ARTHRITIS & RHEUMATOLOGY
Vol. 69, No. 4, April 2017, pp 742–749
DOI 10.1002/art.39974

Phenome-Wide Association Study of Autoantibodies to Citrullinated and Noncitrullinated Epitopes in Rheumatoid Arthritis

Katherine P. Liao,[1] Jeffrey A. Sparks,[1] Boris P. Hejblum,[2] I-Hsin Kuo,[3] Jing Cui,[1] Lauren J. Lahey,[4] Andrew Cagan,[5] Vivian S. Gainer,[5] Weidong Liu,[6] T. Tony Cai,[7] Jeremy Sokolove,[4] and Tianxi Cai[2]

Objective. **Patients with rheumatoid arthritis (RA) develop autoantibodies against a spectrum of antigens, but the clinical significance of these autoantibodies is unclear. Using a phenome-wide association study (PheWAS) approach, we examined the association between autoantibodies and clinical subphenotypes of RA.**

Methods. **This study was conducted in a cohort of RA patients identified from the electronic medical records (EMRs) of 2 tertiary care centers. Using a published multiplex bead assay, we measured 36 autoantibodies targeting epitopes implicated in RA. We extracted all International Classification of Diseases, Ninth Revision (ICD-9) codes for each subject and grouped them into disease categories (PheWAS codes), using a published method. We tested for the association of each autoantibody (grouped by the targeted protein) with PheWAS codes. To determine significant associations (at a false discovery rate [FDR] of ≤0.1), we reviewed the medical records of 50 patients with each PheWAS code to determine positive predictive values (PPVs).**

Results. **We studied 1,006 RA patients; the mean ± SD age of the patients was 61.0 ± 12.9 years, and 79.0% were female. A total of 3,568 unique ICD-9 codes were grouped into 625 PheWAS codes; the 206 PheWAS codes with a prevalence of ≥3% were studied. Using the PheWAS method, we identified 24 significant associations of autoantibodies to epitopes at an FDR of ≤0.1. The associations that were strongest and had the highest PPV for the PheWAS code were autoantibodies against fibronectin and obesity ($P = 6.1 \times 10^{-4}$, PPV 100%), and that between fibrinogen and pneumonopathy ($P = 2.7 \times 10^{-4}$, PPV 96%). Pneumonopathy codes included diagnoses for cryptogenic organizing pneumonia and obliterative bronchiolitis.**

Conclusion. **We demonstrated application of a bioinformatics method, the PheWAS, to screen for the clinical significance of RA-related autoantibodies. Using the PheWAS approach, we identified potentially significant links between variations in the levels of autoantibodies and comorbidities of interest in RA.**

Supported by the Rheumatology Research Foundation (Disease Targeted Pilot Grant) and the NIH (Harvard Clinical and Translational Science Center grant UL1-TR-001102 from the National Center for Advancing Translational Sciences). Dr. Liao's work was supported by the NIH (National Institute of Arthritis and Musculoskeletal and Skin Diseases [NIAMS] grant K08-AR-020657 and National Heart, Lung, and Blood Institute grant R01-HL-127118) and the Harold and Duval Bowen Fund. Dr. Sparks' work was supported by the Rheumatology Research Foundation (Scientist Development Award) and the NIH (NIAMS grants L30-AR-066953 and K23-AR-069688). Drs. Hejblum and Tianxi Cai's work was supported by the NIH (National Human Genome Research Institute grant U54-HG-007963). Dr. Liu's work was supported by the Natural Science Foundation of China (grants 11322107 and 11431006). Dr. T. Tony Cai's work was supported in part by the NSF (grants DMS-1208982 and DMS-1403708) and the NIH (National Cancer Institute grant R01-CA-127334).

[1]Katherine P. Liao, MD, MPH, Jeffrey A. Sparks, MD, MMSc, Jing Cui, MD, PhD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Boris P. Hejblum, PhD, Tianxi Cai, PhD: Harvard T. H. Chan School of Public Health, Boston, Massachusetts; [3]I-Hsin Kuo, PhD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, and Biogen, Cambridge, Massachusetts; [4]Lauren J. Lahey, BSc, Jeremy Sokolove, MD: VA Palo Alto Healthcare System and Stanford University School of Medicine, Palo Alto, California; [5]Andrew Cagan, BS, Vivian S. Gainer, MS: Partners Healthcare, Charlestown, Massachusetts; [6]Weidong Liu, PhD: Shanghai Jiao Tong University, Shanghai, China; [7]T. Tony Cai, PhD: The Wharton School, University of Pennsylvania, Philadelphia.

Drs. Liao and Sparks contributed equally to this work.

Address correspondence to Katherine P. Liao, MD, MPH, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115. E-mail: kliao@partners.org.

Submitted for publication June 28, 2016; accepted in revised form October 27, 2016.

Table 1. Numbers of citrullinated and noncitrullinated autoantibodies targeting epitopes with proteins

Target protein	Citrullinated	Noncitrullinated
Apolipoprotein A-I	2	2
Apolipoprotein E	1	1
Biglycan	1	0
Clusterin	1	1
Enolase	1	0
Fibrinogen	10	1
Fibronectin	2	0
Filaggrin	2	0
Histone	5	2
Vimentin	3	1

A defining feature of rheumatoid arthritis (RA) is a break in tolerance leading to the development of autoantibodies (1,2). The development of antibodies against citrullinated protein epitopes appears to be relatively specific to RA (3), and the presence of anti–citrullinated peptide antibodies (ACPAs) is a component of the American College of Rheumatology/European League Against Rheumatism classification criteria for RA (4). However, RA patients have antibodies against a broad range of epitopes, both citrullinated and noncitrullinated (5). The recent identification of epitopes implicated in the pathogenesis of RA allow investigators to study the types of ACPAs that are present in different patients with RA. Whether these ACPAs can further categorize RA patients into clinically relevant subsets remains unclear. One of the major challenges in determining the clinical relevance of these autoantibodies is the need for comprehensive information regarding a patient's other conditions or comorbidities. Large observational cohort studies include adequate numbers of patients with blood samples but vary in terms of details of the clinical data for patients with conditions not related to RA.

The use of electronic medical record (EMR) data as research databases may facilitate comprehensive studies of relationships between biomarkers and disease phenotypes. In particular, EMRs linked to biorepositories allow for the development of EMR-based research platforms, which can support studies examining clinical subsets of disease. Combined with bioinformatics methods such as the phenome-wide association study (PheWAS) method (6,7), these platforms are ideal for examining the potential clinical relevance of novel biomarkers.

The PheWAS method was initially developed for use with EMRs; this approach is used to screen for important associations between biomarkers or genetic variants of interest and a broad range of phenotypes (6,8–10). Phenotypes are typically defined using structured data such as diagnosis billing codes. Although the PheWAS was initially designed for application in cohorts in which clinical EMR data are linked with genetic data, our group previously applied the PheWAS to screen for potential associations between clinically available biomarkers of autoimmunity and phenotypes based on EMR data (7).

The objective of this study was to use the PheWAS approach to screen for antibodies against RA-related epitopes in a large cohort of RA patients identified from EMRs. The PheWAS method generates hypotheses on potential relationships between RA autoantibodies and subphenotypes of RA. As part of this study, we also provide a "road map" demonstrating how to apply updated biostatistics methods for PheWAS and outline the bioinformatics approach, from performing a screen to evaluating the strength of association.

PATIENTS AND METHODS

Study population. We conducted this study in a cohort of RA patients identified from the EMRs of large academic institutions, Brigham and Women's Hospital and Massachusetts General Hospital. This cohort was identified using a validated algorithm with a positive predictive value (PPV) of 94% to classify patients with RA (11,12). Discarded blood samples from Partners Healthcare (Boston) clinical laboratories were collected from 2009 to 2010, using an institutional review board–approved process (13). The result was a deidentified data set containing clinical data from the EMRs until 2009, with linked biospecimens. Due to data suggesting differences in autoantibody expression across ethnicities (14), the analyses were limited to individuals of European ancestry (13). Ancestry data were obtained from a previous analysis of ancestry using ancestry-informative markers.

Laboratory studies. Using a validated multiplex bead assay (5), we measured 36 autoantibodies targeting epitopes of 10 antigens (citrullinated and noncitrullinated) found in the synovial tissue lysates of patients with RA. Autoantibody data were grouped according to epitope (Table 1). Antibodies against second-generation cyclic citrullinated peptide (anti–CCP-2) were measured using a BioPlex system (Bio-Rad).

Statistical analysis. Using EMR-derived data, we extracted all unique International Classification of Diseases, Ninth Revision (ICD-9) codes associated with each patient any time during follow-up. The ICD-9 codes were grouped into disease categories, called PheWAS codes, using a previously published method (6,9). The method for mapping and grouping ICD-9 codes into PheWAS codes is freely available (https://phewas.mc.vanderbilt.edu/). This method was created to group ICD-9 codes in a manner that would facilitate phenotype–genotype studies in linked EMR data sets. For example, "Rheumatoid arthritis and other inflammatory arthropathies" is the name of a PheWAS code. This PheWAS code for RA includes Felty's syndrome (ICD-9 714.1) along with RA (ICD-9 714.0) but does not include juvenile idiopathic arthritis (JIA) (ICD-9 714.3); the subtypes of JIA are considered genetically distinct from those of RA (15). We considered only PheWAS codes that are present in at least 3% of the population, to ensure that the effective sample size (number of cases) was at least 30. This is the smallest sample size needed to ensure stability for fitting the logistic regression model, adjusting for age, sex, and race and deriving a stable score statistic. For each PheWAS code, patients

Table 2. Clinical characteristics of the 1,006 RA patients*

Age, mean ± SD years	61.0 ± 12.9
Female, no. (%)	795 (79.0)
Anti–CCP-2 positive, no. (%)	725 (72.0)
Follow-up for RA, mean ± SD years	8.9 ± 5.3
Ever DMARD use	
Methotrexate, no. (%)	773 (76.8)
TNF inhibitor, no. (%)	458 (45.5)
Erosions on radiography, no. (%)	694 (69.0)
CRP, median (IQR) mg/liter	5.9 (2.7–13.3)

* RA = rheumatoid arthritis; anti–CCP-2 = anti–cyclic citrullinated peptide 2; DMARD = disease-modifying antirheumatic drug; TNF = tumor necrosis factor; CRP = C-reactive protein; IQR = interquartile range.

with ≥1 code were considered as having the phenotype. Data for these binary traits were linked to the autoantibody data obtained from the laboratory studies described above.

Because this study involved multiple testing of a large number of hypotheses, *P* values were considered significant if they were lower than a threshold selected to control a desired false discovery rate (FDR) of 10% (16). An FDR of 10% was considered, because we expected the rate of Type I errors to be <10% among the rejected hypotheses. PheWAS codes, which are derived from ICD-9 codes, are highly correlated. The standard FDR-controlling method that accounts for the correlation tends to be overly conservative (17,18). Thus, to account for the high degree of correlation without requiring strong assumptions about the correlation structure, we applied a modified Benjamini and Hochberg method, which allows for efficient and simultaneous testing of associations between a large number of PheWAS codes and multiple autoantibodies (Cai T, et al: unpublished observations).

For the primary analysis, we summarized the overall association between each PheWAS code and each autoantibody group according to the epitope target (Table 1). The score test statistic was calculated from fitting the logistic regression model, adjusting for age, sex, and race. For example, we calculated the score test statistic for the association between fibrinogen as a group and each PheWAS code. The *P* value from the score test statistic accounts for the size of the autoantibody group. This approach reduces bias for autoantibodies with a higher number of the same protein targets. For example, fibrinogen had 11 different targeted autoantibodies compared with enolase, for which only a single autoantibody targeting it was measured. As a secondary analysis, we studied the association between the PheWAS codes and autoantibodies, stratified according to whether the targets were citrullinated epitopes. The direction of the effect for a group was determined by averaging the direction of the effect (positive or negative) over all epitopes in the group, and then taking the sign (plus or minus) of this average.

We reported the significant associations for protein targets identified at an FDR level of ≤0.1, and ranked the autoantibody/PheWAS code pairs according to *P* values. For the 15 strongest associations, we selected a random sample of 50 patients who had the corresponding PheWAS code and reviewed their medical records to determine the accuracy of the code. The reviews excluded PheWAS codes based on ICD-9 codes describing nonspecific conditions, e.g., nonspecific abnormal results of function studies.

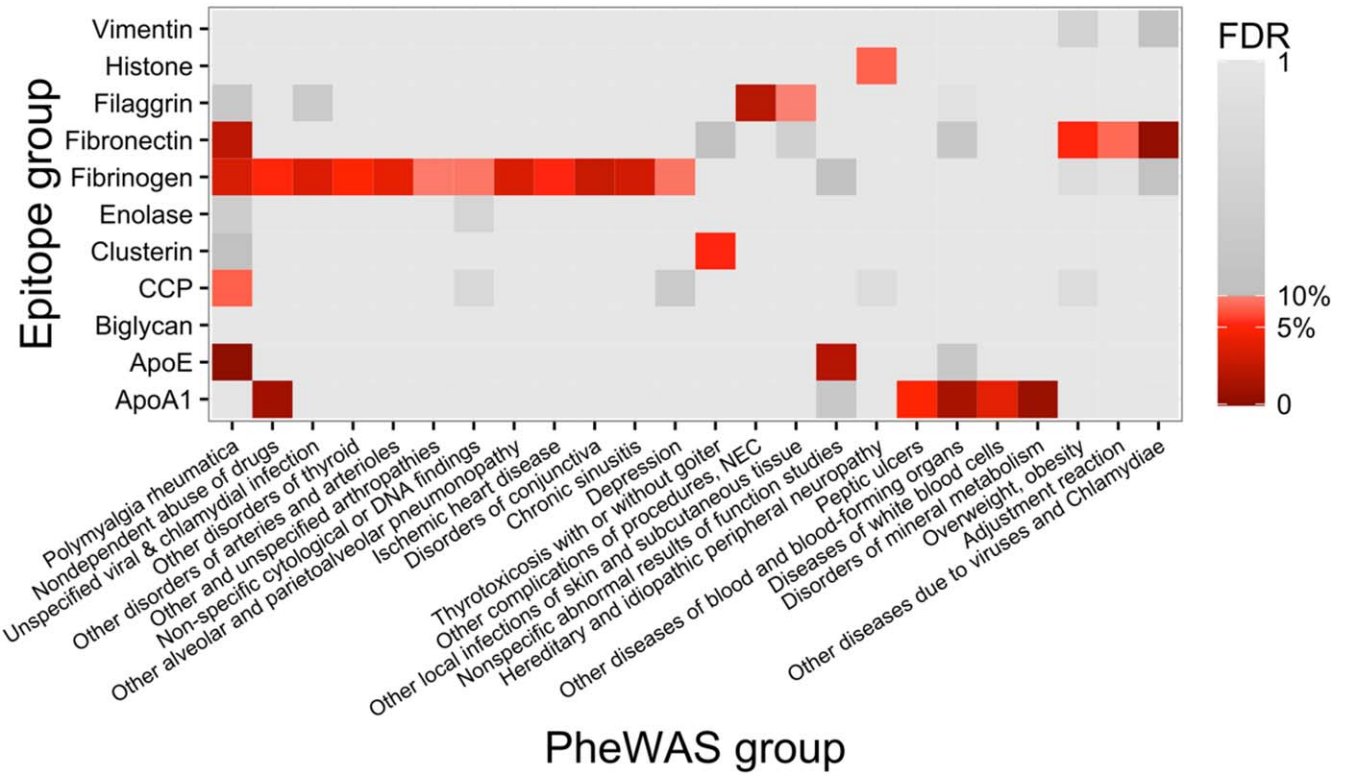

Figure 1. Phenome-wide significant associations between autoantibodies (grouped by epitope target) and phenome-wide association study (PheWAS) codes (false discovery rate [FDR] ≤10%). CCP = cyclic citrullinated peptide; Apo E = apolipoprotein E; NEC = not elsewhere classified.

Table 3. Phenome-wide significant associations between autoantibodies targeting epitope protein groups, and PheWAS codes ranked by the PPV of the PheWAS codes*

PheWAS code description, autoantibodies targeting proteins	P†	Direction of mean effect	PPV of PheWAS code, %
Diseases of white blood cells			100
Apo A-I	3.47×10^{-4}	+	
Overweight, obesity			100
Fibronectin	6.14×10^{-4}	−	
Other alveolar and parietoalveolar pneumonopathy			96
Fibrinogen	2.73×10^{-4}	+	
Peptic ulcers			94
Apo A-I	5.69×10^{-4}	+	
Chronic sinusitis			82
Fibrinogen	2.18×10^{-4}	−	
Disorders of the conjunctiva			76
Fibrinogen	1.69×10^{-4}	+	
Other disorders of the thyroid			72
Fibrinogen	5.64×10^{-4}	−	
Other diseases due to viral or chlamydial infection			72
Fibronectin	3.98×10^{-6}	−	
Ischemic heart disease			68
Fibrinogen	6.20×10^{-4}	−	
Thyrotoxicosis with or without goiter			66
Clusterin	6.17×10^{-4}	−	
Disorders of mineral metabolism			62
Apo A-I	1.26×10^{-5}	−	
Other disorders of arteries and arterioles			56
Fibrinogen	3.76×10^{-4}	−	
Viral and chlamydial infection			52
Fibrinogen	2.91×10^{-4}	−	
Polymyalgia rheumatica			44
Apo E	5.36×10^{-7}	−	
Fibronectin	1.07×10^{-4}		
Fibrinogen	2.53×10^{-4}	−	
CCP	9.21×10^{-4}	−	
Other diseases of blood and blood-forming organs			42
Apo A-I	3.93×10^{-5}	−	

* PPV = positive predictive value; Apo A-I = apolipoprotein A-I; CCP = cyclic citrullinated peptide.
† Value for the association between the autoantibody and the phenome-wide association study (PheWAS) code.

The accuracy of the PheWAS codes was reported as the PPV, calculated by dividing the number of patients confirmed by medical record review to have the condition described by the PheWAS code by the number of patients reviewed. The reviews were performed by 3 of the authors (KPL, IK, and JS). For each PheWAS code, the first 10 patients were reviewed by 2 reviewers (JS and IK, or KPL and IK), and all cases were discussed. Subsequently, any ambiguous cases were discussed between the reviewers until consensus was reached (<10% cases reviewed).

This study was approved by the Partners Institutional Review Board. Analyses were conducted using SAS version 9.3 (SAS Institute) and R Project for Statistical Computing (http://www.r-project.org).

RESULTS

The clinical characteristics of patients in the RA cohort are shown in Table 2. The cohort included 1,006 RA patients. The mean ± SD age of the patients was 61.0 ± 12.9 years, 79% were female, and 72% were anti–CCP-2 positive (as determined by clinical assay) (Table 2).

In this cohort, 3,568 unique ICD-9 codes were grouped into 625 PheWAS codes, of which 206 PheWAS codes had a prevalence of ≥3%. We observed 28 associations that achieved phenome-wide significance (FDR ≤0.1) between autoantibody groups and PheWAS codes (Figure 1). The majority of associations were observed with antibodies against fibrinogen epitopes.

The 3 strongest significant associations were between apolipoprotein E (Apo E) and "polymyalgia rheumatica" ($P = 5.36 \times 10^{-7}$), between fibronectin and "other diseases due to viruses and *Chlamydiae*" ($P = 3.98 \times 10^{-6}$), and between Apo A-I and "other disease of blood and blood-forming organs" ($P = 3.93 \times 10^{-5}$) (Figure 1 and Table 3).

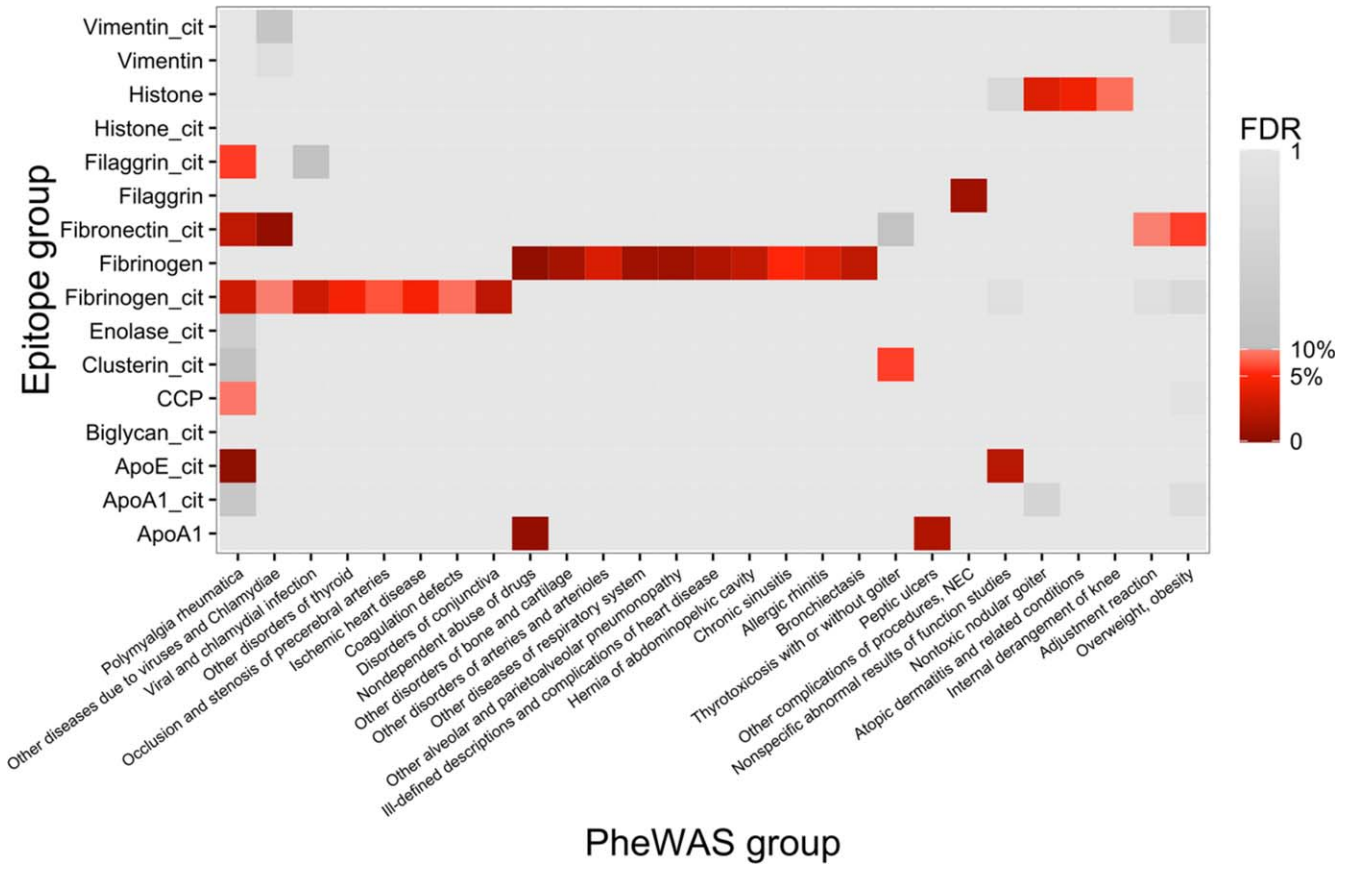

Figure 2. Phenome-wide significant associations between autoantibodies (stratified by citrullination status of the epitope target) and PheWAS codes (FDR ≤10%). See Figure 1 for definitions.

Autoantibodies targeting enolase, vimentin, or biglycan had no significant associations with PheWAS codes.

Among the significant associations (FDR ≤0.10), the PheWAS codes with the highest accuracy were "diseases of white blood cells," "overweight, obesity," and "other alveolar and parietoalveolar pneumonopathy." Physicians used the pneumonopathy code (ICD-9 code 516.x) for the following conditions: cryptogenic organizing pneumonia, obliterative bronchiolitis, lipoid pneumonia, and alveolitis. Table 3 shows the associations between autoantibody epitope targets and PheWAS codes, ranked by PPV. The mean direction of effect for the association was also calculated. For associations with a positive direction, an increasing concentration of titers was associated with an increased likelihood of having the PheWAS code.

As a secondary analysis, we investigated phenome-wide significant associations between autoantibodies, stratified by those targeting citrullinated versus noncitrullinated proteins (Figure 2). For example, the associations between anti–cit-fibrinogen and PheWAS codes did not generally overlap with those between anti-fibrinogen and PheWAS codes. Several of the associations between anti-fibrinogen and PheWAS codes included diseases of the respiratory system, e.g., "other alveolar and parietoalveolar pneumonopathy," "chronic sinusitis," "allergic rhinitis," and "other diseases of respiratory system." In contrast, conditions related to vascular occlusion such as "ischemic heart disease," "occlusion and stenosis of precerebral arteries," and "coagulation defects," had stronger associations with anti–cit-fibrinogen.

In the majority of cases, the associations were driven by autoantibodies targeting either the citrullinated or the noncitrullinated epitopes. The directions of the mean effects, stratified by citrullination status, are shown in Table 4.

DISCUSSION

In this study, we applied the PheWAS method to examine potential associations between RA autoantibodies and RA subphenotypes identified from the EMRs of the cohort. Furthermore, we applied novel biostatistical methods aimed at accounting for correlations between diagnoses and correlations between autoantibodies in our data set. Using

Table 4. Phenome-wide significant associations between autoantibodies targeting epitope groups stratified by antibodies against citrullinated or noncitrullinated epitopes and PheWAS group*

PheWAS code description, autoantibody target epitope	Direction of mean effect		
	All antibodies against epitope targets	Antibodies against citrullinated targets	Antibodies against noncitrullinated targets
Diseases of white blood cells			
Apo A-I	+	NS	NS
Overweight, obesity			
Fibronectin	−	−	NA
Other alveolar and parietoalveolar pneumonopathy			
Fibrinogen	+	NS	+
Peptic ulcers			
Apo A-I	+	NS	+
Chronic sinusitis			
Fibrinogen	−	NS	−
Disorders of conjunctiva			
Fibrinogen	+	+	NS
Other disorders of thyroid			
Fibrinogen	−	−	NS
Other diseases due to viral or chlamydial infection			
Fibronectin	−	−	NA
Ischemic heart disease			
Fibrinogen	−	−	NS.
Thyrotoxicosis with or without goiter			
Clusterin	−	NS	NA
Disorders of mineral metabolism			
Apo A-I	−	NS	NS
Other disorders of arteries and arterioles			
Fibrinogen	−	NS	+
Viral and chlamydial infection			
Fibrinogen	−	−	NS
Polymyalgia rheumatica			
Apo E	−	−	NA
Fibronectin	−	−	NA
Fibrinogen	−	−	NS
CCP	−	−	NA
Other diseases of blood and blood-forming organs			
Apo A-I	−	NS	NS

* PheWAS = phenome-wide association study; Apo A-I = apolipoprotein A-I; NS = not significant; NA = not applicable (due to no autoantibodies in the group); CCP = cyclic citrullinated peptide.

this data-driven approach, we identified significant associations with corresponding PheWAS codes, with an accuracy of $\geq$80%. These include associations between anti-fibronectin antibodies and obesity, between anti–Apo A-I antibodies and peptic ulcers, and between anti-fibrinogen antibodies and inflammatory lung conditions and chronic sinusitis. All of these associations require formal validation in an independent cohort to directly test these hypotheses.

Notably, there are existing data that already support the associations observed in this study. In the current study, we observed an association between higher titers of antibodies against fibrinogen with "other alveolar and parietoalveolar pneumonopathy," describing diagnoses including cryptogenic organizing pneumonia, obliterative bronchiolitis, lipoid pneumonia, and alveolitis. These findings are consistent with those from a previous study examining the association between ACPA titers and interstitial lung disease (ILD) (19,20). In that study of 177 RA patients who had lung imaging, pulmonary function tests, and ACPA testing, a higher titer of all ACPAs was associated with worse ILD. However, an analysis focused on antibodies targeting specific antigens was not performed. Additionally, a recently published study used a PheWAS to evaluate clinical differences between seropositive and seronegative RA (20). That study showed that seropositive RA was associated with lung abnormalities, specifically chronic airway obstruction. Because the study used measurements of rheumatoid factor and anti-CCP, ordered as part of

routine care, to define seropositivity, finer subphenotyping of RA based on specific ACPAs was not performed.

The results of the current study also demonstrated a potential link between autoantibodies against fibrinogen and ischemic heart disease. These findings are consistent with a previously published finding that higher levels of anti–cit-fibrinogen were associated with a higher plaque burden, as measured by the aortic calcium score (21). In a previous study designed for hypothesis testing, we observed that higher levels of anti–cit-fibrinogen were associated with coronary artery disease, as verified by medical record review (Hejblum BP, et al: unpublished observations). However, one study that used a different platform to detect ACPAs did not show an association between anti–cit-fibrinogen levels and subclinical atherosclerosis (22).

Additionally, we highlighted a recent advance in analyzing the clinical importance of groups of biomarkers. In previous studies examining the link between autoantibodies and phenotypes, each association was tested individually. Because ACPAs are highly correlated, testing individual autoantibody/phenotype pairs reduced the power to detect an association. Testing the association between autoantibodies grouped by epitope target and phenotypes has not yet been performed, partially due to the lack of suitable existing methods. Summarizing the overall effect of autoantibody groups on individual phenotypes via a score statistic, and performing multiple testing correction in the presence of complex yet-unknown correlation structures allowed, for the first time, assessment of the overall association between groups of autoantibodies and phenotypes while controlling for a desired FDR. This approach is substantially more powerful than the Bonferroni correction. Additionally, this method allowed us to incorporate information on the concentration of antibodies into the analyses rather than using cutoffs for positive and negative results.

Some notable limitations to this study include the accuracy of ICD-9 codes, which can limit interpretation of the associations. To address this issue, we reviewed the EMRs of patients with the diagnosis codes to determine the accuracy of the codes. As with any EMR study, the data may not reflect complete capture of a patient's diagnoses, because he or she may seek care outside of our healthcare system. On the other hand, we used information derived from a patient's entire follow-up in the medical records, reflecting diagnosis assigned by other healthcare providers not limited to rheumatologists. Because autoantibodies were measured in all patients in the study, we avoided the bias that can occur in studies using measurements that are performed as part of routine care. Future studies are needed to directly test the associations identified in this study and control for additional potential confounders.

In summary, using an EMR-based research platform, we performed a PheWAS screen to examine the potential significance of autoantibodies in patients with RA. Our findings suggest that these biomarkers may help identify subsets of disease. Additionally, the results of our PheWAS screen supported previous findings that autoantibodies may be useful markers for pulmonary conditions and ischemic heart disease in RA. Comprehensive PheWAS screening for ACPAs can provide data for ongoing as well as future studies in RA. Finally, we demonstrate methods that can be applied to any EMR-based population in which biomarkers may be helpful in understanding disease subsets to inform screening, management, and treatment.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Liao, Sparks, and T. Cai had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Liao, Sparks, Hejblum, Cui, Liu, T. T. Cai, Sokolove, T. Cai.

Acquisition of data. Liao, Sparks, Kuo, Lahey, Cagan, Gainer, Sokolove, T. Cai.

Analysis and interpretation of data. Liao, Sparks, Hejblum, Kuo, Cui, Lahey, Cagan, Gainer, Liu, T. T. Cai, Sokolove, T. Cai.

ADDITIONAL DISCLOSURES

Author Luo is an employee of Biogen.

REFERENCES

1. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med 2011;365:2205–19.
2. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet 2009;373:659–72.
3. 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 Rheum 2009;61:1472–83.
4. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62: 2569–81.
5. Sokolove J, Bromberg R, Deane KD, Lahey LJ, Derber LA, Chandra PE, et al. Autoantibody epitope spreading in the preclinical phase predicts progression to rheumatoid arthritis. PLoS One 2012;7:e35296.
6. 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 2010;26:1205–10.
7. Liao KP, Kurreeman F, Li G, Duclos G, Murphy S, Guzman R, et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non–rheumatoid arthritis controls. Arthritis Rheum 2013;65:571–81.

8. Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, et al. Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. PLoS Comput Biol 2013;9: e1003405.
9. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 2013;31:1102–10.
10. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation 2013;127:1377–85.
11. 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:e162–9.
12. 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 Res (Hoboken) 2010;62:1120–7.
13. 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. Am J Hum Genet 2011;88:57–69.
14. Kurreeman FA, Stahl EA, Okada Y, Liao K, Diogo D, Raychaudhuri S, et al. Use of a multiethnic approach to identify rheumatoid arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet 2012;90:524–32.
15. Prahalad S, Glass DN. A comprehensive review of the genetics of juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2008;6:11.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57:289–300.
17. Cai TT, Liu W. Large-scale multiple testing of correlations. J Am Stat Assoc 2016;111:229–40.
18. Yekutieli D, Benjamini Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Stat Plan Inference 1999;82:171–96.
19. Giles JT, Danoff SK, Sokolove J, Wagner CA, Winchester R, Pappas DA, et al. Association of fine specificity and repertoire expansion of anticitrullinated peptide antibodies with rheumatoid arthritis associated interstitial lung disease. Ann Rheum Dis 2014; 73:1487–94.
20. Sokolove J, Brennan MJ, Sharpe O, Lahey LJ, Kao AH, Krishnan E, et al. Citrullination within the atherosclerotic plaque: a potential target for the anti–citrullinated protein antibody response in rheumatoid arthritis. Arthritis Rheum 2013;65:1719–24.
21. Doss J, Mo H, Carroll RJ, Crofford LJ, Denny JC. Phenome-wide association study of rheumatoid arthritis subgroups identifies association between seronegative disease and fibromyalgia. Arthritis Rheumatol 2017;69:291–300.
22. Montes A, Corrales A, Calaza M, Lopez-Mejias R, Parra JA, Gonzalez-Gay MA, et al. Lack of replication of an association between anti–citrullinated fibrinogen and subclinical atherosclerosis in patients with rheumatoid arthritis. Arthritis Rheumatol 2015;67: 2861–5.