



Article

Pervasive Sharing of Causal Genetic Risk Factors Contributes to Clinical and Molecular Overlap between Sjögren's Disease and Systemic Lupus Erythematosus

Karen Chau ¹, Yanint Raksadawan ² , Kristen Allison ¹, John A. Ice ³, Robert Hal Scofield ^{3,4,5,6}, Iouri Chepelev ⁷ and Isaac T. W. Harley ^{1,8,*}

¹ Division of Rheumatology, Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045, USA

² Internal Medicine Residency Program, Louis A. Weiss Memorial Hospital, Chicago, IL 60640, USA

³ Research Service, Oklahoma City US Department of Veterans Affairs Medical Center, Oklahoma City, OK 73104, USA

⁴ Medicine Service, Oklahoma City US Department of Veterans Affairs Medical Center, Oklahoma City, OK 73104, USA

⁵ Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

⁶ Arthritis & Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA

⁷ Research Service, Cincinnati US Department of Veterans Affairs Medical Center, Cincinnati, OH 45220, USA

⁸ Rheumatology Section, Medicine Service, Eastern Colorado Healthcare System, US Department of Veterans Affairs Medical Center, Aurora, CO 80045, USA

* Correspondence: isaac.harley@cuanschutz.edu



Citation: Chau, K.; Raksadawan, Y.; Allison, K.; Ice, J.A.; Scofield, R.H.; Chepelev, I.; Harley, I.T.W. Pervasive Sharing of Causal Genetic Risk Factors Contributes to Clinical and Molecular Overlap between Sjögren's Disease and Systemic Lupus Erythematosus. *Int. J. Mol. Sci.* **2023**, *24*, 14449. <https://doi.org/10.3390/ijms241914449>

Academic Editor: Sharmila Masli

Received: 7 July 2023

Revised: 1 September 2023

Accepted: 5 September 2023

Published: 22 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: SjD (Sjögren's Disease) and SLE (Systemic Lupus Erythematosus) are similar diseases. There is extensive overlap between the two in terms of both clinical features and pathobiologic mechanisms. Shared genetic risk is a potential explanation of this overlap. In this study, we evaluated whether these diseases share causal genetic risk factors. We compared the causal genetic risk for SLE and SjD using three complementary approaches. First, we examined the published GWAS results for these two diseases by analyzing the predicted causal gene protein–protein interaction networks of both diseases. Since this method does not account for overlapping risk intervals, we examined whether such intervals also overlap. Third, we used two-sample Mendelian randomization (two sample MR) using GWAS summary statistics to determine whether risk variants for SLE are causal for SjD and vice versa. We found that both the putative causal genes and the genomic risk intervals for SLE and SjD overlap 28- and 130-times more than expected by chance ($p < 1.1 \times 10^{-24}$ and $p < 1.1 \times 10^{-41}$, respectively). Further, two sample MR analysis confirmed that alone or in aggregate, SLE is likely causal for SjD and vice versa. [SjD variants predicting SLE: OR = 2.56; 95% CI (1.98–3.30); $p < 1.4 \times 10^{-13}$, inverse-variance weighted; SLE variants predicting SjD: OR = 1.36; 95% CI (1.26–1.47); $p < 1.6 \times 10^{-11}$, inverse-variance weighted]. Notably, some variants have disparate impact in terms of effect size across disease states. Overlapping causal genetic risk factors were found for both diseases using complementary approaches. These observations support the hypothesis that shared genetic factors drive the clinical and pathobiologic overlap between these diseases. Our study has implications for both differential diagnosis and future genetic studies of these two conditions.

Keywords: Systemic Lupus Erythematosus (SLE); Sjögren's Disease (SjD); Sjögrens syndrome; overlap; GWAS; Mendelian randomization; causal gene networks

1. Introduction

The advent of omics-based approaches to disease characterization has prompted the reconsideration of traditional phenotype-based disease classification. SjD (Sjögren's

Disease) and SLE (Systemic Lupus Erythematosus) are two diseases worth considering for reclassification. The observations that these diseases share many features [1–3] and overlap substantially [4,5] are not new. A recent focus of the field has been to define a classification criteria that distinguish between these two diseases [6,7]. The goal is to improve the homogeneity of cohorts examined in clinical trials and thereby improve trial success [8]. Nonetheless, several important clinical features overlap. Maternal seropositivity for anti-Ro/SS-A, the characteristic autoantibody for SjD, is the most common cause of neonatal lupus [5]. Many patients share features of both diseases [9], and secondary SjD is found in up to 52% of SLE patients [4,6,10,11]. The clinical diagnostic categories are blurry. Initial diagnosis with another autoimmune disease, such as SLE that is then revised to SjD, is common [12].

In addition to clinical overlap, several pathobiologic mechanisms are shared as well. These shared mechanisms include complement [13], interferon signaling [14,15] and extrafollicular B cell development [16]. Within both diseases, patients fit into stable molecularly defined disease sub-clusters [14,17–20]. However, a study by the PRECISEADS consortium performed longitudinal multi-omic profiling of persons with diverse systemic autoimmune diseases, including SLE and SjD [21]. Analysis of these data defined four stable clusters of multi-omic signatures for systemic autoimmune diseases that were stable over time regardless of therapy. Importantly, these clusters do not align with traditional disease classification. That is, persons diagnosed with either SjD or SLE were found in similar proportions across all clusters with a preponderance for the interferon signature cluster. Similarly, a persistent anti-Ro (or anti-SSA) + molecular signature of inflammatory gene expression spanning systemic autoimmune diseases has been described [22].

There is also evidence of genetic overlap. Both in family-based studies and nation-wide health records studies, the risk of SLE and SjD in first-degree relatives of patients with either is 5–18× higher [23–25]. We previously described rs2431697, a risk variant that disrupts an enhancer for microRNA-146A, as a shared causal risk factor for SLE and SjD [26]. Similarly, SLE and SjD share X chromosome copy number as a major genetic contributor to disease risk [27], and several of the genetic risk loci in SLE and SjD overlap [28,29].

In this paper, we sought to better understand the extent of genetic overlap between these two disease phenotypes. To do so, we collated recent GWAS studies of SjD and SLE. First, we analyzed predicted causal genes and found extensive overlap. Second, we analyzed genomic risk intervals and found extensive overlap. Finally, Mendelian randomization analysis of variants from several published GWAS studies from each phenotype revealed that SLE is likely causal for SjD and vice versa. To our knowledge, we are the first to apply these analytic techniques to evaluate the overlap between SjD and SLE. Our results reveal pervasive sharing of causal genetic risk factors between these two disease phenotypes. They also suggest a revised nosology that either considers SjD as a form fruste of SLE, SLE as a severe form of SjD or that both diseases exist together on a spectrum. This novel paradigm may prove useful as a conceptual approach to these commonly overlapping diseases.

2. Results

To address the extent of genetic overlap between SjD and SLE, we applied three methods to compare GWAS results from these diseases. Both detailed description of our methods (putative causal gene network overlap, association interval overlap and Mendelian randomization) and a schematic of our overall analysis are presented in the Section 4.

2.1. Sharing of GWAS Identified Genes and Loci between Sjögren's Disease and Systemic Lupus Erythematosus

2.1.1. SjD Putative Causal Gene Predicted Protein–Protein Interaction Networks

We used a published approach to annotate the predicted causal genes for SjD genetic risk loci from the EBI GWAS catalog [30] (Table S1A). One SjD GWAS study, that of Khatri et al., has not yet been incorporated into the L2G Open Targets Genetics database [31,32]. There-

fore, we manually annotated these loci using proxies (Table S1B). We then queried the STRING database [<http://string-db.org>, accessed on 11 April 2023] using this list to obtain a predicted protein–protein interaction network for SjD (Figure 1 and Table S2).

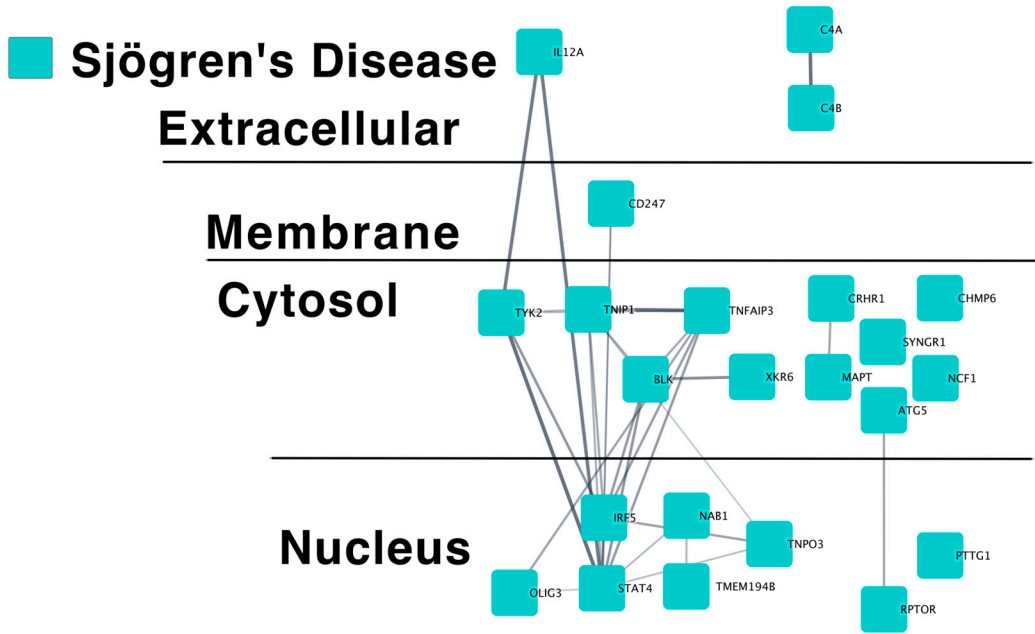


Figure 1. Sjögren’s Disease predicted causal genes for a distinct biological network. The predicted protein–protein interaction network from the STRING database of Sjögren’s Disease GWAS-identified putative causal genes is shown. (<https://www.ndexbio.org/viewer/networks/99b77fa6-478b-11ee-aa50-005056ae23aa>, accessed on 3 June 2023).

Like the published causal gene networks for monogenic and polygenic SLE [33–35] (Tables 1 and S2), the SjD genes are predicted to interact more significantly than would be expected by chance ($p < 5.5 \times 10^{-16}$). The SjD genes also show a relatively high degree of clustering. Taken together, this suggests that these genes regulate the common pathways within a shared molecular network. This finding also supports the annotation of the individual genes as putatively causal.

Table 1. Predicted causal gene network characteristics and explanations from the STRING user documentation [https://version11.string-db.org/help/getting_started/, accessed on 22 May 2023].

Network ¹	Nodes ²	Edges ³	Degree ⁴	Clustering ⁵	Exp. Edges ⁶	p ⁷
Monogenic SLE	54	169	6	0.65	33	1.0×10^{-16}
Polygenic SLE	127	497	8	0.44	107	1.0×10^{-16}
Sjögren’s Disease	23	29	3	0.68	4	5.5×10^{-16}

¹ Putative causal gene network. ² Number of nodes in the network. ³ Number of edges in the network. ⁴ The average node degree is a number of how many interactions (at the score threshold) that a protein has on average in the network. ⁵ The expected number of edges indicates how many edges is to be expected if the nodes were to be selected at random. The clustering coefficient is number between 0 and 1 that measures how connected the nodes in the network are. Highly connected networks have high values. ⁶ The expected number of edges indicates how many edges is to be expected if the nodes were to be selected at random. ⁷ A small PPI enrichment p -value indicates that the nodes are not random and that the observed number of edges is significant.

2.1.2. Pathway Analysis of SjD Putative Causal Genes

We then performed pathway analysis to further characterize the list of putatively causal genes responsible for SjD development. Links to pathway analyses can be found in (Table S2). These analyses further validated the contribution of this gene network to known SjD pathobiology. Specifically, these analyses found pathway enrichment for IFN signaling (both type I and interferon-gamma), transcription by NF- κ B and autoimmune diseases where secondary SjD or overlap is common (systemic sclerosis, SLE, rheumatoid arthritis

and primary biliary cholangitis/primary biliary cirrhosis). One finding of particular interest is the overrepresentation of the IL-35 signaling pathway in this network. This pathway has been found to be dysregulated in several studies of patients with SjD [36–38].

2.1.3. Extensive Overlap between SLE and SjD Putative Causal Gene Networks

To determine the extent of overlap between our SjD risk network and SLE risk genes, we compared the overlap between putative causal genes for SLE [33–35] (Table S2) and SjD (Table S1). This analysis revealed extensive overlap between the predicted causal genes for these two diseases (Figure 2).

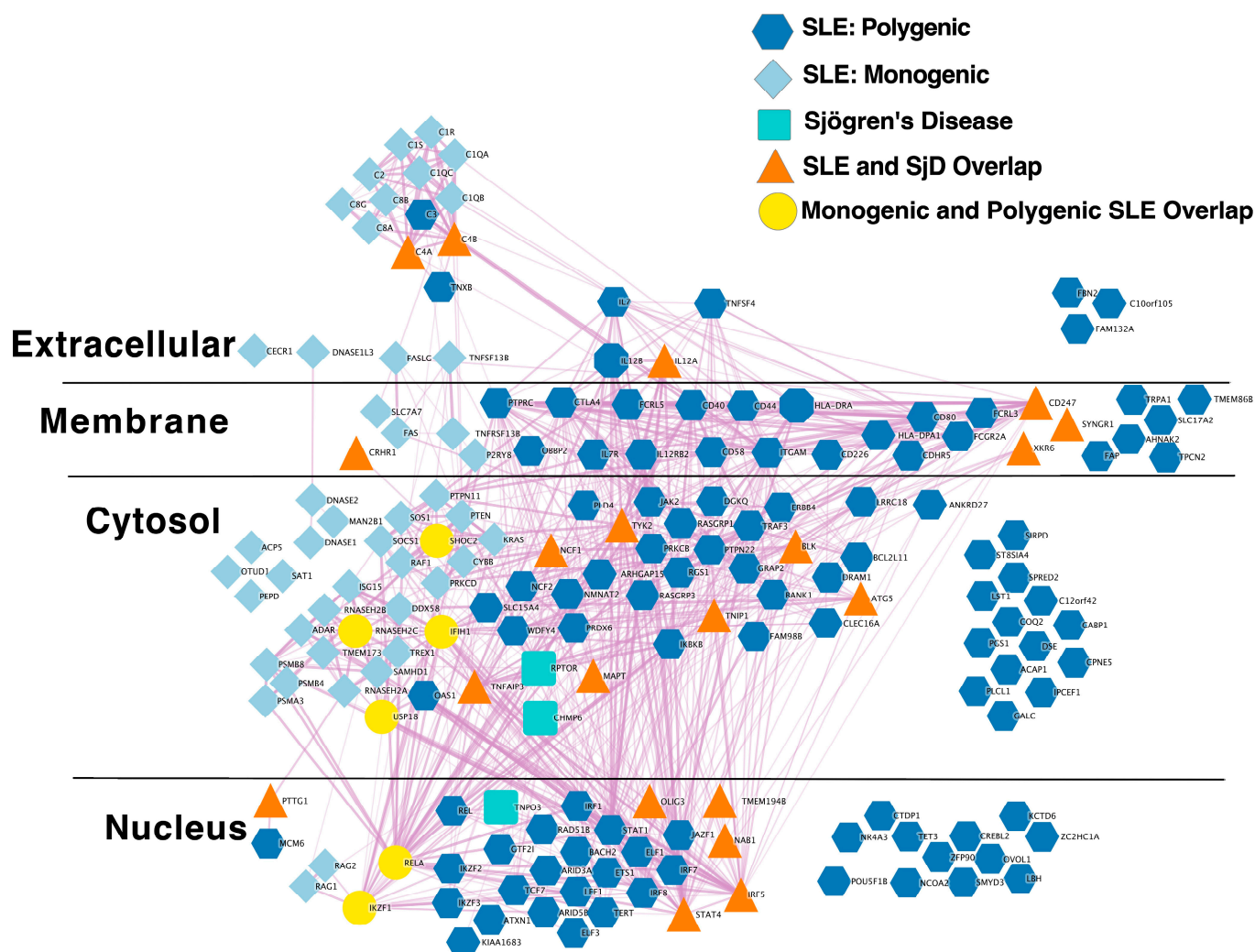


Figure 2. The Sjögren’s Disease putative causal genes network exhibits substantial overlap with that of Systemic Lupus Erythematosus. The predicted protein–protein interaction network from the STRING database of Sjögren’s disease GWAS-identified putative causal genes (teal rounded rectangles) is shown together with those of the polygenic human SLE (dark blue hexagons) and monogenic human SLE (light blue diamonds). Overlaps of SjD with either of the SLE networks are indicated with orange triangles and take priority over overlaps between monogenic and polygenic SLE (yellow circles).

Indeed, this overlap was much greater than expected by chance (Table 2).

Table 2. Overlaps of SLE putative causal gene networks and GWAS loci with SjD.

Network ¹	Exp. Overlap ²	Fold Enriched ³	<i>p</i> ⁴
Monogenic SLE	0.06	46	3.6×10^{-05}
Polygenic SLE	0.15	130	1.1×10^{-41}
All SLE	0.21	96	7.6×10^{-39}
Locus Overlap	0.63	28	1.1×10^{-24}

¹ Either putative causal gene network or genetic risk loci. ² The number of expected overlapping nodes assuming similar length lists were randomly selected from the genome (unassociated). ³ The fold over-represented compared to expectation. ⁴ *p*-value for hypergeometric distribution.

2.1.4. Extensive Overlap between SLE and SjD GWAS Loci

To address the limitations of the putative causal gene network approach (see Section 3), we also calculated the overlap between GWAS loci. We defined a locus as the 500,000 bp window centered on a lead variant from the EBI/NHGRI GWAS catalog, similar to a published approach [39]. We found extensive overlap of SjD GWAS loci with SLE GWAS loci (Tables 2 and S1A). However, examining overlapping loci results in a more moderate estimate of enrichment over expectation as compared to putative causal genes. This may be due to, in part, the way we define regions. This definition results in four SjD-associated regions within the extended MHC locus, of which none are within 500,000 bp of the MHC signal for SLE (Table S1A). A prior study indicates that the major association of MHC with SLE and SjD is the same and is driven by alleles determining the complement factor C4 copy number [13].

In sum, these analyses revealed extensive overlap at both the causal gene and locus level between SjD and SLE.

2.2. Mendelian Randomization Analysis of Variants Associated with Sjögren's Disease and Systemic Lupus Erythematosus

The causal gene overlap, but interval overlap approaches do not necessarily imply sharing of risk alleles. Thus, we also applied an analysis technique that evaluates causal relationship between phenotypes with respect to shared risk alleles, Mendelian randomization (MR).

2.2.1. Analysis 1: The Causal Effect of SjD Variants on SLE in European Ancestry GWAS

We started with lists of genome-wide significant ($p < 5.0 \times 10^{-8}$) SjD risk variants from European ancestry GWAS studies (Table S3A–D). We used the 'TwoSampleMR' R package to analyze the effect of these variants in the summary statistics of a publicly available SLE GWAS in a cohort of persons with European ancestry [40]. Inverse-variance weighted (IVW) MR analysis was the primary inference. We supplemented this with weighted median, weighted mode, and MR Egger analyses. These additional analyses were used to evaluate the robustness of the IVW results. These analyses revealed evidence for a causal relationship between these two phenotypes mediated by these variants or linked variants. Importantly, each of the four MR methods broadly agreed in terms of direction and magnitude of effect size (Table 3). This agreement indicates a stable and robust causal relationship between SjD and SLE in this population.

Table 3. MR results with SjD as exposure and SLE as outcome reveals a causal relationship in persons with European Ancestry.

Method ¹	#SNPs ²	β ³	se ⁴	<i>p</i> ⁵
Inverse-variance weight	12	0.94	0.13	1.4×10^{-13}
Weighted median	12	0.91	0.11	2.4×10^{-15}
Weighted mode	12	1.15	0.12	9.7×10^{-07}
MR Egger	12	1.18	0.27	0.0014

¹ MR analysis method. ² Number of variants used in the analysis. ³ MR effect size, beta. ⁴ standard error of beta. ⁵ *p*-value for causal association.

The forest plot of effect sizes (Figure 3A) shows this consistent effect of SjD on SLE estimated using these risk variants. To ensure that the causal relationship between SjD and SLE was not overly influenced by any single variant, we performed the leave-one-out sensitivity analysis. This analysis did not show that removal of a single variant markedly changes the effect size estimate (Supplementary Figure S1). In the scatter plot of SjD and SLE effect estimates (Figure 3B), the slope of the regression line for each method indicates a consistent estimate of overall effect size across methods (Figure 3B). The plot also revealed an MR Egger regression intercept of -0.072 that does not significantly differ from zero (Figure 3B and Table S4). This finding is consistent with the absence of directional pleiotropy [41] and the InSIDE assumption. Taken together, there is a strong dose-dependent relationship between the genetic associations for SjD and SLE.

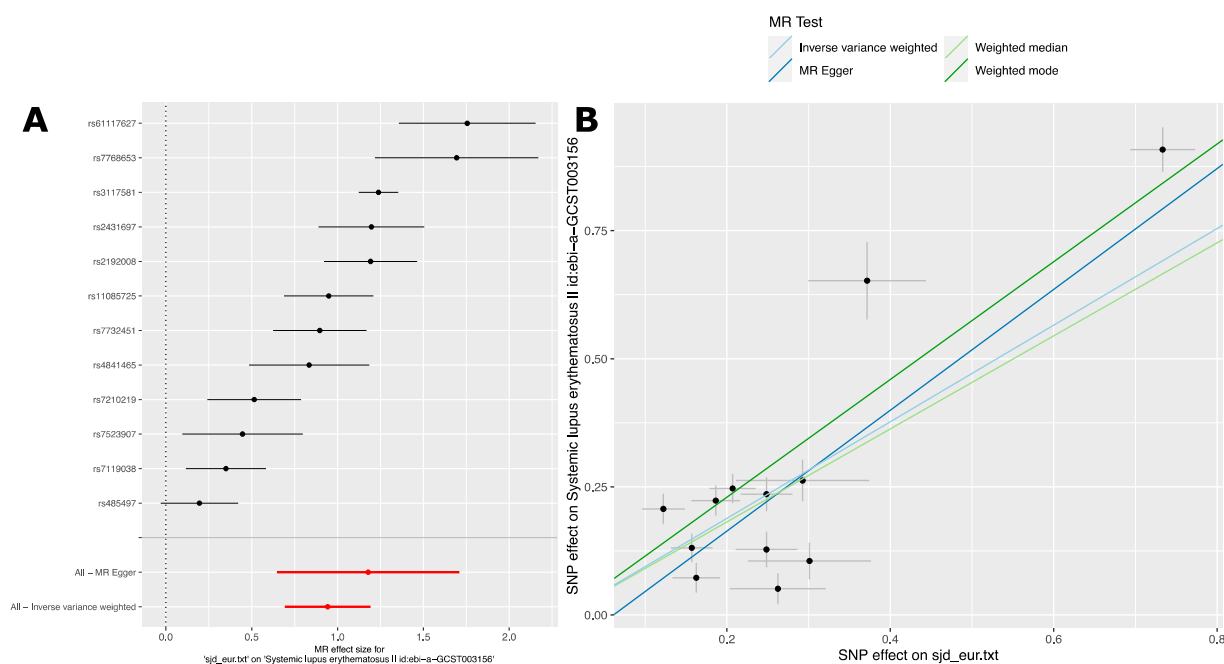


Figure 3. MR analysis reveals a robust and stable causal relationship between Sjögren’s Disease and Systemic Lupus Erythematosus in a GWAS of persons with European ancestry. **(A)** Forest plot of SjD effect estimate on SLE in the Bentham J. et al., GWAS [40] for individual risk variants. MR Egger and inverse-variance weighted overall estimates are also shown. Data are MR effect sizes ($\beta \pm se$). Black dots and black lines indicate MR effect estimates and confidence intervals for individual variants. Red dots and red lines indicate aggregate MR effect estimates and confidence intervals as indicated for MR Egger and Inverse variance weighted analyses. **(B)** scatter plot of MR effect sizes ($\beta \pm se$).

A funnel plot did not reveal marked asymmetry. This is consistent with the absence of effect heterogeneity [42]. However, despite the consistency in estimated MR effect size across variants and methods, there is evidence of significant effect heterogeneity for both IVW and MR Egger analyses (Table S4 and Figure S1). We suspect that this heterogeneity arises from the broad range of estimated odds ratios for the variants selected for both SjD [1.13–2.08] and SLE [1.05–2.48]. Additional results and sensitivity analyses are presented in (Table S4, and Figures S1 and S2).

Taken together, these analyses indicate a causal relationship between SjD and SLE in these cohorts of persons with European ancestry.

2.2.2. Analysis 2: The Causal Effect of SLE Variants on SjD in European Ancestry GWAS

We next performed the reciprocal analysis by using SLE associated variants to evaluate the causal relationship between SLE and SjD. To do this, we used the genome-wide significant ($p < 5.0 \times 10^{-8}$) SLE risk variants from a large GWAS of SLE in a cohort of persons

with European ancestry [40] (see Section 4 for details) (Table S5A). For the outcome phenotype, we used the GWAS summary statistics found by searching for “sjogren syndrome” in the MR-Base platform [43]. These summary statistics are derived from the FinnGen Biobank GWAS (M13_Sjogrens) [44]. MR analyses were carried out as in Section 2.2.3. These analyses revealed evidence for a causal relationship between SjD and SLE mediated by SLE variants as well. Importantly, each of the four MR methods broadly agreed in terms of direction and magnitude of effect size (Table 4). This agreement indicates a stable and robust causal relationship between SjD and SLE in this population.

Table 4. MR results with SLE as exposure and SjD as outcome reveals causal relationship in persons with European Ancestry.

Method ¹	#SNPs ²	β ³	se ⁴	p ⁵
Inverse-variance weight	32	0.31	0.04	1.6×10^{-11}
Weighted median	32	0.28	0.05	2.4×10^{-08}
Weighted mode	32	0.30	0.10	0.0034
MR Egger	32	0.554	0.08	2.7×10^{-07}

¹ MR analysis method. ² Number of variants used in the analysis. ³ MR effect size, beta. ⁴ standard error of beta. ⁵ p -value for causal association.

The forest plot of MR effect sizes (Figure 4A) shows a more moderate effect of SLE variants on SjD. This is consistent with the overall estimates (Table 4). This contrasts with the estimated effect of SjD on SLE (Table 3). Notably, the Finnish population is something of an outlier in comparison to other European populations [45]. This difference in effect sizes may have to do with different LD patterns across populations. It may also be that the SLE on SjD effects are, in fact, less than the effects of SjD on SLE. There was also evidence of effect heterogeneity for both IVW and MR Egger analyses, but less significant than in the reciprocal analysis (Section 2.2.3).

We performed sensitivity analysis as in the reciprocal analysis (Section 2.2.3). The leave-one-out sensitivity analysis does not show that any single variant markedly changes the overall effect size estimate (Figure S3). Like the previous analysis, in the scatter plot of SjD and SLE effect estimates, the slope of the regression line for each method indicates a consistent estimate of overall effect size across methods (Figure 4B). This analysis also revealed an MR Egger regression intercept of -0.089 . This intercept was of a similar size to that seen in the reciprocal analysis (Section 2.2.3). Due to the larger number of SNPs used in this analysis, the standard error of this intercept is smaller, leading to a nominally significant p -value for directional pleiotropy (Table S5B). This finding is consistent with the presence of directional pleiotropy or violation of the Instrument Strength Independent of Direct Effect (InSIDE) assumption [41]. However, the deviation from zero is small in absolute terms. Despite this, the MR Egger-derived effect estimate is significantly positive. While there is evidence of directional pleiotropy, once it is accounted for, a significant causal relationship remains. Taken together, there is a strong dose-dependent relationship that holds for most of the genetic associations for SLE and SjD.

One variant merits separate consideration. The SNP, rs150180633, sticks out in the scatter plot as it does not follow the rest of the variants. The effect direction comparing SjD and SLE is opposite, and the standard error is relatively large (Figure 4B). This variant represents a distant secondary effect in the extended HLA region from the GWAS of SLE performed in a cohort of persons with European ancestry. It is nearly 1 megabase away from the main HLA-region effect, rs389884 [46]. Thus, due to this long distance, this variant may be particularly susceptible to differences in LD pattern between populations. Importantly, this variant is also uncommon in the Finnish population (minor allele frequency, $MAF < 1\%$). Additional supportive data and sensitivity analyses are presented in (Table S5, and Figures S3 and S4). Consistent with the foregoing discussion, removal of this variant led to larger causal effect estimates of SLE for SjD, more significant association and decreased heterogeneity.

Taken together, these results indicate a causal relationship between SjD and SLE in these cohorts of persons with European ancestry.

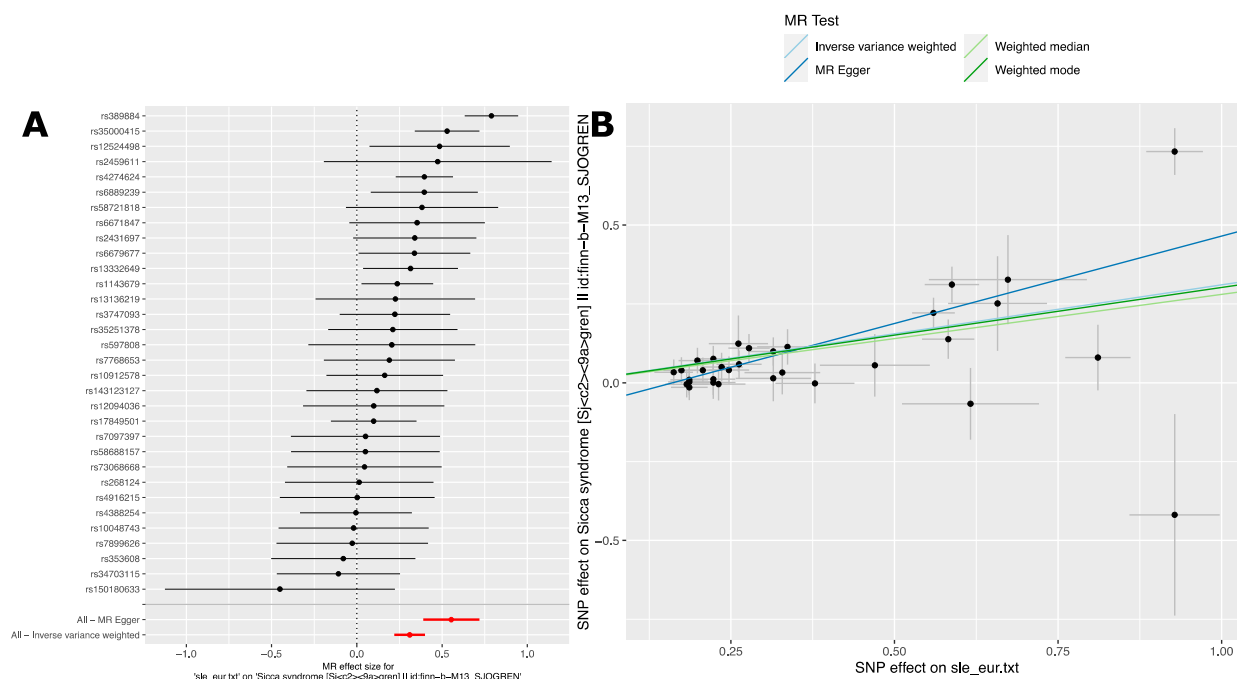


Figure 4. MR analysis reveals a robust and stable causal relationship between Systemic Lupus Erythematosus and Sjögren’s Disease in a GWAS of persons with European ancestry. (A) Forest plot of effect estimate of SLE on SjD (M13_Sjogrens) in the FinnGen GWAS [44] for individual risk variants. MR Egger and inverse-variance weighted overall estimates are also shown. Data shown are MR effect sizes ($\beta \pm se$). Black dots and black lines indicate MR effect estimates and confidence intervals for individual variants. Red dots and red lines indicate aggregate MR effect estimates and confidence intervals as indicated for MR Egger and Inverse variance weighted analyses. (B) scatter plot of MR effect sizes ($\beta \pm se$).

2.2.3. Analysis 3: The Causal Effect of SjD Variants on SLE in East Asian Ancestry GWAS

The results in Sections 2.2.3 and 2.2.4 point toward a causal relationship between SjD and SLE mediated in European populations. The SjD GWAS studies published to date in East Asian ancestry population have smaller sample sizes than European ancestry SjD GWAS studies. As a result, the associated variants available for MR analysis are fewer and have less precise estimates of effect size (Table S3B). In other words, for any given genetic association, smaller sample size results in a larger standard error. In addition to smaller sample size, the full, genome-wide summary statistics from East Asian ancestry GWAS of SjD have not been made publicly available. This precludes examination of SLE effects in SjD GWAS studies of persons with East Asian ancestry. Nonetheless, we wondered whether the causal relationship between SjD and SLE that we identified could be independently validated. We started with the list of genome-wide significant ($p < 5.0 \times 10^{-8}$) SjD risk variants reported within the manuscripts reporting several East Asian ancestry SjD GWAS studies (Table S3B). We again used the ‘TwoSampleMR’ R package to analyze the effect of SjD on SLE using the summary statistics of a large, publicly available SLE GWAS in a cohort of persons with East Asian ancestry [47]. MR analysis parameters were the same as above. The analyses revealed inconsistent evidence for a causal relationship between these two phenotypes estimated by these variants. Importantly, IVW and the outlier selection and removal-based methods (weighted median and weighted mode) agreed broadly in terms of direction and magnitude of effect size (Table 5). However, MR Egger did not agree with the other methods. MR Egger resulted in a highly imprecise estimate of the MR effect size that could not be distinguished from the null (Figure 5A). Therefore, this result

was not significant. The purpose of MR Egger is to serve as a kind of sensitivity analysis that was developed to model directional pleiotropy [48]. It applies Egger regression to the estimation of the MR effect size. Egger regression was originally developed to detect small study bias in meta-analysis. In the MR scenario, it detects bias toward either high or low values from weaker genetic variants. To do so, it performs regression analysis of the variant effects between the exposure and outcome phenotypes. In our case, this is depicted in the scatter plots (Figure 5B). Thus, an MR Egger intercept term that significantly differs from 0 provides evidence of directional pleiotropy. Like the non-significant MR Egger estimate (Table 5), the MR Egger intercept could not be distinguished from 0, as the standard error was larger than the estimate by a considerable margin (Table S6). Visual inspection of the scatterplot provides an explanation for this. First, there are few variants ($n = 4$) to fit a regression line, as evident from the results of few SJD GWAS studies having been performed with a large sample size. Thus, the regression is more likely to be influenced by outliers or extreme values. Second, outliers are present: both the weighted median and weighted mode MR methods provide lower estimates of MR effect size than IVW. Therefore, the IVW estimate is being biased upwards by outliers. MR Egger tests for a dose–response relationship between exposure effect size and outcome effect size. The precision of the MR Egger estimate relies both on the proportion of variance in the exposure explained by the genetic variants and on the variability between genetic associations with both the exposure and the outcome [41]. For these four variants, the proportion of the variance of SJD explained by each is small since they are the first loci identified as risk loci in cohorts of persons with East Asian ancestry. Likewise, there is considerable variability in these associations both with the exposure and the outcome. Thus, both the estimated MR effect and the estimated intercept produced via MR Egger for these variants are highly imprecise. Thus, the MR Egger analysis indicates either the presence of pleiotropy or a causal effect. Together with the consistent estimates of a causal effect via the IVW, weighted median and weighted mode methods, a causal effect seems more likely. Additional supportive data and sensitivity analyses are presented in (Table S6, and Figures S5 and S6).

Table 5. MR results with SJD as exposure and SLE as outcome reveals either a causal relationship in persons with East Asian Ancestry or pleiotropy.

Method ¹	#SNPs ²	β ³	se ⁴	p ⁵
Inverse-variance weight	4	0.55	0.19	0.004
Weighted median	4	0.38	0.05	2.5×10^{-12}
Weighted mode	4	0.35	0.05	0.003
MR Egger	4	−0.45	0.89	0.911

¹ MR analysis method. ² Number of variants used in the analysis. ³ MR effect size, beta. ⁴ standard error of beta. ⁵ p -value for causal association.

Taken together, our analyses suggest a causal relationship between SJD and SLE in these cohorts of persons with East Asian ancestry. However, a scenario where each trait is independently influenced by the same genetic factor cannot be excluded. This scenario is termed as horizontal pleiotropy.

2.2.4. Analysis 4: The Causal Effect of SLE Variants from East Asian Ancestry GWAS on SJD in European Ancestry GWAS

Next, we explored whether the causal relationship we found between SLE and SJD in European ancestry populations was more generalizable. To do so, we evaluated the causal relationship between SLE and SJD using GWAS of SLE in persons with East Asian ancestry [47] and a GWAS of SJD in persons of European ancestry [44]. We again restricted the variants in question to those that passed the genome-wide significance ($p < 5.0 \times 10^{-8}$) in the exposure GWAS of SLE. We also applied clumping and pruning to remove LD between variants using the ‘TwoSampleMR’ R package. Analysis parameters were the same as the other analyses. This analysis revealed consistent evidence for a causal relationship between these two phenotypes mediated by these variants (Table 6 and Figure 6). In single variant

analysis, several of the SLE risk variants (rs16870693, rs3800387, rs55701306 and rs16869875) show consistent evidence against a causal relationship between SLE and SjD (Table S8). Notably, each of these variants is uncommon in the European ancestry population examined. The minor allele frequencies (MAFs) are all less than 3.5%. In contrast, the MAF for these variants in East Asian ancestry samples (EAS) from the 1000 Genomes Project ranges from 11.41 to 42.76% [49,50]. This suggests that these variants may be ill-suited to detect casual relationships across populations. That is, they may be particularly susceptible to differences in local LD structure varying across populations. As a result, the relevance condition for instrumental variables may not hold. Despite the challenge with comparing causal effects across populations, the general agreement of the four analyses presented in this manuscript argues for a stable and robust causal relationship between SLE and SjD. Additional supportive data and sensitivity analyses are presented in (Tables S7 and S8, and Figures S7 and S8).

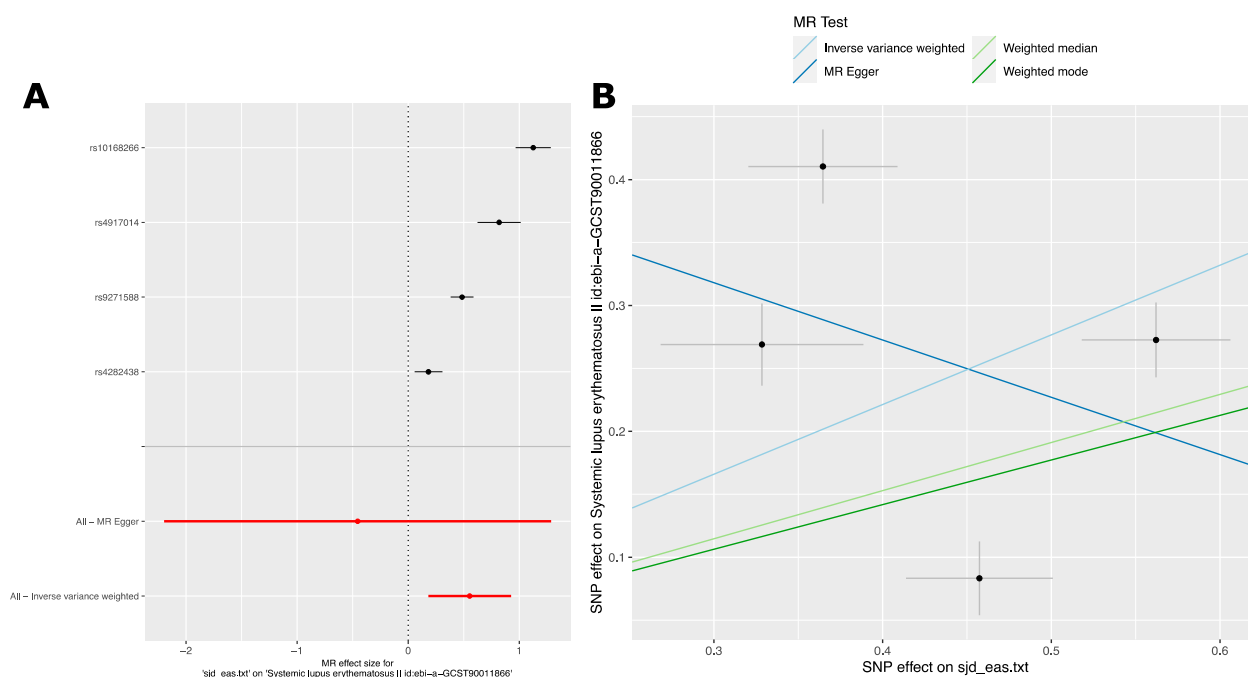


Figure 5. MR analysis reveals a causal relationship or pleiotropy between Sjögren’s Disease and Systemic Lupus Erythematosus in a GWAS of persons with East Asian ancestry. **(A)** Forest plot of SjD on SLE in the Wang, Y.F. et al., GWAS [47] for individual risk variants. MR Egger and Inverse-variance weighted overall estimates are also shown. Data shown are MR effect sizes (beta ± se). Black dots and black lines indicate MR effect estimates and confidence intervals for individual variants. Red dots and red lines indicate aggregate MR effect estimates and confidence intervals as indicated for MR Egger and Inverse variance weighted analyses. **(B)** scatter plot of MR effect sizes (beta) ± standard error (se).

Table 6. MR results with SLE in persons with East Asian Ancestry as exposure and SjD in persons with European ancestry as outcome reveals a causal relationship.

Method ¹	#SNPs ²	β ³	se ⁴	p ⁵
Inverse-variance weight	30	0.29	0.06	3.5×10^{-05}
Weighted median	30	0.23	0.06	5.6×10^{-06}
Weighted mode	30	0.23	0.08	0.006
MR Egger	30	0.42	0.17	0.019

¹ MR analysis method. ² Number of variants used in the analysis. ³ MR effect size, beta. ⁴ standard error of beta.

⁵ p-value for causal association.

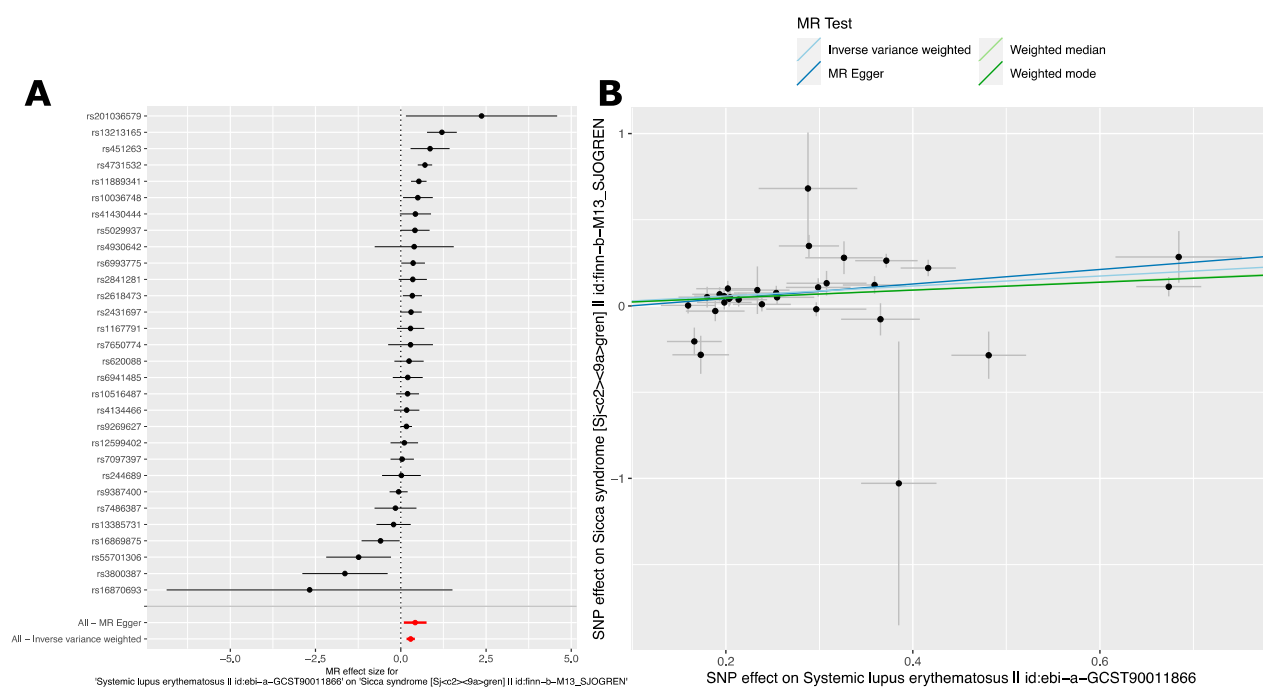


Figure 6. MR analysis reveals a causal relationship between Systemic Lupus Erythematosus in persons of East Asian ancestry and Sjögren's Disease in GWAS of persons with European ancestry. (A) Forest plot of SLE from the Wang, Y.F. et al. GWAS [47] on SjD in FinnGen [44] for individual risk variants. MR Egger and Inverse-variance weighted overall estimates are also shown. Data shown are MR effect sizes (beta ± se). Black dots and black lines indicate MR effect estimates and confidence intervals for individual variants. Red dots and red lines indicate aggregate MR effect estimates and confidence intervals as indicated for MR Egger and Inverse variance weighted analyses. (B) scatter plot of MR effect sizes (beta) ± standard error (se) across the two conditions/populations.

Taken together, our analyses suggest a causal relationship between SLE in persons with East Asian ancestry and SjD in persons with European ancestry.

3. Discussion

The causal gene network and gene interval results provide strong support of a causal relationship between SjD and SLE mediated by shared genetic risk factors. Prior studies investigating the clinical, serologic, immunologic, and molecular overlap between these two conditions also support this hypothesis. Thus, our conclusion may not be surprising. However, the degree of overlap perhaps is. Only three predicted causal genes do not overlap between SLE and SjD (Figure 2). One of the genes, *TNPO3*, is adjacent to the shared gene, *IRF5*. While this region is shared, this gene was not algorithmically assigned as a predicted causal gene for SLE. Notably, a prior study indicated that there are multiple independent genetic associations with SLE in this region [51]. The other two genes that are not shared reside in the same SjD-associated locus on chromosome 17. We did include the variants at this locus in the SjD for SLE MR analysis. However, the variants at this locus were excluded by the algorithm in the 'TwoSampleMR' R package from the final analysis due to allele harmonization issues. Full analysis parameters and logs can be found at the link in the Data Availability section. This SjD-associated locus has not been reported as being associated with SLE and encodes the genes *RPTOR* and *CHMP6*. Understanding how this risk locus confers risk for SjD will be important to comprehend how the etiology of SjD is distinct from that of SLE.

The Mendelian randomization results also support a strong causal link between these two disease phenotypes mediated by causal genetic variants. Two observations strengthen this causal link. First, we observed that this relationship is reciprocal: SjD is causal for SLE and vice versa. Second, this relationship is evident in both European and East Asian

ancestry populations with caveats (discussed below). Despite these interpretive caveats, the causal link between SjD and SLE revealed via MR is both stable and robust.

One general limitation of our study is that the common genetic contribution to SLE has been more extensively studied than that of SjD. Thus, there are more putative causal genes, associated intervals, and variants for SLE than for SjD. This could cause us to have an under or overestimation of the true extent of overlap between SLE and SjD, depending on what future GWAS of SjD reveal. In addition, there are limitations of the putative causal gene and gene interval approaches. First, two drawbacks of the putative causal gene network approach should be noted. One is that the causal gene at many risk loci cannot be confidently assigned. If these loci are shared, this could cause the overlapping causal gene network to underestimate the true extent of overlap. Another is that as the number of potential overlaps in a population gets larger, so does the number of expected overlaps. Applied to our situation, since 55 SLE risk loci cannot be confidently assigned a putative causal gene, a gene-based approach likely over-estimates the true extent of overlap. As for limitations of MR, one limitation is that there are only five genetic risk variants identified to date for SjD in East Asian ancestry populations. Of these, four remained after LD pruning (Table 5 and Figure 5). As a result, MR Egger cannot distinguish between true causal association and directional pleiotropy. The second caveat is that we have not been able to test whether SLE is directly causal for SjD in GWAS from East Asian ancestry populations. This is because genome-wide summary statistics for SjD GWAS studies in East Asian ancestry cohorts are not available within the MR-Base repository. Instead, we tested whether SLE risk variants identified in a cohort of persons with East Asian ancestry were causally linked to SjD in a cohort of persons with European (Finnish) ancestry (Figure 6). Despite the application of MR across ancestries, our results generally reinforce MR analysis of SjD and SLE in European ancestry populations. Importantly, our analyses of causal links between SjD and SLE used complementary approaches (overlap of putative causal gene networks, intervals and MR analysis), each of which addressed the limitations of the others. Also important, they all led to the same broad conclusion: genetic risk factors are extensively shared between SjD and SLE.

There are several interesting questions regarding the causal relationship between SjD and SLE. Are the effect sizes the same? The MR analyses generate a wide range of causal estimates. Most fell within the range [0.23–1.18]. In both analyses evaluating the impact of SLE variants on SjD risk, the estimated effect sizes in SjD were lower [0.23–0.55] (Tables 4 and 6). These analyses suggest that the causal effect of SjD variants on SLE is larger. However, several points should be considered in their interpretation. First, the number of GWAS-identified variants is smaller for SjD as compared to SLE. Second, in both instances the comparison GWAS for SjD used to determine the effect of SLE on SjD is derived from the FinnGen study. Amongst European ancestry populations, the Finnish population does show some inter-population differences [45]. Third, the FinnGen study ascertains disease status based on electronic health record (EHR) diagnosis codes. The cohort of participants selected via these codes could include persons who do not, in fact, meet classification criteria for SjD. If so, we would expect smaller effect size estimates for SjD in the FinnGen population relative to a cohort composed exclusively of SjD patients who meet the classification criteria. Other potential explanations include the impact of factors (genetic, environmental, or stochastic immune response), or the so-called “Winner’s Curse”. That is, by nature of their discovery, disease-associated variants tend to systematically overestimate the effect size on a given phenotype relative to the true effect size in a population [52,53]. Thus, several factors may contribute to a downward bias of SjD effect size estimates. Future investigation will be necessary to resolve these possibilities.

Another question is related to the portion of genetic risk for these diseases that is shared. Is only a core network shared between SjD and SLE or is there a broader sharing of causal risk variants? The genomic risk variant footprint for most common human genetic phenotypes comprises two parts. The first part is composed of risk variants identified in initial GWAS scans. These variants constitute a core network of disease risk alleles. The

second part is composed of the aggregate effects of thousands to millions of risk alleles spanning the entire genome. Each of these variants add or subtract a small amount of risk [54]. In some instances, gene–gene interactions will likely synergistically contribute to this risk. Our study, by necessity, can only evaluate GWAS studies where there is publicly available data. For SjD, this results in a relatively small number of risk loci. Thus, we may have only addressed the core genetic risk network for SjD. Two methods that may be able to address this, at least in part, are genetic correlation and polygenic risk score calculation. These methods do not only evaluate the effect of the most strongly associated variants but can also include the contribution of risk variants across the entire genome. The SjD GWAS by Khatri et al. [29] did compare genetic correlation, r_g , between GWAS of SLE and SjD in cohorts of persons with European ancestry. This metric calculates the correlation between effect size of variants spanning the entire genome and not just those that are strongly associated with disease. Their analysis found a high degree of genomic correlation between GWAS for these two phenotypes as compared to other disease phenotypes. Likewise, some approaches to polygenic risk scores (PRS) calculation use the entire genome or include a broader subset of weakly associated (i.e., $p < 1.0 \times 10^{-5}$) variants. Calculating cross-disease PRS would be another approach to address shared causality between phenotypes. There are tools that allow for the derivation of PRS from summary statistics [55]. However, to our knowledge, current tools do not allow for the estimation of individual-level PRS based on summary statistics. This question warrants further study with individual level genotype data.

A related question also arises from the above consideration: is there a subtype or subtypes of SjD and SLE that exhibit a greater degree of sharing? For example, do anti-Ro/SS-A+ or anti-La/SS-B+ SLE patients exhibit a greater degree of shared genetic risk to SjD patients that exhibit these seropositivities? It may be that loss of immune tolerance to self within these autoantibody systems is under particular genetic control. Indeed, this appears to be the case for anti-Sm autoimmunity, which is restrained by CD72 under normal circumstances [56]. If this is generally true, then it may be that sharing of genetic risk between SjD and SLE is more pronounced in genes that also affect this shared pathway. Studies examining this question will require GWAS of large sample sizes to compensate for the loss of power with substratification of these phenotypes. A related question is how do other phenotypes associated with SjD compare in terms of causal risk variant sharing? Secondary SjD is found with several other autoimmune disease phenotypes, including primary biliary cholangitis, rheumatoid arthritis, systemic sclerosis, and autoimmune thyroid disease. Future studies examining the causal links between SjD and these diseases will be of interest.

The proponents of precision or personalized medicine envision a world where treatment of disease is personalized to the characteristic features of an individual person's disease process. The hope is that with knowledge of the human genome and the advent of other omics technologies, healthcare practitioners will be able to precisely stratify diseases to smaller, molecularly defined subsets. Of course, delivering the correct therapy to the correct patient at the correct time is a worthy improvement goal for all of healthcare. However, a reductionistic focus on precision, allowed by increasingly precise technical advancements could carry costs in terms of accuracy of disease categorization. The history of revising disease understanding when old paradigms were overturned by new technologies is instructive [57]. Whereas dropsy was once considered a single disease process, most medical students can facilely review a differential diagnosis of common causes for this condition (heart, kidneys, liver, etc.). This understanding of dropsy, comprising distinct entities, is rooted in a mechanistic understanding of how hemodynamics and protein production impact oncotic pressure [58,59]. Similarly, advances leading to the germ theory of disease and molecular genetics led to the reclassification of diseases. For example, diverse clinical presentations including variable presentation of a honey-colored skin rash, inflammatory arthritis, frothy urine, sore throat and chorea could be reclassified as being related to group A streptococcal infection [60]. In the same vein, Mendelian genetic syndromes could be

reclassified as related based on action within a defined biochemical pathway. At first glance, abnormally dark urine, liver failure, relative absence of skin pigmentation and intellectual disability are not necessarily related symptoms. However, each of these corresponds to a genetic syndrome caused by loss of function in phenylalanine metabolism (homogentisic acid oxidase, fumarylacetoacetate hydrolase, tyrosinase, and phenylalanine hydroxylase, respectively) [61–64]. Thus, prior paradigm shifts leading to the reclassification of disease entities improved our understanding and treatment of disease in both directions. Lumping similar entities based on shared mechanisms and splitting based on new mechanism-defined subphenotypes both led to improved disease understanding.

Our findings are consistent with a developing notion that targeting common pathways shared across autoimmune diseases may lead to improved therapeutic responses. One example is the recent application of low-dose IL-2 to expand T regulatory cells across multiple autoimmune and inflammatory disease states [65]. This trial of persons diagnosed with 1 of 11 autoimmune was geared towards demonstrating the safety of this approach in terms of effector T cell activation. However, these data also demonstrated significant improvement in aggregate measures of disease activity from baseline. There is an increasingly precise understanding of the shared pathobiology underlying the clinical overlap between SjD and SLE. Accurately understanding the shared pathobiology of these diseases presents an opportunity to advance therapeutically relevant understanding of both. Since SjD and SLE are diseases with polygenic contributions to disease risk, we wondered what degree of shared genetic risk might be present.

Finally, our findings have implications for future studies of SLE and SjD. They suggest that application of meta-analysis methods to jointly model shared phenotypes may be a way to increase power to detect genetic variants in both SjD and SLE [66–69]. Combining summary statistics from individual disease GWAS or using individual level genotype data to substratify individuals based on shared phenotype (i.e., combining persons with primary SjD together with persons with SLE and secondary SjD) will increase the effective sample size. In this way, discovery of genetic risk factors for SjD may be accelerated. There are also potential implications impacting clinical care. Most trials in SjD (and SLE for that matter) have not met their primary endpoint. Indeed, it is noteworthy when a trial of therapy for SjD does meet its primary endpoint [70–72]. As we have noted, there is clinical and pathobiologic similarity between SjD and other diseases, like SLE and RA. Similarly, some persons are affected by SjD and another autoimmune disease, like SLE or RA (secondary Sjögren's Disease). Because of this overlap, clinical treatment guidelines of constitutional and non-sicca organ-based manifestations of SjD rely on extrapolation from diseases like SLE and RA [73,74]. Our finding that SLE and SjD are causally linked add credence to this clinical practice. Future studies to better define shared pathways between SjD and other diseases outlined above may further inform therapeutic selection for these manifestations in SjD.

4. Materials and Methods

4.1. Overview of Analysis Scheme

A schematic of our analysis is presented below (Figure 7).

4.2. Determining the Overlap of GWAS-Associated Loci in SLE and SjD

4.2.1. Defining SLE and SjD Loci

We applied a previously described approach [30,75] to define disease-associated genetic loci that have been identified in published genome-wide genetic association scan (GWAS) with Sjögren's Disease. To do this, we first searched the EBI/NHGRI GWAS catalog [76] using search term "Sjogren syndrome" for all variants associated with Sjögren's syndrome $p < 5.0 \times 10^{-8}$. We also included a novel loci from a recently published GWAS of SjD [29] that has not yet been incorporated into the GWAS catalog. A locus was defined as any sequence of adjacent associated variants within 250,000 base pairs. We used the list

of SLE-associated loci from a previously published paper that was derived using the same method [30]. The SjD and corresponding SjD association intervals are listed in (Table S1A).

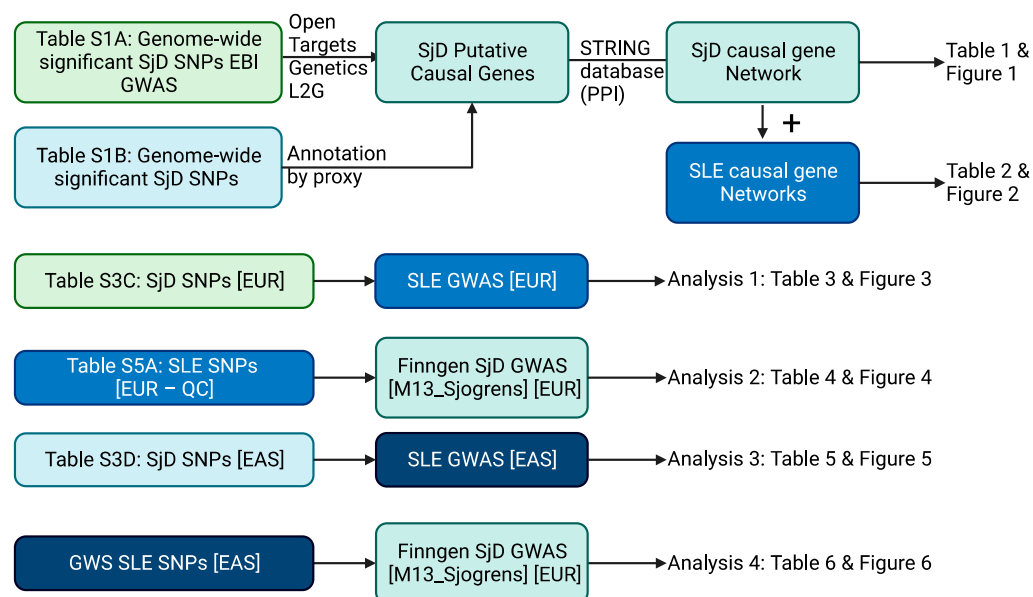


Figure 7. Analysis scheme. [EUR] indicates European ancestry GWAS. [EAS] indicates East Asian ancestry GWAS. M13_Sjogren's indicates the phenotype code from FinnGen used to define Sjögren's for that GWAS. [EUR–QC] indicates genome-wide significant variants after removal of variants likely violating the relevance condition as described in the methods. Created with [BioRender.com](https://www.biorender.com) (accessed on 6 June 2023).

4.2.2. Pathway Analysis of Putatively Causal SjD Risk Genes

We performed pathway analysis on the list of putatively causal genes for SjD using the STRING database [77], Enrichr [78], ToppGene [79] and the Network Data Exchange at [NDExBio.org](https://ndexbio.org) (accessed on 6 June 2023) gene set enrichment analysis tools [80]. The links to these queries can be found in (Table S2).

4.2.3. Calculating Locus-Level Overrepresentation between SLE and SjD

Using the two lists of disease-associated loci defined in this way, we then calculated whether the overlaps between these sets of loci in terms of genomic interval was greater than expected by chance. To do this, we wrote a python script that counts the number of overlapping loci and between these diseases (https://github.com/harleyi/sjd_sle_overlap, accessed on 24 May 2023). We then used these overlaps to calculate the fold overrepresentation between these two lists across the genome. The total number of base pairs in the genome was derived from the genome reference consortium, Human Genome Assembly GRCh38.p14 (<https://www.ncbi.nlm.nih.gov/grc/human/data>, accessed on 3 June 2023). This indicates that there are 3,298,912,062 bases in the human genome. When it is divided by 500,000 bp blocks, this yields 6598 such loci. We then used the hypergeometric calculator from the Graeber lab (<https://systems.crump.ucla.edu/hypergeometric/index.php>, accessed on 3 June 2023) to calculate the fold overrepresentation of these genomic intervals.

4.3. Predicted Causal Polygenic Risk Gene Annotation, Network Construction and Overlap

4.3.1. Calculating Locus-Level Overrepresentation between SLE and SjD

To build the list of predicted causal SjD risk genes, we applied the same approach that we previously used with modification [30]. Once the list of genomic loci was defined, we annotated this list of SjD-associated loci using the output form of the Locus2Gene (L2G) algorithm [31] from Open Targets Genetics [32] as previously described. However, two loci were clearly misannotated using this approach based on published data, NCF1 [81] and

C4A, C4B [13]. Therefore, they were manually reassigned. In addition, one large GWAS study of Sjögren's Disease, that of Khatri et al. [29], has not been incorporated into either of these data sources. Therefore, we manually reviewed the unannotated variants from this study. We then assigned a predicted causal gene for each locus. To do so, we used a combination of manual annotation from the literature and selected a proxy from the list of SLE loci. We then assigned the predicted causal gene as the likely causal gene for SjD. Details of proxy selection are in (Table S1B). Detailed instructions can be found in the Supplementary Materials (Supplementary Note). These instructions include step-by-step instructions on navigating the EBI GWAS Catalog [76]. They also include links to: resources to interpret EBI GWAS Catalog data tables [82]; documentation for the L2G algorithm [83]; and documentation for the open targets genetics database [84].

4.3.2. Building the Predicted Causal Gene Network for SjD

To build the predicted protein–protein interaction network of causal SjD risk genes, we used the STRING database version 11.5 (<http://string-db.org>, accessed on 22 May 2023) [77]. We entered the list of SjD-predicted causal genes, ensured correct identifier mapping and then exported this network to Cytoscape v3.10.0 for further annotation and analysis. Default settings were used to generate the predicted protein–protein interaction network. The predicted causal genes for SjD can be found in Table S1A. Detailed instructions can be found in the Supplementary Materials (Supplementary Note). They include links to resources to interpret STRING networks [85] and various tutorials and guides for navigating and using Cytoscape [86–91].

4.3.3. Calculating Overlap

As above (Section 4.2.3), the Graeber lab hypergeometric calculator was used to calculate the degree of overlap in terms of risk genes with the following parameters: 19 genes overlapping networks, 23 predicted causal SjD genes, 131 predicted causal SLE genes and 19,303 genes in the STRING v11.5 network for *Homo sapiens* (https://string-db.org/organism_overview.html, accessed on 22 May 2023). Detailed instructions can be found in the Supplementary Materials (Supplementary Note) and includes a link to the Graeber lab hypergeometric calculator [92].

4.4. Mendelian Randomization Analysis

4.4.1. Overview of Mendelian Randomization and Application of Instrumental Variable Analysis to Genetic Studies

Mendelian randomization (MR) is a causal inference method that has been developed to estimate and test causal effects between phenotypes [93]. MR applies instrumental variable analysis techniques, using genetic variants as the instrumental variables. In the example of two-sample MR, causality between an exposure phenotype and an outcome phenotype is assessed by evaluating the effect of the variants associated with the exposure phenotype on the outcome phenotype. Mendel's laws of segregation and independent assortment ensure that non-linked variants (that is, independent variants not in linkage disequilibrium with one another) are a source variation potentially impacting the outcome phenotype. Inheritance of independent variants is a source of random variation unrelated to confounding factors. Furthermore, germline genetic variants are not susceptible to issues with reverse causality [94,95]. Therefore, the effect of variants on the exposure and the outcome phenotypes can be used to determine causality of the exposure on the outcome if certain conditions are met (Figure 8).

The main three conditions of MR are those of (1) relevance, (2) independence, and (3) exclusion restriction [96,97]. The relevance condition requires that a variant needs to be robustly associated with the exposure of interest. When MR is applied to GWAS studies, this often results in limiting variants under examination to those that pass the commonly accepted threshold for genome-wide significance, $p < 5.0 \times 10^{-8}$. The independence condition holds that variants are not associated with confounding factors impacting the

association between the exposure phenotype and the outcome phenotype. The exclusion restriction condition holds that each variant is associated with the outcome only through the exposure of interest. In sum, MR is a powerful method to determine causal relationships between phenotypes that rely on specific conditions to be met regarding the relationship of genetic variants, confounding factors, and exposure/outcome phenotypes.

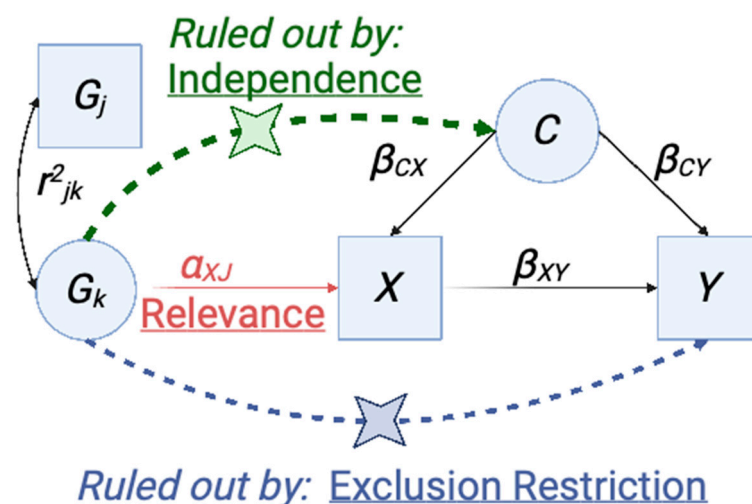


Figure 8. Mendelian randomization causal graph and conditions. In our study, the most likely scenario is that the chosen variant, G_k , is in linkage disequilibrium with a causal variant, G_j . In the event that this LD is perfect ($r^2 = 1$) and the conditions of MR are met, then the association between G_j and the outcome is fully mediated by the exposure. Therefore, the ratio of marginal effects of G_j on X and G_j on Y is equal to β_{xy} . Effects of the variant G_j through any confounder, C , or on Y are both ruled out through the Independence and Exclusion Restriction conditions. The relevance condition is that G_j (through G_k in this case) is truly associated with the effect, X . In situations of comparison across populations with different fine-scale LD patterns, the relevance assumption is likely to not be met. [CreatedwithBioRender.com](https://www.createdwithbio.com/) (accessed on 6 June 2023).

The putative causal gene-level overlap between SjD and SLE suggests that both phenotypes are mediated by genetic factors that impact the same unmeasured, intermediate phenotype. We posit that persistent loss of tolerance to nuclear autoantigens is a likely intermediate. With this framework, which one of the instrumental variable conditions is most important to attend to in our study of whether SjD and SLE are causally related? We have ensured that condition 1 (relevance) is met by setting a threshold for robust association of variants. Condition 2 (independence) seems unlikely as this would invoke a confounding factor that impacts both the genotype at these variants and the outcome phenotype. In the Introduction, we briefly reviewed the extensive body of literature supporting shared phenotypic, clinical, and molecular features of SjD and SLE. Considering these data, violation of the independence condition, seems less plausible. If independence were violated, a genetic risk factor impacts a confounding phenotype, C , that in turn impacts both SjD and SLE by invoking them. While the extensive phenotypic sharing makes violation of the independence condition somewhat less likely, correlated pleiotropy remains possible in any MR analysis evaluating causal relationships between phenotypes [98]. Thus, of these three conditions, the most relevant to our study is that of exclusion restriction. When this condition is violated, the genetic variant impacts the outcome phenotype independently of the exposure phenotype or the unmeasured phenotype that causes it.

Importantly, there are several MR analysis methods that relax the exclusion restriction condition/assumption [93,97]. The basic approach to MR is similar to inverse-variance weighted meta-analysis. It involves weighting the ratio of the effect size in the exposure and the outcome by their associated uncertainty (inverse-variance weighted). This approach is valid under the assumption of balanced pleiotropy, where pleiotropic effects are equally as likely to be positive as negative. One method that allows for balanced pleiotropy in

summary statistic MR analyses is the weighted median approach that selects and removes outliers. This approach assumes that the majority of the variants tested are valid [99]. Another approach, weighted mode, also selects and removes outliers. This approach remains valid even if a majority of the variants are valid but is overly conservative and sensitive to the addition/removal of variants [100]. Importantly, the weighted mode allows for directional pleiotropy. Another method that allows for directional pleiotropy is MR Egger, which also puts constraints on the pleiotropy. MR Egger assumes that the magnitude of pleiotropy across all variants is unrelated to the strength of association between the variant and the phenotype of interest. This assumption is termed the Instrument Strength Independent of Direct Effect or InSIDE assumption. The InSIDE assumption requires that the association between the genetic variant and the exposure is not correlated with a path from the variant to the outcome that is independent of the exposure of interest. When this assumption is violated and unbalanced pleiotropy is present, this will bias the MR Egger estimate of causal effect. This will occur when genetic variants influence the outcome via pleiotropy or through confounder impact on both the exposure and outcome [41,101]. While the outlier selection and removal methods attempt to limit the degree of heterogeneity between variants, the premise of MR Egger is for a set of relevant variants; the causal effects average out to a reasonable estimate of the true causal effect between the exposure and outcome phenotypes [41]. We refer the interested reader to five recent publications further explaining these methods [93,96,97,99].

4.4.2. Evaluating the Causal Relationship between SjD and SLE

To evaluate if SjD and SLE are causally related through individual causal risk variants, we performed Mendelian randomization analysis. Because each of the four methods discussed above (Section 4.2.1) has different sensitivity and conditions, we applied all of them separately. With genetic variant data, a common mode of violation of instrumental variable conditions/assumptions is that the variant in question is not causal but is in linkage disequilibrium with a causal variant. The fine-scale linkage disequilibrium patterns differ between continental ancestral populations [49]. Therefore, wherever possible, we restricted our initial analyses to comparisons between SjD and SLE GWAS studies carried out within the same continental ancestral population. Specifically, we performed trans-ancestral MR in Analysis 4. Several GWAS of SjD have been carried out in cohorts of persons with East Asian ancestry. The most significant findings were reported in the respective publications, and we were able to use these data to select instrumental variables for Analysis 3. However, summary statistics for these studies are not currently available in a harmonized form in the 2-sample MR R library/MRC IEU open GWAS catalog. Thus, given this limitation, we performed trans-ancestral MR in Analysis 4. Our analysis scheme is shown (Figure 7).

4.4.3. Selection of Instrumental Variables for SLE and SjD

We used lists of variants associated with SjD or SLE by GWAS. For SjD, these variants and links to the GWAS studies are listed in (Table S3B). For SLE, we used harmonized GWAS data sets from Bentham et al. [40] and Wang et al. [47] as well as from FinnGen [44]. All of these were accessed through the MR-Base web app v1.4.3 [43,102] interface for the ‘TwoSampleMR’ v0.5.5 R package [43,103]. A full listing of the GWAS studies used for the putative causal gene network and mendelian randomization analyses is listed below (Table 7).

In terms of summary statistics or the primary study, there was a great deal of variability. Some studies reported odds ratio and 95% confidence interval. Some studies reported the natural log of the odds ratio and its standard error. Other studies did not consistently indicate the effect allele for the odds ratio. We used the following equations to convert between odds ratio and beta/effect size scale.

$$Z = \beta / se \quad (1)$$

Z is the Z score of a normal distribution, β is the natural logarithm of the odds ratio, and $\ln(\text{OR})$ and se is the standard error of β .

$$[\text{lower 95\% CI of OR-upper 95\% CI of OR}] = [e^{(\beta - 1.96 \text{ se})} - e^{(\beta + 1.96 \text{ se})}] \quad (2)$$

where the upper and lower CI are the limits of the 95% confidence interval of the OR and e is Euler's number, the sum of the reciprocals of $n!$ for all numbers, n , between 0 and infinity.

$$P = 2 \phi(|-Z|) \quad (3)$$

where P is the p -value and ϕ is the cumulative distribution function for a standard normal distribution.

Table 7. GWAS or other studies used in analysis or referenced in Supplementary Materials.

1st Author [Ref] ¹	$N_{\text{case}}:N_{\text{ctl}}$ ²	PMID ³	Other IDs ⁴	Note ⁵
Lessard [104]	1638:6754	24097067		EUR SjD GWAS
Li [105]	1845:3757	24097066		EAS SjD GWAS
Khatri [29]	3851:23652	35896530	phs002723.v1.p1	EUR SjD GWAS
Song [106]	420:15876	27503288		EAS SjD GWAS
Qu [107]	665:863	28552951		EAS SjD GWAS
Kurki [44]	2735:416383	36653562	M13_Sjogren finn-b-M13_SJOGREN	EUR SjD GWAS URL in ST3A
Sakaue [108]	303:175599	34594039	Sjogren_Syndrome	EAS SjD GWAS URL in ST3A
Bentham [40]	5201:9066	26502338	GCST003156 ebi-a-GCST003156	EUR SLE GWAS
Wang [47]	4222:8431	33536424	GCST90011866 ebi-a-GCST90011866	EAS SLE GWAS

¹ 1st Author [Ref]—Name of first author as well as reference within this report. ² $N_{\text{case}}:N_{\text{ctl}}$, number of persons with disease (cases) and persons without disease or population controls, separated by “:”. ³ The pubmed identifier for the primary publication. ⁴ Other study identifiers from dbGAP, FinnGen, Biobank Japan, or the EBI GWAS Catalog. ⁵ Notes on the individual study.

4.4.4. Quality Control of Instrumental Variables for SLE and SjD

One of the three assumptions of Mendelian randomization is that the instrumental variables are robustly associated with the exposure. To ensure this was the case for SjD, we used the conservative cutoff for genetic association at the genome-wide significance threshold of $p < 5.0 \times 10^{-8}$ in any single study (Table S3B). We sought to ensure that this was the case for the instrumental variables from SLE GWAS studies through MRBase. To do so, we visually inspected the regional association plots from GWAS in LocusZoom (<https://my.locuszoom.org/>, accessed on 30 May 2023) [109]. There were numerous variants that were robustly associated with SLE in Bentham et al. GWAS summary statistics [40] but looked suspicious for artefacts. These variants are in strong linkage disequilibrium (according to 1000 genomes project data) with one or more nearby variants that do not exhibit robust evidence for association. Review of the primary publication [40] revealed that these association signals were not replicated. Therefore, these variants were removed from our list of instrumental variables, as we deemed them to likely violate the relevance condition. Importantly, this problem of quality control has also been noted for Bentham et al. GWAS summary statistics available on dbGAP [110]. We used the same procedure to ensure that Wang et al. GWAS summary statistics [47] available in the EBI GWAS catalog did not harbor similar potential artifacts. We did not find any variants of this nature with this procedure.

4.4.5. General Analysis Approach

Fine-scale linkage disequilibrium patterns differ between continental ancestral populations [49]. If the associated variants are not in linkage disequilibrium in a different ancestral population, this could represent a violation of the MR instrumental variable assumptions. So, where data were available, we compared variants within the ancestral population where the original GWAS was performed. This allows for the possibility that the variants

identified are not bona fide disease causal variants but are in linkage disequilibrium with bona fide disease causal variants.

4.4.6. MR Analysis Parameters

Linkage disequilibrium between variants is one of the ways in which MR conditions/assumptions can be violated for some MR analysis methods. Therefore, for each MR analysis presented, we used LD-based clumping to prune variants in LD with one another. Since our comparisons were against variants where we were confident regarding strand and effect allele as detailed in the Supplementary Tables, we did allow for palindromic LD-based proxies to be used. We set a minor allele frequency cutoff for assigning palindromic variants of 0.3. The GWAS studies in 'TwoSampleMR' and MRBase.org underwent automated harmonization to a particular strand of the reference genome. Since the inaccurate assignment of palindromic variants could change the direction of an MR result, we excluded palindromic variants. For each analysis, we used the four MR analysis methods described: inverse-variance weighted, weighted median, weighted mode and MR Egger.

5. Conclusions

Our complementary analyses revealed pervasive sharing of causal genetic risk factors between Systemic Lupus Erythematosus and Sjögren's Disease. This shared causal genetic risk may partly explain the clinical, immunologic, and molecular overlap between these diseases. Our study has implications for future genetic studies. Jointly modeling these two disease phenotypes as a single phenotype will likely increase power to detect shared and divergent risk factors. Likewise, there are also implications for clinical care that include the differential diagnosis of these two disease phenotypes and therapeutic selection where features of more than one disease are present.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241914449/s1>.

Author Contributions: Conceptualization, Y.R., J.A.I. and I.T.W.H.; Data curation, K.C., Y.R., K.A. and I.T.W.H.; Formal analysis, I.T.W.H.; Funding acquisition, I.T.W.H.; Methodology, K.C., K.A., I.C. and I.T.W.H.; Resources, I.T.W.H.; Software, K.C. and I.T.W.H.; Supervision, R.H.S. and I.T.W.H.; Visualization, K.C., K.A. and I.T.W.H.; Writing—original draft, I.T.W.H.; Writing—review and editing, K.C., Y.R., K.A., J.A.I., R.H.S., I.C. and I.T.W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Rheumatology Research Foundation Scientist Development Award (AWD-195543) and K Bridge Award (AWD-231383) to ITWH. The APC was funded by RRF grants to ITWH.

Institutional Review Board Statement: The analyses in the present study consisted of re-analysis of publicly available summary data. The study was conducted in accordance with the Declaration of Helsinki. Please refer to original individual studies for Institutional Review Board approval information. Links to the relevant publications are found in Tables 7 and S2).

Informed Consent Statement: Not applicable.

Data Availability Statement: Public data from the original studies analyzed in this manuscript can be found in the original publications or via the identifiers listed in (Table 7) and (Table S2). Further details of the analyses that we presented, including supporting logs, analysis R and python scripts and output files can be found at (https://github.com/harleyi/sjd_sle_overlap, accessed on 24 May 2023). Any pertinent information that may have been omitted will be furnished by the corresponding author upon reasonable request.

Acknowledgments: We want to acknowledge participants and investigators of the FinnGen study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The contents of this manuscript do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

References

1. Moutsopoulos, H.M.; Mann, D.L.; Johnson, A.H.; Chused, T.M. Genetic Differences between Primary and Secondary Sicca Syndrome. *N. Engl. J. Med.* **1979**, *301*, 761–763. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Grennan, D.M.; Ferguson, M.; Ghobarey, A.E.; Williamson, J.; Dick, W.C.; Buchanan, W.W. Sjogren's Syndrome in SLE: Part 2. An Examination of the Clinical Significance of Sjogren's Syndrome by Comparison of Its Frequency in Typical and Atypical Forms of SLE, Overlap Syndromes and Scleroderma. *N. Z. Med. J.* **1977**, *86*, 376–379. [\[PubMed\]](#)
3. Heaton, J.M. Sjögren's Syndrome and Systemic Lupus Erythematosus. *Br. Med. J.* **1959**, *1*, 466–469. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Provost, T.T.; Watson, R.; Simmons-O'Brien, E. Anti-Ro(SS-A) Antibody Positive Sjögren's/Lupus Erythematosus Overlap Syndrome. *Lupus* **1997**, *6*, 105–111. [\[CrossRef\]](#)
5. Alexander, E.L.; McNicholl, J.; Watson, R.M.; Bias, W.; Reichlin, M.; Provost, T.T. The Immunogenetic Relationship between Anti-Ro(SS-A)/La(SS-B) Antibody Positive Sjögren's/Lupus Erythematosus Overlap Syndrome and the Neonatal Lupus Syndrome. *J. Invest. Dermatol.* **1989**, *93*, 751–756. [\[CrossRef\]](#)
6. Gianordoli, A.P.E.; Laguardia, R.V.R.B.; Santos, M.C.F.S.; Jorge, F.C.; da Silva Salomão, A.; Caser, L.C.; Moulaz, I.R.; Serrano, É.V.; Miyamoto, S.T.; Machado, K.L.L.L.; et al. Prevalence of Sjögren's Syndrome According to 2016 ACR-EULAR Classification Criteria in Patients with Systemic Lupus Erythematosus. *Adv. Rheumatol.* **2023**, *63*, 11. [\[CrossRef\]](#)
7. Aringer, M.; Costenbader, K.H.; Daikh, D.I.; Brinks, R.; Mosca, M.; Ramsey-Goldman, R.; Smolen, J.S.; Wofsy, D.; Boumpas, D.; Kamen, D.L.; et al. 2019 EULAR/ACR Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol.* **2019**, *71*, 1400–1412. [\[CrossRef\]](#)
8. Aringer, M.; Costenbader, K.; Dörner, T.; Johnson, S.R. Response to: 'New 2019 SLE EULAR/ACR Classification Criteria Are Valuable for Distinguishing Patients with SLE from Patients with PSS' by Assan et Al. *Ann. Rheum. Dis.* **2021**, *80*, e123. [\[CrossRef\]](#)
9. Pasoto, S.G.; Adriano de Oliveira Martins, V.; Bonfa, E. Sjögren's Syndrome and Systemic Lupus Erythematosus: Links and Risks. *Open Access Rheumatol. Res. Rev.* **2019**, *11*, 33–45. [\[CrossRef\]](#)
10. Pan, H.F.; Ye, D.Q.; Wang, Q.; Li, W.X.; Zhang, N.; Li, X.P.; Xu, J.H.; Dai, H. Clinical and Laboratory Profiles of Systemic Lupus Erythematosus Associated with Sjögren Syndrome in China: A Study of 542 Patients. *Clin. Rheumatol.* **2008**, *27*, 339–343. [\[CrossRef\]](#)
11. Lockshin, M.D.; Levine, A.B.; Erkan, D. Patients with Overlap Autoimmune Disease Differ from Those with "pure" Disease. *Lupus Sci. Med.* **2015**, *2*, e000084. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Rasmussen, A.; Radfar, L.; Lewis, D.; Grundahl, K.; Stone, D.U.; Kaufman, C.E.; Rhodus, N.L.; Segal, B.; Wallace, D.J.; Weisman, M.H.; et al. Previous Diagnosis of Sjögren's Syndrome as Rheumatoid Arthritis or Systemic Lupus Erythematosus. *Rheumatology* **2016**, *55*, 1195–1201. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Kamitaki, N.; Sekar, A.; Handsaker, R.E.; de Rivera, H.; Tooley, K.; Morris, D.L.; Taylor, K.E.; Whelan, C.W.; Tombleson, P.; Loohuis, L.M.O.; et al. Complement Genes Contribute Sex-Biased Vulnerability in Diverse Disorders. *Nature* **2020**, *582*, 577–581. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Soret, P.; Le Dantec, C.; Desvaux, E.; Foulquier, N.; Chassagnol, B.; Hubert, S.; Jamin, C.; Barturen, G.; Desachy, G.; Devauchelle-Pensec, V.; et al. A New Molecular Classification to Drive Precision Treatment Strategies in Primary Sjögren's Syndrome. *Nat. Commun.* **2021**, *12*, 3523. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Caielli, S.; Wan, Z.; Pascual, V. Systemic Lupus Erythematosus Pathogenesis: Interferon and Beyond. *Annu. Rev. Immunol.* **2023**, *41*, 533–560. [\[CrossRef\]](#)
16. Rincon-Arevalo, H.; Wiedemann, A.; Stefanski, A.-L.; Lettau, M.; Szelinski, F.; Fuchs, S.; Frei, A.P.; Steinberg, M.; Kam-Thong, T.; Hatje, K.; et al. Deep Phenotyping of CD11c+ B Cells in Systemic Autoimmunity and Controls. *Front. Immunol.* **2021**, *12*, 635615. [\[CrossRef\]](#)
17. Guthridge, J.M.; Lu, R.; Tran, L.T.-H.; Arriens, C.; Aberle, T.; Kamp, S.; Munroe, M.E.; Dominguez, N.; Gross, T.; DeJager, W.; et al. Adults with Systemic Lupus Exhibit Distinct Molecular Phenotypes in a Cross-Sectional Study. *EclinicalMedicine* **2020**, *20*, 100291. [\[CrossRef\]](#)
18. Banchereau, R.; Hong, S.; Cantarel, B.; Baldwin, N.; Baisch, J.; Edens, M.; Cepika, A.-M.; Acs, P.; Turner, J.; Anguiano, E.; et al. Personalized Immunomonitoring Uncovers Molecular Networks That Stratify Lupus Patients. *Cell* **2016**, *165*, 551–565. [\[CrossRef\]](#)
19. James, J.A.; Guthridge, J.M.; Chen, H.; Lu, R.; Bourn, R.L.; Bean, K.; Munroe, M.E.; Smith, M.; Chakravarty, E.; Baer, A.N.; et al. Unique Sjögren's Syndrome Patient Subsets Defined by Molecular Features. *Rheumatology* **2020**, *59*, 860–868. [\[CrossRef\]](#)
20. Teruel, M.; Barturen, G.; Martínez-Bueno, M.; Castellini-Pérez, O.; Barroso-Gil, M.; Povedano, E.; Kerick, M.; Català-Moll, F.; Makowska, Z.; Buttgereit, A.; et al. Integrative Epigenomics in Sjögren's Syndrome Reveals Novel Pathways and a Strong Interaction between the HLA, Autoantibodies and the Interferon Signature. *Sci. Rep.* **2021**, *11*, 23292. [\[CrossRef\]](#)
21. Barturen, G.; Babaei, S.; Català-Moll, F.; Martínez-Bueno, M.; Makowska, Z.; Martorell-Marugán, J.; Carmona-Sáez, P.; Toro-Domínguez, D.; Carnero-Montoro, E.; Teruel, M.; et al. Integrative Analysis Reveals a Molecular Stratification of Systemic Autoimmune Diseases. *Arthritis Rheumatol.* **2021**, *73*, 1073–1085. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Foulquier, N.; Le Dantec, C.; Bettacchioli, E.; Jamin, C.; Alarcón-Riquelme, M.E.; Pers, J.-O. Machine Learning for the Identification of a Common Signature for Anti-SSA/Ro 60 Antibody Expression Across Autoimmune Diseases. *Arthritis Rheumatol.* **2022**, *74*, 1706–1719. [CrossRef] [PubMed]
23. Kuo, C.-F.; Grainge, M.J.; Valdes, A.M.; See, L.-C.; Luo, S.-F.; Yu, K.-H.; Zhang, W.; Doherty, M. Familial Risk of Sjögren's Syndrome and Co-Aggregation of Autoimmune Diseases in Affected Families: A Nationwide Population Study. *Arthritis Rheumatol.* **2015**, *67*, 1904–1912. [CrossRef]
24. Aggarwal, R.; Anaya, J.-M.; Koelsch, K.A.; Kurien, B.T.; Scofield, R.H. Association between Secondary and Primary Sjögren's Syndrome in a Large Collection of Lupus Families. *Autoimmune Dis.* **2015**, *2015*, 298506. [CrossRef]
25. Kuo, C.-F.; Grainge, M.J.; Valdes, A.M.; See, L.-C.; Luo, S.-F.; Yu, K.-H.; Zhang, W.; Doherty, M. Familial Aggregation of Systemic Lupus Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families. *JAMA Intern. Med.* **2015**, *175*, 1518–1526. [CrossRef]
26. Hou, G.; Harley, I.T.W.; Lu, X.; Zhou, T.; Xu, N.; Yao, C.; Qin, Y.; Ouyang, Y.; Ma, J.; Zhu, X.; et al. SLE Non-Coding Genetic Risk Variant Determines the Epigenetic Dysfunction of an Immune Cell Specific Enhancer That Controls Disease-Critical MicroRNA Expression. *Nat. Commun.* **2021**, *12*, 135. [CrossRef]
27. Liu, K.; Kurien, B.T.; Zimmerman, S.L.; Kaufman, K.M.; Taft, D.H.; Kottyan, L.C.; Lazaro, S.; Weaver, C.A.; Ice, J.A.; Adler, A.J.; et al. X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased Prevalence of 47,XXX in Systemic Lupus Erythematosus and Sjögren's Syndrome. *Arthritis Rheumatol.* **2016**, *68*, 1290–1300. [CrossRef] [PubMed]
28. Barturen, G.; Beretta, L.; Cervera, R.; Van Vollenhoven, R.; Alarcón-Riquelme, M.E. Moving towards a Molecular Taxonomy of Autoimmune Rheumatic Diseases. *Nat. Rev. Rheumatol.* **2018**, *14*, 75–93. [CrossRef]
29. Khatri, B.; Tessneer, K.L.; Rasmussen, A.; Aghakhanian, F.; Reksten, T.R.; Adler, A.; Alevizos, I.; Anaya, J.-M.; Aqrawi, L.A.; Baecklund, E.; et al. Genome-Wide Association Study Identifies Sjögren's Risk Loci with Functional Implications in Immune and Glandular Cells. *Nat. Commun.* **2022**, *13*, 4287. [CrossRef]
30. Harley, I.T.W.; Allison, K.; Scofield, R.H. Polygenic Autoimmune Disease Risk Alleles Impacting B Cell Tolerance Act in Concert across Shared Molecular Networks in Mouse and in Humans. *Front. Immunol.* **2022**, *13*, 953439. [CrossRef]
31. Mountjoy, E.; Schmidt, E.M.; Carmona, M.; Schwartzentruber, J.; Peat, G.; Miranda, A.; Fumis, L.; Hayhurst, J.; Buniello, A.; Karim, M.A.; et al. An Open Approach to Systematically Prioritize Causal Variants and Genes at All Published Human GWAS Trait-Associated Loci. *Nat. Genet.* **2021**, *53*, 1527–1533. [CrossRef] [PubMed]
32. Ghousaini, M.; Mountjoy, E.; Carmona, M.; Peat, G.; Schmidt, E.M.; Hercules, A.; Fumis, L.; Miranda, A.; Carvalho-Silva, D.; Buniello, A.; et al. Open Targets Genetics: Systematic Identification of Trait-Associated Genes Using Large-Scale Genetics and Functional Genomics. *Nucleic Acids Res.* **2021**, *49*, D1311–D1320. [CrossRef] [PubMed]
33. Harley, I.T.W. Monogenic SLE (8829bcb2-C7d9-11ec-B397-0ac135e8bacf). Available online: <https://www.ndexbio.org/viewer/networks/8829bcb2-c7d9-11ec-b397-0ac135e8bacf> (accessed on 6 June 2023).
34. Harley, I.T.W. MonogenicPolygenicSLEUnion (Aa310899-C7de-11ec-B397-0ac135e8bacf). Available online: <https://www.ndexbio.org/viewer/networks/aa310899-c7de-11ec-b397-0ac135e8bacf> (accessed on 6 June 2023).
35. Harley, I.T.W. Polygenic SLE (26855444-C7db-11ec-B397-0ac135e8bacf). Available online: <https://www.ndexbio.org/viewer/networks/26855444-c7db-11ec-b397-0ac135e8bacf> (accessed on 6 June 2023).
36. Guo, J.; Gu, M.; Zhang, W.; Liu, Y.; Qian, C.; Deng, A. Aberrant IL-35 Levels in Patients with Primary Sjögren's Syndrome. *Scand. J. Immunol.* **2018**, *88*, e12718. [CrossRef] [PubMed]
37. Han, M.; Li, Y.; Liu, S.; Jiang, S.; Yuan, L.; Xia, L.; Shen, H.; Lu, J. Elevation of Serum IL-35 in Patients with Primary Sjögren's Syndrome. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **2018**, *38*, 452–456. [CrossRef] [PubMed]
38. Fogel, O.; Rivière, E.; Seror, R.; Nocturne, G.; Boudaoud, S.; Ly, B.; Gottenberg, J.-E.; Le Guern, V.; Dubost, J.-J.; Nititham, J.; et al. Role of the IL-12/IL-35 Balance in Patients with Sjögren Syndrome. *J. Allergy Clin. Immunol.* **2018**, *142*, 258–268.e5. [CrossRef] [PubMed]
39. Yin, X.; Kim, K.; Suetsugu, H.; Bang, S.-Y.; Wen, L.; Koido, M.; Ha, E.; Liu, L.; Sakamoto, Y.; Jo, S.; et al. Meta-Analysis of 208370 East Asians Identifies 113 Susceptibility Loci for Systemic Lupus Erythematosus. *Ann. Rheum. Dis.* **2021**, *80*, 632–640. [CrossRef]
40. Bentham, J.; Morris, D.L.; Graham, D.S.C.; Pinder, C.L.; Tomblason, P.; Behrens, T.W.; Martín, J.; Fairfax, B.P.; Knight, J.C.; Chen, L.; et al. Genetic Association Analyses Implicate Aberrant Regulation of Innate and Adaptive Immunity Genes in the Pathogenesis of Systemic Lupus Erythematosus. *Nat. Genet.* **2015**, *47*, 1457–1464. [CrossRef]
41. Burgess, S.; Thompson, S.G. Interpreting Findings from Mendelian Randomization Using the MR-Egger Method. *Eur. J. Epidemiol.* **2017**, *32*, 377–389. [CrossRef]
42. Sterne, J.A.C.; Sutton, A.J.; Ioannidis, J.P.A.; Terrin, N.; Jones, D.R.; Lau, J.; Carpenter, J.; Rücker, G.; Harbord, R.M.; Schmid, C.H.; et al. Recommendations for Examining and Interpreting Funnel Plot Asymmetry in Meta-Analyses of Randomised Controlled Trials. *BMJ* **2011**, *343*, d4002. [CrossRef]
43. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome. *eLife* **2018**, *7*, e34408. [CrossRef]
44. Kurki, M.I.; Karjalainen, J.; Palta, P.; Sipilä, T.P.; Kristiansson, K.; Donner, K.M.; Reeve, M.P.; Laivuori, H.; Aavikko, M.; Kaunisto, M.A.; et al. FinnGen Provides Genetic Insights from a Well-Phenotyped Isolated Population. *Nature* **2023**, *613*, 508–518. [CrossRef]
45. Lao, O.; Lu, T.T.; Nothnagel, M.; Junge, O.; Freitag-Wolf, S.; Caliebe, A.; Balascakova, M.; Bertranpetit, J.; Bindoff, L.A.; Comas, D.; et al. Correlation between Genetic and Geographic Structure in Europe. *Curr. Biol.* **2008**, *18*, 1241–1248. [CrossRef]

46. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Wang, Y.-F.; Zhang, Y.; Lin, Z.; Zhang, H.; Wang, T.-Y.; Cao, Y.; Morris, D.L.; Sheng, Y.; Yin, X.; Zhong, S.-L.; et al. Identification of 38 Novel Loci for Systemic Lupus Erythematosus and Genetic Heterogeneity between Ancestral Groups. *Nat. Commun.* **2021**, *12*, 772. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Bowden, J.; Davey Smith, G.; Burgess, S. Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression. *Int. J. Epidemiol.* **2015**, *44*, 512–525. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [\[CrossRef\]](#)
50. Machiela, M.J.; Chanock, S.J. LDlink: A Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants. *Bioinformatics* **2015**, *31*, 3555–3557. [\[CrossRef\]](#)
51. Kottyan, L.C.; Zoller, E.E.; Bene, J.; Lu, X.; Kelly, J.A.; Rupert, A.M.; Lessard, C.J.; Vaughn, S.E.; Marion, M.; Weirauch, M.T.; et al. The IRF5-TNPO3 Association with Systemic Lupus Erythematosus Has Two Components That Other Autoimmune Disorders Variably Share. *Hum. Mol. Genet.* **2015**, *24*, 582–596. [\[CrossRef\]](#)
52. Göring, H.H.; Terwilliger, J.D.; Blangero, J. Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. *Am. J. Hum. Genet.* **2001**, *69*, 1357–1369. [\[CrossRef\]](#)
53. Ioannidis, J.P.A. Why Most Discovered True Associations Are Inflated. *Epidemiol. Camb. Mass* **2008**, *19*, 640–648. [\[CrossRef\]](#)
54. Harley, I.T.W.; Sawalha, A.H. Systemic Lupus Erythematosus as a Genetic Disease. *Clin. Immunol.* **2022**, *236*, 108953. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Choi, S.W.; Mak, T.S.-H.; O'Reilly, P.F. Tutorial: A Guide to Performing Polygenic Risk Score Analyses. *Nat. Protoc.* **2020**, *15*, 2759–2772. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Akatsu, C.; Shinagawa, K.; Numoto, N.; Liu, Z.; Ucar, A.K.; Aslam, M.; Phoon, S.; Adachi, T.; Furukawa, K.; Ito, N.; et al. CD72 Negatively Regulates B Lymphocyte Responses to the Lupus-Related Endogenous Toll-like Receptor 7 Ligand Sm/RNP. *J. Exp. Med.* **2016**, *213*, 2691–2706. [\[CrossRef\]](#) [\[PubMed\]](#)
57. McKusick, V.A. On Lumpers and Splitters, or the Nosology of Genetic Disease. *Perspect. Biol. Med.* **1969**, *12*, 298–312. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Eknoyan, G. A History of Edema and Its Management. *Kidney Int. Suppl.* **1997**, *59*, S118–S126. [\[PubMed\]](#)
59. Peitzman, S.J. Framing Disease: Studies in Cultural History. From Bright's Disease to End-Stage Renal Disease. *Hosp. Pract. Off. Ed.* **1992**, *27*, 192, 201–204, 207.
60. CDC. Diseases Caused by Group A Strep. Available online: <https://www.cdc.gov/groupastrep/diseases-public/index.html> (accessed on 5 June 2023).
61. Entry—#203100—ALBINISM, OCULOCUTANEOUS, TYPE IA; OCA1A—OMIM. Available online: <https://omim.org/entry/203100?search=203100&highlight=203100> (accessed on 5 June 2023).
62. Entry—#203500—ALKAPTONURIA; AKU—OMIM. Available online: <https://omim.org/entry/203500?search=203500&highlight=203500> (accessed on 5 June 2023).
63. Entry—#261600—PHENYLKETONURIA; PKU—OMIM. Available online: <https://omim.org/entry/261600?search=261600&highlight=261600> (accessed on 5 June 2023).
64. Entry—#276700—TYROSINEMIA, TYPE I; TYRSN1—OMIM. Available online: <https://omim.org/entry/276700?search=276700&highlight=276700> (accessed on 5 June 2023).
65. Rosenzweig, M.; Lorenzon, R.; Cacoub, P.; Pham, H.P.; Pitoiset, F.; El Soufi, K.; Rllet, C.; Bernard, C.; Aractingi, S.; Banneville, B.; et al. Immunological and Clinical Effects of Low-Dose Interleukin-2 across 11 Autoimmune Diseases in a Single, Open Clinical Trial. *Ann. Rheum. Dis.* **2019**, *78*, 209–217. [\[CrossRef\]](#)
66. Khunsriraksakul, C.; Li, Q.; Markus, H.; Patrick, M.T.; Sauteraud, R.; McGuire, D.; Wang, X.; Wang, C.; Wang, L.; Chen, S.; et al. Multi-Ancestry and Multi-Trait Genome-Wide Association Meta-Analyses Inform Clinical Risk Prediction for Systemic Lupus Erythematosus. *Nat. Commun.* **2023**, *14*, 668. [\[CrossRef\]](#)
67. Shirai, Y.; Nakanishi, Y.; Suzuki, A.; Konaka, H.; Nishikawa, R.; Sonehara, K.; Namba, S.; Tanaka, H.; Masuda, T.; Yaga, M.; et al. Multi-Trait and Cross-Population Genome-Wide Association Studies across Autoimmune and Allergic Diseases Identify Shared and Distinct Genetic Component. *Ann. Rheum. Dis.* **2022**, *81*, 1301–1312. [\[CrossRef\]](#)
68. Demela, P.; Pirastu, N.; Soskic, B. Cross-Disorder Genetic Analysis of Immune Diseases Reveals Distinct Gene Associations That Converge on Common Pathways. *Nat. Commun.* **2023**, *14*, 2743. [\[CrossRef\]](#)
69. Turley, P.; Walters, R.K.; Maghazian, O.; Okbay, A.; Lee, J.J.; Fontana, M.A.; Nguyen-Viet, T.A.; Wedow, R.; Zacher, M.; Furlotte, N.A.; et al. Multi-Trait Analysis of Genome-Wide Association Summary Statistics Using MTAG. *Nat. Genet.* **2018**, *50*, 229–237. [\[CrossRef\]](#) [\[PubMed\]](#)
70. van der Heijden, E.H.; Blokland, S.L.M.; Hillen, M.R.; Lopes, A.P.P.; van Vliet-Moret, F.M.; Rosenberg, A.J.W.P.; Janssen, N.G.; Welsing, P.M.J.; Iannizzotto, V.; Tao, W.; et al. Leflunomide–Hydroxychloroquine Combination Therapy in Patients with Primary Sjögren's Syndrome (RepurpSS-I): A Placebo-Controlled, Double-Blinded, Randomised Clinical Trial. *Lancet Rheumatol.* **2020**, *2*, e260–e269. [\[CrossRef\]](#)

71. Bowman, S.J.; Fox, R.; Dörner, T.; Mariette, X.; Papas, A.; Grader-Beck, T.; Fisher, B.A.; Barcelos, F.; De Vita, S.; Schulze-Koops, H.; et al. Safety and Efficacy of Subcutaneous Ianalumab (VAY736) in Patients with Primary Sjögren's Syndrome: A Randomised, Double-Blind, Placebo-Controlled, Phase 2b Dose-Finding Trial. *Lancet* **2022**, *399*, 161–171. [\[CrossRef\]](#)
72. He, J.; Chen, J.; Miao, M.; Zhang, R.; Cheng, G.; Wang, Y.; Feng, R.; Huang, B.; Luan, H.; Jia, Y.; et al. Efficacy and Safety of Low-Dose Interleukin 2 for Primary Sjögren Syndrome: A Randomized Clinical Trial. *JAMA Netw. Open* **2022**, *5*, e2241451. [\[CrossRef\]](#)
73. Carsons, S.E.; Vivino, F.B.; Parke, A.; Carteron, N.; Sankar, V.; Brasington, R.; Brennan, M.T.; Ehlers, W.; Fox, R.; Scofield, H.; et al. Treatment Guidelines for Rheumatologic Manifestations of Sjögren's Syndrome: Use of Biologic Agents, Management of Fatigue, and Inflammatory Musculoskeletal Pain. *Arthritis Care Res.* **2017**, *69*, 517–527. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Baer, A.N.; Vivino, F.B. Treatment of Sjögren's Syndrome: Constitutional and Non-Sicca Organ-Based Manifestations—UpToDate. Available online: <https://www.uptodate.com/contents/treatment-of-sjogrens-syndrome-constitutional-and-non-sicca-organ-based-manifestations> (accessed on 6 June 2023).
75. Harley, I.T.W. Repository Contains a Tutorial with Instructions for Synthesizing GWAS Results as a Putative Causal Gene Network. Available online: https://github.com/harleyi/gwas_catalog2polygenic_risk_network (accessed on 6 June 2023).
76. Sollis, E.; Mosaku, A.; Abid, A.; Buniello, A.; Cerezo, M.; Gil, L.; Groza, T.; Güneş, O.; Hall, P.; Hayhurst, J.; et al. The NHGRI-EBI GWAS Catalog: Knowledgebase and Deposition Resource. *Nucleic Acids Res.* **2023**, *51*, D977–D985. [\[CrossRef\]](#)
77. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [\[CrossRef\]](#)
78. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [\[CrossRef\]](#)
79. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for Gene List Enrichment Analysis and Candidate Gene Prioritization. *Nucleic Acids Res.* **2009**, *37*, W305–W311. [\[CrossRef\]](#)
80. Pillich, R.T.; Chen, J.; Churas, C.; Fong, D.; Gyori, B.M.; Ideker, T.; Karis, K.; Liu, S.N.; Ono, K.; Pico, A.; et al. NDEx IQuery: A Multi-Method Network Gene Set Analysis Leveraging the Network Data Exchange. *Bioinformatics* **2023**, *39*, btad118. [\[CrossRef\]](#)
81. Zhao, J.; Ma, J.; Deng, Y.; Kelly, J.A.; Kim, K.; Bang, S.-Y.; Lee, H.-S.; Li, Q.-Z.; Wakeland, E.K.; Qiu, R.; et al. A Missense Variant in NCF1 Is Associated with Susceptibility to Multiple Autoimmune Diseases. *Nat. Genet.* **2017**, *49*, 433–437. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Available online: <https://www.ebi.ac.uk/gwas/docs> (accessed on 17 March 2023).
83. Available online: <https://genetics-docs.opentargets.org/our-approach/prioritising-causal-genes-at-gwas-loci-l2g> (accessed on 17 March 2023).
84. Available online: <https://genetics.opentargets.org/> (accessed on 17 March 2023).
85. 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. The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [\[CrossRef\]](#)
86. Available online: <https://string-db.org/> (accessed on 17 March 2023).
87. Available online: <https://www.youtube.com/c/LarsJuhlJensen> (accessed on 17 March 2023).
88. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [\[CrossRef\]](#)
89. Available online: https://cytoscape.org/what_is_cytoscape.html (accessed on 17 March 2023).
90. Available online: <https://github.com/cytoscape/cytoscape-tutorials/wiki> (accessed on 17 March 2023).
91. Available online: <https://manual.cytoscape.org/en/stable/> (accessed on 17 March 2023).
92. Available online: <https://systems.crump.ucla.edu/hypergeometric/index.php> (accessed on 17 March 2023).
93. Sanderson, E.; Glymour, M.M.; Holmes, M.V.; Kang, H.; Morrison, J.; Munafò, M.R.; Palmer, T.; Schooling, C.M.; Wallace, C.; Zhao, Q.; et al. Mendelian Randomization. *Nat. Rev. Methods Primer* **2022**, *2*, 6. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Davey Smith, G. Use of Genetic Markers and Gene-Diet Interactions for Interrogating Population-Level Causal Influences of Diet on Health. *Genes Nutr.* **2011**, *6*, 27–43. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Smith, G.D.; Ebrahim, S. “Mendelian Randomization”: Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease? *Int. J. Epidemiol.* **2003**, *32*, 1–22. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Lassi, G.; Taylor, A.E.; Timpson, N.J.; Kenny, P.J.; Mather, R.J.; Eisen, T.; Munafò, M.R. The CHRNA5-A3-B4 Gene Cluster and Smoking: From Discovery to Therapeutics. *Trends Neurosci.* **2016**, *39*, 851–861. [\[CrossRef\]](#)
97. de Leeuw, C.; Savage, J.; Bucur, I.G.; Heskes, T.; Posthuma, D. Understanding the Assumptions Underlying Mendelian Randomization. *Eur. J. Hum. Genet. EJHG* **2022**, *30*, 653–660. [\[CrossRef\]](#)
98. Morrison, J.; Knoblauch, N.; Marcus, J.H.; Stephens, M.; He, X. Mendelian Randomization Accounting for Correlated and Uncorrelated Pleiotropic Effects Using Genome-Wide Summary Statistics. *Nat. Genet.* **2020**, *52*, 740–747. [\[CrossRef\]](#)
99. Burgess, S.; Davey Smith, G.; Davies, N.M.; Dudbridge, F.; Gill, D.; Glymour, M.M.; Hartwig, F.P.; Holmes, M.V.; Minelli, C.; Relton, C.L.; et al. Guidelines for Performing Mendelian Randomization Investigations. *Wellcome Open Res.* **2019**, *4*, 186. [\[CrossRef\]](#)

100. Hartwig, F.P.; Davey Smith, G.; Bowden, J. Robust Inference in Summary Data Mendelian Randomization via the Zero Modal Pleiotropy Assumption. *Int. J. Epidemiol.* **2017**, *46*, 1985–1998. [[CrossRef](#)]
101. Bowden, J.; Del Greco, M.F.; Minelli, C.; Zhao, Q.; Lawlor, D.A.; Sheehan, N.A.; Thompson, J.; Davey Smith, G. Improving the Accuracy of Two-Sample Summary-Data Mendelian Randomization: Moving beyond the NOME Assumption. *Int. J. Epidemiol.* **2019**, *48*, 728–742. [[CrossRef](#)] [[PubMed](#)]
102. Elsworth, B.; Lyon, M.; Alexander, T.; Liu, Y.; Matthews, P.; Hallett, J.; Bates, P.; Palmer, T.; Haberland, V.; Smith, G.D.; et al. The MRC IEU OpenGWAS Data Infrastructure. *bioRxiv* **2020**. [[CrossRef](#)]
103. Hemani, G.; Tilling, K.; Smith, G.D. Orienting the Causal Relationship between Imprecisely Measured Traits Using GWAS Summary Data. *PLoS Genet.* **2017**, *13*, e1007081. [[CrossRef](#)]
104. Lessard, C.J.; Li, H.; Adrianto, I.; Ice, J.A.; Rasmussen, A.; Grundahl, K.M.; Kelly, J.A.; Dozmorov, M.G.; Miceli-Richard, C.; Bowman, S.; et al. Variants at Multiple Loci Implicated in Both Innate and Adaptive Immune Responses Are Associated with Sjögren's Syndrome. *Nat. Genet.* **2013**, *45*, 1284–1292. [[CrossRef](#)]
105. Li, Y.; Zhang, K.; Chen, H.; Sun, F.; Xu, J.; Wu, Z.; Li, P.; Zhang, L.; Du, Y.; Luan, H.; et al. A Genome-Wide Association Study in Han Chinese Identifies a Susceptibility Locus for Primary Sjögren's Syndrome at 7q11.23. *Nat. Genet.* **2013**, *45*, 1361–1365. [[CrossRef](#)]
106. Song, I.-W.; Chen, H.-C.; Lin, Y.-F.; Yang, J.-H.; Chang, C.-C.; Chou, C.-T.; Lee, M.-T.M.; Chou, Y.-C.; Chen, C.-H.; Chen, Y.-T.; et al. Identification of Susceptibility Gene Associated with Female Primary Sjögren's Syndrome in Han Chinese by Genome-Wide Association Study. *Hum. Genet.* **2016**, *135*, 1287–1294. [[CrossRef](#)]
107. Qu, S.; Du, Y.; Chang, S.; Guo, L.; Fang, K.; Li, Y.; Zhang, F.; Zhang, K.; Wang, J. Common Variants near IKZF1 Are Associated with Primary Sjögren's Syndrome in Han Chinese. *PLoS ONE* **2017**, *12*, e0177320. [[CrossRef](#)]
108. Sakaue, S.; Kanai, M.; Tanigawa, Y.; Karjalainen, J.; Kurki, M.; Koshiba, S.; Narita, A.; Konuma, T.; Yamamoto, K.; Akiyama, M.; et al. A Cross-Population Atlas of Genetic Associations for 220 Human Phenotypes. *Nat. Genet.* **2021**, *53*, 1415–1424. [[CrossRef](#)]
109. Boughton, A.P.; Welch, R.P.; Flickinger, M.; VandeHaar, P.; Taliun, D.; Abecasis, G.R.; Boehnke, M. LocusZoom.js: Interactive and Embeddable Visualization of Genetic Association Study Results. *Bioinformatics* **2021**, *37*, 3017–3018. [[CrossRef](#)]
110. Chepelev, I.; Harley, I.T.W.; Harley, J.B. Modeling of Horizontal Pleiotropy Identifies Possible Causal Gene Expression in Systemic Lupus Erythematosus. *Front. Lupus* **2023**, *1*, 1234578. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.