Research paper

Cohort profile of a US military population for evaluating pre-disease and disease serological biomarkers in rheumatoid and reactive arthritis: Rationale, organization, design, and baseline characteristics

Chad K. Porter [a],[*], Mark S. Riddle [a], Renee M. Laird [b], Matthew Loza [c], Suzanne Cole [c], Christina Gariepy [b], Ashley Alcala [b], Ramiro Gutierréz [a], Frédéric Baribaud [c], Navin L. Rao [c], Sunil Nagpal [c]

[a] Enteric Diseases Department, Naval Medical Research Center, Silver Spring, MD, USA
[b] Henry M. Jackson Foundation for Military Medicine, Bethesda, MD, USA
[c] Immunology, Janssen Research & Development, Spring House, PA, USA

ARTICLE INFO

ABSTRACT

Purpose: The etiology of several autoimmune disorders, including rheumatoid arthritis, remains unknown. While there are clear phases of disease progression, the mechanisms of transition between these phases are poorly understood. Additionally, treatment focuses on an alteration of the biological processes to prevent joint damage and functional decline. A goal is to potentially treat the disease during the preclinical phase to mitigate the disease process. Reactive arthritis is another rheumatologic condition known to be secondary to a distal infection. While prevention of infection would mitigate risk, serologic profiling patients with the disease may assist in the elucidation of potential disease risk factors. This study was initiated to enable an assessment of pre-disease biomarkers in patients newly diagnosed with rheumatoid arthritis and reactive arthritis.
Participants: A retrospective cohort of 500 rheumatoid and 500 reactive arthritis cases with 500 matched controls was drawn from a population of active component US military personnel. Appropriate inclusion criteria limited subject selection. Additionally, 4 serum samples (3 pre-disease and 1 disease-associated) were obtained for each case and control.
Findings to date: The established cohort provides the framework for novel exploration of the host response through serum profiling and seroepidemiology prior to disease onset.
Future plans: This study establishes the framework for the evaluation of novel serum biomarkers enabling the identification of signals prior to clinical disease that may enable disease prediction, elucidate disease pathogenesis and identify novel exposures leading to increased disease risk and/or disease severity.

1. Introduction

Rheumatoid Arthritis (RA) is a relatively common, chronic, systemic inflammatory disease of unknown aetiology that affects approximately 1% of the population worldwide [1–3]. RA is three times more common in women than men [4]. There is regional variation in the prevalence of RA. The incidence appears to be highest in Pima Indians (5.3%) and Chippewa Indians (6.8%), and lowest in people from China and Japan (0.2%–0.3%), suggesting the possibility that genetic factors contribute to RA [5,6]. These differences in regional RA prevalence also may suggest a role for environmental factors. It predominantly targets joints and

can lead to significant long-term disability, secondary to cartilage and bone erosion. These long term effects are reflected in the work disability secondary to RA, which can be as high as 59%, with 90% of patients with RA retiring from work prematurely [7].

Reactive arthritis (ReA) is a well-known sequelae following gastrointestinal (*Salmonella, Yersinia, Shigella,* and *Campylobacter* spp) and urogenital (*Chlamydia trachomatis, neisseria gonorrhoeae*) infections. The maximum interval between the preceding infection and the arthritis is generally accepted as 4 weeks. Typical features include asymmetric oligoarthritis, most often in the large joints of the lower extremities. Up to half of patients also have arthritis in their upper limbs. A recent study

has estimated that the annual incidence of enteric-infection-related ReA within a population may approach one per 1000 [8]. The frequency of ReA or new-onset joint pain following a gastrointestinal infection varies widely among studies likely confounded by variable study design and outcome definitions [9–23].

2. Study rationale

Initial development of RA autoimmunity likely occurs in the pre-clinical/asymptomatic period of inflammatory arthritis and it is important to study the biology in this early period to understand the progression of the disease from asymptomatic to symptomatic phase and ultimately to clinical RA. RA patients show antibodies reactive against several citrullinated peptides or proteins, e.g., anti-citrullinated protein antibodies (ACPAs) against enolase, vimentin, fibrinogen, and type II collagen. Since these autoantibodies develop in asymptomatic individuals much before the onset of inflammatory arthritis and RA, ACPAs are excellent serological tools to understand disease progression and to explore the biology associated with pre-RA stage. ACPAs develop 5–10 years before the clinical onset of the disease and their sub-specificities may help us identify appropriate diagnostic, prognostic and patient stratification markers. However, there is a paucity and limited understanding of biomarkers of disease progression, specifically those which are altered in patients 6–24 months before the clinical diagnosis of RA.

Unlike the serological markers of RA, there is a significant gap in our understanding of the markers of ReA. The presence and identity of citrullinated antigens, carbamylated antigens or other autoantigens and their autoantibodies is an open question and worthy of study. Londono et al. described the association between several potential serum biomarkers and disease prognosis in patients with ReA and other spondyloarthritidies [24]. Specifically, an increase in several cytokines, including IL-6, IL-1α as well as C-reactive protein was associated with poorer disease prognosis. It is unclear if these changes are temporally associated with disease onset or if this variability is present prior to initiating infection.

The goal of this effort was to establish a cohort that could provide insights into the pathophysiological mechanisms that operate before the clinical diagnosis of RA/ReA. Insights and predictions gleaned from such a cohort has the potential to change the treatment paradigm focused on the remedial treatment of established and essentially incurable disease to a future treatment paradigm of prevention or early intervention. This early pre-diagnosis phase may be the ideal opportunity to restore immune tolerance and homeostasis, something that is difficult to achieve in established RA/ReA. Additionally, identification of novel immune biomarkers may point to diagnostic and prognostic assays and new drug targets worthy of subsequent preclinical and clinical development.

Another goal of this cohort is to enable the comparison of serologic and serum proteome associated with ReA to those in RA. ReA is an acute disease, whereas RA is a chronic disease. We hypothesize that initiating triggers and disease pathways might be shared between the two forms of arthritis. This has important ramifications not only in understanding disease processes but also in the identification of novel targets for treatment, prevention and cure of RA. In this project, we propose to identify the acute pathways or networks that are modulated in reactive arthritis compared to healthy controls. We will then interrogate the status of these pathways or networks in RA serial samples. Therefore, using reactive and rheumatoid arthritis serum samples, we hope to identify pathways involved in the manifestation of ReA disease as well as pathways or networks that may be the initial triggers of autoimmunity in RA.

In brief, this study enables a unique opportunity to compare reactive and rheumatoid arthritis prior to diagnosis. Data resulting from the analysis of samples from this cohort have the potential to not only identify the earliest triggering pathways and networks leading to disease, but may also facilitate the identification of novel drug and therapeutic targets as well as diagnostic and prognostic markers associated with these two forms of arthritis.

3. Cohort description

3.1. Study design and participants

This study was designed as a nested case-control study that incorporates non-diseased subjects as a comparator population using age and gender frequency matching. Subjects with RA/ReA between 1998 and 2012 were identified from the Defense Medical Surveillance System (DMSS), the main data repository for all US armed forces and contains relevant data from more than 7 million members having served in the armed forces since 1990, documenting their military and medical experiences throughout their career [25].

### 3.2. Case and control selection

Cases were defined as subjects with at least 2 medical encounters with any of the relevant International Classification of Diseases 9 - Clinical Modification (ICD9-CM) codes for reactive arthritis (099.3, 711.1, 711.3) or rheumatoid arthritis (714.0). The date of onset was defined as the first medical encounter with a relevant ICD9-CM code, confirmed by ≥ 1 additional encounter. A matched control population based on similar age-group and gender distributions of the cases subjects was obtained from a random sample of subjects in the DMSS with no medical encounters for reactive or rheumatoid arthritides. Additionally, all subjects with documented history of non-specific arthropathies and arthralgias were excluded.

3.3. Serum sampling

For each case and control subject, up to four serum samples were obtained from the US Department of Defense (DoD) Serum Repository as per Fig. 1. For RA subjects, sample A was the first sample prior to the first RA diagnosis, unless a sample was available within 30 days of the first diagnosis. In this case, the sample 30 days subsequent to diagnosis was obtained. Sample D for RA cases was the first sample available in the repository, and samples B and C were interim samples approximately evenly distributed across the service member's duty time. For ReA, samples A and B represent the first available samples after and before an initial ReA diagnosis, respectively. Sample D was the first sample available in the repository, and sample C was an interim sample approximately evenly distributed between samples B and D. For control subjects, sample A was matched by year (± 1 year) to ReA and RA cases (50% of controls matched to RA cases and 50% matched to ReA cases) such that an equal proportion of Sample A from the controls were collected at approximately the same time (± 1 year) as Sample A from the cases. Samples B and C for control subjects represented the two preceding serum samples available from the repository and sample D was the earliest serum sample available.

3.4. Biorepository description

Similar to our previously described repository [26], we have established a centralized data and sample storage at the Naval Medical Research Center (NMRC). All data are linked by a unique identifier to subject samples. Samples were accessioned upon arrival and are maintained at NMRC at $-80\,°C$. For testing, the parent samples were thawed, sub-aliquoted and multiple daughter aliquots made to meet the volume requirements of the planned testing as well as anticipated future testing. Aliquots were refrozen and sent to partner laboratories. All aliquots are maintained in bar-code labeled cryovials with a desiccation proof seal without additives. All aliquots are linked to the parent sample through the bar coding and a laboratory information management system.
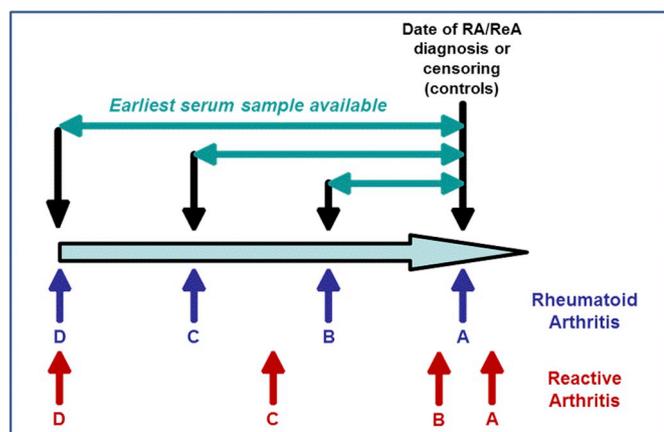
Fig. 1. Planned serum sample selection concurrent with and prior to rheumatoid or reactive arthritis diagnosis

Fig. 1 Legend: For RA subjects, sample A was the first sample prior to the first RA diagnosis unless a sample was available within 30 days of the first diagnosis. In this case, the sample 30 days subsequent to diagnosis was obtained. Sample D for RA cases was the first sample available in the repository, and samples B and C were interim samples approximately evenly distributed across the service member's duty time.

For ReA, samples A and B represent the first available sample after and before an initial ReA diagnosis, respectively. Sample D was the first sample available in the repository, and sample C was an interim sample approximately evenly distributed between samples B and D.

For control subjects, sample A was matched by year (±1 year) to ReA and RA cases (50% of controls matched to RA cases and 50% matched to ReA cases) such that an equal proportion of Sample A from the controls were collected at approximately the same time (±1 year) as Sample A from the cases. Samples B and C for control subjects represented the two preceding serum samples available from the repository and sample D was the earliest serum sample available.

3.5. Sample size

This study focused on a total of 500 cases of RA, 500 cases of ReA and 500 healthy controls {250 age-, gender- and time- (based on Sample A collection year) matched to the RA cases and 250 age-, gender- and time-matched to the ReA cases}. The power to detect a significant difference in a single serologic marker is predicated on the sero-prevalence of that marker in cases and controls. For example, several studies to date have assessed the sensitivity and specificity of ACPAs in RA with high specificity (approximately 90%) yet relatively lower sensitivity (approximately 50%) [27]. Presuming comparable estimates of sensitivity, the proposed sample size would ensure adequate power (>80%) to detect a significant difference in the presence of ACPAs between cases and controls if the specificity is as low as 39% (two-group continuity-adjusted chi-square; two sided $alpha = 0.05$). Importantly, adjustments for multiple comparisons may be needed to control the Type I error rate which may yield a lower overall statistical power. Furthermore, sero-prevalence of ACPAs in 50% of the RA cases would enable an estimate of seroprevalence in the general RA population ± 4%.

Given the utilization of network analyses to maximize the probability of identifying statistical data patterns, it is anticipated that this study is powered adequately to identify significant differences in serum profiles across the study populations assuming there are in fact differences in the studied groups. For these algorithms, models will be developed on a subset of the study population (approximately 1/3) tested on another subset (approximately 1/3) and validated on a third subset (approximately 1/3).

3.6. Clinical and covariate data

The study population was active component US military personnel,

which constitute a large cohort of generally young people without severe co-morbidities. These characteristics are particularly appropriate for studies of RA/ReA, as these diseases mainly affect young adults. Moreover, regarding studies on impact of GI infections on further development of chronic conditions (such as ReA), cohorts of military members are of great interest as these subjects are exposed to high rates of GI infections due to deployment to endemic regions where exposure risk is high [28,29]. It is important to note there are some differences between this population and the general population that may be important in interpreting the findings from this cohort.

Demographic data were obtained from personnel data and deployment data were derived from deployment rosters and deployment health assessments. All medical encounters with a RA/ReA diagnosis at any diagnostic position were obtained from AFHSB. Additionally, any medical encounters associated with a procedural code associated with RA/ReA diagnosis were also obtained. Data on infectious gastroenteritis episodes were obtained based on ICD-9 codes for specific bacterial and viral pathogens as were events with no pathogens identified. Medical encounters for *Chlamydia* and gonorrhea were also obtained.

This study was reviewed and approved by the Institutional Review Board at the NMRC in compliance with all regulations.

4. Baseline description of study population

The demographics of the study population are shown in Table 1. In brief, the population was broadly reflective of active duty US military personnel in that it is predominantly young, male, with a high school (or equivalent) education. Subjects with RA were more commonly female than were those with ReA (p < 0.001). Additionally, subjects with RA represented an older subset of the active component population compared to those with ReA with means of 37 and 30, respectively (p < 0.001). This corresponded to RA subjects that were of more advanced rank (p < 0.001) and more commonly married (p < 0.001) than the ReA subjects. Subjects with reactive arthritis were more commonly white (70.4%) compared to those with RA (56.4%) (p < 0.001). The majority of subjects (65.9%) had at least one operational deployment prior to being included as a case or control. Control subject demographics were comparable to the cases. An approximately equal proportion of subjects was selected from each year of study from 1998 through 2012 (Fig. 2).

4.1. Findings to date

Serum samples were identified for all subjects and represented the planned sampling scheme (Fig. 3). Samples after ReA (sample A) were collected a mean of 293 days (std dev: 308) post-diagnosis, while the last sample prior to diagnosis was collected a mean of 295 days pre-diagnosis (std dev: 288). Antecedent samples were a mean of 4.6 years and 7.9 years prior to diagnosis. In contrast, the disease associated sample for RA cases were collected a mean of 289 days (std dev: 286) prior to diagnosis while the antecedent samples were a mean of 4.7, 8.7 and 12.7 years pre-diagnosis. Samples from control subjects were from a comparable distribution relative to their matched case sample. It should be noted that future analyses should assess for the influence of sample age which may bias results in one way or another; however, it is anticipated that such an 'age effect' would be non-differential in cases and controls.

Procedural codes highlight additional data on the clinical evaluation of the cases (Table 2). Approximately one third of all cases had a documented arthrocentesis and a comparable proportion had an MRI of the affected joint(s). Compared to ReA cases, those with RA more commonly had procedural codes documenting the assessment for actin smooth muscle antibody (p < 0.001), anti-nuclear antibody (p = 0.001), extractable nuclear antigen antibody (p < 0.001) and rheumatoid factor (RF; p < 0.001); however, the proportion of cases tested was low. A higher proportion of RA cases had a documented medical encounters in a rheumatology clinic (p < 0.001); however, among those with documented rheumatology clinic visits, RA subjects had a high frequency of

Table 1
Demographics of study population.

	Rheumatoid Arthritis (RA) (N = 500)		RA Controls (N = 250)		Reactive Arthritis (ReA) (N = 500)		ReA Controls (N = 250)	
	N	%	N	%	N	%	N	%
Gender								
Male	306	61.2	210	84.0	461	92.2	244	97.6
Female	194	38.8	40	16.0	39	7.8	6	2.4
Age								
≤20	0	0.0	0	0.0	7	1.40	0	0.0
20–29	74	14.8	55	22.0	260	52.0	169	67.6
30–39	247	49.4	145	58.0	167	33.4	68	27.20
40–49	149	29.8	49	19.6	58	11.6	13	5.2
≥50	30	6.0	1	0.4	8	1.6	0	0.0
Education [a]								
Less Than High School	1	0.2	2	0.8	5	1.0	4	1.6
High school (or equivalent)	279	55.8	140	56.0	331	66.2	191	76.4
Some college	90	18.0	23	9.2	48	9.6	15	6.0
Bachelor's Degree	76	15.2	48	19.2	61	12.2	26	10.4
Master's degree	42	8.4	30	12	34	6.8	11	4.4
Doctorate Degree	4	0.8	2	0.8	3	0.6	1	0.4
Marital status [b]								
Married	389	77.8	186	74.4	278	55.6	139	55.6
Single	67	13.4	54	21.6	192	38.4	106	42.4
Other	43	8.6	10	4.0	29	5.8	5	2.0
Military Rank								
Jr. Enlisted	28	5.6	20	8.0	184	36.8	108	43.2
Sr. Enlisted	368	73.6	151	60.4	217	43.4	105	42.0
Jr. Officer	78	15.6	73	29.2	82	16.4	35	14.0
Sr. Officer	11	2.2	1	0.4	9	1.8	0	0.0
Warrant Officer	15	3.0	5	2.0	8	1.6	2	0.8
Branch of Service								
Army	201	40.2	50	20.0	168	33.6	61	24.4
Air Force	118	23.6	54	21.6	116	23.2	30	12.0
Coast Guard	5	1.0	6	2.4	9	1.8	3	1.2
Marines	30	6.0	30	12.0	69	13.8	48	19.2
Navy	146	29.2	110	44.0	138	27.6	108	43.2
Race/Ethnicity [c]								
American Indian/ Alaskan Native	7	1.4	5	2.0	8	1.6	5	2.0
Asian/Pacific Islander	17	3.4	8	3.2	11	2.2	7	2.8
White	282	56.4	157	62.8	352	70.4	163	65.2
Black	126	25.2	43	17.2	57	11.4	38	15.2
Hispanic	51	10.2	29	11.6	48	9.6	26	10.4
Other	6	1.2	3	1.2	7	1.40	5	2.0
Operational Deployment	317	63.4	176	70.4	317	63.4	179	71.6

[a] There are a total of 33 subjects (8 RA cases; 5 RA controls; 18 ReA cases; 2 ReA controls) where education information was not provided.

[b] There are 2 individuals (1 Rheumatoid Arthritis; 1 Reactive Arthritis) in which marital status information was not provided.

[c] There were 39 individuals (11 Rheumatoid Arthritis; 5 RA controls; 17 Reactive Arthritis; 6 ReA Controls) where race/ethnicity information was not provided.

encounters. For all other associated encounters, the median number of visits per subject was low and comparable between reactive and rheumatoid arthritis cases (see Table 3).

In addition to the rheumatology-related ICD9-CM codes, concurrent diagnoses were also obtained, and in general, across all of the diagnostic categories, both reactive and rheumatoid arthritis cases were more commonly likely to have a related ICD9-CM code than were the healthy controls. Unsurprisingly, the most common concurrent diagnoses for rheumatoid and reactive arthritis cases included diseases of the musculoskeletal system and connective tissue (73.6% and 74.0%, respectively) with a high median number of related medical encounter visits (32 and 24, respectively). In contrast, only 5.6% (p < 0.0001) of the healthy controls had any medical encounters associated with

diseases of the musculoskeletal system and connective tissue with 3–4 fold fewer encounters per subject. The next most prevalent ICD9-CM code among ReA cases was infectious and parasitic diseases (31.0%), higher (p = 0.01) than in RA cases (23.8%) and almost 3-times more common than the healthy controls (p < 0.0001). Similarly, symptoms, signs and ill-defined conditions were more common among RA cases (41.0%) than among those with ReA (25.4%) or the healthy controls (2.2%). The next most prevalent conditions were endocrine, nutritional, and metabolic diseases and immunity disorders occurring in 25.8% of the RA cases and 13.4% of the ReA cases (p < 0.0001) with also a higher number of encounters per diagnosis.

5. Discussion

There is now increasing evidence that RA develops in phases and its underlying pathogenesis begins many years before clinical diagnosis [30]. The natural course of RA evolution involves genotype-environment interaction in the susceptible individuals resulting in the development of asymptomatic autoimmunity (first phase), where subjects are healthy but for the presence of autoantibodies, namely ACPA and/or rheumatoid factors (RFs). Over time these autoantibodies gain access to joints and asymptomatic subjects transition to an arthralgia phase (second phase), where they start experiencing non-specific joint symptoms without any clinical signs of synovitis. As more autoantibodies and immune cells enter joints, arthralgia patients progress to an undifferentiated/inflammatory arthritis (UA/IA) phase, characterized by clinical signs of synovitis (third phase). Finally, with increased systemic and joint inflammation and ACPA epitope spreading, UA/IA patients convert to clinically identifiable RA (fourth phase). Since UA/IA is characterized by tenderness and swelling of joints, this stage may indicate an imminent RA diagnosis, and thus it may be difficult to clearly identify a population of these patients for a prevention/interception clinical trial. Therefore, asymptomatic autoimmunity and arthralgia are referred to as two pre-RA phases [30], wherein basic pathogenic processes of inflammation, namely T-B cell interaction, production of autoantibodies and infiltration of a small number of autoimmune cells into joints, have begun but the inflammatory burden is low and subjects have no (or minimal) joint inflammation, synovial hyperplasia, tissue damage or joint destruction, hallmarks of established RA. The current therapeutic paradigm limits patient treatment until after diagnosis with full-fledged RA. However, due to heterogeneity of the disease and the irreversible tissue damage that has occurred by the time of RA diagnosis, the remission rates with the current generation of disease modifying anti-rheumatic drugs (DMARDs), including biological DMARDs (bDMARDs) encompassing various therapeutic mechanisms (anti-TNFα, anti-IL6, and IL-1 receptor antagonist), are abysmally low. Although ~60% of patients on bDMARDs show some improvement, only a small number achieve remission or low disease activity, which is also lost overtime [31]. A large number of patients are either incomplete or non-responders to bDMARDs, e.g., approximately 40% of RA patients do not achieve even ACR20 response (minimal response) and only 40% of the patient achieve ACR50 response with anti-TNF therapies [31]. In view of the association of significant co-morbidities with RA and reduced life-expectancy of patients compared to matched controls [32], coupled with the observation that bDMARDs exhibit poor or negligible disease remission rates, it is time to take a cue from cardiovascular field, where statin-mediated down-regulation of serum cholesterol levels is an accepted paradigm for the prevention of cardiovascular diseases. Therefore, halting the progression of disease from pre-RA phases to RA may prevent the down-stream sequalae of irreversible pathogenic tissue changes that lead to inadequate drug responses and therapy-resistant RA in patient populations with established disease.

In order to develop a successful paradigm of disease prevention or disease interception, one needs to identify biomarkers of disease progression from pre-RA to RA that can classify subjects at risk of
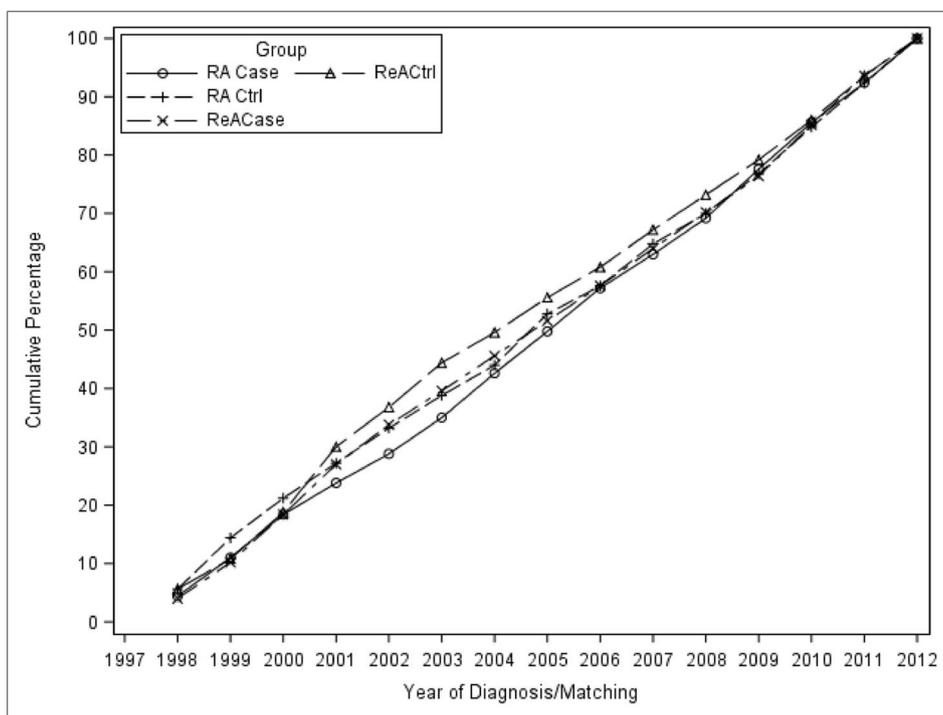
Fig. 2. Year of incident diagnosis (or matching)
Fig. 2 legend: The cumulative percent of subjects by the first year in which the subject had their incident medical encounter for the outcome of interest (based on ICD9-CM code) or the year in which a control was identified for either the matched RA case or ReA case.
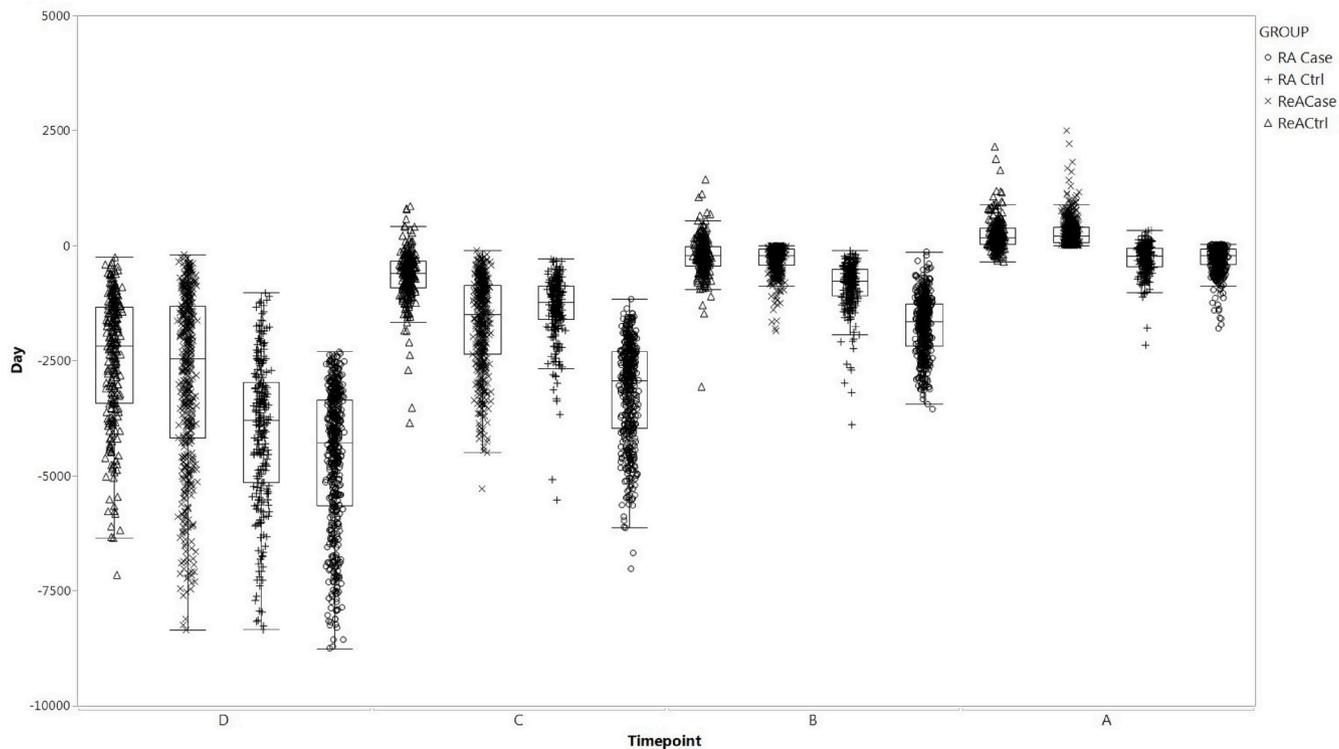

Fig. 3. Box and Whisker Plots of Serum Samples Timing Stratified by RA, ReA, and Healthy ControlsReACase, reactive arthritis cases; RACase, rheumatoid arthritis cases; ReACtrl, reactive arthritis healthy controls; RACtrl, rheumatoid arthritis healthy controls. The box and whisker plots represents the median (mid-line) 25% and 75% quartiles (boxes) and the 1st and 3rd quartiles + 1.5 * the interquartile range (whiskers).

developing RA. These biomarkers may also provide insight into the pathways involved in disease evolution and identification of potential targets for disease interception. A number of studies have reported that the presence of autoantibodies (ACPAs, RFs, anti-peptidyl arginine deiminase antibodies and anti-carbamylated protein antibodies) pre-date clinical RA by several years [30,33–41]. In a prospective study of

Table 2
Additional Information Related to medical RA- or ReA-associated medical encounters.

	N (%) with RA/ReA-associated medical encounters		Median (IQR) Visits Per Subject	
	RA	ReA	RA	ReA
Arthrocentesis	153 (30.6)	174 (34.8)	1.0 (1.0, 3.0)	1.0 (1.0, 3.0)
Magnetic Resonance Imaging (MRI)	187 (37.4)	112 (22.4)	2.0 (1.0, 3.0)	1.0 (1.0, 2.0)
Mitochondrial Antibody	10 (2.0)	4 (0.8)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Actin Smooth Muscle Antibody	14 (2.8)	1 (0.2)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Anti-nuclear Antibody	24 (4.8)	6 (1.2)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Extractable Nuclear Antigen Antibody	23 (4.6)	5 (1.0)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Rheumatoid Factor	33 (6.6)	4 (0.8)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Rheumatology Clinic	179 (35.8)	255 (51.0)	6.0 (2.0, 12.0)	3.0 (2.0, 5.0)

354 first-degree unaffected relatives of RA patients, 39% were found to be RF positive [42], demonstrating RF autoantibodies in the pre-RA stage. A study conducted on a Swedish biobank samples of RA patients (n = 83) collected before disease diagnosis, revealed the presence of anti-cyclic citrullinated protein (CCP) antibodies (33.7%), IgG-RF (16.9%), IgM-RF (19.3%) and IgA-RF (33.7%) in pre-RA samples, thus demonstrating that both ACPA and RF predate RA onset by several years and showed that anti-CCP and IgA-RF predicted the RA development [33]. In the study by Rantapaa-Dahlqvist et al., a majority of the RA patients had only one pre-disease sample. In contrast, in our cohort we have obtained 3 pre-disease samples from each RA patient, potentially providing a more comprehensive view of the evolution of autoantibodies and serum protein analytes. Sokolove et al., provided the proof of concept that pre-RA biomarkers can be identified from a military cohort. Using a biobank of 81 subjects with pre-RA and RA samples, increased levels of ACPA specificities indicating epitope spreading and elevated

levels of serum cytokines were observed as seropositive subjects approached the time of disease diagnosis [43]. We have assembled a comparable but much larger cohort of RA patients. The current generation of biomarkers (ACPAs, RF and cytokines) that define pre-RA phases, develop 5–10 years before RA diagnosis. However, from a practical point of view, an interception clinical trial with a novel agent could be of maximum 2–3 year duration and therefore, one needs to identify at risk subjects 2–3 years before the development of the disease. The DoD cohort of pre-RA serum samples from RA patients described herein provides a unique opportunity to identify not only ACPA sub-specificities and cytokines but other protein analytes (using global proteomics profiling) as novel biomarkers for more precise classification of at-risk subjects. Since autoantibodies develop as a result of breakdown of tolerance and given the observations that therapeutic anti-PD-1 antagonistic antibodies have shown remarkable efficacy in human tumors along with the emergence of autoimmunity in responsive cancer patients due to loss of immune tolerance [44], we also plan to examine the levels of soluble PD-1 (sPD-1) protein in pre-RA samples [45].

Many autoimmune diseases, including RA, show gender differences in their prevalence. RA is three times more common in females than males with women having worse symptoms of disease [46]. However, the pre-RA and RA cohort assembled herein is more biased towards male subjects (61% males; Table 1). The large collection of male pre-RA and RA samples also provide an opportunity to understand pre-disease and disease differences between males and females in terms of their biomarkers, disease pathways and underlying mechanisms.

In contrast to RA, biomarkers of ReA arthritis disease pre-disposition and progression are not very well known. HLA-B27 has been shown to be linked with disease susceptibility in ReA patients with exposure to gastrointestinal or urinogenital bacterial infections *(Campylobacter, Chlamydia, Salmonella, Shigella, and Yersinia)* via mucosal surfaces [47]. HLA-B27 positive ReA patients exhibit more severe disease and the HLA positivity is more common in chronic or relapsing arthritis, uveitis, aortitis, sacroiliitis and spondylitis [47]. Since both RA and ReA show some level of HLA restriction (HLA Class II for RA and HLA class I ReA), and both types of arthritis show synovial inflammation, global proteomics, cytokine/chemokine and autoantibody profiling of the two cohorts will provide a unique opportunity to delineate inflammatory pathways and biomarkers that are common to both RA and ReA as well

Table 3
ICD-9 CM codes corresponding with RA/ReA medical encounters.

ICD-9-CM Description	ICD-9-CM code	Number (%) of subjects with Diagnosis			Median (IQR) of separate medical encounters.		
		RA (N = 500)	ReA (N = 500)	HC (N = 500)	RA	ReA	HC
Infectious and parasitic diseases	001–139	119 (23.8)	155 (31.0)	58 (11.6)	8 (8, 16)	8 (8, 16)	8 (8, 14)
Neoplasms	140–239	35 (7.0)	11 (2.2)	0 (0.0)	8 (7, 8)	8 (7, 8)	–
Endocrine, nutritional, and metabolic diseases and immunity disorders	240–279	129 (25.8)	67 (13.4)	6 (1.2)	12 (7, 24)	8 (6, 14)	7 (6, 7)
Diseases of the blood and blood forming organs	280–289	57 (11.4)	23 (4.6)	1 (0.2)	8 (7, 21)	7 (6, 13)	24 (24, 24)
Mental Disorders	290–319	81 (16.2)	55 (11.0)	0 (0.0)	11 (7, 18)	7 (6, 13)	–
Diseases of the Nervous System	320–359	97 (19.4)	27 (5.4)	1 (0.2)	10 (7, 22)	8 (7, 16)	7 (7, 7)
Disease of the Sense Organs	360–389	49 (9.8)	81 (16.2)	0 (0.0)	13 (7, 21)	14 (7, 24)	–
Diseases of the Circulatory System	390–459	79 (15.8)	46 (9.2)	0 (0.0)	18 (7, 29)	8 (7, 19)	–
Diseases of the Respiratory System	460–519	92 (18.4)	38 (7.6)	2 (0.4)	8 (7, 20)	7 (7, 13)	7 (7, 7)
Diseases of the digestive system	520–579	96 (19.2)	48 (9.6)	3 (0.6)	8 (7, 15)	8 (7, 15)	7 (7, 8)
Diseases of the genitourinary system	580–629	76 (15.2)	35 (7.0)	1 (0.2)	8 (6, 15)	7 (7, 10)	65 (65, 65)
Complications of pregnancy, childbirth, and the puerperium	630–679	7 (1.4)	0 (0.0)	0 (0.0)	20 (7, 62)	–	–
Diseases of the skin and subcutaneous tissue	680–709	70 (14.0)	50 (10.0)	2 (0.4)	8 (7, 11)	7 (6, 10)	6.5 (5, 8)
Diseases Of The Musculoskeletal System And Connective Tissue	710–739	368 (73.6)	370 (74.0)	28 (5.6)	32 (15, 70)	24 (15, 46)	8 (8, 20)
Congenital Anomalies	740–759	10 (2.0)	5 (1.0)	1 (0.2)	7 (6, 8)	7 (7, 14)	6 (6, 6)
Certain Conditions Originating In The Perinatal Period	760–779	0 (0.0)	0 (0.0)	0 (0.0)	–	–	–
Symptoms, Signs, And Ill-Defined Conditions	780–799	205 (41.0)	127 (25.4)	11 (2.2)	13 (7, 25)	8 (7, 14)	13 (8, 16)
Injury And Poisoning	800–999	95 (19.0)	65 (13.0)	3.0 (0.6)	8 (7, 16)	8 (7, 12)	8 (5, 16)

RA, rheumatoid arthritis; ReA, reactive arthritis; HC, healthy control.

as those that are specific to these two diseases. To the best of our knowledge, the ReA cohort described herein will be the first systematic attempt to identify biomarkers associated with the development of ReA as well as those associated with disease susceptibility.

Herein we describe the establishment of a large cohort of adult subjects with RA or ReA as well as a matched cohort of control subjects. Serial serum samples starting up to 8–12 years prior to diagnosis have been obtained from subjects to enable an assessment of evolving proteomic markers of disease to potential elucidate novel mechanisms of disease and/or novel treatments that could be incorporated prior to full-fledged disease. Similar to the PREDICTS cohort for IBD [26], this first-of-its-kind cohort is well-suited to refine our understanding of two important rheumatological conditions with significant global disease burden.

Author contributions

Study concept and design: Chad K. Porter, Mark S. Riddle, Sunil Nagpal;

Acquisition of data: Chad K. Porter, Ramiro L. Gutierrez, and Mark S. Riddle;

Analysis and data interpretation: Chad K. Porter, Ashley N. Alcala;

Sample repository archiving: Renee M. Laird, Christian Gariepy;

Drafting of manuscript: Chad K. Porter, Sunil Nagpal;

Critical revision of manuscript for important intellectual content: Chad K. Porter, Mark S. Riddle, Renee M. Laird, Matthew Loza, Suzanne Cole, Christina Gariepy, Ashley Alcala, Ramiro Gutierrez, Frédéric Baribaud, Navin L. Rao, Sunil Nagpal;

Study supervision: Chad K. Porter, Mark S. Riddle, Renee M. Laird, Sunil Nagpal;

Funding

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Human subjects research and data protections

The human subjects' research (NMRC.2014.0012) under which the data and samples were obtained were approved as 'Exempt' research by the Naval Medical Research Center Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects. In addition, this study is conducted under a support agreement with the AFHSB. All data are were de-identified by personnel at AFHSB who removed the list of 18 identifiers outlined in 45 CFR 164.514(b)(2) and DOD 6025.18-R (DoD Health Information Privacy Regulation) prior to providing the data to the investigators.

Acknowledgements

RLG is a military service member and CKP is an employee of the U.S. Government. This work was prepared as part of their official duties. Title 17 U S C. §105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 U S C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

References

[1] J.J. Sacks, Y.H. Luo, C.G. Helmick, Prevalence of specific types of arthritis and other rheumatic conditions in the ambulatory health care system in the United States, 2001-2005, Arthritis Care Res. 62 (4) (2010) 460–464.

[2] R.C. Lawrence, et al., Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II, Arthritis Rheum. 58 (1) (2008) 26–35.

[3] C.G. Helmick, et al., Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I, Arthritis Rheum. 58 (1) (2008) 15–25.

[4] E.M. Badley, N.M. Kasman, The impact of arthritis on Canadian women, BMC Women's Health 4 (Suppl 1) (2004) S18.

[5] E.D. Ferucci, D.W. Templin, A.P. Lanier, Rheumatoid arthritis in American Indians and Alaska Natives: a review of the literature, Semin. Arthritis Rheum. 34 (4) (2005) 662–667.

[6] A.J. Silman, J.E. Pearson, Epidemiology and genetics of rheumatoid arthritis, Arthritis Res. 4 (Suppl 3) (2002) S265–S272.

[7] N. Bansback, et al., Factors associated with absenteeism, presenteeism and activity impairment in patients in the first years of RA, Rheumatology 51 (2) (2012) 375–384.

[8] J.S. Hill Gaston, M.S. Lillicrap, Arthritis associated with enteric infection, Best Pract. Res. Clin. Rheumatol. 17 (2) (2003) 219–239.

[9] M. Leirisalo-Repo, Reactive arthritis, Scand. J. Rheumatol. 34 (4) (2005) 251–259.

[10] T. Hannu, et al., Campylobacter-triggered reactive arthritis: a population-based study, Rheumatology 41 (3) (2002) 312–318.

[11] H. Locht, K.A. Krogfelt, Comparison of rheumatological and gastrointestinal symptoms after infection with Campylobacter jejuni/coli and enterotoxigenic Escherichia coli, Ann. Rheum. Dis. 61 (5) (2002) 448–452.

[12] T. Hannu, et al., Reactive arthritis attributable to Shigella infection: a clinical and epidemiological nationwide study, Ann. Rheum. Dis. 64 (4) (2005) 594–598.

[13] I. Paronen, Reiter's disease, A study of 344 cases observed in Finland, Acta Med. Scand. 130 (1948) 1–112.

[14] H. Locht, E. Kihlstrom, F.D. Lindstrom, Reactive arthritis after Salmonella among medical doctors–study of an outbreak, J. Rheumatol. 20 (5) (1993) 845–848.

[15] H. Locht, K. Mølbak, K.A. Krogfelt, Reactive arthritis (ReA) after an outbreak of Salmonella enteriditis, Scand. J. Rheumatol. 29 (suppl 114) (2000).

[16] A. Keat, Reiter's syndrome and reactive arthritis in perspective, N. Engl. J. Med. 309 (26) (1983) 1606–1615.

[17] J.R. Rees, et al., Persistent diarrhea, arthritis, and other complications of enteric infections: a pilot survey based on California FoodNet surveillance, 1998-1999, Clin. Infect. Dis. 38 (Suppl 3) (2004) S311–S317.

[18] M.S. Dworkin, et al., Reactive arthritis and Reiter's syndrome following an outbreak of gastroenteritis caused by Salmonella enteritidis, Clin. Infect. Dis. 33 (7) (2001) 1010–1014.

[19] R.D. Inman, et al., Postdysenteric reactive arthritis. A clinical and immunogenetic study following an outbreak of salmonellosis, Arthritis Rheum. 31 (11) (1988) 1377–1383.

[20] T. Hannu, et al., Reactive arthritis following an outbreak of Salmonella typhimurium phage type 193 infection, Ann. Rheum. Dis. 61 (3) (2002) 264–266.

[21] H.R. Noer, An "experimental" epidemic of Reiter's syndrome, J. Am. Med. Assoc. 198 (7) (1966) 693–698.

[22] R. Tertti, et al., An outbreak of Yersinia pseudotuberculosis infection, J. Infect. Dis. 149 (2) (1984) 245–250.

[23] J.E. Pope, et al., Campylobacter reactive arthritis: a systematic review, Semin. Arthritis Rheum. 37 (1) (2007) 48–55.

[24] J. Londono, et al., The association between serum levels of potential biomarkers with the presence of factors related to the clinical activity and poor prognosis in spondyloarthritis, Rev. Bras. Reumatol. 52 (4) (2012) 536–544.

[25] M.V. Rubertone, J.F. Brundage, The Defense Medical Surveillance System and the Department of Defense serum repository: glimpses of the future of public health surveillance, Am. J. Public Health 92 (12) (2002) 1900–1904.

[26] C.K. Porter, et al., Cohort profile of the PRoteomic evaluation and discovery in an IBD cohort of tri-service subjects (PREDICTS) study: rationale, organization, design, and baseline characteristics, Contemp Clin Trials Commun 14 (2019) 100345.

[27] S. Farid, G. Azizi, A. Mirshafiey, Anti-citrullinated protein antibodies and their clinical utility in rheumatoid arthritis, Int J Rheum Dis 16 (4) (2013) 379–386.

[28] C.K. Porter, et al., Travelers' diarrhea: an update on the incidence, etiology, and risk in military deployments and similar travel populations, Mil. Med. 182 (S2) (2017) 4–10.

[29] S. Olson, et al., Travelers' diarrhea: update on the incidence, etiology and risk in military and similar populations - 1990-2005 versus 2005-2015, does a decade make a difference? Trop Dis Travel Med Vaccines 5 (2019) 1.

[30] K.D. Deane, V.M. Holers, The natural history of rheumatoid arthritis, Clin. Ther. 41 (7) (2019) 1256–1269.

[31] S. Siebert, et al., Cytokines as therapeutic targets in rheumatoid arthritis and other inflammatory diseases, Pharmacol. Rev. 67 (2) (2015) 280–309.

[32] K. Loppenthin, et al., Morbidity and mortality in patients with rheumatoid arthritis compared with an age- and sex-matched control population: a nationwide register study, J. Comorbidity 9 (2019), 2235042X19853484.

[33] S. Rantapaa-Dahlqvist, et al., Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis, Arthritis Rheum. 48 (10) (2003) 2741–2749.

[34] M.M. Nielen, et al., Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors, Arthritis Rheum. 50 (2) (2004) 380–386.

[35] D.S. Majka, et al., Duration of preclinical rheumatoid arthritis-related autoantibody positivity increases in subjects with older age at time of disease diagnosis, Ann. Rheum. Dis. 67 (6) (2008) 801–807.

[36] J.R. Kolfenbach, et al., Autoimmunity to peptidyl arginine deiminase type 4 precedes clinical onset of rheumatoid arthritis, Arthritis Rheum. 62 (9) (2010) 2633–2639.

[37] W.H. Bos, et al., Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study, Ann. Rheum. Dis. 69 (3) (2010) 490–494.

[38] L.A. van de Stadt, et al., Monoclonal anti-citrullinated protein antibodies selected on citrullinated fibrinogen have distinct targets with different cross-reactivity patterns, Rheumatology 52 (4) (2013) 631–635.

[39] C. Rakieh, et al., Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study, Ann. Rheum. Dis. 74 (9) (2015) 1659–1666.

[40] J. Shi, et al., Anti-carbamylated protein (anti-CarP) antibodies precede the onset of rheumatoid arthritis, Ann. Rheum. Dis. 73 (4) (2014) 780–783.

[41] R.W. Gan, et al., Anti-carbamylated protein antibodies are present prior to rheumatoid arthritis and are associated with its future diagnosis, J. Rheumatol. 42 (4) (2015) 572–579.

[42] A.J. Silman, B. Ollier, R.A. Mageed, Rheumatoid factor detection in the unaffected first degree relatives in families with multicase rheumatoid arthritis, J. Rheumatol. 18 (4) (1991) 512–515.

[43] J. Sokolove, et al., Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis, PLoS One 7 (5) (2012), e35296.

[44] D. Benfaremo, et al., Musculoskeletal and rheumatic diseases induced by immune checkpoint inhibitors: a review of the literature, Curr. Drug Saf. 13 (3) (2018) 150–164.

[45] Y. Guo, et al., Immune checkpoint inhibitor PD-1 pathway is down-regulated in synovium at various stages of rheumatoid arthritis disease progression, PLoS One 13 (2) (2018), e0192704.

[46] R.F. van Vollenhoven, Sex differences in rheumatoid arthritis: more than meets the eye, BMC Med. 7 (2009) 12.

[47] I. Colmegna, R. Cuchacovich, L.R. Espinoza, HLA-B27-associated reactive arthritis: pathogenetic and clinical considerations, Clin. Microbiol. Rev. 17 (2) (2004) 348–369.