Autoimmune diseases and risk of non-Hodgkin lymphoma: A Mendelian randomisation study

Xiaoting Shi, PhD Candidate\textsuperscript{1,4}; Joshua D. Wallach, Assistant Professor\textsuperscript{2}; Xiaomei Ma, Professor\textsuperscript{3}; Tormod Rogne, Assistant Professor\textsuperscript{3,4}

\textsuperscript{1} Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA

\textsuperscript{2} Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

\textsuperscript{3} Department of Chronic Diseases Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

\textsuperscript{4} Yale Center for Perinatal, Pediatric, and Environmental Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

\textbf{Corresponding author:}

Xiaoting Shi, MPhil
Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA
xiaoting.shi@yale.edu

ORCID number: Xiaoting Shi https://orcid.org/0000-0002-5256-0646

Joshua Wallach https://orcid.org/0000-0002-2816-6905

Tormod Rogne: https://orcid.org/0000-0002-9581-7384

Twitter: @Xiaoting_yale, @JoshuaDWallach, @TormodRogne

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Objective: To examine whether genetically predicted susceptibility to ten autoimmune diseases (Behçet's disease, coeliac disease, dermatitis herpetiformis, lupus, psoriasis, rheumatoid arthritis, sarcoidosis, Sjögren's syndrome, systemic sclerosis, and type 1 diabetes) is associated with risk of non-Hodgkin lymphoma (NHL).

Design: Two sample Mendelian randomization (MR) study.

Setting: Genome wide association studies (GWASs) of ten autoimmune diseases, NHL, and four NHL subtypes (i.e., follicular lymphoma, mature T/natural killer-cell lymphomas, non-follicular lymphoma, and other and unspecified types of NHL).

Analysis: We used data from the largest publicly available GWASs of European ancestry for each autoimmune disease, NHL, and NHL subtypes. For each autoimmune disease, we extracted single nucleotide polymorphisms (SNPs) strongly associated ($P < 5 \times 10^{-8}$) with that disease and that were independent of one another ($R^2 < 1 \times 10^{-3}$) as genetic instruments. SNPs within the human leukocyte antigen region were not considered due to potential pleiotropy. Our primary MR analysis was the inverse-variance weighted analysis. Additionally, we conducted MR-Egger, weighted mode, and weighted median regression to address potential bias due to pleiotropy, and robust adjusted profile scores to address weak instrument bias. We carried out sensitivity analysis limited to the non-immune pathway for nominally significant findings. To account for multiple testing, we set the thresholds for statistical significance at $P < 5 \times 10^{-3}$.

Participants: The number of cases and controls identified in the relevant GWASs were 437 and 3,325 for Behçet's disease, 4,918 and 5,684 for coeliac disease, 435 and 341,188 for dermatitis herpetiformis, 4,576 and 8,039 for lupus, 11,988 and 275,335 for psoriasis, 22,350 and 74,823 for rheumatoid arthritis, 3,597 and 337,121 for sarcoidosis, 2,735 and 332,115 for Sjögren's
syndrome, 9,095 and 17,584 for systemic sclerosis, 18,942 and 501,638 for type 1 diabetes, 2,400 and 410,350 for NHL; and 296 to 2,340 cases and 271,463 controls for NHL subtypes.

**Exposures:** Genetic variants predicting ten autoimmune diseases: Behçet's disease, coeliac disease, dermatitis herpetiformis, lupus, psoriasis, rheumatoid arthritis, sarcoidosis, Sjögren's syndrome, systemic sclerosis, and type 1 diabetes.

**Main outcome measures:** Estimated associations between genetically predicted susceptibility to ten autoimmune diseases and the risk of NHL.

**Results** The variance of each autoimmune disease explained by the SNPs ranged from 0.3% to 3.1%. Negative associations between type 1 diabetes and sarcoidosis and the risk of NHL were observed (odds ratio [OR] 0.95, 95% confidence interval [CI]: 0.92 to 0.98, \( P = 5 \times 10^{-3} \), and OR 0.92, 95% CI: 0.85 to 0.99, \( P = 2.8 \times 10^{-2} \), respectively). These findings were supported by the sensitivity analyses accounting for potential pleiotropy and weak instrument bias. No significant associations were found between the other eight autoimmune diseases and NHL risk. Of the NHL subtypes, type 1 diabetes was most strongly associated with follicular lymphoma (OR 0.91, 95% CI: 0.86 to 0.96, \( P = 1 \times 10^{-3} \)), while sarcoidosis was most strongly associated with other and unspecified NHL (OR 0.86, 95% CI: 0.75 to 0.97, \( P = 1.8 \times 10^{-2} \)).

**Conclusions** These findings suggest that genetically predicted susceptibility to type 1 diabetes, and to some extent sarcoidosis, might reduce the risk of NHL. However, future studies with different datasets, approaches, and populations are warranted to further examine the potential associations between these autoimmune diseases and the risk of NHL.

**Keywords:** Autoimmune diseases; Mendelian randomisation; Non-Hodgkin lymphoma; Type 1 diabetes; Sarcoidosis.
WHAT IS ALREADY KNOWN ON THIS TOPIC

1. The etiology of non-Hodgkin lymphoma, a common hematological malignancy, is not fully understood.

2. Observational studies have reported statistically significant associations between ten autoimmune diseases (Behçet's disease, coeliac disease, dermatitis herpetiformis, lupus, psoriasis, rheumatoid arthritis, sarcoidosis, Sjögren's syndrome, systemic sclerosis, and type 1 diabetes) and risk of non-Hodgkin lymphoma, but these studies may be susceptible to residual confounding and reverse causation.

WHAT THIS STUDY ADDS

1. Genetically predicted susceptibility to type 1 diabetes, and to some extent sarcoidosis, may be associated with a reduced risk of non-Hodgkin lymphoma, while no clear associations were observed between the other eight autoimmune diseases and risk of non-Hodgkin lymphoma or its subtypes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

1. Using an approach that seeks to address residual confounding and reverse causation, these findings contradict previously reported associations between autoimmune diseases and risk of non-Hodgkin lymphoma from traditional observational studies.

2. Future studies with different datasets, approaches, and populations are warranted to further examine the potential associations between these autoimmune diseases and the risk of NHL.
INTRODUCTION

Non-Hodgkin lymphoma (NHL), a hematological malignancy that arises from lymphocytes, is one of the most common cancers, with more than 544,000 new cases every year worldwide. Despite substantial efforts to identify risk factors for NHL, the exact etiology remains elusive. According to a recent umbrella review (i.e., systematic review of meta-analyses) evaluating the associations between 134 unique environmental risk factors and the risk of NHL, ten autoimmune diseases - Behçet's disease, coeliac disease, dermatitis herpetiformis, psoriasis, rheumatoid arthritis, sarcoidosis, systemic lupus erythematosus (SLE), Sjögren's syndrome, systemic sclerosis, and type 1 diabetes (T1D) - were identified to be statistically significantly associated with an increased risk of NHL. Of these, coeliac disease, rheumatoid arthritis, Sjögren's syndrome, and SLE, were classified as presenting highly suggestive (i.e., \( P < 1 \times 10^{-6} \), at least 1000 NHL cases, and largest study in the review reporting a nominally significant result) or convincing evidence (i.e., \( P < 1 \times 10^{-6} \), at least 1000 NHL cases, largest study in the review reporting a nominally significant result, minimal between-study heterogeneity, and no evidence of publication bias) evidence for NHL risk. Autoimmune diseases have long been considered potential risk factors for NHL. Proposed mechanisms for the associations between autoimmune diseases and NHL include chronic inflammation, antigen stimulation, and overlapping genetic susceptibility. However, the associations identified by the umbrella review were from systematic reviews and meta-analyses with various study design and reporting limitations. Furthermore, the systematic reviews and meta-analyses included only case-control and cohort studies, which are susceptible to multiple biases that limit their ability to evaluate causal relationships. For instance, confounding factors (e.g., socioeconomic status, family history of lymphoma, and infectious diseases) and reverse causation were often not considered by the
previous studies.¹²⁻¹⁸ For some autoimmune diseases, such as sarcoidosis, it is often unclear if the autoimmune disease precedes or develops after NHL.¹⁹⁻²⁴

One way of addressing the issue of residual confounding and reverse causation is through instrumental variable analyses with genetic instruments, often called Mendelian randomisation (MR) analyses.²⁵ Because genetic variants are randomly assigned at conception and are not affected by external factors such as chronic diseases and lifestyle factors, MR analyses mimic randomized experiments and are less susceptible to confounding and reverse causation compared with conventional observational studies.²⁶,²⁷

Therefore, the aim of this study was to use the MR design to evaluate the associations between genetically predicted susceptibility to ten autoimmune diseases and the risk of NHL and NHL subtypes (follicular lymphoma, mature T/NK-cell lymphomas, non-follicular lymphoma, and other and unspecified types of NHL).⁴ Because nearly all observational studies on the associations between autoimmune diseases and NHL suggested a positive association,⁴,⁹,¹²,¹³ our hypothesis was that the genetically predicted susceptibility to each autoimmune disease was associated with an increased risk of NHL and at least one of the NHL subtypes.
METHODS

This study is reported following the Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomisation guidelines (STROBE-MR, Supplement 1). Although there is no pre-registered protocol for this study, the analyses were designed prior to the conduct of the study. Data were retrieved between June 2022 and June 2023, and analyses were conducted between June 2022 and October 2023. The manuscript was posted on medRxiv.

Study design

Figure 1 shows a schematic summary of this two-sample MR study. Overall, we extracted summary level data from genome-wide association studies (GWASs) to find genetic instruments for our exposures of interest (i.e., exposure GWASs) and then investigated the associations of these genetic instruments with the outcomes of interest (i.e., outcome GWASs). For the genetic instruments to be valid, three core assumptions must be met: (1) the instruments are associated with the exposure of interest (the relevance assumption), (2) the instruments are not associated with any confounders of the exposure-outcome relationship (the independence assumption), and 3) the instruments are associated with the outcome only through the exposure (the exclusion restriction assumption).

The analyses were carried out in three steps: first, genetic instruments for each exposure of interest (i.e., autoimmune disease) were selected from relevant exposure GWASs. Second, for these genetic instruments, the genetic instrument-outcome associations were extracted from the relevant outcome GWASs. For each autoimmune disease, the exposure-outcome association (i.e., the Wald ratio) was estimated by dividing the genetic instrument-outcome association by the
genetic instrument-exposure association. Finally, the Wald ratios were summarised using different techniques (See Statistical analyses). Due to data availability and in order to minimize confounding due to population stratification, we only considered subjects of European ancestry.

Instrumental variable selection for autoimmune diseases

To fulfill the relevance assumption, we used single nucleotide polymorphisms (SNPs) as genetic instruments and selected those from GWASs that were (1) strongly associated with the specific autoimmune disease at genome-wide significance (i.e., $P < 5 \times 10^{-8}$) and (2) independent of one-another (i.e., an $R^2 < 1 \times 10^{-3}$). We used a less stringent criterion for significance level, $P < 5 \times 10^{-6}$ for dermatitis herpetiformis and Sjögren’s syndrome to ensure that we had at least five SNPs as genetic instruments for each autoimmune disease.

We evaluated the ten autoimmune diseases that were at least nominally significantly associated (i.e., $P < 5 \times 10^{-2}$) with risk of NHL according to a recent umbrella review: Behçet disease, coeliac disease, dermatitis herpetiformis, psoriasis, rheumatoid arthritis, sarcoidosis, Sjögren's syndrome, SLE, systemic sclerosis, and T1D. These autoimmune diseases were all associated with an increased risk of NHL. Of these, coeliac disease, rheumatoid arthritis, Sjögren's syndrome, and SLE, were classified as having highly suggestive (i.e., $P < 1 \times 10^{-6}$, at least 1000 NHL cases, and largest study in the review reporting a nominally significant result) or convincing evidence (i.e., $P < 1 \times 10^{-6}$, at least 1000 NHL cases, largest study in the review reporting a nominally significant result, minimal between-study heterogeneity, and no evidence of publication bias) in the umbrella review. For each autoimmune disease, we selected the
GWAS that had the largest sample size. In situations where no distinct GWAS publication for an autoimmune disease was available, we used summary statistics from genome-wide association analyses from FinnGen Release 8 (https://r8.finngen.fi/, Supplementary 2). Study characteristics of all summary level GWAS datasets for the autoimmune diseases used in main analyses are shown in Table 1.

Because human leukocyte antigen (HLA) region is highly predictive of both autoimmune diseases and NHL, we excluded SNPs from this locus from all analyses in order to prevent potential bias due to genetic pleiotropy.31,32 We defined the HLA region as location from 28,477,797 to 33,448,354 on chromosome 6 for Genome Reference Consortium Human Build 37.33

Outcomes

Our primary outcome was risk of NHL. We selected the largest GWAS on NHL, which evaluated subjects from UK Biobank and Kaiser Permanente (Table 2 and Supplementary Text).34 In secondary analyses for the genetically-predicted autoimmune diseases found to be nominally significantly associated with NHL (see Statistical analyses), we further evaluated the associations between these autoimmune diseases and NHL subtypes. We identified the NHL subtypes from FinnGen Release 8 and we selected four major NHL subtypes, consistent with the FinnGen classification system: follicular lymphoma, mature T/natural killer (NK)-cell lymphomas, non-follicular lymphoma, and other and unspecified types of NHL (Table 2).30 Of note, as the data on the NHL subtypes were extracted from a different dataset from the NHL...
GWAS, the secondary analyses also served as replication analyses using an independent dataset.\cite{30}

204

**Statistical analyses**

205 **Main analyses and sensitivity analyses**

For each autoimmune disease, we calculated the SNP-specific Wald ratio, defined as $\beta_{\text{EXP-OUT}} = \beta_{\text{SNP-OUT}} / \beta_{\text{SNP-EXP}}$ (Figure 1). We used inverse-variance weighted (IVW) analysis as our main analyses to sum the Wald ratios, which assign weights to each SNP in inverse proportion to the variance of the $\beta_{\text{SNP-OUT}}$, assuming all instruments to be valid.\cite{35,36} However, the IVW analysis may be biased if any of the included instruments are invalid (e.g., if the genetic instruments affect multiple traits, which is known as horizontal pleiotropy).\cite{37} Therefore, we carried out three sensitivity analyses that provide unbiased estimates even in the presence of some invalid instruments:\cite{25,38} MR-Egger regression,\cite{36,39} weighted mode estimator analysis,\cite{40} and weighted median estimator analysis (Supplementary Text).\cite{41} Furthermore, to address potential weak instrument bias, which may be introduced when the genetic variants explain a very small proportion of the variation in the exposure,\cite{29} we included robust adjusted profile scores (RAPS) as an additional sensitivity analysis (Supplementary Text).\cite{25,42}

We evaluated the strength of the instruments using $R^2$ (i.e., proportion of variance of the exposure explained by the genetic instrument) and $F$ statistics. In particular, we used the \texttt{get_r_from_bsen} function in the \textit{TwoSampleMR} package in R (version 4.3.0), and summed the absolute values across the independent SNPs to estimate the composite $R^2$ for each autoimmune
disease. $F$ statistics were calculated using the formula $F = \left( \frac{\hat{\beta}_X}{\text{se}(\hat{\beta}_X)} \right)^2$, where $\hat{\beta}_X$ refers to the genetic association between the instrument $X$ with the exposure, and $\text{se}(\hat{\beta}_X)$ refers to the standard error of $\hat{\beta}_X$.

Bidirectional analyses

To help establish the direction of effects between two traits, we carried out bidirectional analyses of genetically predicted NHL and risk of each autoimmune disease. Evidence of an effect in both directions could suggest that an effect acts in both directions between two traits so that changing one will change the other (i.e., the true bidirectional relationship). Otherwise it may suggest that a biasing pathway may be present. Evidence of an effect in one direction but not the other supports that a biasing pathway is less likely to be present.

Secondary analyses

To further validate the principal findings from the main analyses, we carried out two sets of secondary analyses for the autoimmune diseases that were nominally significantly associated with risk of NHL (i.e., T1D and sarcoidosis, see Results): (1) evaluating the risk of four NHL subtypes using a dataset independent from the main analyses and (2) restricting the genetic instruments to those that are less likely to affect immune function. The analyses of the NHL subtypes were carried out to identify which of the NHL subtypes were driving the observed association with NHL, and also to serve as a replication using a separate outcome study population not overlapping with the NHL study population of the other analyses. The analysis involving selected pathways were carried out to address horizontal pleiotropy. In particular, given the shared role of immune function in the pathway to developing both autoimmune diseases and NHL, we sought to identify biological pathways for the autoimmune diseases that
were to a lesser degree linked to immune function (Supplementary Table 1). For T1D, we restricted this secondary analysis to SNPs linked to insulin production, while for sarcoidosis we selected SNPs that were not directly linked to immune function. The KEGG and GeneCards databases were used to identify the function of all T1D and sarcoidosis SNPs. For all secondary analyses, only IVW analyses were carried out.

Software

We used TwoSampleMR package in R (version 4.3.0) to run the two-sample MR analyses, and forpl to generate the forest plots. A $P < 5 \times 10^{-2}$ was considered nominally significant. To correct for multiple testing of the ten autoimmune diseases in the main analyses, the level for statistical significance was set at $P < 5 \times 10^{-2}/10 = 5 \times 10^{-3}$. To make the results more interpretable, all causal estimates were multiplied by 0.693 ($= \log(2)$) and next exponentiated in order to represent the odds ratios (ORs) for NHL per doubling in the prevalence of the autoimmune disease under study.

Ethics

Only summary-level data from published studies with relevant ethical approvals were used in this study so approval from institutional review board was not necessary.

Patient and public involvement

No patients or members of the public were involved in the conception of the study, interpretation of the results, or drafting of the manuscript. We do not have plans to disseminate the results to research participants or relevant patient communities.
RESULTS

The number of cases and controls identified in the relevant GWASs were 437 and 3,325 for Behçet’s disease, 4,918 and 5,684 for coeliac disease, 435 and 341,188 for dermatitis herpetiformis, 4,576 and 8,039 for lupus, 11,988 and 275,335 for psoriasis, 22,350 and 74,823 for rheumatoid arthritis, 3,597 and 337,121 for sarcoidosis, 2,735 and 332,115 for Sjögren’s syndrome, 9,095 and 17,584 for systemic sclerosis, 18,942 and 501,638 for type 1 diabetes, and 2,400 and 410,350 for NHL (Table 1 and Table 2). The variance in the exposure explained by the genetic instruments ranged from 0.3% for Behçet disease to 3.1% for T1D (Table 1).

Primary analyses

A doubling in the genetically-predicted prevalence of T1D was associated with an OR for NHL of 0.95 (95% confidence interval [CI]: 0.92 to 0.98, \( P = 5 \times 10^{-3} \)), while a doubling in the genetically-predicted prevalence of sarcoidosis was associated with an OR for NHL of 0.92 (95% CI: 0.85 to 0.99, \( P = 2.8 \times 10^{-2} \)) (Figure 2). We did not observe significant associations between the other eight autoimmune diseases and risk of NHL (Figure 2).

Sensitivity analyses

MR-Egger, weighted mode and weighted median yielded ORs comparable to those from the main analyses, and with overlapping CIs, indicating little presence of pleiotropy (Figure 2). Furthermore, the RAPS sensitivity analyses did not suggest bias due to weak instruments. For the bidirectional analyses we did not observe significant associations between NHL and the risk of any of the autoimmune diseases (Supplementary Table 2).
Secondary analyses

In the secondary analyses, we evaluated the autoimmune diseases that were at least nominally significantly associated with risk of NHL, i.e., T1D and sarcoidosis. In the analyses restricted to non-immune pathways, we observed ORs of 0.96 (95% CI: 0.87 to 1.07) for T1D and 0.96 (95% CI: 0.88 to 1.06) for sarcoidosis, respectively, supporting the main analyses.

Figure 3 shows the IVW analyses of genetically predicted susceptibility to T1D and sarcoidosis, and the risk of NHL subtypes. For T1D, the association with composite NHL appeared to be driven by follicular lymphoma, with an OR of 0.91 (95% CI: 0.86 to 0.96, \( P = 1 \times 10^{-3} \)).

Sarcoidosis was most strongly associated with other and unspecified types of NHL, with an OR of 0.86 (95% CI: 0.75 to 0.97, \( P = 1.8 \times 10^{-2} \)).
DISCUSSION

Principal findings

In this MR study of ten autoimmune diseases previously linked to an increased risk of NHL, we found that genetically predicted susceptibility to T1D, and to some extent, genetically predicted susceptibility to sarcoidosis, were associated with a reduced risk of NHL. While these findings were consistent across a wide range of sensitivity analyses, no clear associations were observed between the other eight autoimmune diseases and risk of NHL. Using an approach that attempts to address potential residual confounding and reverse causation, our findings contradict those reported in previous traditional observational studies. This highlights the need for future studies with different datasets, approaches, and populations to further examine the potential associations between these autoimmune diseases and the risk of NHL.

Context of primary findings

Autoimmune diseases have long been considered potential risk factors for NHL, especially rheumatoid arthritis, SLE, and Sjögren's syndrome. A recent large-scale prospective cohort study found that among all cancers, lymphoma demonstrated the most extensive associations with different immune-mediated diseases. According to a previous umbrella review evaluating the associations between any environmental risk factors and the risk of NHL reported in published meta-analyses, there were consistent statistically significant associations between autoimmune diseases and NHL risk. Although the exact mechanisms for positive associations between autoimmune diseases and NHL remain unclear, there are several mechanisms proposed, including chronic inflammation, antigen stimulation, overlapping genetic susceptibility, and
dysfunction of certain protein families. Furthermore, the observed associations between autoimmune diseases and NHL may be attributable to the immunosuppressants that are used as treatments as well.

In our study, two autoimmune diseases - T1D and sarcoidosis - were found to be associated with NHL. However, contrary to our pre-specified hypotheses, which were based on the summary relative risks from an umbrella review (relative risk 1.55 (95% CI: 1.15 to 2.08) for associations between T1D and NHL, and relative risk 1.43 (95% CI: 1.03 to 1.99) for sarcoidosis and NHL), we found a statistically significant negative association between genetically predicted susceptibility to T1D and NHL. Although we also observed a negative association between genetically predicted susceptibility to sarcoidosis and NHL, the association was no longer significant after accounting for multiple testing. While no previous epidemiologic studies have reported statistically significant negative associations for either T1D or sarcoidosis with NHL, it has been suggested that the pathways involved in autoimmune disease and cancer development may work in opposite directions. Moreover, it remains unclear if sarcoidosis precedes or follows NHL, as several observational studies reported occurrence of sarcoidosis among NHL patients after treatment of NHL or concomitant occurrence of two diseases.

There are several potential explanations for the discrepancy between our findings and those of the previous umbrella review. First, meta-analyses of observational studies and MR analyses have different bias structures. While traditional observational studies often suffer from confounding, misclassification, and selection bias, and meta-analyses are often susceptible to publication bias (i.e., the lack of publishing certain findings, often null findings), horizontal pleiotropy is of particular concern for MR studies. Second, many of the traditional observational
studies and meta-analyses have relatively small sample sizes. In particular, the previous meta-analyses for T1D only included three individual observational studies with 1155 NHL cases and the meta-analysis for sarcoidosis only included seven studies with 150 sarcoidosis cases.\textsuperscript{4,53} Furthermore, according to a formal critical appraisal tool for systematic reviews and meta-analyses (i.e., AMSTAR 2), both meta-analyses were found to have at least one critical weakness.\textsuperscript{4,5} Third, given that genetic variants are fixed at conception, the exposure measured in MR analyses are typically interpreted as the lifelong exposure. In observational studies on the association between autoimmune diseases and NHL, the exposure is usually measured and interpreted in a defined time period from disease onset of autoimmune diseases. Finally, we conducted our study in population of European ancestry, while some meta-analyses investigated the associations in multiple populations other than European ancestry.\textsuperscript{53,54} However, the findings from populations with different ancestries did not differ drastically in the umbrella review.\textsuperscript{53,54}

**Strengths and limitations of this study**

The MR design is a major strength of our study, as it reduces the influence of residual confounding and reverse causation. In particular, confounding due to infectious diseases (an important risk factor for NHL and for autoimmune diseases) is minimized in this study. Moreover, to our knowledge, this is the first MR study that investigates the association between multiple autoimmune diseases and NHL. Our study had several limitations. First, this study is limited to populations of European ancestry, which have the highest incidence of NHL, and it is unclear whether the findings can be generalised beyond this population. However, only GWAS data for populations of European
ancestry are publicly available for all the autoimmune diseases, NHL, and NHL subtypes that we investigated. Furthermore, using the same population also ensures that the exposure dataset and outcome dataset are as similar as they could be. Second, NHL is a heterogenous group of haematologic disorders, with more than 20 subtypes. Although we used the most robust dataset for NHL subtypes (i.e., FinnGen), different categorisation of NHL subtypes is possible and may yield different results. Third, one autoimmune disease, Behçet disease, had suboptimal statistics for its genetic instruments, which may cause weak instrument bias. However, when we conducted RAPS analysis to address potential bias due to weak instruments, the observed findings were similar to the main analyses. Lastly, given the shared genetic basis of autoimmune diseases and NHL, we were particularly cautious about potential bias due to horizontal pleiotropy by immune function. To minimize the impact of this bias, we excluded SNPs in the HLA region and carried out various sensitivity analyses. We also carried out secondary analyses for T1D and sarcoidosis restricted to the insulin and non-immune pathways, respectively. Although our sensitivity analyses did not suggest important bias due to genetic pleiotropy, it is still possible that potential pleiotropy may partially affect our study findings.

While our study addressed potential residual confounding that may have been present in previous observational studies, the MR design may be biased if the instrumental variable assumptions do not hold. Thus, the association between autoimmune diseases and risk of NHL needs further investigation with evidence triangulation using different datasets, populations, and approaches.

Conclusions
This MR analysis found that genetically predicted susceptibility to T1D, and to some extent genetically predicted susceptibility to sarcoidosis, were associated with a lower risk of NHL. Future research using different datasets, approaches, and populations is necessary to develop a more comprehensive understanding of the associations between T1D and NHL, and sarcoidosis and NHL.
Acknowledgements

Contributors: XS, JDW and TR originally conceived this study. XS and TR designed this study. XS and TR acquired the data. XS and TR conducted the statistical analysis. XS drafted the manuscript. All authors participated in the interpretation of the data. All authors and critically revised the manuscript for important intellectual content. XS and TR had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. TR provided supervision. XS and TR are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding/Support: Dr. Rogne received funding from CTSA Grant Number UL1 TR001863 from the National Center for Advancing Translational Science (NCATS), a component of the NIH. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of NIH. The NIH had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: In the past 36 months, Dr. Wallach reported receiving grant support from the FDA, Arnold Ventures, Johnson & Johnson through Yale University, and the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health (NIH) under award 1K01AA028258; serving as a consultant for Hagens Berman Sobol Shapiro LLP and Dugan Law Firm APLC; and serving as a medRxiv affiliate. Dr. Ma received research funding from the NIH and the Frederick A. DeLuca Foundation and served as a consultant for Bristol Myers Squibb.
Patient consent: Not required

Ethical approval: Not required

Data sharing: The data will be made available via a publicly accessible repository on publication. The code for analyses is included as supplementary information.

Transparency: The senior author (manuscript guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant registered) have been explained.

License: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide license to the Publishers and its licensees in perpetuity, in all forms, formats and median (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) license any third party to do any or all of the above. The default license, a CC BY NC license, is needed. This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

See: http://creativecommons.org/licenses/by-nc/4.0/.
REFERENCES


<table>
<thead>
<tr>
<th>Trait</th>
<th>Study</th>
<th>Countries</th>
<th>Cohorts</th>
<th>Number of cases</th>
<th>Number of controls</th>
<th>Number of SNPs</th>
<th>R² (% of variance explained)</th>
<th>Median F statistics (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behçet disease</td>
<td>Fernández et al 2021¹</td>
<td>West Europe, Italy, Spain</td>
<td>Meta-analysis of three independent cohorts</td>
<td>437</td>
<td>3,325</td>
<td>7</td>
<td>0.3</td>
<td>1 (0.1-7.9)</td>
</tr>
<tr>
<td>Coeliac disease</td>
<td>Dubois et al 2010²</td>
<td>The UK, the USA, Finland, Italy, Netherlands, Poland, Hungary, Spain</td>
<td>Meta-analysis of seven independent cohorts</td>
<td>4,918</td>
<td>5,684</td>
<td>12</td>
<td>1.2</td>
<td>53 (30-96)</td>
</tr>
<tr>
<td>Dermatitis herpetiformis</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen</td>
<td>435</td>
<td>341,188</td>
<td>8*</td>
<td>1.3</td>
<td>22 (20-35)</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Tsoi et al 2017⁴</td>
<td>North America, Sweden</td>
<td>Meta-analysis of eight independent cohorts</td>
<td>11,988</td>
<td>275,335</td>
<td>32</td>
<td>1.8</td>
<td>59 (30-310)</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Ishigaki et al 2022²</td>
<td>The US, the UK, Canada, Spain, Sweden, the Netherlands, France</td>
<td>Meta-analysis of 25 independent cohorts</td>
<td>22,350</td>
<td>74,823</td>
<td>59</td>
<td>2.4</td>
<td>42 (30-762)</td>
</tr>
<tr>
<td>Sarcoidosis</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>3,597</td>
<td>337,121</td>
<td>9</td>
<td>0.7</td>
<td>34 (32-116)</td>
</tr>
<tr>
<td>Sjögren's syndrome</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>2,735</td>
<td>332,115</td>
<td>11*</td>
<td>0.8</td>
<td>23 (21-73)</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>Wang et al 2021¹</td>
<td>Spain, northern and western Europe, Italy</td>
<td>Meta-analysis of three independent cohorts</td>
<td>4,576</td>
<td>8,039</td>
<td>24</td>
<td>2.5</td>
<td>44 (30-268)</td>
</tr>
<tr>
<td>Systemic sclerosis</td>
<td>López-Isac et al 2019⁷</td>
<td>Spain, Germany, the Netherlands, USA, France, Spain, Italy, UK, Sweden, Norway, Australia/UK</td>
<td>Meta-analysis of 14 independent cohorts</td>
<td>9,095</td>
<td>17,584</td>
<td>21</td>
<td>1.3</td>
<td>41 (30-100)</td>
</tr>
<tr>
<td>Type I diabetes</td>
<td>Chiou et al 2021¹</td>
<td>The USA, the UK, Ireland, Finland</td>
<td>Meta-analysis of nine independent cohorts including FinnGen Study</td>
<td>18,942</td>
<td>501,638</td>
<td>72</td>
<td>3.1</td>
<td>45 (30-1077)</td>
</tr>
</tbody>
</table>

Footnote: GWAS: genome-wide association study; ICD-10: International Classification of Diseases, Tenth Revision; NA: not available; the UK: the United Kingdoms; the USA: the United States.

* P < 5 × 10⁻⁶ used as criteria to select genetic instruments.
<table>
<thead>
<tr>
<th>Trait</th>
<th>GWAS data source</th>
<th>Countries</th>
<th>Cohorts</th>
<th>Case definition</th>
<th>Control definition</th>
<th>Number of cases</th>
<th>Number of controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hodgkin lymphoma</td>
<td>Rashkin et al 2020⁹</td>
<td>The UK, the USA</td>
<td>UK Biobank, Kaiser Permanente cohorts</td>
<td>ICD-O-3 codes and SEER cite recode paradigm¹⁰</td>
<td>Individuals with no record of any cancer in the relevant registries</td>
<td>2,400</td>
<td>410,350</td>
</tr>
<tr>
<td>Follicular lymphoma</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>ICD-10: C82</td>
<td>Population controls</td>
<td>955</td>
<td>271,463</td>
</tr>
<tr>
<td>Mature T/NK-cell lymphomas</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>ICD-10: C84</td>
<td>Population controls</td>
<td>296</td>
<td>271,463</td>
</tr>
<tr>
<td>Non-follicular lymphoma</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>ICD-10: C83</td>
<td>Population controls</td>
<td>2,340</td>
<td>271,463</td>
</tr>
<tr>
<td>Other and unspecified types of non-Hodgkin lymphoma</td>
<td>Kurki et al 2023³</td>
<td>Finland</td>
<td>FinnGen Study</td>
<td>ICD-10: C85</td>
<td>Population controls</td>
<td>982</td>
<td>271,463</td>
</tr>
</tbody>
</table>

Footnote: GWAS: genome-wide association study; ICD-10: International Classification of Diseases, Tenth Revision; NA: not available; UK: United Kingdoms; the USA: the United States.
Figure 1. A schematic overview of the study design

Abbreviations: EXP: exposure; HLA: human leukocyte antigen; MR: mendelian randomisation; OUT: outcome; RAPS: robust adjusted profile scores; SNP: single nucleotide polymorphism.

Wald ratio $\beta_{\text{EXP-OUT}} = \frac{\beta_{\text{SNP-OUT}}}{\beta_{\text{SNP-EXP}}}$. The analyses were carried out in three steps: Step 1, genetic instruments for each autoimmune disease were selected from relevant GWASs (i.e., $P < 5 \times 10^{-8}$ and an $R^2 < 0.001$), with $\beta_{\text{SNP-EXP}}$ (i.e., association between genetic instrument and the exposure) extracted. Step 2, for these genetic instruments, the genetic instrument-outcome associations $\beta_{\text{SNP-OUT}}$ were extracted from the outcome GWAS. Step 3, for each autoimmune disease, the Wald ratio (i.e., the causal estimate) was estimated for $k$ number of SNPs by the formula $\beta_{\text{EXP-OUT}} = \frac{\beta_{\text{SNP-OUT}}}{\beta_{\text{SNP-EXP}}}$, and next summarised using the listed MR analyses.

In the secondary analyses, only inverse-variance weighted method was applied.
Figure 2a. Genetically predicted susceptibility to ten autoimmune diseases and risk of non-Hodgkin lymphoma overall

Abbreviations: CI: confidence interval; IVW: inverse-variance weighted; MR: mendelian randomisation; OR: odds ratio; RAPS: robust adjusted profile scores.

Legend: all causal estimates were multiplied by 0.693 and next exponentiated in order to represent the odds ratios (ORs) for NHL per doubling in the prevalence of the autoimmune disease under study.

Figure 2b. Genetically predicted susceptibility to ten autoimmune diseases and risk of non-Hodgkin lymphoma overall
Abbreviations: CI: confidence interval; IVW: inverse-variance weighted; MR: mendelian randomisation; OR: odds ratio; RAPS: robust adjusted profile scores.

Legend: all causal estimates were multiplied by 0.693 and next exponentiated in order to represent the odds ratios (ORs) for NHL per doubling in the prevalence of the autoimmune disease under study.
Figure 3. Inverse-variance weighted analyses of genetically predicted susceptibility to type 1 diabetes, and sarcoidosis, and risk of non-Hodgkin lymphoma subtypes

Abbreviations: CI: confidence interval; NHL: non-Hodgkin lymphoma; OR: odds ratio.

Legend: 1. the analyses were based on inverse-variance weighted approach. 2. all causal estimates were multiplied by 0.693 and next exponentiated in order to represent the odds ratios (ORs) for NHL per doubling in the prevalence of the autoimmune disease under study. 3. As mentioned in Table 2, we used different datasets for NHL and NHL subtypes.
References
