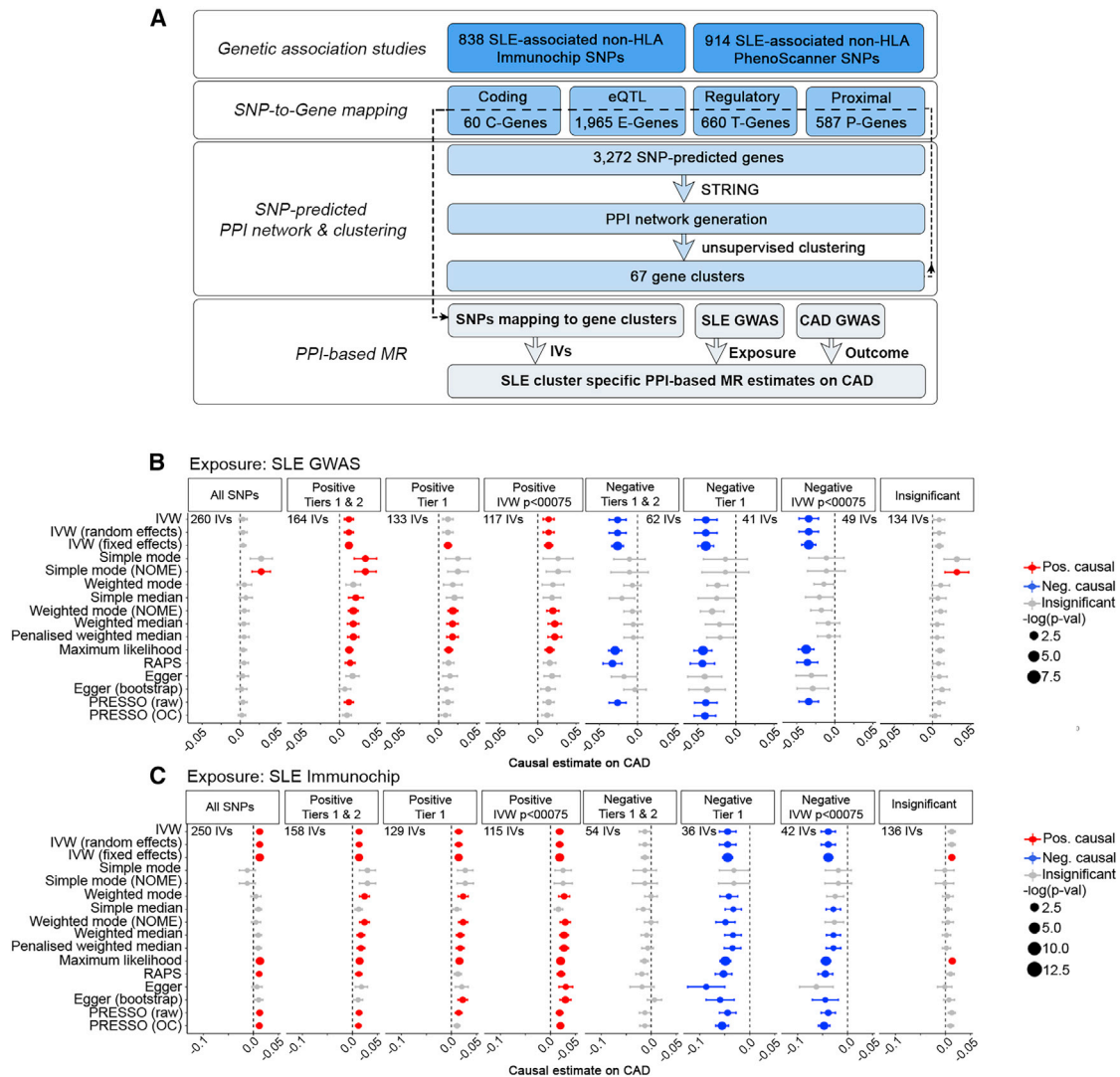
**Figure 6. PPI-based MR identifies SLE SNPs with positive and negative causal effects on CAD**

(A) Workflow depicting PPI-based MR.

(B and C) PPI-based MR validation. Forest plots (β ± standard error) from 16 MR methods using summary statistics from the SLE GWAS (B) or SLE Immunochip (C) as the exposure and CAD GWAS as the outcome. SLE-associated non-HLA SNPs mapping to positive and negative clusters, separately (by tier) and together ("All SNPs") were used as IVs after excluding CVD and confounder-associated SNPs followed by stringent LD clumping ($R^2 = 0.001$) and harmonization. The number of SNPs used as IVs for each SNP set are indicated in the plots. For results, gray indicates insignificant (p > 0.05), dark red indicates positive (p < 0.00075), red indicates positive (p < 0.05), dark blue indicates negative (p < 0.00075), and blue indicates negative (p < 0.05) by each MR method.

be as comprehensive as possible. Our subsequent PPI-based clustering elucidated a broad range of biologically relevant molecular networks within the diverse set of implicated genes and importantly served to filter out noise. Furthermore, our PPI-based MR approach served to highlight SNP-to-gene modules contributing most to the causal effects of SLE on CAD. Together, these results demonstrate how SLE genetics can be used to identify both known and novel loci and pathways with causal implications for CAD.

Numerous biologically relevant SNP-to-gene modules were determined to have positive causal effects on CAD through SLE by MR, spanning inflammatory factors, adaptive and innate immunity, intracellular signaling, cell differentiation, microRNA and mRNA processing, mitochondrial function, and more. A wide range of enrichments among positive causal clusters have been hypothesized and/or demonstrated to contribute to CVD in SLE patients, including glucocorticoids, neutrophil cell death (NETosis) and degranulation, tumor necrosis factor-like weak inducer of apoptosis (TWEAK) signaling, canonical and alternative complement pathways, Th1 differentiation, and lipid and lipoprotein metabolism, among others.

Considering the drastically increased prevalence and mortality of CAD in SLE, the considerable portion of SLE-associated risk variants with negative causal effects on CAD was unexpected

A

Our approach also has the advantage of identifying "actionable" points of therapeutic intervention with the potential to impact the inflammatory environment associated with CAD in SLE. This is especially important given that CAD risk in SLE cannot be fully accounted for by the increased prevalence of traditional atherosclerotic risk factors. SLE subjects therefore may derive particular benefit from treatments that mitigate inflammatory intermediates such as type I interferons with anifrolumab. Our findings also highlight additional putative targets, including PCSK9 involved in LDL receptor recycling. Inhibitors of PCSK9 activity, such as alirocumab and evolocumab, are approved by the Food and Drug Administration to treat hyperlipidemia and may prove to be effective in controlling atherosclerosis in chronic inflammatory conditions.[51] Finally, recent reports also support targeting oxidized LDL molecules (anti-oxLDL, orticumab) for the prevention of cardiovascular events in SLE.[52]

and suggested that numerous variants contributing to SLE have atheroprotective effects. Further SNP-to-gene mapping and detailed pathway analyses revealed that these variants are involved in various processes, predominantly related to oxidative stress and cholesterol homeostasis, whose atheroprotective effects have been found to be impaired in certain disease-related contexts such as SLE. For example, the enzyme responsible for maintaining cholesterol homeostasis though lipoprotein lipase synthesis, cholesterol 27-hydroxylase, has been shown to be decreased in human monocytes and aortic endothelial cells of SLE patients and is thought to impair the protective mechanism of efflux of cellular cholesterol.[48] Cyp27a1 is the gene that encodes cholesterol 27-hydroxylase and is a liver X receptor (LXR) target activated by oxysterols as well as a target of retinoid X receptor (RXR) and peroxisome proliferator activated receptor (PPAR) in human macrophages.[49] LXR activation has additional proatherogenic and atheroprotective effects, as LXR activation in the liver promotes atherosclerosis via excess lipogenesis whereas LXR activation in macrophages and dendritic cells has anti-inflammatory effects, linking lipid metabolism, immune cell function, and inflammation.[50]

## Limitations of the study

Limitations of this study include those related to the data integrated in our pipeline. First, SLE genetic association studies have been restricted in size and scope, yielding limited power and genomic coverage, especially considering the extensive heterogeneity and polygenicity of lupus. To maximize both power and scope, we used the largest genetic association study for SLE, which is limited to Immunochip SNPs, the largest SLE GWAS, as well as SLE-associated SNPs pooled from the Phenoscanner platform. However, most genetic association studies, including the multi-ancestral data used in this study, are heavily biased toward European ancestries[53]. This is especially problematic given the increased CVD morbidity and mortality in SLE patients of African ancestry[54] in addition to the ancestry-dependent disparities observed in both SLE and CAD. It is also of note that certain risk factors leading to distinct phenotypic outcomes such as CAD are likely to be impacted by environmental factors that cannot be accounted for by genetics alone.

This is important with respect to the higher disease burden observed in patients of African ancestry, where barriers to treatment (such as delayed diagnosis and/or limited access to a specialist) may contribute to elevated mortality in this population and further underscores the importance of generating large datasets with diverse patient populations. In addition, the ability to map genetic variants to implicated genes is limited to known SNP-to-gene relationships included in Ensembl's variant effect predictor (VEP), Geno-type-Tissue Expression (GTEx), and Human ACtive Enhancer to interpret Regulatory variants (HACER) databases. Although putative causal pathways associated with the HLA region are intriguing, mapping of the SNPs within the HLA region to genes is challenging because of the extensive LD across the region. Additionally, genes included in our PPI networks and clusters are limited to protein-coding genes and interactions included in STRINGdb. This is a potential shortcoming of our pipeline, especially considering the large number of non-coding genes implicated in our SNP-to-gene predictions in addition to the growing evidence highlighting the contributions of non-coding long RNAs and microRNAs in both SLE and CAD.[55,56] Similarly, the ability to annotate gene clusters functionally is limited and potentially biased by the data underlying the numerous enrichment platforms used in our pathway analyses. Ingenuity Pathway Analysis (IPA), EnrichR, which pools a myriad of public databases, and cell and functional analytic tools were all utilized to obtain orthogonal and reproducible annotations. Ultimately, however, our robust SNP-to-gene mapping approach, which included multiple sources of information in combination with biologically informed clustering employing numerous sources of annotation, enabled comprehensive analysis of both small and large sets of genetic variants to specific pathways with excellent reproducibility.

In summary, we have employed various approaches to clearly identify shared genetic risk factors for SLE and CAD. These results have provided new information about common molecular pathways in SLE and CAD, as well as the genetic and molecular information needed to consider novel therapeutic interventions in these conditions.

## STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
- METHOD DETAILS
  - Identification of SLE- and CAD-associated SNPs and overlap
  - Identification of SNP-predicted genes
  - Network analysis and visualization
  - Functional gene set analysis
  - Drug candidate identification
- QUANTIFICATION AND STATISTICAL ANAYSIS

  - Linkage disequilibrium score regression (LDSC) genetic correlations
  - Stratified linkage disequilibrium score regression (S-LDSC)
  - Selection of valid, independent instrumental variables for traditional MR analysis
  - Mendelian randomization (MR)
  - PPI-based MR
  - Monte Carlo simulations for expected MR results using random sets of Immunochip-derived SNP-to-gene modules

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2022.100805.

### AUTHOR CONTRIBUTIONS

Conceptualization, P.E.L., A.C.G., and K.A.O.; methodology, P.E.L. and J.K.; software, formal analysis, and investigation, J.K.; validation, C.D.L. and M.C.M.; writing – original draft, J.K., P.E.L., and K.A.O.; writing – review & editing, C.D.L., M.C.M., P.E.L., and K.A.O.; visualization, J.K. and K.A.O.; funding acquisition, P.E.L., A.C.G., and C.D.L.; supervision, P.E.L. and A.C.G.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

1. Yen, E.Y., and Singh, R.R. (2018). Brief report: lupus—an unrecognized leading cause of death in young females: a population-based study using nationwide death certificates, 2000–2015. Arthritis Rheumatol. *70*, 1251–1255.

2. Liu, Y., and Kaplan, M.J. (2018). Cardiovascular disease in systemic lupus erythematosus: an update. Curr. Opin. Rheumatol. *30*, 441–448.

3. Gupta, S., and Kaplan, M.J. (2021). Bite of the wolf: innate immune responses propagate autoimmunity in lupus. J. Clin. Invest. *131*. e144918.

4. Thomas, G., Mancini, J., Jourde-Chiche, N., Sarlon, G., Amoura, Z., Harlé, J.R., Jougla, E., and Chiche, L. (2014). Mortality associated with systemic lupus erythematosus in France assessed by multiple-cause-of-death analysis. Arthritis Rheumatol. *66*, 2503–2511.

5. Teixeira, V., and Tam, L.S. (2017). Novel insights in systemic lupus erythematosus and atherosclerosis. Front. Med. *4*, 262.

6. Reppe Moe, S., Haukeland, H., Molberg, Ø., and Lerang, K. (2021). Long-term outcome in systemic lupus erythematosus; knowledge from population-based cohorts. J. Clin. Med. *10*, 4306. https://doi.org/10.3390/jcm10194306.

7. Katz, G., Smilowitz, N.R., Blazer, A., Clancy, R., Buyon, J.P., and Berger, J.S. (2019). Systemic lupus erythematosus is associated with increased prevalence of atherosclerotic cardiovascular disease in hospitalized patients. Mayo Clin. Proc. *94*, 1436–1443.

8. Lopez, E.O., Ballard, B.D., and Jan, A. (2021). Cardiovascular disease. In StatPearls [Internet], editorial board (StatPearls Publishing). PMID: 30571040; Bookshelf ID: NBK535419.

9. Petri, M., Perez-Gutthann, S., Spence, D., and Hochberg, M.C. (1992). Risk factors for coronary artery disease in patients with systemic lupus erythematosus. Am. J. Med. *93*, 513–519.

10. Moghaddam, B., Marozoff, S., Li, L., Sayre, E.C., and Zubieta, J.A.A. (2021). All-cause and cause-specific mortality in systemic lupus erythematosus: a population-based study. Rheumatology *61*, 367–376.

11. Bernatsky, S., Boivin, J.F., Joseph, L., Manzi, S., Ginzler, E., Gladman, D.D., Urowitz, M., Fortin, P.R., Petri, M., Barr, S., et al. (2006). Mortality in systemic lupus erythematosus. Arthritis Rheum. *54*, 2550–2557.

12. Wu, G.C., Liu, H.R., Leng, R.X., Li, X.P., Li, X.M., Pan, H.F., and Ye, D.-Q. (2016). Subclinical atherosclerosis in patients with systemic lupus erythematosus: a systemic review and meta-analysis. Autoimmun. Rev. *15*, 22–37.

13. Yazdany, J., Pooley, N., Langham, J., Nicholson, L., Langham, S., Embleton, N., Wang, X., Desta, B., Barut, V., and Hammond, E. (2020). Original research: systemic lupus erythematosus; stroke and myocardial infarction risk: a systematic review and meta-analysis. RMD Open *6*. e001247.

14. Magder, L.S., and Petri, M. (2012). Incidence of and risk factors for adverse cardiovascular events among patients with systemic lupus erythematosus. Am. J. Epidemiol. *176*, 708–719.

15. Petri, M.A., Barr, E., and Magder, L.S. (2019). Development of a systemic lupus erythematosus cardiovascular risk equation. Lupus Sci. Med. *6*. e000346.

16. Levinson, D.J., Abugroun, A., Daoud, H., and Abdel-Rahman, M. (2020). Coronary artery disease (CAD) risk factor analysis in an age-stratified hospital population with systemic lupus erythematosus (SLE). Int. J. Cardiol. Hypertens. *7*. 100056.

17. Ajeganova, S., Hafström, I., and Frostegård, J. (2021). Patients with SLE have higher risk of cardiovascular events and mortality in comparison with controls with the same levels of traditional risk factors and intima-media measures, which is related to accumulated disease damage and antiphospholipid syndrome: a case–control study over 10 years. Lupus Sci. Med. [Internet] *8*. e000454.

18. Bakshi, J., Segura, B.T., Wincup, C., and Rahman, A. (2018). Unmet needs in the pathogenesis and treatment of systemic lupus erythematosus. Clin. Rev. Allergy Immunol. *55*, 352–367.

19. Calvo-Alén, J., Alarcón, G.S., Campbell, R., Fernández, M., Reveille, J.D., and Cooper, G.S. (2005). Lack of recording of systemic lupus erythematosus in the death certificates of lupus patients. Rheumatology *44*, 1186–1189.

20. Falasinnu, T., Rossides, M., Chaichian, Y., and Simard, J.F. (2018). Do death certificates underestimate the burden of rare diseases? The example of systemic lupus erythematosus mortality, Sweden, 2001-2013. Public Health Rep. *133*, 481–488.

21. Ashley, E.A., Hershberger, R.E., Caleshu, C., Ellinor, P.T., Garcia, J.G.N., Herrington, D.M., Ho, C.Y., Johnson, J.A., Kittner, S.J., Macrae, C.A., et al. (2012). Genetics and cardiovascular disease: a policy statement from the American Heart Association. Circulation *126*, 142–157.

22. Leonard, D., Svenungsson, E., Dahlqvist, J., Alexsson, A., Ärlestig, L., Taylor, K.E., Sandling, J.K., Bengtsson, C., Frodlund, M., Jönsen, A., et al. (2018). Novel gene variants associated with cardiovascular disease in systemic lupus erythematosus and rheumatoid arthritis. Ann. Rheum. Dis. *77*, 1063–1069.

23. Ramirez, G.A. (2018). Genetics in systemic lupus erythematosus: entering the borough of cardiovascular risk. Ann. Transl. Med. *6* (*Suppl 1*), S14.

24. Svenungsson, E., Gustafsson, J., Leonard, D., Sandling, J., Gunnarsson, I., Nordmark, G., Jönsen, A., Bengtsson, A.A., Sturfelt, G., Rantapää-Dahlqvist, S., et al. (2010). A STAT4 risk allele is associated with ischaemic cerebrovascular events and anti-phospholipid antibodies in systemic lupus erythematosus. Ann. Rheum. Dis. *69*, 834–840.

25. Leonard, D., Svenungsson, E., Sandling, J.K., Berggren, O., Jönsen, A., Bengtsson, C., Wang, C., Jensen-Urstad, K., Granstam, S.-O., Bengtsson, A.A., et al. (2013). Coronary heart disease in systemic lupus erythematosus is associated with interferon regulatory factor-8 gene variants. Circ. Cardiovasc. Genet. *6*, 255–263.

26. Owen, K.A., Price, A., Ainsworth, H., Aidukaitis, B.N., Bachali, P., Catalina, M.D., Dittman, J.M., Howard, T.D., Kingsmore, K.M., Labonte, A.C., et al. (2020). Analysis of trans-ancestral SLE risk loci identifies unique biologic networks and drug targets in African and European ancestries. Am. J. Hum. Genet. *107*, 864–881.

27. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. *47*, 1457–1464.

28. Lessard, C.J., Sajuthi, S., Zhao, J., Kim, K., Ice, J.A., Li, H., Ainsworth, H., Rasmussen, A., Kelly, J.A., Marion, M., et al. (2016). Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. Arthritis Rheumatol. *68*, 1197–1209.

29. Morris, D.L., Sheng, Y., Zhang, Y., Wang, Y.-F., Zhu, Z., Tombleson, P., Chen, L., Cunninghame Graham, D.S., Bentham, J., Roberts, A.L., et al. (2016). Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat. Genet. *48*, 940–946.

30. Sun, C., Molineros, J.E., Looger, L.L., Zhou, X.-J., Kim, K., Okada, Y., Ma, J., Qi, Y.-Y., Kim-Howard, X., Motghare, P., et al. (2016). High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. Nat. Genet. *48*, 323–330.

31. Langefeld, C.D., Ainsworth, H.C., Cunninghame Graham, D.S., Kelly, J.A., Comeau, M.E., Marion, M.C., Howard, T.D., Ramos, P.S., Croker, J.A., Morris, D.L., et al. (2017). Transancestral mapping and genetic load in systemic lupus erythematosus. Nat. Commun. *8*. 16021.

32. Van Der Harst, P., and Verweij, N. (2018). Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ. Res. *122*, 433–443.

33. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

34. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat. Genet. *49*, 403–415.

35. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat. Genet. *43*, 333–338.

36. Howson, J.M.M., Zhao, W., Barnes, D.R., Ho, W.K., Young, R., Paul, D.S., Waite, L.L., Freitag, D.F., Fauman, E.B., Salfati, E.L., et al. (2017). Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat. Genet. *49*, 1113–1119.

37. Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E.M. (2018). A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int. J. Methods Psychiatr. Res. *27*. e1608.

38. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.

39. Hartiala, J.A., Han, Y., Jia, Q., Hilser, J.R., Huang, P., Gukasyan, J., Schwartzman, W.S., Cai, Z., Biswas, S., Trégouët, D.A., et al. (2021). Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. Eur. Heart J. *42*, 919–933.

40. Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.-K., van der Laan, S.W., Gretarsdottir, S., et al. (2018). Multiancestry genome-wide association study of 520, 000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat. Genet. *50*, 524–537.

41. Nielsen, J.B., Thorolfsdottir, R.B., Fritsche, L.G., Zhou, W., Skov, M.W., Graham, S.E., Herron, T.J., McCarthy, S., Schmidt, E.M., Sveinbjornsson, G., et al. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat. Genet. *50*, 1234–1239.

42. Raj, P., Rai, E., Song, R., Khan, S., Wakeland, B.E., Viswanathan, K., Arana, C., Liang, C., Zhang, B., Dozmorov, I., et al. (2016). Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. Elife *5*. e12089.

43. Davies, R.W., Wells, G.A., Stewart, A.F.R., Erdmann, J., Shah, S.H., Ferguson, J.F., Hall, A.S., Anand, S.S., Burnett, M.S., Epstein, S.E., et al. (2012). A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. Circ. Cardiovasc. Genet. *5*, 217–225.

44. Trowsdale, J., and Knight, J.C. (2013). Major histocompatibility complex genomics and human disease. Annu. Rev. Genomics Hum. Genet. *14*, 301–323.

45. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife *7*. e34408.

46. Piranavan, P., and Perl, A. (2020). Management of cardiovascular disease in patients with systemic lupus erythematosus. Expert Opin. Pharmacother. *21*, 1617–1628.

47. Burgess, S., Zuber, V., Valdes-Marquez, E., Sun, B.B., and Hopewell, J.C. (2017). Mendelian randomization with fine-mapped genetic data: choosing from large numbers of correlated instrumental variables. Genet. Epidemiol. *41*, 714–725.

48. Reiss, A.B., Anwar, K., Merrill, J.T., Chan, E.S.L., Awadallah, N.W., Cronstein, B.N., Michael Belmont, H., Belilos, E., Rosenblum, G., Belostocki, K., et al. (2010). Plasma from systemic lupus patients compromises cholesterol homeostasis: a potential mechanism linking autoimmunity to atherosclerotic cardiovascular disease. Rheumatol. Int. *30*, 591–598.

49. Quinn, C.M., Jessup, W., Wong, J., Kritharides, L., and Brown, A.J. (2005). Expression and regulation of sterol 27-hydroxylase (CYP27A1) in human macrophages: a role for RXR and PPARgamma ligands. Biochem. J. *385* (*Pt 3*), 823–830.

50. Bilotta, M.T., Petillo, S., Santoni, A., and Cippitelli, M. (2020). Liver X receptors: regulators of cholesterol metabolism, inflammation, autoimmunity, and cancer. Front. Immunol. *11*. 584303.

51. Liu, A., Rahman, M., Hafström, I., Ajeganova, S., and Frostegård, J. (2020). Proprotein convertase subtilisin kexin 9 is associated with disease activity and is implicated in immune activation in systemic lupus erythematosus. Lupus *29*, 825–835.

52. Yao Mattisson, I., Rattik, S., Björkbacka, H., Ljungcrantz, I., Terrinoni, M., Lebens, M., Holmgren, J., Fredrikson, G.N., Gullstrand, B., Bengtsson, A.A., et al. (2021). Immune responses against oxidized LDL as possible targets for prevention of atherosclerosis in systemic lupus erythematosus. Vascul. Pharmacol. *140*. 106863.

53. Demirci, F.Y., Wang, X., Kelly, J.A., Morris, D.L., Barmada, M.M., Feingold, E., Kao, A.H., Sivils, K.L., Bernatsky, S., Pineau, C., et al. (2016). Identification of a new susceptibility locus for systemic lupus erythematosus on chromosome 12 in individuals of European ancestry. Arthritis Rheumatol. *68*, 174–183.

54. Barnado, A., Carroll, R.J., Casey, C., Wheless, L., Denny, J.C., and Crofford, L.J. (2018). Phenome-wide association study identifies marked increased in burden of comorbidities in African Americans with systemic lupus erythematosus. Arthritis Res. Ther. *20*, 69.

55. Esteller, M. (2011). Non-coding RNAs in human disease. Nat. Rev. Genet. *12*, 861–874.

56. Hrdlickova, B., de Almeida, R.C., Borek, Z., and Withoff, S. (2014). Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease. Biochim. Biophys. Acta *1842*, 1910–1922.

57. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics *32*, 3207–3209.

58. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics *35*, 4851–4853.

59. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen Consortium; Psychiatric Genomics Consortium; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3; Duncan, L., Perry, J.R.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

60. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software Environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

61. Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., and Ferrin, T.E. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinf. *12*, 436.

62. Wang, J., Dai, X., Berry, L.D., Cogan, J.D., Liu, Q., and Shyr, Y. (2019). HACER: an atlas of human active enhancers to interpret regulatory variants. Nucleic Acids Res. *47*, D106–D112.

63. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. Genome Biol. *17*, 122.

64. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. *49*, D605–D612.

65. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128.

66. Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., Von Mering, C., Jensen, L.J., Beyer, A., and Bork, P. (2010). Stitch 2: an interaction network database for small molecules and proteins. Nucleic Acids Res. *38*, D552–D556.

67. Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A., et al. (2018). The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. Cell Syst. *6*, 13–24.

68. The GTEx Consortium (2013). The genotype-tissue expression (GTEx) project. Nat. Gen. *45*, 580–585.

69. Labonte, A.C., Kegerreis, B., Geraci, N.S., Bachali, P., Madamanchi, S., Robl, R., Catalina, M.D., Lipsky, P.E., and Grammer, A.C. (2018). Identification of alterations in macrophage activation associated with disease activity in systemic lupus erythematosus. PLoS One *13*. e0208132.

70. Catalina, M.D., Bachali, P., Geraci, N.S., Grammer, A.C., and Lipsky, P.E. (2019). Gene expression analysis delineates the potential roles of multiple interferons in systemic lupus erythematosus. Commun. Biol. *2*, 140.

71. Catalina, M.D., Owen, K.A., Labonte, A.C., Grammer, A.C., and Lipsky, P.E. (2020). The pathogenesis of systemic lupus erythematosus: harnessing big data to understand the molecular basis of lupus. J. Autoimmun. *110*. 102359.

72. Ren, J., Catalina, M.D., Eden, K., Liao, X., Read, K.A., Luo, X., McMillan, R.P., Hulver, M.W., Jarpe, M., Bachali, P., et al. (2019). Selective histone deacetylase 6 inhibition normalizes b cell activation and germinal center formation in a model of systemic lupus erythematosus. Front. Immunol. *10*, 2512.

73. Yang, W., Shen, N., Ye, D.Q., Liu, Q., Zhang, Y., Qian, X.X., Hirankarn, N., Ying, D., Pan, H.-F., Mok, C.C., et al. (2010). Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet. *6*. e1000841.

74. Chung, S.A., Taylor, K.E., Graham, R.R., Nititham, J., Lee, A.T., Ortmann, W.A., Jacob, C.O., Alarcón-Riquelme, M.E., Tsao, B.P., Harley, J.B., et al. (2011). Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. PLoS Genet. *7*. e1001323.

75. Yang, J., Yang, W., Hirankarn, N., Ye, D.Q., Zhang, Y., Pan, H.F., Mok, C.C., Chan, T.M., Wong, R.W.S., Mok, M.Y., et al. (2011). ELF1 is associated with systemic lupus erythematosus in Asian populations. Hum. Mol. Genet. *20*, 601–607.

76. Lee, Y.H., Bae, S.C., Choi, S.J., Ji, J.D., and Song, G.G. (2012). Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol. Biol. Rep. *39*, 10627–10635.

77. Okada, Y., Shimane, K., Kochi, Y., Tahira, T., Suzuki, A., Higasa, K., Takahashi, A., Horita, T., Atsumi, T., Ishii, T., et al. (2012). A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. PLoS Genet. *8*. e1002455.

78. Yang, W., Tang, H., Zhang, Y., Tang, X., Zhang, J., Sun, L., Yang, J., Cui, Y., Zhang, L., Hirankarn, N., et al. (2013). Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am. J. Hum. Genet. *92*, 41–51.

79. Armstrong, D.L., Zidovetzki, R., Alarcón-Riquelme, M.E., Tsao, B.P., Criswell, L.A., Kimberly, R.P., Harley, J.B., Sivils, K.L., Vyse, T.J., Gaffney, P.M., et al. (2014). GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. Genes Immun. *15*, 347–354.

80. Alarcón-Riquelme, M.E., Ziegler, J.T., Molineros, J., Howard, T.D., Moreno-Estrada, A., Sánchez-Rodríguez, E., Ainsworth, H.C., Ortiz-Tello, P., Comeau, M.E., Rasmussen, A., et al. (2016). Genome-wide association study in an amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of European admixture. Arthritis Rheumatol. *68*, 932–943.

81. Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V.K., et al. (2008). Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N. Engl. J. Med. *358*, 900–909.

82. Graham, R.R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C.J., Leon, J.M., Burtt, N.P., Guiducci, C., Parkin, M., Gates, C., et al. (2008). Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat. Genet. *40*, 1059–1061.

83. International Consortium for Systemic Lupus Erythematosus Genetics SLEGEN; Harley, J.B., Alarcón-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., et al. (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat. Genet. *40*, 204–210.

84. Kozyrev, S.V., Abelson, A.K., Wojcik, J., Zaghlool, A., Linga Reddy, M.V.P., Sanchez, E., Gunnarsson, I., Svenungsson, E., Sturfelt, G., Jönsen, A., et al. (2008). Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat. Genet. *40*, 211–216.

85. Oishi, T., Iida, A., Otsubo, S., Kamatani, Y., Usami, M., Takei, T., Uchida, K., Tsuchiya, K., Saito, S., Ohnisi, Y., et al. (2008). A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population. J. Hum. Genet. *53*, 151–162.

86. Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R.C., Nordmark, G., et al. (2009). A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat. Genet. *41*, 1228–1233.

87. Han, J.W., Zheng, H.F., Cui, Y., Sun, L.D., Ye, D.Q., Hu, Z., Xu, J.-H., Cai, Z.-M., Huang, W., Zhao, G.-P., et al. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat. Genet. *41*, 1234–1237.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Deposited data** | | |
| GTEX v.6.8 | The Geno-type-Tissue Expression (GTEx) Project | GTEXportal.org |
| Pre-processed summary statistics | The Broad Institute | https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/ |
| UK Biobank | Biobank UK | https://www.ukbiobank.ac.uk/ |
| Cardiomyopathy GWAS | FinnGen Biobank | Finn-a-I9_CARDMYO; https://www.finngen.fi/fi |
| MR-Base | Hemani et al., 2018[45] | https://www.mrbase.org |
| Phenoscanner | Staley et al., 2016[57]; Kamat et al., 2019[58] | www.phenoscanner.medschl.cam.ac.uk |
| **Software and algorithms** | | |
| Bioconductor (R) | Open source | https://www.bioconductor.org |
| Ingenuity pathway analysis (IPA) | Qiagen | https://www.qiagenbioinformatics.com |
| S-LDSC | Finucane et al., 2015[33] | https://github.com/bulik/ldsc |
| LDSC | Bulik-Sullivan et al., 2015[59] | https://github.com/bulik/ldsc |
| Cytoscape v.3.9.1 | Shannon et al., 2003[60] | https://cytoscape.org |
| Clustermaker2 v.1.2.1 | Morris et al., 2011[61] | https://apps.cytoscape.org |
| TwoSample MR | Hemani et al., 2018[45] | https://github.com/MRCIEU/TwoSampleMR |
| PPI-based MR | This manuscript | https://doi.org/10.6084/Lm9.figshare.21225251 |
| **Other** | | |
| Human Active Enhancers to interpret Regulatory variants (HACER) | Wang et al., 2019[62] | http://bioinfo.vanderbilt.edu/AE/HACER |
| Variant Effect Predictor (VEP) | McLaren et al., 2016[63] | ensembl.org/info/docs/tools/vep |
| Search Tool for the Retrieval of Interacting Genes/proteins (STRING) v. 11.0b | Szklarczyk et al., 2021[64] | https://string-db.org |
| EnrichR | Chen et al., 2013[65] | https://maayanlab.cloud/Enrichr/ |
| Search Tool for Interacting Chemicals (STITCH) | Kuhn et al., 2009[66] | http://stitch.embl.de |
| Library of Integrated Network-based Cellular Signatures (LINCS) | Keenan et al., 2018[67] | http://www.lincs.hms.harvard.edu/db/ |

## RESOURCE AVAILABILITY

### Lead contact
Further information and requests should be directed to the lead contact, Katherine A. Owen (kate.owen@ampelbiosolutions.com).

### Materials availability
This study did not generate new unique reagents.

### Data and code availability
- This paper analyzes existing, publicly available data. All GWAS and Immunochip studies are referenced.
- Original software code and documentation to conduct PPI-based MR analyses have been deposited on figshare (www.figshare.com; https://doi.org/10.6084/m9.figshare.21225251) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this report is available from the lead contact upon request.

## EXPERIMENTAL MODELS AND SUBJECT DETAILS

This study did not use any experimental models or enroll human subjects.

## METHOD DETAILS

### Identification of SLE- and CAD-associated SNPs and overlap

SNPs associated with each disease were obtained from previous GWAS and Immunochip studies. For CAD, we used a comprehensive multi-ancestral meta-analysis of GWAS.[32] For SLE, we included results of multiple GWAS and Immunochip studies to account for as many ancestries as possible.[27–31] In total, 7,222 and 16,163 unique SNPs were significantly ($p < 10^{-6}$) associated with SLE and CAD, respectively, and were employed in these studies. A full list of the SNPs, chromosome locations, positions and sources used are detailed in Table S9.

### Identification of SNP-predicted genes

Expression quantitative trait loci (eQTLs) were identified using GTEx[68] version 6.8 (GTEXportal.org) and mapped to their associated eQTL expression genes (E-Genes). To find SNPs in enhancers and promoters, and their associated transcription factors and downstream target genes (T- Genes), we queried the atlas of Human Active Enhancers to interpret Regulatory variants[62] (HACER, http://bioinfo.vanderbilt.edu/AE/HACER). To find SNPs in exons of protein-coding genes (C-Genes) and include proximal genes (P-Genes, within 5kb), we queried the human Ensembl genome browser's variant effect predictor[63] (VEP, ensembl.org/info/docs/tools/vep, GRCh38.p12).

### Network analysis and visualization

Protein-protein interaction (PPI) networks of SNP-predicted protein-coding genes were generated by STRING[64] (https://string-db.org, version 11.0b), and resulting networks were imported into Cytoscape[60] (version 3.6.1) for visualization and partitioned with MCODE via the clusterMaker2[61] (version 1.2.1) plugin. Metastructures are based on PPI networks. For all metastructures, node gradient shading is proportional to intra-cluster connectivity, cluster size indicates number of genes per cluster and edge weight indicates inter-cluster connections.

### Functional gene set analysis

Predicted genes were examined using Biologically Informed Gene Clustering (BIG-C; version 4.4.). BIG-C is a custom functional clustering tool developed to annotate the biological meaning of large lists of genes and has been previously described.[69–71] I-Scope is a custom clustering tool used to identify immune cell types in large gene datasets.[72] The Ingenuity Pathway Analysis (IPA; https://www.qiagenbioinformatics.com) platform and EnrichR[65] (https://maayanlab.cloud/Enrichr/) web server provided additional molecular pathway enrichment analysis.

### Drug candidate identification

Drug candidates were identified using LINCS,[67] STITCH[66] (v5.0), IPA and literature mining. Each of the database tools includes either a programmatic method of matching existing therapeutics to their targets or else is a list of drugs and targets for achieving the same end.

## QUANTIFICATION AND STATISTICAL ANAYSIS

### Linkage disequilibrium score regression (LDSC) genetic correlations

LDSC[59] was used to estimate genome-wide genetic correlations between traits using GWAS summary statistics. Pre-processed summary statistics from SLE, CAD and CVD GWAS were obtained from the Broad webpage (https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/). Using the LDSC software provided on github (https://github.com/bulik/ldsc) and reference data on the Broad webpage (https://alkesgroup.broadinstitute.org/LDSCORE/), including European LD scores 'eur_w_ld_chr' or 'weights_hm3_no_hla' as weights for analyses excluding the HLA region. Using standard parameters, the "ldsc.py" (with the "–rg" flag) script was used to generate genome-wide genetic correlation estimates between SLE and CVD or CAD.

### Stratified linkage disequilibrium score regression (S-LDSC)

S-LDSC[33] was used to obtain gene-set specific disease-heritability estimates using GWAS summary statistics. Pre-processed summary statistics from SLE, CAD and CVD GWAS were obtained from Broad webpage (https://alkesgroup.broadinstitute.org/LDSCORE/all_sumstats/). Using the S-LDSC software provided on github (https://github.com/bulik/ldsc) and reference data on the Broad webpage (https://alkesgroup.broadinstitute.org/LDSCORE/), annotation and LD score files were generated for each SNP-predicted gene- and protein-set, separately. Using standard parameters, the "make_annot.py" and "ldsc.py" (with the "–l2" flag) scripts were first used to generate the gene-set-specific annotation and LD files, then the "ldsc.py" (with the "--h2-cts" flag) script was used to generate stratified heritability scores for each GWAS.

### Selection of valid, independent instrumental variables for traditional MR analysis

Traditional MR methods, such as MR-IVW, operate under three strict assumptions for instrumental variable (IV) validity: 1) the relevance assumption, 2) the exclusion restriction criteria assumption, and 3) the independence assumption. To satisfy the relevance

assumption, SNPs significantly (genome-wide significance p value $< 5 \times 10^{-8}$) associated with SLE[27–29,73–87] were obtained from the Phenoscanner database (www.phenoscanner.medschl.cam.ac.uk)(82,83) (Table S9). To satisfy the exclusion restriction criteria and independence assumptions, 89,336 SNPs weakly associated (p value $< 1 \times 10^{-5}$) with CVD and confounders including cholesterol, obesity, blood pressure, insulin resistance, smoking, age-related diseases, and many more, were excluded from being IVs for SLE-exposure (see Table S2 for the full list of excluded traits). HLA-region SNPs were conservatively removed from MR analyses by excluding the short-arm of chromosome 6. Stringent LD clumping[37] was employed using the clump data ($R^2 = 0.001$, 100kb window, 1000G EA reference population) function to generate an independent set of 60 SLE-IVs harmonized for each GWAS.

### Mendelian randomization (MR)

MR was used to test for causal relationships between SLE and CAD using the MR-Base[45] (https://www.mrbase.org) Two SampleMR[45] package in R (https://github.com/MRCIEU/TwoSampleMR). Various sets of SLE-associated genetic variants used as instrumental variables (IVs) and summary statistics for SLE-exposure were manually imported into R and summary statistics were carried out for MR-base compatibility using the 'format data' command. All effect sizes and standard errors were obtained from the exposure summary statistics used in each analysis, regardless of the study in which each IV was associated with the exposure. Given the availability of well-powered CAD/MI GWAS on MR-Base, IVs for CAD and MI were directly obtained from each exposure GWAS using the 'extract instruments' command for the bidirectional analyses. Data from the SLE and all CVD-related GWAS studies used in our MR analyses are publicly available and also accessible through the MR-Base software, which was used to obtain the outcome summary statistics via the 'extract outcome data' command. The 'allele harmonization' command was used to ensure the effect estimates of the exposure and outcome are based on matching alleles, excluding SNPs with completely mismatching alleles from the MR analysis or reversing the effect and non-effect alleles along with the effect estimates when applicable. Because of the allele harmonization step and because some SNPs are absent from the available summary statistics, a small proportion of SNPs used as IVs are absent from the final MR calculations. Up to sixteen individual MR methods were carried out through the TwoSampleMR package, including inverse variance weighted (IVW), simple mode, weighted mode, simple median, weighted median (WMedian), MR-Egger, MR-PRESSO (raw and outlier-corrected), MR-RAPS, and two sample maximum likelihood (ML). The 'MR report' function was used to generate a summary containing heterogeneity and directional pleiotropy tests and scatterplots (Figure S2). MR-IVW and MR-Egger heterogeneity test results (Q-value) indicate whether significant heterogeneity was detected, which does not necessarily indicate biased causal estimates. MR-Egger intercept indicate whether significant *directional* horizontal pleiotropy was detected, which usually indicates biased causal estimates. For single-SNP MR, the 'MR single-SNP' function was also carried out using the Wald Ratio method. Full details of all MR results are included in Table S2 and a summary of all the main findings are included in Table S10.

### PPI-based MR

SLE-associated variants from the Immunochip[31] and Phenoscanner database[57,58] were linked to their most likely genes, and the genes used to generate PPI-informed gene clusters. The SLE-associated SNPs mapping to genes in each of PPI-based clusters were then extracted to "reverse engineer" subsets of SNPs that could be used separately as SLE-IVs for MR to independently estimate the causal effects of each PPI-informed SNP-to-Gene module on CAD. Up to sixteen MR methods were carried out for each SNP-to-gene module through the TwoSampleMR package.

In additional analyses (related to Figure 6) using the combined Immunochip and Phenoscanner SNP dataset, filtering eliminated SNPs weakly associated (p value $< 1 \times 10^{-5}$) with CVD and confounders including cholesterol, obesity, blood pressure, insulin resistance, smoking, age-related diseases, and many more (Table S2). HLA-region SNPs were conservatively removed from MR analyses by excluding the short-arm of chromosome 6. Stringent LD clumping[37] was employed using the clump_data ($R^2 = 0.001$, 100kb window, 1000G EA reference population) function to generate an independent set of SLE-IVs. Various analyses were performed using independent, valid IVs derived from 'All SNPs', SNPs mapping to 'Insignificant' clusters, 'Positive Tier 1' clusters, 'Positive Tier 1 and Tier 2' clusters, 'Positive MR-IVW p value<0.00075' clusters, 'Negative MR-IVW p value<0.00075' clusters, 'Negative Tier 1' clusters, and 'Negative Tier 1 and Tier 2' clusters.

### Monte Carlo simulations for expected MR results using random sets of Immunochip-derived SNP-to-gene modules

Monte Carlo simulations were implemented and performed to estimate the false discovery rate with respect to significant PPI-based MR causal estimates. 120,026 Immunochip SNPs included in the SLE summary statistics were mapped to putative genes using the VEP, including regulatory effects, to generate an Immunochip SNP-to-Gene library with 67,211 unique SNPs mapping to 7,602 STRINGdb proteins. In each simulation, a random set of 3–152 SNP-predicted proteins were selected from the 7,602 proteins and used to extract up to 400 Immunochip SNPs. MR-IVW was then performed for SLE on CAD using harmonized, non-HLA SNPs (via removal of the entire short-arm of chromosome 6) from the simulated set of Immunochip SNPs as IVs. By using our Immunochip derived SNP-to-Gene dictionary for random selection of protein clusters and associated SNPs to generate random sets of IVs, our simulations account for both a high degree of LD and pleiotropy, especially considering the major influence of loci associated with diabetes in development of the Immunochip.