Published in final edited form as:
Arthritis Rheum. 2013 March ; 65(3): 571–581. doi:10.1002/art.37801.

Autoantibodies, autoimmune risk alleles and clinical associations in rheumatoid arthritis cases and non-RA controls in the electronic medical records

Katherine P. Liao, MD, MPH[1], **Fina Kurreeman, PhD**[2,3,4], **Gang Li, PhD**[2], **Grant Duclos, BS**[2], **Shawn Murphy, MD, PhD**[5,7], **P Raul Guzman, BS**[5], **Tianxi Cai, ScD**[6], **Namrata Gupta, PhD**[3], **Vivian Gainer, MS**[5], **Peter Schur, MD**[1], **Jing Cui, PhD**[1], **Joshua C. Denny, MD, MS**[8], **Peter Szolovits, PhD**[9], **Susanne Churchill, PhD**[5,10], **Isaac Kohane, MD, PhD**[10,11], **Elizabeth W. Karlson, MD**[1], and **Robert M. Plenge, MD, PhD**[1,2,3]

[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital [2]Division of Genetics, Brigham and Women's Hospital [3]Program in Medical and Population Genetics, The Broad Institute, Cambridge, MA [4]Department of Rheumatology, Leiden University Medical Center, Leiden, Netherlands [5]Information Systems, Partners Healthcare, Charlestown, MA [6]Department of Biostatistics, Harvard School of Public Health [7]Department of Neurology, Massachusetts General Hospital, Boston, MA [8]Departments of Biomedical Informatics and Medicine, Vanderbilt University, Nashville, TN [9]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA [10]I2b2 Center for Biomedical Computing, Boston, MA [11]Center for Biomedical Informatics, Harvard Medical School, Boston, MA

Abstract

Objectives—The significance of non-RA autoantibodies in patients with rheumatoid arthritis (RA) is unclear. We studied associations between autoimmune risk alleles and autoantibodies in RA cases and non-RA controls, and autoantibodies and clinical diagnoses from the electronic medical records (EMR).

Methods—We studied 1,290 RA cases and 1,236 non-RA controls of European genetic ancestry from the EMR from two large academic centers. We measured antibodies to citrullinated peptides (ACPA), anti-nuclear antibodies (ANA), antibodies to tissue transglutaminase (anti-tTG), antibodies to thyroid peroxidase (anti-TPO). We genotyped subjects for autoimmune risk alleles, and studied the association between number of autoimmune risk alleles and number of types of autoantibodies present. We conducted a phenome-wide association study (PheWAS) to study potential associations between autoantibodies and clinical diagnoses among RA cases and controls.

Results—Mean age was 60.7 in RA and 64.6 years in controls, and both were 79% female. The prevalence of ACPA and ANA was higher in RA cases compared to controls ($p<0.0001$, both); we observed no difference in anti-TPO and anti-tTG. Carriage of higher numbers of autoimmune risk alleles was associated with increasing types of autoantibodies in RA cases ($p=4.4\times10^{-6}$) and controls ($p=0.002$). From the PheWAS, ANA was significantly associated with Sjogren's/siccain RA cases.

Conclusion—The increased frequency of autoantibodies in RA cases and controls was associated with the number of autoimmune risk alleles carried by an individual. PheWAS analyses

Corresponding author: Katherine P. Liao, MD, MPH, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 75 Francis St, PBB-B3, Boston, MA 02115, kliao@partners.org, Phone: 617-732-5325, Fax : 617-713-3030.

within the EMR linked to blood samples provide a novel method to test for the clinical significance of biomarkers in disease.

INTRODUCTION

Presence of anti-citrullinated peptide/protein antibodies (ACPA) is a component of the rheumatoid arthritis (RA) classification criteria(1). ACPA can be found in patient sera more than 10 years prior to the diagnosis of disease(2–5) and portends higher disease severity(6, 7). Other autoantibodies such as anti-nuclear antibodies (ANA), associated with systemic lupus erythematosus (SLE), are found more frequently in RA patients compared to controls(8). However, the prevalence and clinical significance of these autoantibodies in RA is not well established. Moreover, little is known about the factors associated with autoantibody production(9).

The majority of studies utilizing autoantibodies focus on the association between genetic risk alleles and risk of having a disease (e.g., human leukocyte antigen (HLA) genes and ACPA positive RA vs controls)(10, 11), rather than autoantibody formation itself as the outcome of interest. Genetics studies of individuals with ANA, antibodies to thyroid peroxidase (anti-TPO), and antibodies to tissue transglutaminase (anti-tTG) are conducted within the associated autoimmune diseases: systemic lupus erythematosus (SLE), autoimmune thyroiditis (AITD) and celiac disease, respectively. SLE risk alleles are associated with ANA autoantibodies in patients with SLE(12), but the effect of SLE risk alleles on ANA formation in non-SLE patients has not been investigated. The diagnosis of AITD and celiac disease are tightly linked to the presence of anti-TPO and anti-tTG antibodies, and no study has investigated whether AITD and celiac disease risk alleles are associated with the formation of these autoantibodies.

In this study we hypothesize that there is not only a genetic basis for autoantibody formation in patients with RA, but also in controls without clinical evidence of a rheumatic disease. Further, integrating clinical, genetic, and autoantibody data may provide additional information regarding clinical subsets of RA. Specifically, we hypothesize that (1) RA patients will carry more types of autoantibodies than controls and (2) disease-specific risk alleles will be associated with disease-specific autoantibodies (e.g., SLE risk alleles are associated with ANA autoantibodies) in RA patients and controls without having the associated autoimmune disease. Furthermore, we employ a Phenome-wide association study (PheWAS)(13) as an exploratory analysis where we investigate associations between the presence of autoantibodies and autoimmune risk alleles with clinical diagnoses codes ascertained through the EMR.

PATIENTS AND METHODS

Study population

Our study was conducted on 1,290 RA subjects and 1,236 non-RA controls of European ancestry identified from the electronic medical records (EMR) of Brigham and Women's Hospital (BWH) and Massachusetts General Hospital (MGH), characterized in previously published studies from our group(10, 14). RA subjects were identified using a published RA EMR phenotype algorithm with a positive predictive value (PPV) of 94%(14) trained on a gold standard set of subjects classified either as RA or not RA cases by 3 board certified rheumatologists (KPL, EWK, RMP) using the 1987 American College of Rheumatology Classification Criteria for RA(15) as the benchmark. This algorithm was further validated at two other academic institutions(16). Please refer to Liao, et al., 2010 for details on development, training, and validation of this RA phenotype algorithm(14).

The non-RA control group was created from the EMR by excluding all subjects with an *International Classification of Diseases 9th Revision* (ICD9) code for any rheumatic disease in the EMR (this excluded all subjects in the RA cohort); please refer to Kurreeman, et al., 2011 for details(10). The remaining subjects were matched to RA cases (3:1) by age, gender, self-reported ethnicity, and level of health care utilization (represented by the number of 'facts,' or contacts with the health care system, i.e. office visits, laboratory blood draws) (17). For both RA cases and controls, information regarding age, gender, ICD9, laboratory test results and electronic prescriptions for medications were extracted from structured EMR data. Bone erosion information was obtained using natural language processing (NLP) on bone radiology reports from RA cases and controls using Health Information Text Extraction (HITex) system(14, 18).

Discarded blood samples from five clinical laboratories at Partners Healthcare (Boston, USA) were collected by the BWH Clinical Specimen Bank from 2009–2010, using an Institutional Review Board (IRB) approved process, as described in Kurreeman, et al., 2010(10). The final RA cases and non-RA control populations analyzed for this study were conducted in those where blood samples were obtained and were of European ancestry determined by ancestry informative markers (AIMs). As a result the RA cases and controls were no longer perfectly matched.

Genotyping

Detailed methods for genotyping and assigning genetic ancestry for the RA case and the non-control groups can be found in Kureeman, et al., 2010(10). Briefly, processing and genotyping of the discarded blood samples was performed at the Broad Institute Broad Institute (Cambridge, MA, USA). We genotyped 192 ancestry informative markers (AIMs), 28 single nucleotide polymorphisms (SNPs) associated with RA, 33 SNPs associated with SLE, and 16 SNPs associated with celiac disease (Supplementary Table 2)(19–24). For quality control, we removed SNPs with missing genotype rate >10% and minor allele frequency <1%. Genetic ancestry using the AIMs was determined using the Bayes classifier and principal components analysis.

Aggregate Genetic Risk Scores (GRS)

We calculated a cumulative aggregate genetic risk score for RA, SLE and celiac for each individual using the following formula(10, 25, 26):

$$GRS = \sum_{i}^{n} X_i$$

Where n is the number of SNPs for the particular disease (RA, SLE, celiac) (Supplementary Table 1), i is the SNP, X_i is the number of risk alleles (0, 1, or 2).

The RA GRS excludes the *HLA-DRB1*04* tag SNP because we were interested in understanding the effects of non-HLA risk alleles and production of ACPA in RA. In addition, the associations in HLA region are complex and require dense genotyping not available in this study(27). We created a combined autoimmune (AI) GRS which consists of all risk alleles in the study with the exception of SNPs in linkage disequilibrium with another SNP (Supplementary Table 1). All GRSs were unweighted due to absence of information on the strength of association for any individual risk allele and autoantibody outcome. The literature for AITD was less definitive(28) and we therefore did not construct a GRS for AITD.

Autoantibody measurement

We measured ACPA using the INOVA CCP3 IgG ELISA, ANA using INOVA Quanta-Lite ANA, anti-TPO using INOVA Quanta-Lite TPO, and anti-tTG IgA using the INOVA Quanta-Lite IgA TTG kits. We determined positivity of an autoantibody based on the manufacturer cut-offs: ACPA ≥ 20 units, ANA ≥ 20 units (high titer positive (ANAht) >60 units), anti-TPO >100 WHO units, anti-tTG ≥ 20 units. These autoantibodies were selected because of the relationship between each autoimmune disease and RA in both epidemiologic(29, 30) and genetic studies(31–33). ANA, anti-TPO and anti-tTG antibodies were measured in all RA cases and controls. ACPA was measured in all RA cases and 202 controls which provided adequate power to detect a difference in ACPA prevalence between the two groups (Supplementary Table 2).

Statistical analyses

Autoantibody prevalence—We first determined the distribution and overlap of ACPA, ANA, anti-TPO, and anti-tTG among RA cases and non-RA controls. We then constructed separate logistic regression models for each of the autoantibodies to study the association between each autoantibody (ACPA, ANA, anti-TPO, anti-tTG) and RA case or control status, adjusted by age, gender and health care utilization (represented by the number of facts). We then conducted an RA case-only analysis examining whether the prevalence of ANA, anti-TPO and anti-tTG differed with ACPA status using logistic regression models, adjusted by age, gender and health care utilization.

Association between GRSs and autoantibody positivity—We constructed separate logistic regression models to determine the association between the disease specific GRSs (RA, SLE, celiac) and their related autoantibodies (ACPA, ANA, and anti-tTG, respectively) for RA cases. In controls, we tested only the association between the SLE GRS and ANA, due to the low prevalence of ACPA and anti-tTG (post-hoc power calculations, Supplementary Table 2). Linear regression was used to assess the association between the AI GRS and the number of types autoantibodies present (ANA, anti-TPO, anti-tTG) for both RA cases and controls.

We employed multivariate logistic regression models to determine the relative importance of age, gender and disease specific GRS in predicting autoantibody positivity among RA cases. In controls we tested only the association between the SLE GRS and ANA, due to the power issues stated above.

To test whether SLE risk alleles have the same effect on ANA production on RA cases and controls, we categorized all subjects into SLE GRS tertiles (tertile 1: SLE GRS ≥0 and ≤33.36, tertile 2: SLE GRS>33.36 and ≤40.90, tertile 3: SLE GRS ≥ 40.9). Within each tertile, we fitted logistic regression models to estimate the odds of ANA positivity in RA cases compared to non-RA controls given a similar number of SLE risk alleles. Each model was adjusted by age, gender and health care utilization. A subgroup analysis was conducted only in subjects without a prescription for a TNFi because TNFi's have been associated with conversion to ANA positivity(34).

Phenome-wide analysis (PheWAS) for association of autoantibodies and autoimmune risk alleles with clinical diagnoses—To discover potential associations between non-RA autoantibody positivity and a clinical diagnosis, and the AI GRS and a clinical diagnosis, we conducted a PheWAS. Clinical diagnoses were defined using ICD9 codes manually grouped into clinically relevant diseases ("PheWAS code"; e.g. ICD9 codes 401–405, representing different types of hypertension and complications are grouped into a "hypertension" PheWAS code) by a team of physicians in a previously published

PheWAS(13). We included all PheWAS codes used with a prevalence of ≥1% in each group, yielding 512 diagnoses codes for RA cases and 698 diagnoses codes for controls for analysis.

We tested for potential associations between an autoantibody and PheWAS codes in RA cases and non-RA controls separately by fitting logistic regression models using ACPA, ANA, or anti-TPO status as the predictor (adjusted by age, gender) and the presence or absence (0 or 1) of the PheWAS code as the outcome. We did not study anti-tTG in cases and controls and ACPA in controls due to the low prevalence (≤1%) which would lead to unstable estimates. To study the associations between the AI GRS and PheWAS codes, we fitted logistic regression models using the AI GRS as the predictor (adjusted by age, gender) and the presence or absence of the PheWAS code as the outcome. Since multiple comparisons are applied in the PheWAS, suggestive associations can be observed due to chance. Therefore, we set the p-value cut-off for a significant association using the Bonferroni correction: RA cases, significant p-value, $0.05/512=9.76\text{x}10^{-5}$; non-RA control group, significant p-value, $0.05/698=7.16\text{x}10^{-5}$.

For significant associations between an autoantibody and PheWAS code after Bonferroni correction, we reviewed the medical records to determine the diagnostic accuracy of the code. For example, if autoantibody_b was significantly associated with PheWAS_code_b in RA cases, we randomly selected 20 RA cases who had 1 PheWAS_code_b and reviewed their medical record for clinical documentation of this diagnosis by their treating physician (reviewed by KPL). This method was applied for each PheWAS code with a significant association after Bonferroni correction with an autoantibody. We reported the percentage of confirmed diagnoses as the PPV of the PheWAS code.

This study was approved by the Partners' IRB. Statistical analyses were conducted using the SAS and R 2.10 software packages.

RESULTS

Study population and autoantibody prevalence

We studied 1,290 RA cases and 1,236 non-RA controls of European ancestry with both EMR clinical data and DNA/plasma from discarded blood samples. The mean age was 60.7 in the RA cases and 64.6 years in the controls. Both groups were 79% female. The clinical characteristics of RA subjects included: 70.3% ACPA positive, 61.6% bone erosions, 62.3% with ≥1 methotrexate and 40.9% with ≥1 tumor necrosis inhibitor (TNFi) electronic prescription at some point during their treatment (Table 1). ACPA was the most prevalent autoantibody in RA (70%) and the least prevalent among the controls (0.5%) (Table 1). Figure 1 demonstrates the distribution and overlap autoantibodies in RA cases and controls.

RA cases were more likely to be ANA positive compared to non-RA controls (Figure 2A) without significant differences in the prevalence of anti-TPO and anti-tTG antibodies in RA cases compared to controls. In an RA case only analysis, we found that ACPA-positive RA cases were more likely to be ANA positive than ACPA-negative cases without significant differences in prevalence for anti-TPO and anti-tTG antibodies (Figure 2B).

Association between GRSs and autoantibody positivity

To determine the genetic basis of autoantibodies in RA cases, we tested RA(23), SLE(19, 35) and celiac disease(24, 36) genetic risk alleles for association with disease-specific autoantibodies (ACPA, ANA and anti-tTG, respectively). We found that the disease-specific GRSs predicted the presence of the disease-specific autoantibodies among RA cases (Figure 3A–C). As we have shown previously for ACPA positivity in RA cases(10) the mean RA

GRS was significantly higher in ACPA positive than in ACPA negative subjects (P=2.5x10^{-9}) (Figure 3A). Notably, this RA GRS does not include the HLA shared epitope alleles. As shown in Figure 3B, we found that ANA positive RA cases had a significantly higher mean SLE GRS than ANA negative cases (P=8.0x10^{-4}). Anti-tTG RA cases also had a significantly higher mean celiac GRS than anti-tTG negative subjects (P=0.03) (Figure 3C).

In RA cases, the associations between the disease specific GRSs and autoantibody status did not change after adjusting for age, gender and health care utilization in a multivariable logistic regression model, nor were age and gender significantly associated with autoantibody status in the model (data not shown).

In controls, there was no significant association between the SLE GRS and ANA status, although the trend was similar to that in RA cases: ANA positive controls had a trend for higher mean SLE GRS than ANA negative controls (P=0.31) (Supplementary Figure 1A). In contrast to RA cases, age and gender were significantly associated with ANA positivity in controls (age per year: OR 1.01 (95% CI 1.002, 1.03, p=0.02); female gender: OR 1.7, (95% CI 1.2, 2.75, p=0.007) while the SLE GRS was not (p=0.34).

We also combined all non-RA autoimmune risk alleles into a single GRS (AI GRS), and tested for association with the number of autoantibodies present (with and without ACPA). In RA cases, the increasing count of any autoimmune risk allele (AI GRS) was significantly associated with increasing types of any autoantibody (range 0–4, representing ACPA, ANA, anti-TPO, or anti-tTG) (P= 3.2x10^{-8}). The association remained with the removal of ACPA as a potential outcome: an increasing AI GRS was associated with increasing types of any autoantibody (range 0–3, representing ANA, anti-TPO and anti-tTG) in RA cases (P= 2.6x10^{-5}) (Figure 3D). In controls, increasing types of any autoantibody (range 0–3, representing ANA, anti-TPO, anti-tTG) was also significantly associated with the AI GRS (P=4.0x10^{-3}) (Supplementary Fig 1B). The association between the AI GRS and types of autoantibodies (range 0–3, representing ANA, anti-TPO, anti-tTG) remained significant in RA cases (P= 2.1x10^{-5}) and controls (p=5.0x10^{-3}) after adjusting for age, gender and health care utilization.

One possible explanation for the association between ANA-positivity and RA case status (Figure 2A) is that RA cases carry more SLE risk alleles than controls. While RA cases carried significantly more SLE risk alleles than non-RA controls [mean SLE GRS: RA cases 38.6 (SD 5.3); controls 37.8 (SD 5.4), p=4.0x10^{-4}], we found that RA cases still had a higher rate of ANA-positivity than non-RA controls after controlling for the effect of SLE risk alleles (Figure 4). More specifically, when comparing RA cases and controls with similar numbers of SLE risk alleles (grouped by tertile of SLE GRS), RA cases were still significantly more likely to be ANA positive. A subgroup analysis comparing RA cases and controls without a TNFi prescription electronically prescribed through the EMR (RA cases, n=762; controls, n=1228) resulted in similar findings: RA subjects were significantly more likely to be ANA positive given the same number of SLE risk alleles as controls (p<0.0001, all tertiles).

Phenome-wide analysis (PheWAS) for association of autoantibodies and autoimmune risk alleles with clinical diagnoses

To determine whether autoantibodies or a higher number of autoimmune risk alleles were associated with codified clinical diagnoses, we performed an exploratory PheWAS in our EMR. From the PheWAS, we observed that in both RA cases and controls, the presence of anti-TPO was significantly associated with hypothyroidism (RA cases, p=1.22x10^{-16}; controls, p=9.22x10^{-10}), which was expected given the known association between anti-

TPO and autoimmune thyroid disease (Table 2). The PPV of a PheWAS code for 'acquired hypothyroidism' was 75% for RA cases and 100% for controls on medical record review.

Two PheWAS codes were significantly associated with ANA high titer positivity in RA cases or in controls. In RA cases, the PheWAS code of Sjogren's or sicca syndrome diagnoses was associated with high-titer ANA positivity ($p=8.59x10^{-6}$) (Table 2). The PPV of the PheWAS code for 'Sjogren's/sicca' was 70%. In controls, we observed a significant association between ANA high titer positivity and chronic non-alcoholic liver disease ($p=2.9x10^{-6}$) (Table 2). The PPV of the PheWAS code for 'other chronic non-alcoholic liver disease' was 75% based on medical record review for documentation of liver disease not associated with alcohol.

We found no significant association between ACPA-positivity and PheWAS categories among RA cases (Supplementary Table 3). Furthermore, we observed no significant associations between carriage of higher numbers of autoimmune risk alleles in aggregate using the AI GRS, and specific a PheWAS code corresponding to a clinical diagnosis in the RA cases or controls.

DISCUSSION

RA diagnosis and classification is based on a distinct clinical phenotype(37). Underlying the clinical phenotype of RA are variations in numbers of autoimmune risk alleles and differential production of RA and non-RA autoantibodies. We observed: (1) RA patients have a higher prevalence of ANA, with a trend towards higher anti-tTG (but not anti-TPO) compared to non-RA controls; (2) SLE and celiac disease genetic risk alleles influence the production of the SLE and celiac disease-related autoantibodies (ANA and anti-tTG, respectively) among RA cases; (3) the number of autoimmune risk alleles carried by a patient influence the number of types of autoantibodies carried in both RA cases and controls; (4) given a similar number of SLE risk alleles, RA subjects are more likely to be ANA positive than controls; and (5) applying the PheWAS, we observed that ANA positivity was associated with a diagnosis of Sjogen's syndrome or sicca among RA cases.

To our knowledge, this is the largest study that systematically tested and compared the prevalence of clinical autoantibodies (ACPA, ANA, anti-TPO and anti-tTG) in RA cases and non-RA controls. For the purposes of our study, we defined autoantibodies as being "disease-specific", linking ACPA with RA and linking other "non-RA autoantibodies" with specific autoimmune diseases: ANA (SLE), anti-tTG (celiac disease), and anti-TPO (AITD). However, we show that this definition is not precise, as "non-RA autoantibodies" are actually more common in RA patients compared to controls (i.e., ANA).

We propose two explanations regarding why RA patients are at an increased risk of disease-specific autoantibodies compared to non-RA controls. Under a purely genetic model, RA patients have a higher prevalence of autoantibodies because they have a higher burden of autoimmune risk alleles. The increased frequency of ANA in RA may be due to the fact that RA and SLE share risk alleles(31, 33, 38), leading to a larger burden of SLE risk alleles in RA patients. Supporting this, we find that RA patients have a significantly higher mean SLE GRS than controls. A consequence of the increased burden of autoimmune risk alleles is that these risk alleles may also have general effects on self-tolerance. For example, the *PTPN22* missense allele (*R620W),* a shared genetic risk factor for RA, SLE and AITD, has been shown to cause defects in the counterselection of autoreactive B cells(31, 39, 40) that lead to production of autoantibodies.

The second explanation, not mutually exclusive with the genetic model, is one of differential environmental and endogenous (immune dysregulation) exposures associated with RA(41).

These exposures may render RA patients more susceptible to the downstream effects of carrying additional autoimmune risk alleles by having a lower threshold for producing autoantibodies than controls. However, if general immune dysregulation alone were responsible for autoantibody formation, we would not expect the specific associations we observed between the SLE and celiac disease risk alleles with ANA and anti-tTG autoantibodies, respectively, in RA cases. In addition, we would expect to see an increased frequency of all autoantibodies; anti-TPO had the same prevalence in RA cases and non-RA controls.

Integrating the two explanations above, we interpret the similarities and differences between autoantibody prevalence in RA cases compared to non-RA controls in the following way. The common theme shared by RA cases and controls was that an increasing burden of autoimmune risk alleles was associated with carriage of more types of any autoantibody (Figure 3D, Supplementary Figure 1B). Autoimmune risk alleles therefore provide the substrate for autoantibody production. However, it is clear that additional factors contribute to increased autoantibody production. For example, among subjects carrying a similar number of SLE risk alleles, RA cases were more likely to be ANA positive than controls (Figure 4). We hypothesize that in RA cases, the effect of inherited and environmental factors that leads to a breakdown in self-tolerance and clinical symptoms of RA also predisposes to the development of additional autoantibodies such as ANA. In controls, there is less background immune dysregulation, and the same autoimmune risk alleles, are less important for autoantibody production. Thus, we posit that autoimmune risk alleles predispose both RA cases and controls to developing autoantibodies, but that environmental factors and endogenous exposures (immune dysregulation) that are also important contributors.

A prediction of our integrated model is that carriage of autoimmune risk alleles leading to immune dysregulation will have phenotypic consequences beyond RA case-control status. To investigate this hypothesis systematically, we utilized EMR clinical data to determine which phenotypes may be associated with specific autoantibodies or higher numbers of autoimmune risk alleles. This approach, termed PheWAS, has been applied to test associations between SNPs and diagnosis codes(13, 42). However, this approach has not yet been used to test for clinical associations with autoantibodies or GRSs. From our exploratory PheWAS analysis, we observed, as expected, that the presence of anti-TPO was strongly associated with hypothyroidism in both RA cases and non-RA controls. We also observed associations between ANA and Sjogren's/sicca in RA cases. While this finding is consistent with clinical teaching regarding the association between RA and Sjogren's syndrome(43), this association has not been definitively shown until now. Among RA cases, ACPA status was not associated with RA because virtually all patients had a diagnosis code for RA.

Our study design represents a novel approach to translational research. The RA case and non-RA control study population was created with anonymized clinical data from the EMR(14) and linked to a biospecimen repository of discarded blood samples. Clinical information was obtained from the structured EMR (i.e. age, gender, diagnosis codes). This EMR platform afforded us the opportunity to understand the relationship between clinical, genetic factors and autoantibodies in RA, and to compare these findings with a control group. Moreover, we had the ability to integrate clinical EMR data and our research laboratory based autoantibody and genetic data, for thousands of patients to determine significant associations between autoimmune risk alleles and autoantibodies, and the potential clinical relevance of autoantibodies or a high number of autoimmune risk alleles.

There are important limitations of our study. We focused our study on a subset of autoimmune diseases, limiting the generalizeability of our conclusions. In the selection of our controls, we excluded those with any ICD9 code for a rheumatic disease(10). Thus the controls may also have a lower prevalence of autoantibodies than general population controls, however we found that the prevalence of ACPA, ANA and anti-TPO observed in our study were comparable to controls in other studies(8, 44–46). Because we do not have temporal information relating to when a patient developed an autoantibody, we cannot state that an autoantibody predicted a particular outcome. Similarly, we cannot study whether an environmental exposure (i.e., smoking) contributed to autoantibody development.

We chose to use ACPA to subset RA rather than RF. The correlation between these two autoantibodies is high but discrepancies are known(47–49). Therefore, whether these results would directly apply to RF positive vs RF negative RA cases is unclear. There is concern that RF can lead to false positive results in multiplex antibody assays(50), however we found no evidence that RF results in false positive readings for the standard commercial ELISAs used in this study. Our findings suggest that RF interference was not a major issue; among RA cases, the prevalence of autoantibodies in ACPA positive compared to ACPA negative subjects was similar with the exception of ANA. Finally, results stemming from the PheWAS approach are exploratory and require further refinement in terms of validation of outcomes (i.e. ICD9 diagnostic codes), covariates and determining the metrics of what constitutes a true association (e.g., p-value threshold, effect size).

In conclusion, we employed a novel EMR-based approach to test genetic risk factors for association with clinical autoantibodies. Further, we provide insight into the immunologic heterogeneity underlying the clinical entity of RA. In the context of a linked clinical EMR-research laboratory database, we believe the PheWAS approach can be a powerful hypothesis generating tool for uncovering associations that are not readily apparent with our current knowledge of pathways and mechanisms of disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Kurt Bloch, Massachusetts General Hospital, for his advice on selection and testing of autoantibodies in this study, and Dr. Lynn Bry, Director of the BWH Clinical Specimen Bank.

This project was funded by the NIH grant U54-LM008748. KPL is supported by the NIH K08-AR060257, Katherine Swan Ginsburg Fund. EWK is supported by K24-AR052403, R01-AR049880 and P60-AR047782. RMP is supported by grants from the NIH (R01-AR057108, R01-AR056768, U01-GM092691, R01-AR059648), and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund.

References

1. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis and rheumatism. 2010; 62(9):2569–81. [PubMed: 20872595]
2. Chibnik LB, Mandl LA, Costenbader KH, Schur PH, Karlson EW. Comparison of threshold cutpoints and continuous measures of anti-cyclic citrullinated peptide antibodies in predicting future rheumatoid arthritis. The Journal of rheumatology. 2009; 36(4):706–11. [PubMed: 19228654]
3. Nielen MM, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MH, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of

serial measurements in blood donors. Arthritis and rheumatism. 2004; 50(2):380–6. [PubMed: 14872479]

4. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet. 2009; 373(9664):659–72. [PubMed: 19157532]
5. Majka DS, Deane KD, Parrish LA, Lazar AA, Baron AE, Walker CW, et al. Duration of preclinical rheumatoid arthritis-related autoantibody positivity increases in subjects with older age at time of disease diagnosis. Annals of the rheumatic diseases. 2008; 67(6):801–7. [PubMed: 17974596]
6. Forslind K, Ahlmen M, Eberhardt K, Hafstrom I, Svensson B. Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: role of antibodies to citrullinated peptides (anti-CCP). Annals of the rheumatic diseases. 2004; 63(9):1090–5. [PubMed: 15308518]
7. van der Helm-van Mil AH, Detert J, le Cessie S, Filer A, Bastian H, Burmester GR, et al. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: moving toward individualized treatment decision-making. Arthritis and rheumatism. 2008; 58(8): 2241–7. [PubMed: 18668546]
8. Solomon DH, Kavanaugh AJ, Schur PH. Evidence-based guidelines for the use of immunologic tests: antinuclear antibody testing. Arthritis and rheumatism. 2002; 47(4):434–44. [PubMed: 12209492]
9. Plagnol V, Howson JM, Smyth DJ, Walker N, Hafler JP, Wallace C, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. PLoS genetics. 2011; 7(8):e1002216. [PubMed: 21829393]
10. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V, et al. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. American journal of human genetics. 2011; 88(1):57–69. [PubMed: 21211616]
11. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010; 376(9746):1094–108. [PubMed: 20870100]
12. Chung SA, Taylor KE, Graham RR, Nititham J, Lee AT, Ortmann WA, et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. PLoS genetics. 2011; 7(3):e1001323. [PubMed: 21408207]
13. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics (Oxford, England). 2010; 26(9):1205–10.
14. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis care & research. 2010; 62(8):1120–7. [PubMed: 20235204]
15. Arnett FC. Revised criteria for the classification of rheumatoid arthritis. Bulletin on the rheumatic diseases. 1989; 38(5):1–6. [PubMed: 2679945]
16. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc. 2012; 19(e1):e162–e9. [PubMed: 22374935]
17. Sperrin, M.; Thew, S.; Weahterall, J.; Dixon, W.; Buchan, I. Quantifying the longitudinal value of healthcare record collections for pharmacoepidemiology. American Medical Informatics Association (AMIA) 35th Annual Symposium; 2011. p. 1318-25.
18. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R, et al. Instrumenting the health care enterprise for discovery research in the genomic era. Genome research. 2009; 19(9):1675–81. [PubMed: 19602638]
19. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nature genetics. 2009; 41(11):1228–33. [PubMed: 19838195]
20. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nature genetics. 2008; 40(2): 204–10. [PubMed: 18204446]

21. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40(2):211–6. [PubMed: 18204447]
22. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008; 359(26):2767–77. [PubMed: 19073967]
23. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature genetics. 2010; 42(6):508–14. [PubMed: 20453842]
24. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nature genetics. 2007; 39(7):827–9. [PubMed: 17558408]
25. Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, Ding B, et al. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Annals of the rheumatic diseases. 2010; 69(6):1077–85. [PubMed: 20233754]
26. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V, et al. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. American journal of human genetics. 88(1):57–69. [PubMed: 21211616]
27. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature genetics. 2012; 44(3):291–6. [PubMed: 22286218]
28. Tomer Y. Genetic susceptibility to autoimmune thyroid disease: past, present, and future. Thyroid. 2010; 20(7):715–25. [PubMed: 20604685]
29. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. Journal of autoimmunity. 2009; 33(3–4):197–207. [PubMed: 19819109]
30. Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune diseases co-occurring within individuals and within families: a systematic review. Epidemiology (Cambridge, Mass. 2006; 17(2):202–17.
31. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. American journal of human genetics. 2005; 76(4):561–71. [PubMed: 15719322]
32. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews. 2009; 10(1):43–55.
33. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS genetics. 7(8):e1002254. [PubMed: 21852963]
34. De Rycke L, Baeten D, Kruithof E, Van den Bosch F, Veys EM, De Keyser F. Infliximab, but not etanercept, induces IgM anti-double-stranded DNA autoantibodies as main antinuclear reactivity: biologic and clinical implications in autoimmune arthritis. Arthritis and rheumatism. 2005; 52(7): 2192–201. [PubMed: 15986349]
35. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS genetics. 2011; 7(2):e1001311. [PubMed: 21379322]
36. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nature genetics. 2008; 40(4):395–402. [PubMed: 18311140]
37. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis and rheumatism. 2010; 62(9):2569–81. [PubMed: 20872595]
38. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes and immunity. 2005; 6(6):459–61. [PubMed: 15875058]

39. Menard L, Saadoun D, Isnardi I, Ng YS, Meyers G, Massad C, et al. The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans. The Journal of clinical investigation.
40. Reveille JD. The genetic basis of autoantibody production. Autoimmunity reviews. 2006; 5(6): 389–98. [PubMed: 16890892]
41. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011; 365(23): 2205–19. [PubMed: 22150039]
42. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. American journal of human genetics. 2011; 89(4): 529–42. [PubMed: 21981779]
43. Theander E, Jacobsson LT. Relationship of Sjogren's syndrome to other connective tissue and autoimmune disorders. Rheumatic diseases clinics of North America. 2008; 34(4):935–47. viii–ix. [PubMed: 18984413]
44. Aggarwal R, Liao K, Nair R, Ringold S, Costenbader KH. Anti-citrullinated peptide antibody assays and their role in the diagnosis of rheumatoid arthritis. Arthritis and rheumatism. 2009; 61(11):1472–83. [PubMed: 19877103]
45. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). The Journal of clinical endocrinology and metabolism. 2002; 87(2):489–99. [PubMed: 11836274]
46. Li QZ, Karp DR, Quan J, Branch VK, Zhou J, Lian Y, et al. Risk factors for ANA positivity in healthy persons. Arthritis research & therapy. 2011; 13(2):R38. [PubMed: 21366908]
47. Bizzaro N, Tonutti E, Tozzoli R, Villalta D. Analytical and diagnostic characteristics of 11 2nd- and 3rd-generation immunoenzymatic methods for the detection of antibodies to citrullinated proteins. Clin Chem. 2007; 53(8):1527–33. [PubMed: 17586589]
48. Schellekens GA, Visser H, de Jong BA, van den Hoogen FH, Hazes JM, Breedveld FC, et al. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis and rheumatism. 2000; 43(1):155–63. [PubMed: 10643712]
49. Lee DM, Schur PH. Clinical utility of the anti-CCP assay in patients with rheumatic diseases. Annals of the rheumatic diseases. 2003; 62(9):870–4. [PubMed: 12922961]
50. Todd DJ, Knowlton N, Amato M, Frank MB, Schur PH, Izmailova ES, et al. Erroneous augmentation of multiplex assay measurements in patients with rheumatoid arthritis due to heterophilic binding by serum rheumatoid factor. Arthritis and rheumatism. 2011; 63(4):894–903. [PubMed: 21305505]

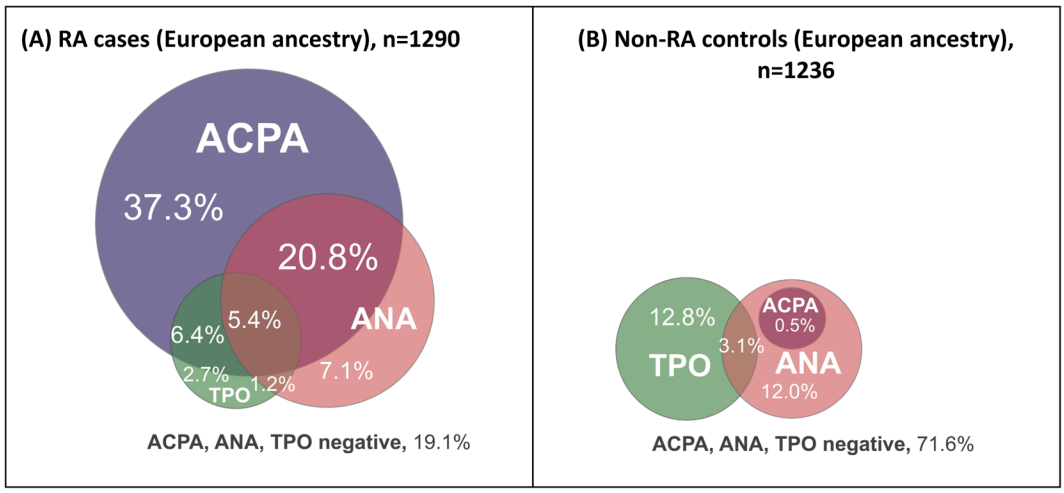
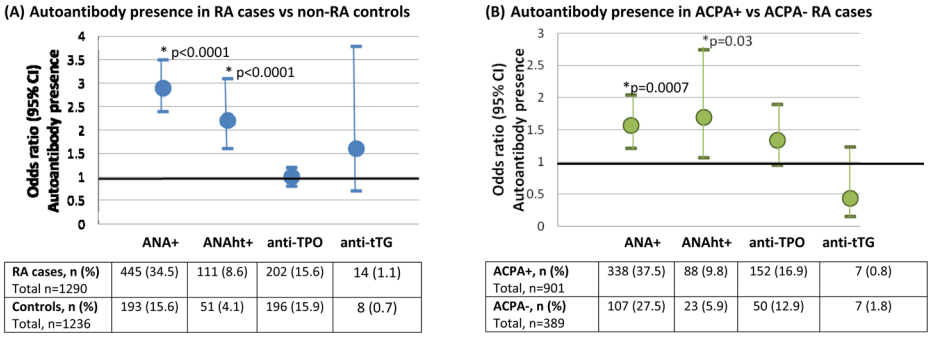


Figure 1.

(A) Distribution of autoantibodies (ACPA, ANA, anti-TPO) among RA cases of EU ancestry; **(B)** Distribution of autoantibodies (ACPA*, ANA, anti-TPO) among non-RA controls of EU ancestry. Anti-tTG (not shown), RA cases: 14 subjects (1.1%); non-RA controls: 8 subjects (0.6%). [*ACPA checked in n=202 controls with prevalence of 0.5%].



RA cases, n (%) Total n=1290	445 (34.5)	111 (8.6)	202 (15.6)	14 (1.1)
Controls, n (%) Total, n=1236	193 (15.6)	51 (4.1)	196 (15.9)	8 (0.7)

ACPA+, n (%) Total, n=901	338 (37.5)	88 (9.8)	152 (16.9)	7 (0.8)
ACPA-, n (%) Total, n=389	107 (27.5)	23 (5.9)	50 (12.9)	7 (1.8)

Figure 2.

Comparison of autoantibody presence between **(A)** RA cases and non-RA controls, and **(B)** ACPA positive compared to ACPA negative RA cases (adjusted by age, gender and health care utilization; European ancestry).

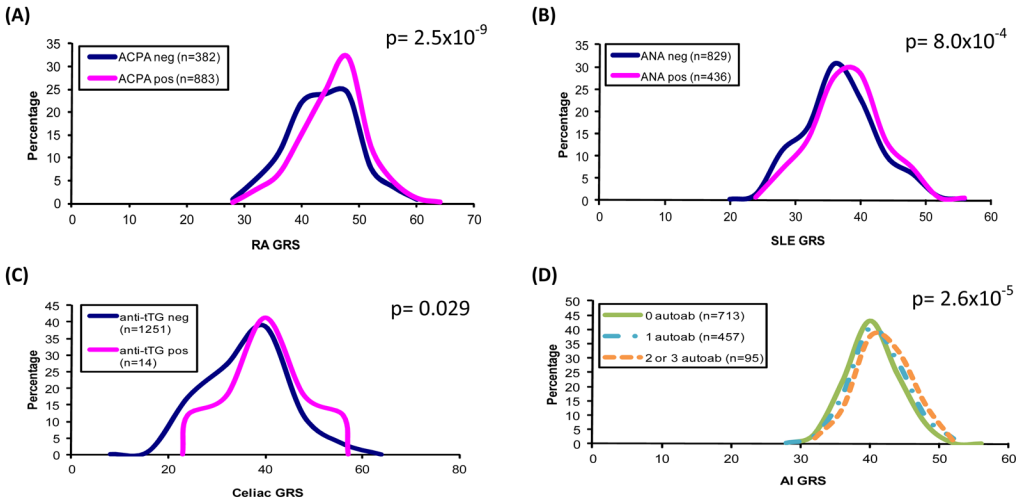
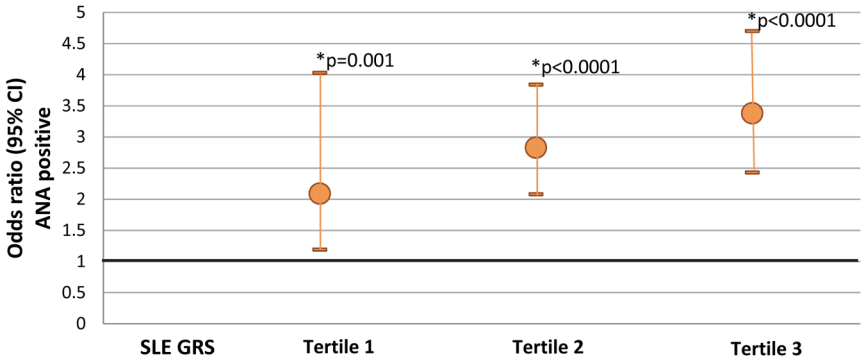


Figure 3.

Distribution of disease specific genetic risk scores and associated autoantibodies among RA cases (European ancestry, n=1265). **(A)** Distribution of RA GRS among ACPA positive and ACPA negative RA cases, **(B)** Distribution of the SLE GRS among ANA positive and ANA negative RA cases, **(C)** Distribution of the celiac GRS among anti-tTG positive and anti-tTG negative RA cases, **(D)** Distribution of AI GRS with increasing count of autoantibodies in RA cases (ANA, anti-TPO or anti-tTG).



ANA positive (%) (RA cases / non-RA controls)	27.4 / 14.2	32.8 / 15.4	40.2 / 16.9
RA cases (n)	245	555	465
Non-RA controls (n)	302	520	403

Figure 4.

ANA positivity in RA cases compared to non-RA controls among individuals with similar numbers of SLE risk alleles (SLE GRS categorized by tertiles from lowest, tertile 1 to highest, tertile 3; ORs adjusted by age, gender, and health care utilizationin subjects of European ancestry).

Table 1

Characteristics of RA cases and non-RA controls.

Characteristics	RA cases, n=1290	Controls, n=1236
Mean age (SD)	60.7 (13.6)	64.6 (13.4)
Female, n (%)	1024 (79.4)	965 (78.7)
Methotrexate, ever use, n (%)	803 (62.3)	15 (1.2)
TNFi, ever use, n (%)✦	528 (40.9)	8 (0.65)
Bone erosions, n (%)	795 (61.6)	69 (5.6)
Measured autoantibodies, n (%)		
ACPA	901 (70.3)	1* (0.5)
ANA	445 (34.5)	193 (15.6)
ANAht✦	111 (8.6)	51 (4.1)
Ant-TPO	202 (15.6)	196 (15.9)
anti-tTG	14 (1.1)	8 (0.7)

*out of 202 controls

✦ANAht= manufacturer high titer cutoff >60 units

Table 2

Results of PheWAS analyses with significant associations between autoantibodies and EMR diagnoses among **(A)** RA cases (n=1265) and **(B)** non-RA controls (n=1225) of European ancestry.*

(A) RA cases

Autoantibody	Outcome/diagnosis	Prevalence of diagnosis		Odds ratio	(95% CI)	p-value
		Autoab pos(%)	Autoab neg (%)			
Anti-TPO	Acquired hypothyroidism	40.1	14.3	4.2	3.0, 5.9	$1.2x10^{-16}$
ANAht⅄	Sjogren's/sicca	13.5	3.5	4.2	2.2, 7.9	$8.6x10^{-6}$

(B) Non-RA controls

Autoantibody	Outcome/diagnosis	Prevalence of diagnosis		Odds ratio	(95% CI)	p-value
		Autoab pos(%)	Autoab neg (%)			
Anti-TPO	Acquired hypothyroidism	48.5	24.1	2.8	2.0, 3.8	$9.2x10^{-10}$
	Thyroiditis	12.2	2.7	4.7	2.7, 8.4	$1.2x10^{-7}$
ANAht	Other chronic non-alcoholic liver disease	11.7	1.7	8.0	3.0, 21.1	$2.9x10^{-5}$

*Please see Supplementary Table 4 for top 3 results for all analyses

⅄Using ANA high titer manufacturer cutoff

Note: Bonferroni significant p-value in cases, $p<9.76x10^{-5}$, and in non-RA controls, $p<7.16x10^{-5}$