



## Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk

Elizabeth W Karlson, Lori B Chibnik, Peter Kraft, et al.

*Ann Rheum Dis* published online March 16, 2010

doi: 10.1136/ard.2009.120170

---

Updated information and services can be found at:

<http://ard.bmj.com/content/early/2010/03/12/ard.2009.120170.full.html>

---

*These include:*

**References**

This article cites 45 articles, 12 of which can be accessed free at:

<http://ard.bmj.com/content/early/2010/03/12/ard.2009.120170.full.html#ref-list-1>

**P<P**

Published online March 16, 2010 in advance of the print journal.

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

**Notes**

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To order reprints of this article go to:

<http://ard.bmj.com/cgi/reprintform>

To subscribe to *Annals of the Rheumatic Diseases* go to:

<http://ard.bmj.com/subscriptions>

# Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk

Elizabeth W Karlson,<sup>1</sup> Lori B Chibnik,<sup>1</sup> Peter Kraft,<sup>2,3</sup> Jing Cui,<sup>1</sup> Brendan T Keenan,<sup>1</sup> Bo Ding,<sup>4</sup> Souyma Raychaudhuri,<sup>1</sup> Lars Klareskog,<sup>5</sup> Lars Alfredsson,<sup>4</sup> Robert M Plenge<sup>1</sup>

<sup>1</sup>Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA  
<sup>2</sup>Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA  
<sup>3</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA  
<sup>4</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden  
<sup>5</sup>Rheumatology Unit, Department of Medicine, Karolinska Institutet/Karolinska Hospital, Stockholm, Sweden

Correspondence to Elizabeth W Karlson, 75 Francis Street, Boston, MA 02115, USA; [ekarlson@partners.org](mailto:ekarlson@partners.org)

Accepted 1 December 2009

## ABSTRACT

**Background** Recent discoveries of risk alleles have made it possible to define genetic risk profiles for patients with rheumatoid arthritis (RA). This study examined whether a cumulative score based on 22 validated genetic risk alleles for seropositive RA would identify high-risk, asymptomatic individuals who might benefit from preventive interventions.

**Methods** Eight human leucocyte antigen (HLA) alleles and 14 single-nucleotide polymorphisms representing 13 validated RA risk loci were genotyped among 289 white seropositive cases and 481 controls from the US Nurses' Health Studies (NHS) and 629 white cyclic-citrullinated peptide antibody-positive cases and 623 controls from the Swedish Epidemiologic Investigation of Rheumatoid Arthritis (EIRA). A weighted genetic risk score (GRS) was created, in which the weight for each risk allele is the log of the published odds ratio (OR). Logistic regression was used to study associations with incident RA. Area under the curve (AUC) statistics were compared from a clinical-only model and clinical plus genetic model in each cohort.

**Results** Patients with GRS >1.25 SD of the mean had a significantly higher OR of seropositive RA in both NHS (OR=2.9, 95%CI 1.8 to 4.6) and EIRA (OR 3.4, 95% CI 2.3 to 5.0) referent to the population average. In NHS, the AUC for a clinical model was 0.57 and for a clinical plus genetic model was 0.66, and in EIRA was 0.63 and 0.75, respectively.

**Conclusion** The combination of 22 risk alleles into a weighted GRS significantly stratifies individuals for RA risk beyond clinical risk factors alone. Given the low incidence of RA, the clinical utility of a weighted GRS is limited in the general population.

Rheumatoid arthritis (RA) is a complex autoimmune disease thought to develop in genetically predisposed individuals when exposed to certain environmental factors. Early diagnosis and treatment strategies are critical to minimise disability from joint destruction.<sup>1</sup> Although epidemiological research has produced convincing data linking cigarette smoking to RA risk,<sup>2–4</sup> and genetic variants associated with RA risk in the major histocompatibility complex (MHC) region were discovered over 30 years ago,<sup>5</sup> these risk factors are not used clinically for behaviour modification, preventive therapy, or in establishing a diagnosis of RA. Similarly, the presence of RA-specific autoantibodies and inflammatory biomarkers appear years before disease onset and predict more severe disease, but are not used in clinical medicine before the onset of symptoms.<sup>6–9</sup>

Advances in human genetics have led to a dramatic increase in the number of validated disease

risk alleles in RA. There are now up to 22 risk alleles that explain approximately one-third of the genetic burden of seropositive RA risk.<sup>5 10–20</sup> Much of the risk is derived from eight alleles that reside within the MHC region,<sup>5</sup> with up to 5% of risk explained by the 14 alleles outside of the MHC.<sup>20</sup> The discoveries of these alleles for RA, and similar discoveries for risk alleles in other diseases, has spurred much discussion about the clinical validity of using genetic results in personalised medicine.<sup>21–24</sup>

Despite these advances, it is not clear how to utilise genetic information for the prediction of RA risk in clinical practice. A critical first step is to understand the role of aggregate genetic risk factors, rather than associations of individual alleles with RA. Towards this end, we used 22 validated RA risk alleles to derive an aggregate genetic risk score (GRS) in seropositive RA patients derived from over 238 000 prospectively followed subjects from the US Nurses' Health Study (NHS) and seropositive RA patients derived from a large case-control study of over 3600 subjects from Sweden (Epidemiologic Investigation of Rheumatoid Arthritis; EIRA). We calculated odds ratios (OR) for seropositive RA relative to the median risk group in these datasets and estimated genotype-specific incidence, which is a more useful measure of risk in a clinical setting. We compared predicted multilocus OR—formed by taking the product of individual-locus OR estimated in a previous meta-analysis<sup>20</sup>—to multilocus OR estimated in this dataset. We included the strongest epidemiological risk factors for RA in the general population in the models (age, sex and smoking) as 'clinical' risk factors. Although the GRS is strongly associated with seropositive RA and adds significantly to the discrimination of a clinical model, the genotype-specific incidence remains low, suggesting that genetic information is not yet clinically useful in an asymptomatic individual patient.

## METHODS

### Study sample

The NHSI is a prospective cohort of 121 700 female nurses, aged 30–55 years in 1976 in which 32 826 (27%) NHSI participants aged 43–70 years provided blood samples for future studies and an additional 33 040 (27%) provided buccal cell samples, a total of 65 866 (54% of the cohort). NHSII is a similar prospective cohort, established in 1989, with 116 609 female nurses aged 25–42 years in which 29 611 (25%) provided blood samples for future studies. In the current study, we combine both NHSI and NHSII, herein referred to simply as 'NHS'. All women in both cohorts completed an initial questionnaire and have been followed

## Extended report

biennially by questionnaire to update exposures and disease diagnoses. The specificity of connective tissue disease detection using a staged series design is very high, reducing the misclassification of healthy individuals.<sup>25</sup> RA cases were validated using previously described methods,<sup>4</sup> in which two board-certified rheumatologists trained in chart abstraction independently conducted a medical record review blinded to the second reviewer's result, examining the charts for the American College of Rheumatology (ACR) classification criteria for RA,<sup>26</sup> date of first RA symptom, evidence of RA-specific medication treatment and the treating physician's diagnosis. Definite RA included subjects with four of the seven ACR criteria documented in the medical record or agreement by two rheumatologists on the diagnosis of RA with three documented ACR criteria for RA and a diagnosis of RA by their physician. Seropositive status was determined by chart review, and in some cases by direct assay, as previously described.<sup>9</sup> Each NHS participant with confirmed incident or prevalent RA was matched by year of birth, race/ethnicity, menopausal status and postmenopausal hormone use to a single healthy woman in the same cohort without RA.

This initial NHS nested case-control dataset consisted of 585 RA cases and 585 matched controls. To minimise potential population stratification, we excluded non-white women (based on self-report), resulting in 564 total RA cases and 571 controls. We restricted our analysis to only seropositive RA, resulting in a sample of 327 seropositive RA cases and 571 controls. Covariate information was collected from the subjects in both cohorts by prospective biennial questionnaires regarding diseases, lifestyle and health practices. All aspects of this study were approved by the Partners' HealthCare Institutional Review Board.

EIRA is a population-based case-control study on incident RA in Sweden. Data on more than 3600 cases and controls were collected between May 1996 and December 2006. As described previously,<sup>3,27</sup> a case is defined as an individual who fulfils ACR 1987 criteria for the classification of RA and had symptoms for less than 1 year. For each potential case, a control subject was randomly selected from the study base, taking into consideration the subject's age, sex and geographical location. In total, 659 confirmed cyclic-citrullinated peptide (CCP) positive RA cases and 650 controls were included. All aspects of the EIRA study were approved by the Karolinska Institutet Institutional Review Board.

## Selection of genetic risk factors and genotyping

We selected all validated seropositive RA susceptibility single-nucleotide polymorphisms (SNP) established before September 2008. We define validated as those alleles demonstrating  $p < 5 \times 10^{-7}$  with evidence of replication at  $p < 0.05$  in at least one independent study.<sup>10-17,20</sup> One locus, *CDK6*, has a strong but not unequivocal evidence of association based on these criteria. In NHS, low resolution *HLA-DRB1* genotyping was performed using PCR with sequence-specific primers (SSP) using OLERUP SSP kits (Qiagen, West Chester, Pennsylvania, USA), as previously described.<sup>28</sup> For samples with positive two-digit human leucocyte antigen (HLA) signals, SSP were used for high-resolution four-digit allele detection of *DRB1*\*0401, \*0404, \*0405, \*0408, \*0101, \*0102, \*09 and \*1001. In EIRA, low-resolution HLA typing was performed using Olerup PCR-SSP (DR low resolution and DR4 kits; Olerup SSP AB, Saltsjöbaden, Sweden). High-resolution typing was performed for positive \*04 samples. Four-digit HLA subtypes were thus available from EIRA for \*0401, \*0404, \*0405, \*0408 and two-digit subtypes were available for other alleles. All non-MHC risk alleles for both NHS and EIRA were genotyped using iPLEX (Sequenom, San Diego, California, USA) at the Broad Institute, as previously described.<sup>20</sup> All SNP had call rates greater than 95% and Hardy-Weinberg equilibrium  $p$  values greater than 0.01.

We filtered our data to account for missing genotype information, dropping individuals with greater than 10% missing SNP data and dropping individuals missing any HLA data. In NHS, among 327 seropositive RA cases, six (2%) were missing HLA data and 32 (10%) were missing greater than 10% SNP information, leaving 289 seropositive RA cases in the analysis (table 1). Among 571 controls, 20 (4%) were missing HLA and 70 (12%) were missing greater than 10% SNP information data, leaving us with 481 controls in the analysis. In EIRA, among 659 cases, three (0.5%) were missing HLA data and 27 (4%) were missing greater than 10% SNP information, leaving 629 cases in the analysis. Among 650 controls, one (0.1%) was missing HLA results and 25 (4%) were missing greater than 10% SNP information, leaving 623 controls in the analysis. The higher rates of genotyping failure in NHS were due primarily to poor quality cheek cell DNA samples. We are confident that this omission is completely at random, and therefore does not bias our results, because the case and control samples were randomly interspersed on the genotyping plate and our resulting OR are consistent with previously published results (see table 2).

**Table 1** Characteristics of seropositive RA cases and matched controls in NHS and CCP-positive RA cases and matched controls in EIRA

	NHS		EIRA	
	RA cases (n=289)	Controls (n=481)	RA cases* (n=629)	Controls† (n=623)
Age, mean (SD)‡	55.3 (±8.1)	55.7 (±7.9)	51.2 (±11.7)§	52.1 (±11.8)¶
Pack-years among smokers, mean (SD)	25.2 (±17.5)	23.0 (±21.4)	18.3 (±15.3)	14.7 (±13.8)
RA features				
Mean age at diagnosis (SD)	56.2 (±9.9)	–	51.2 (±11.7)§	–
Rheumatoid factor, positive (%)	270 (93.4%)	–	523 (87.3%)§	–
CCP** positive (%)	106 (55.5%)	–	629 (100%)	–
Seropositive (%)	289 (100.0%)	–	629 (100%)	–
Rheumatoid nodules (%)	45 (15.6%)	–	–	–
Radiographic changes (%)	95 (32.9%)	–	–	–

\*EIRA: 184 male cases, 446 female cases.

†EIRA: 163 male controls, 460 female controls.

‡Age at blood draw.

§31 EIRA cases missing covariate data.

¶15 EIRA controls missing covariate data.

\*\*CCP assayed in subset of NHS cases (n=191) with stored blood samples collected at different points with respect to RA onset, up to 12 years prior to onset or after diagnosis.

CCP, cyclic-citrullinated peptide antibody; EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; NHS, Nurses' Health Studies; RA, rheumatoid arthritis.

## Statistical methods

Characteristics of RA cases and controls were summarised by means and SD for continuous variables and frequency and percentage for categorical variables. Data for NHS were presented separately from data for EIRA. All analyses were performed using SAS version 9.1 or version 9.2.

## Selection of epidemiological covariates

In NHS and EIRA, lifetime history of smoking was collected at baseline. In the NHS cohorts, data concerning current smoking and number of cigarettes smoked per day were updated in 2-year questionnaire cycles and data on pack-years of smoking (number of packs per day  $\times$  number of years smoking) were selected from the questionnaire cycle before the date of RA diagnosis (or index date in controls). In EIRA, pack-years of smoking were calculated before the onset of RA for cases or index date for controls. We included age, sex, geographical region (in EIRA only) and pack-years of smoking as 'clinical' risk factors in the models.

## Association between genetic risk alleles and RA

We used logistic regression to study the association of each allele with the risk of seropositive RA according to an additive log-odds model in NHS and in EIRA.

## Weighted GRS

We developed a 'weighted GRS' (wGRS) that utilised the allelic OR from published studies to account for the strength of the genetic association within each allele. We calculated a wGRS22 that included eight *HLA-DRB1* 'shared epitope' (*HLA-SE*) alleles and 14 non-MHC risk alleles, and a wGRS14 (no HLA) that included only the 14 non-MHC risk alleles. This is preferred over a simple count GRS, calculated as the sum of the number of risk alleles

carried, as *PTPN22* and *HLA-SE* have substantially higher OR for RA than do the more recently discovered SNP. The weights used in the wGRS were calculated as the natural log of the published OR with respect to the risk allele as presented in table 2. The OR for *HLA-SE* alleles were derived from a recent meta-analysis of all published studies.<sup>29</sup> The OR for the 14 non-MHC alleles were derived from published studies for which results have been extensively replicated, including the following alleles: *PTPN22* (rs2476601),<sup>10</sup> *TRAF1-C5* (rs3761847),<sup>13</sup> *STAT4* (rs7574865),<sup>12</sup> *TNFAIP3* (rs17066662 in linkage disequilibrium with 10499194,  $r^2=1.0$ ),<sup>14</sup> *TNFAIP3* (rs6920220).<sup>14</sup> We also included nine alleles from a meta-analysis of GWAS data for 3393 seropositive cases and 12 462 controls with replication in 3929 seropositive RA cases and 5807 matched controls by Raychaudhuri *et al.*<sup>20</sup> *CD40* (rs4810485), *CCL21* (rs2812378), *CTLA4* (rs3087243), *PADI4* (rs2240340), *CDK6* (rs42041), *TNFRSF14* (rs3890745), *PRKCQ* (rs4750316), *KIF5A* (rs1678542) and *4q27* (rs6822844). For each non-MHC allele, we chose the OR in replication samples to avoid overestimation of the true effect size.<sup>30</sup> In EIRA, we used a proxy SNP for *STAT4* (rs11889341,  $r^2=1.0$  with rs7574865) and a proxy SNP for *KIF5A* (rs775322,  $r^2=1.0$  with rs1678542). For any individual with missing genotype data for a particular SNP, we assigned the expected allele count (twice the risk allele frequency) to that individual. We tested for epistasis and did not find any significant gene-gene interaction, in agreement with our previous studies.<sup>13 14 20</sup> Our results are consistent with a multiplicative genetic model. We did not consider more complex HLA associations, including analysis of compound heterozygotes that have a substantially higher risk such as HLA 0401/0404 (nine cases and three controls in NHS and 52 cases and four controls in EIRA).

To determine the cumulative effect of the 14 or 22 alleles on the risk of RA we first divided wGRS scores into seven categories based on the mean and SD of the wGRS distribution in the

**Table 2** Allele frequencies and association with seropositive RA in NHS and CCP-positive RA in EIRA for 22 alleles

Loci	SNP	Risk allele	Published OR†	NHS			EIRA		
				Cases	Controls	OR (95% CI)	Cases	Controls	OR (95% CI)
				RAF	RAF		RAF	RAF	
DRB*0401	SE	SE	3.30	0.15	0.09	1.70 (1.25 to 2.29)	0.30	0.13	3.04 (2.45 to 3.77)
DRB*0404	SE	SE	1.85	0.07	0.04	1.81 (1.13 to 2.90)	0.10	0.04	2.94 (2.07 to 4.18)
DRB*0405	SE	SE	3.84	0.02	0.003	6.31 (1.75 to 22.81)	0.008	0.004	1.98 (0.67 to 5.83)
DRB*0408	SE	SE	1.04	0.003	0.006	0.55 (0.11 to 2.76)	0.02	0.004	5.07 (1.92 to 13.34)
DRB*0101	SE	SE	1.60	0.11	0.06	1.83 (1.27 to 2.63)	0.14‡	0.10	1.47 (1.15 to 1.89)
DRB*0102	SE	SE	1.10	0.01	0.01	1.69 (0.63 to 4.54)	–	–	–
DRB*1001	SE	SE	2.35	0.01	0.01	2.26 (0.78 to 6.58)	0.02	0.01	1.53 (0.79 to 2.97)
DRB*09	SE	SE	1.48	0.01	0.01	1.34 (0.36 to 5.03)	0.02	0.02	0.94 (0.53 to 1.67)
PTPN22	rs2476601	T	1.75	0.14	0.09	1.75 (1.27 to 2.41)	0.17	0.12	1.49 (1.19 to 1.86)
TRAF1-C5	rs3761847	G	1.32	0.44	0.40	1.18 (0.95 to 1.45)	0.51	0.45	1.29 (1.1 to 1.51)
STAT4	rs7574865	T	1.27	0.24	0.21	1.18 (0.93 to 1.51)	0.24	0.22	1.07 (0.89 to 1.29)
TNFAIP3	rs17066662	C*	1.33	0.73	0.73	1.00 (0.80 to 1.25)	0.80	0.79	1.04 (0.86 to 1.26)
TNFAIP3	rs6920220	A	1.22	0.22	0.21	1.05 (0.82 to 1.35)	0.26	0.22	1.23 (1.02 to 1.48)
CD40	rs4810485	G*	1.15	0.76	0.75	1.05 (0.83 to 1.35)	0.78	0.72	1.36 (1.13 to 1.64)
CCL21	rs2812378	C	1.12	0.35	0.34	1.07 (0.87 to 1.33)	0.36	0.34	1.11 (0.94 to 1.31)
CTLA4	rs3087243	G*	1.11	0.54	0.52	1.06 (0.87 to 1.30)	0.63	0.59	1.20 (1.02 to 1.40)
PADI4	rs2240340	A	1.02	0.43	0.40	1.10 (0.89 to 1.35)	0.42	0.39	1.14 (0.97 to 1.33)
CDK6	rs42041	G	1.11	0.27	0.24	1.21 (0.96 to 1.53)	0.25	0.24	1.02 (0.85 to 1.23)
TNFRSF14	rs3890745	T*	1.12	0.67	0.68	0.93 (0.75 to 1.16)	0.70	0.70	1.02 (0.86 to 1.21)
PRKCQ	rs4750316	C*	1.14	0.85	0.82	1.25 (0.95 to 1.65)	0.82	0.80	1.09 (0.89 to 1.34)
KIF5A	rs1678542	G*	1.12	0.70	0.64	1.28 (1.02 to 1.60)	0.59	0.58	1.06 (0.91 to 1.24)
IL2/IL21	rs6822844	G*	1.09	0.84	0.84	0.99 (0.74 to 1.31)	0.85	0.82	1.29 (1.04 to 1.62)

\*Major allele.

†Published OR with respect to the risk allele.

‡Only two-digit DR1 data available.

CCP, cyclic-citrullinated peptide antibody; EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; NHS, Nurses' Health Studies; OR, odds ratio; RA, rheumatoid arthritis; RAF, risk allele frequency; SE, *HLA-DRB1* shared epitope; SNP, single-nucleotide polymorphism.

## Extended report

controls. Dividing our score into seven categories provided the most robust distribution, allowing us to parse out the highest and lowest risk groups while ensuring that there were sufficient numbers of cases and controls in these extreme categories of interest. Additional details on determination of the groupings are available in the supplementary methods. We used logistic regression models adjusting for age, sex, geographical region (in EIRA) and pack-years of smoking to study the association of wGRS22 with seropositive RA and wGRS14 (no HLA) with seropositive RA (table 2), comparing each group with a referent median group. An ordinal wGRS variable based on our groupings was used to calculate a p value for trend. Finally, we calculated the odds of RA for the top group (group 7) compared with the bottom group (group 1) in two ways. First, we used group 1 as the referent group, similar to the method used in other GRS analyses of complex diseases (eg, macular degeneration,<sup>31</sup> prostate cancer,<sup>32–33</sup> lipid levels and heart disease<sup>34–37</sup> and diabetes<sup>38–40</sup>). Second, because group 1 had few cases and the first method only considers subjects in groups 7 and 1 we compared the median wGRS score in group 7 to the median wGRS score in group 1 using a model derived from an ordinal wGRS variable in which each group was given its median wGRS value as a score.

### Additional statistical analysis

To determine how well our wGRS predictors discriminate between cases and controls, we generated receiver operating characteristic (ROC) curves by plotting the sensitivity of the wGRS22 score (continuous) against 1-specificity and calculated the area under the curve (AUC) for both NHS and EIRA. Because there are few established epidemiological predictors other than age, sex and smoking in the asymptomatic general population, any improvement in the ROC curve contributed by the wGRS may have value in a clinical setting. ROC curves were plotted for a 'clinical' model that included year of birth and pack-years of smoking in NHS and age, sex, pack-years of smoking and geographical region in EIRA, for a 'clinical plus genetic' model based on adding wGRS14 (no HLA) and a full 'clinical plus genetic' model that included clinical factors and wGRS22. The AUC statistics were compared using a non-parametric approach with each 'clinical plus genetic' model compared with the 'clinical' model as described by DeLong *et al.*<sup>41</sup>

To judge how well previously reported association results could be used to distinguish cases and controls in this dataset, using a likelihood ratio test we studied the calibration of a model for the multilocus OR, formed by multiplying the individual-locus OR, from the published OR in table 2 (ie, exponentiating the continuous wGRS) (see supplementary methods).

To determine whether wGRS22 is clinically useful on an individual patient basis, we estimated risk score-specific incidence among US women. We used the average annual incidence estimated from the full NHS cohort:  $\lambda=33/100\ 000$ ; the risk score-specific OR  $OR_G$ ; and one minus the population attributable risk  $(1-PAR)=1/(\sum_G OR_G \pi_G)$ , where  $\pi_G$  is the prevalence of genotype G in the controls. The risk score-specific incidence is then  $\lambda(1-PAR)OR_G \pi_G$ .<sup>42</sup> To estimate risk score-specific absolute risks among Swedish men and women we used data on RA incidence rates in northern Europe from Alamanos *et al.*, and estimated Swedish annual incidence rates of  $\lambda=40/100\ 000$  for women and  $\lambda=20/100\ 000$  for men.<sup>43</sup>

## RESULTS

### Patients

Characteristics of RA cases and controls for NHS and EIRA are presented in table 1. The demographics of both groups are

similar although seropositive status in NHS was defined as either rheumatoid factor or CCP positive and in EIRA as those who were CCP positive; NHS includes patients with new-onset and long-standing disease, whereas EIRA patients are of new onset only; and NHS is all female, whereas EIRA is both female and male (at the expected ratio of approximately 3:1).

### Association between genetic risk alleles and RA

The results for each of the 22 risk alleles with the risk of RA are presented in table 2. The majority of the OR are in the same direction for the risk allele and of the same magnitude as from published discovery studies. Not surprisingly, many of the 95% CI cross 1.0, as might be expected given the modest OR of the non-MHC alleles and the sample sizes of the two cohorts.

### Observed relative risk with GRS

The results for wGRS22 as a predictor of seropositive RA are presented in table 3 and figure 1. For wGRS22, the median level of risk (group 4, containing 20% of controls) was used as the referent group. Those with the highest risk (group 7) had a significantly higher odds of RA compared with group 4 in both NHS (OR 2.85, 95% CI 1.75 to 4.64) and in EIRA (OR 3.36, 95% CI 2.27 to 4.97) (table 3, figure 1A,B). Using group 1 (lowest level of risk) as a reference group, group 7 had a higher odds of RA, 5.61 (95% CI 2.41 to 13.07) in NHS and 8.83 (95% CI 4.77 to 16.32) in EIRA. In the ordinal model that takes into account all data in the model, group 7 had even higher odds of RA, 6.30 (95% CI 3.78 to 10.48) for NHS and 12.31 (95% CI 8.12 to 18.67) for EIRA. The trends across all seven categories of risk were highly significant, with  $p<0.001$  for both NHS and EIRA.

A similar analysis was performed using only the 14 non-HLA risk alleles (table 3, figure 1C,D). For wGRS14 (no HLA), those in group 7 (highest risk) relative to group 4 (median) had an elevated OR of 2.52 (95% CI 1.49 to 4.28) and 2.43 (95% CI 1.62 to 3.63) in both NHS and EIRA, respectively. Using group 1 as the reference, group 7 had a higher odds of RA 3.43 (95% CI 1.74 to 6.74) and 2.81 (95% CI 1.66 to 4.73) in NHS and EIRA, respectively. The OR from an ordinal model for group 7 was 2.39 (95% CI 1.44 to 3.98) in NHS and 3.22 (95% CI 2.14 to 4.86) in EIRA. The trends across all seven categories were highly significant ( $p=0.002$  for NHS,  $p<0.001$  for EIRA).

### Discrimination of cases and controls by GRS scores

The statistics used during the discovery phase of research (such as OR or p values for association) are not the most appropriate measures for evaluating the predictive value of genetic profiles in clinical practice. Other measures—sensitivity, specificity and risk classification—are more useful when proposing a genetic profile for risk prediction.<sup>23–24–44</sup> ROC curves that plot the sensitivity of the GRS score (continuous) against 1-specificity, and calculated the AUC, also known as the c-statistic, for both NHS and EIRA are shown in figure 2. In the NHS, the AUC for the clinical model including age and pack-years of smoking was 0.566. Adding wGRS14 (no HLA) to this model did not significantly improve discrimination (AUC 0.589;  $p=0.31$ ). Adding HLA subtypes to the clinical plus genetic model significantly improved discrimination relative to both the clinical model and the clinical plus wGRS14 model (AUC 0.660;  $p<0.001$  for both comparisons). In EIRA, ROC curves for the clinical model adjusted for age, sex, geographic region and pack-years of smoking demonstrate significant improvements in discrimination with the addition of wGRS14 (no HLA) or wGRS22 scores, with AUC of 0.627, 0.662 and 0.752 (clinical plus wGRS22 vs wGRS14 (no

**Table 3** Weighted GRS scores and OR of seropositive RA in NHS and CCP-positive RA in EIRA

	NHS			EIRA		
	Sero+ cases	Controls	OR (95% CI)*	CCP+ cases	Controls	OR (95% CI)†
wGRS22 groups						
1	8	32 (7%)	0.51 (0.22 to 1.19)	20	54 (9%)	0.38 (0.20 to 0.71)
2	27	79 (16%)	0.67 (0.38 to 1.19)	32	101 (16%)	0.32 (0.19 to 0.53)
3	37	113 (23%)	0.67 (0.40 to 1.13)	54	133 (21%)	0.44 (0.28 to 0.70)
4	49	96 (20%)	1.00 (ref)	96	107 (17%)	1.00 (ref)
5	51	64 (13%)	1.69 (1.01 to 2.82)	78	85 (14%)	1.01 (0.65 to 1.57)
6	32	38 (8%)	1.77 (0.98 to 3.20)	96	65 (10%)	1.52 (0.98 to 2.36)
7	85	59 (12%)	2.85 (1.75 to 4.64)	253	78 (13%)	3.36 (2.27 to 4.97)
7 versus 1‡			5.61 (2.41 to 13.07)			8.83 (4.77 to 6.32)
wGRS22 ordinal model						
7 versus 1§			6.30 (3.78 to 10.48)			12.31 (8.12 to 18.67)
wGRS14 (no HLA) groups						
1	17	47 (10%)	0.74 (0.38 to 1.43)	42	60 (10%)	0.86 (0.53 to 1.42)
2	38	68 (14%)	1.13 (0.66 to 1.94)	61	79 (13%)	1.01 (0.65 to 1.58)
3	46	82 (17%)	1.09 (0.65 to 1.83)	88	116 (19%)	0.95 (0.64 to 1.41)
4	45	90 (19%)	1.00 (ref)	114	141 (23%)	1.00 (ref)
5	55	89 (19%)	1.20 (0.73 to 1.98)	107	94 (15%)	1.62 (1.09 to 2.41)
6	28	57 (12%)	1.00 (0.55 to 1.80)	85	60 (10%)	1.89 (1.22 to 2.94)
7	60	48 (10%)	2.52 (1.49 to 4.28)	132	73 (12%)	2.43 (1.62 to 3.63)
7 versus 1‡			3.43 (1.74 to 6.74)			2.81 (1.66 to 4.73)
wGRS14 (no HLA) ordinal model						
7 versus 1§			2.39 (1.44 to 3.98)			3.22 (2.14 to 4.86)

\*OR adjusted for age and pack-years smoking.

†OR adjusted for age, sex, geographical region and pack-years smoking.

‡OR adjusted for age and pack-years smoking (NHS), sex, geographical region and pack-years smoking (EIRA) using group 1 as referent group.

§Model based estimate of group 7 versus group 1 using an ordinal model and takes into account all data in the model.

CCP, cyclic-citrullinated peptide antibody; EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; GRS, genetic risk score; HLA, human leucocyte antigen; NHS, Nurses' Health Studies; OR, odds ratio; RA, rheumatoid arthritis; wGRS22, weighted genetic risk score with 22 alleles; wGRS14 (no HLA), weighted genetic risk score with 14 alleles, without HLA alleles.

HLA),  $p < 0.001$ ; clinical plus wGRS22 vs clinical,  $p < 0.001$ ; clinical plus wGRS14 (no HLA) vs clinical  $p = 0.002$ ).

### Genotype-specific risk and comparison between predicted and observed OR

Figure 3 plots the distribution of genotype (or genotype category) annual incidence for predicted models based on previous locus-specific OR estimates and the observed categorised wGRS models fit to these datasets. For NHS and women in EIRA, the observed risks from our groupings approximate the predicted risk from a continuous wGRS, except for the lowest risk group (group 1) in which observed risk exceeds predicted risk. For men in EIRA the observed risks from our groupings approximate the predicted risk from a continuous wGRS except for the highest risk group (group 7) in which predicted risk exceeds observed risk, suggesting that in the highest risk group the risk based on grouping the wGRS is biased toward the null or the predicted risk is an overestimate. Figure 3 also shows that despite the statistically significant improvement in the AUC after incorporating the wGRS22, the predicted risks of RA were still small (<1% annual risk) for all of the observed genotypes.

### DISCUSSION

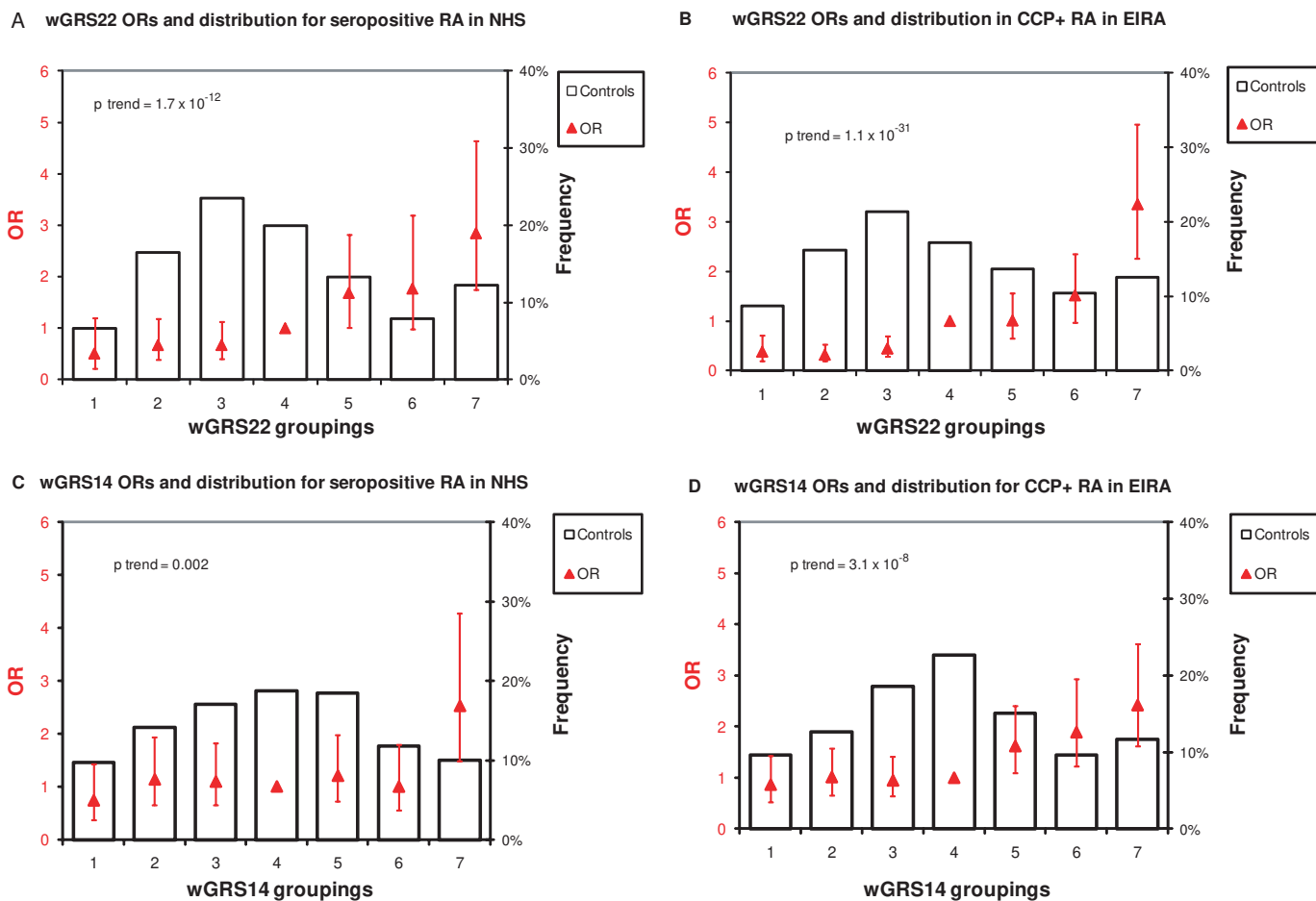
Until 2004, only two genetic loci had been unequivocally associated with the risk of RA susceptibility: *HLA-DRB1* and *PTPN22*.<sup>5 10</sup> Recent large studies using genome-wide scans or related methodologies have discovered and replicated 12 additional non-MHC risk loci.<sup>12–15 20</sup> In the current study, we develop a wGRS including 14 established risk alleles from 13 non-MHC RA loci and eight HLA subtypes based on high-resolution genotyping. We demonstrate that a weighted composite GRS significantly improves the discrimination ability of the model for

seropositive RA compared with no RA when compared with a risk model with epidemiological variables alone when applied in the general population.

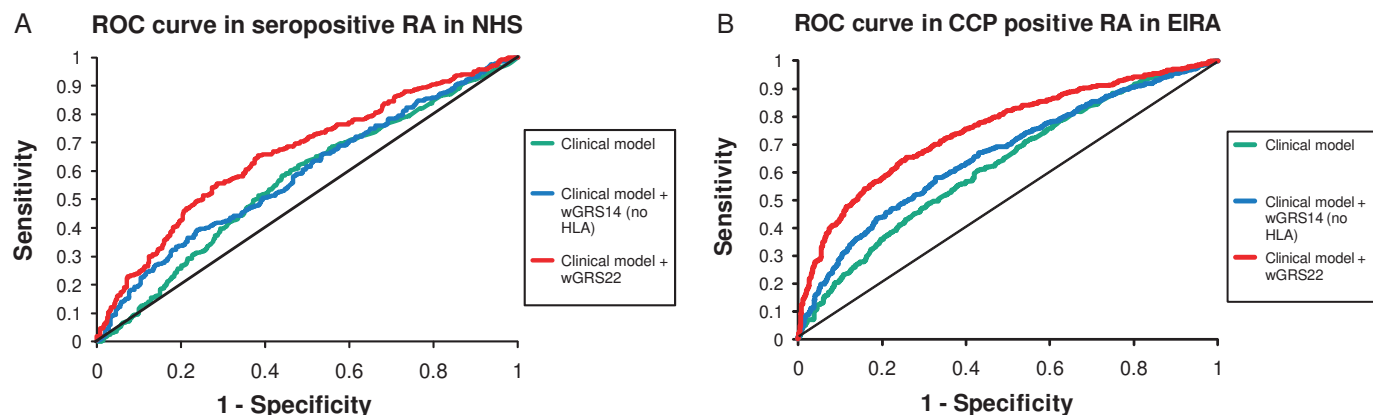
We found that in our top wGRS group with 22 alleles there was a 2.9-fold increase in the odds of seropositive RA compared with the median wGRS group, and a 5.6-fold increase in the odds of RA compared with the wGRS group with the lowest score in the NHS. In EIRA, the top wGRS group with 22 alleles had a higher increase in the odds of RA than in the US cohort, with a 3.4-fold increase compared with the median wGRS group and an 8.8-fold increase compared with the lowest wGRS group. However, comparing results from the cumulative score with 14 alleles, without the *HLA-SE* alleles, there were similar increased OR for RA in both cohorts (2.5-fold in NHS and 2.4-fold in EIRA). This suggests that the increased risk in the Swedish cohort is primarily due to the higher frequency of *HLA-SE* alleles in that population, which may reflect the higher percentage of patients seropositive for CCP autoantibodies (table 3).

Publications on genetic risks for other complex human diseases and quantitative traits such as macular degeneration,<sup>31</sup> prostate cancer,<sup>32 33</sup> lipid levels and heart disease,<sup>34–37</sup> height<sup>45 46</sup> and diabetes<sup>38–40</sup> have combined risk alleles into a single risk score simply by summing the number of risk alleles carried. Our study extends the methodology by weighting the risk score by the published allelic OR, thus accounting for the different strengths of association for genes such as the *HLA-SE* and *PTPN22*. Although models have been developed to identify which patients presenting with early inflammatory arthritis will progress to RA,<sup>47</sup> this is the first demonstration of risk models that include all known genetic risk factors and the two strongest epidemiological factors, age and smoking, in the prediction of incident seropositive RA among healthy individuals without symptoms.

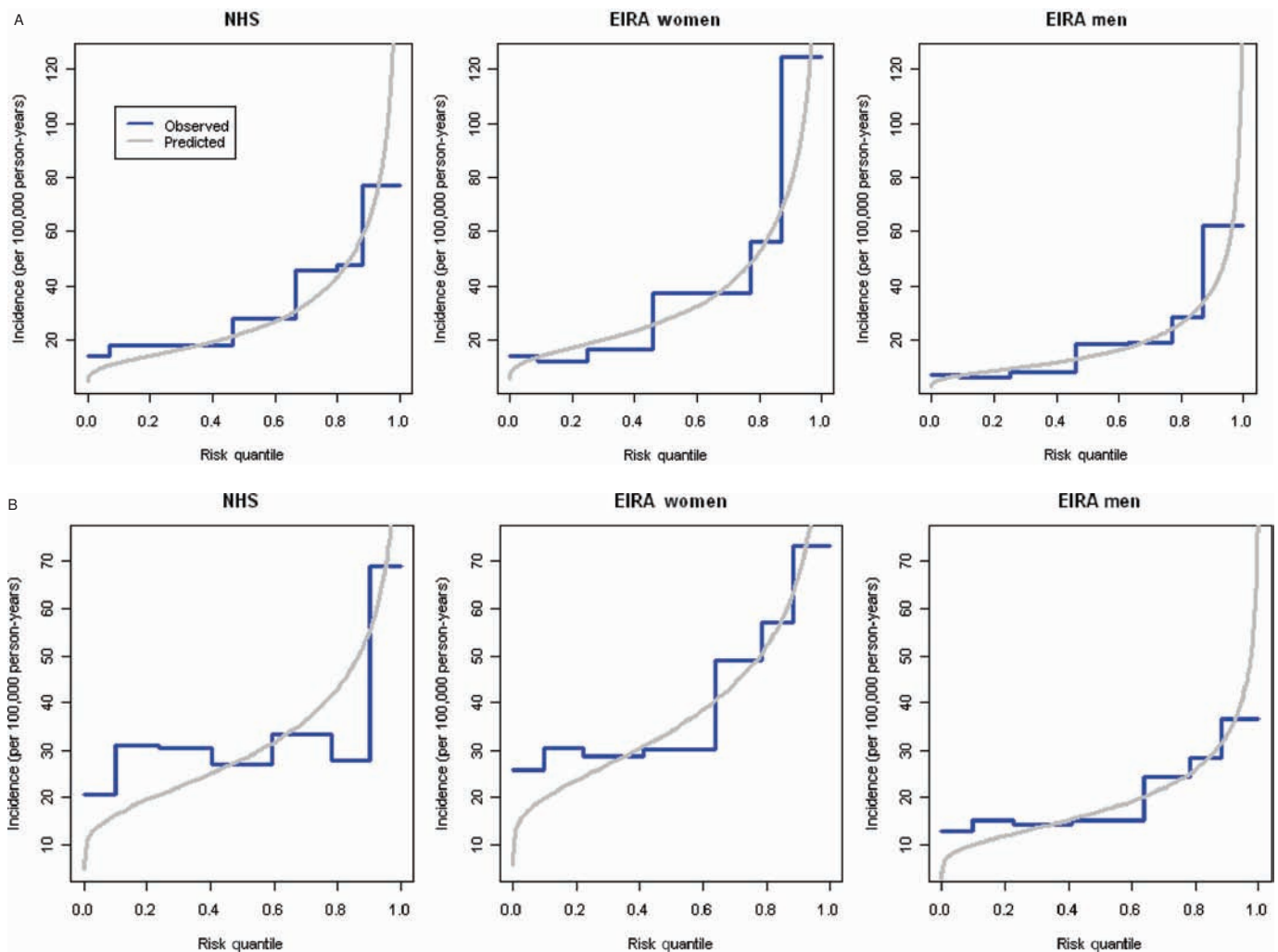
## Extended report



**Figure 1** OR for wGRS22 and wGRS14 (no HLA) in NHS and EIRA. wGRS distribution among controls shown in bars, OR shown in red triangles. (A) OR for wGRS22 and seropositive RA in NHS; (B) OR for wGRS22 and CCP-positive RA in EIRA; (C) OR for wGRS14 (no HLA) and seropositive RA in NHS; (D) OR for wGRS14 (no HLA) and CCP-positive RA in EIRA. CCP, cyclic-citrullinated peptide antibody; EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; HLA, human leucocyte antigen; NHS, Nurses' Health Studies; RA, rheumatoid arthritis; wGRS22, weighted genetic risk score with 22 alleles; wGRS14 (no HLA), weighted genetic risk score with 14 alleles, without HLA alleles.



**Figure 2** ROC curves for predicting seropositive RA in NHS (A) and CCP-positive RA in EIRA (B). The NHS clinical model is adjusted for age and pack-years of smoking. The EIRA clinical model is adjusted for age, sex, geographical region and pack-years of smoking. NHS AUC: clinical model: AUC 0.566; clinical plus wGRS14 (no HLA): AUC 0.589; clinical plus wGRS22: AUC 0.660. NHS AUC comparisons: clinical plus wGRS22 versus clinical plus wGRS14 (no HLA),  $p < 0.001$ ; clinical plus wGRS22 versus clinical,  $p < 0.001$ ; clinical plus wGRS14 versus clinical,  $p = 0.31$ . EIRA AUC: clinical model: AUC 0.626; clinical plus wGRS14 (no HLA): AUC 0.662; clinical plus wGRS22: AUC 0.752. EIRA AUC comparisons: clinical plus wGRS22 versus clinical plus wGRS14 (no HLA),  $p < 0.0001$ ; clinical plus wGRS22 versus clinical,  $p < 0.0001$ ; clinical plus wGRS14 versus clinical,  $p = 0.002$ . AUC, area under the curve; CCP, cyclic-citrullinated peptide antibody; EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; HLA, human leucocyte antigen; NHS, Nurses' Health Studies; RA, rheumatoid arthritis; ROC, receiver operating characteristic; wGRS22, weighted genetic risk score with 22 alleles; wGRS14 (no HLA), weighted genetic risk score with 14 alleles, without HLA alleles.



**Figure 3** (A) Predicted versus observed incidence rates for wGRS22 in NHS women, EIRA women and EIRA men; (B) predicted versus observed incidence rates for wGRS14 (no HLA) in NHS women, EIRA women and EIRA men. EIRA, Epidemiologic Investigation of Rheumatoid Arthritis; HLA, human leucocyte antigen; NHS, Nurses' Health Studies; wGRS22, weighted genetic risk score with 22 alleles; wGRS14 (no HLA), weighted genetic risk score with 14 alleles, without HLA alleles.

Our wGRS is a first step towards the development of RA risk prediction models that incorporate aggregate genetic factors. In contrast to other complex diseases such as diabetes<sup>38,39</sup> and heart disease,<sup>34–37</sup> in which adding genetic markers to clinical risk factors does not add to discrimination, the addition of genetic factors to a clinical model that includes epidemiological risk factors improves discrimination significantly for RA, which supports the clinical validity of this approach. The AUC of 0.566 and 0.627 in NHS and EIRA, respectively, suggest that clinical risk factors alone, in subjects without symptoms, do not provide much discrimination between RA cases and controls. Adding genetic alleles to the aggregate score significantly improves the model AUC to 0.660 in NHS and 0.752 in EIRA. However, there is a variance in risk that remains unexplained, suggesting that further work is needed to incorporate environmental exposure data and gene–environment interactions into risk models and to discover additional genetic variants. We note that in patients with early symptoms consistent with inflammatory arthritis, clinical prediction models that include sex, age, localisation of symptoms, morning stiffness, tender joint count, swollen joint count, C-reactive protein level, rheumatoid factor positivity and the presence of anti-CCP

antibodies accurately predict who will go on to develop RA.<sup>47,48</sup> Under this clinical scenario, it will be important to test whether genetic factors help discriminate which patients will develop RA.

OR alone are difficult to interpret for patients and physicians in a clinical setting.<sup>24</sup> However, as suggested by Kraft *et al*,<sup>24</sup> measures of absolute risk (ie, risk that a disease-free individual will develop disease) such as the results shown in figure 3, provide a more intuitive context of RA risk at the individual level. A strength of our study is that we have data on the entire prospective NHS cohort from which our nested samples were taken, and thus we have an accurate estimate of the population annual incidence. Using data from the full NHS cohort, we see an absolute risk of RA among US women aged 25–50 years of 0.3%, thus a wGRS22 in group 7 increases the absolute risk to 0.7%. In EIRA women, the wGRS22 score in group 7 increases the absolute risk from 0.4% to 1.3%. In EIRA men, the wGRS22 score in group 7 increases the absolute risk from 0.2% to 0.7%. These predictive models demonstrate that there is a small portion of the general population at very high risk.

Although the hope is that we will soon be able to apply genetic information to individual patients, the wGRS for RA



## Extended report

is unlikely to be useful in routine clinical practice for assessing risk among healthy asymptomatic patients. Even the highest risk category, group 7, has a modest absolute risk of RA. It is possible that genetic results might eventually help us to identify subsets of patients who are at substantially elevated absolute risk, and would be willing to undergo potentially toxic therapies to prevent RA. It will be important to perform studies in subsets of patients at higher risk of RA; for example, patients with early undifferentiated arthritis, patients with anti-CCP-positive arthralgia and first-degree relatives of RA patients.<sup>49</sup> We propose that wGRS22 may be clinically useful as part of an overall risk assessment tool among high-risk groups.

We recognise that the ideal setting to perform prognostic modelling analyses is a prospective cohort study, such as the Framingham Heart Study or the full NHS cohorts. However, no such large study has blood samples available on the full dataset and validated RA cases. Instead, we approximated risk by use of the odds, which in a population-based case-control study with a proper sampling of controls approximates relative risk well. We calculated risk score-specific absolute risks using these OR and the average population risk estimated from the full NHS cohort and from the literature for northern Europe. The estimated incidence in NHS is consistent with RA incidence rates observed in other studies in women of northern European ancestry,<sup>43</sup> except for a single study from north America.<sup>50</sup> The NHS dataset is limited by the absence of CCP antibody information on cases that were diagnosed before the widespread use of the test. The phenotype used in NHS analyses is thus seropositive RA, whereas the phenotype used in EIRA analyses is CCP-positive RA, which is more strongly associated with genetic factors such as the HLA-SE. Although stronger associations are demonstrated in EIRA, the results from NHS are very consistent, suggesting that the general category of seropositive RA is associated with these genetic factors.

Despite the rapid advances in our understanding of the genetic basis of complex human diseases such as RA, it is not clear how to utilise this information for clinical care, prediction or prevention. Although a combination of known genetic factors for RA aggregated into a weighted score identifies a high-risk group with a threefold increased odds for the development of seropositive RA, the absolute risk of this disease remains low, suggesting that GRS, calculated as in this paper, have little clinical utility in predicting RA risk in asymptomatic individuals. More research to identify genetic and environmental risk factors, as well as gene-environment interactions, is critical to understanding the determinants of RA risk before this information can be used in patient counselling or preventive trials.

**Acknowledgements** The authors wish to thank the participants, investigators and study staff of the Nurses' Health Studies in the USA and Epidemiologic Investigation of Rheumatoid Arthritis in Sweden for their contributions.

**Funding** The NHS is supported by NIH grants R01 AR49880, CA87969, CA49449, CA67262, CA50385, P60 AR047782, K24 AR0524-01. RMP is supported by grants from NIAMS-NIH (R01-AR056768 and R01 AR057108) and the William Randolph Hearst Fund of Harvard University, and also holds a career award for medical scientists from the Burroughs Wellcome Fund. The EIRA study was supported by grants from the Swedish Medical Research Council, from the Swedish Council for Working life and Social Research, from King Gustaf V's 80-year foundation, from the Swedish Rheumatism Foundation, from Stockholm County Council and from the insurance company AFA.

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** This study was conducted with the approval of the Partners HealthCare Inc Institutional Review Board and Karolinska Institutet Institutional Review Board.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1. **Finckh A**, Liang MH, van Herckenrode CM, *et al*. Long-term impact of early treatment on radiographic progression in rheumatoid arthritis: a meta-analysis. *Arthritis Rheum* 2006;**55**:864-72.
2. **Criswell LA**, Merlino LA, Cerhan JR, *et al*. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. *Am J Med* 2002;**112**:465-71.
3. **Stolt P**, Bengtsson C, Nordmark B, *et al*; EIRA study group. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 2003;**62**:835-41.
4. **Costenbader KH**, Feskanich D, Mandl LA, *et al*. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med* 2006;**119**:503.e1-9.
5. **Gregersen PK**, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;**30**:1205-13.
6. **Rantapää-Dahlqvist S**, de Jong BA, Berglin E, *et al*. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum* 2003;**48**:2741-9.
7. **Nielen MM**, van Schaardenburg D, Reesink HW, *et al*. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum* 2004;**50**:380-6.
8. **Chibnik LB**, Mandl LA, Costenbader KH, *et al*. Comparison of threshold cutpoints and continuous measures of anti-cyclic citrullinated peptide antibodies in predicting future rheumatoid arthritis. *J Rheumatol* 2009;**36**:706-11.
9. **Karlson EW**, Chibnik LB, Tworoger SS, *et al*. Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. *Arthritis Rheum* 2009;**60**:641-52.
10. **Begovich AB**, Carlton VE, Honigberg LA, *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330-7.
11. **Pflege RM**, Padyukov L, Remmers EF, *et al*. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 2005;**77**:1044-60.
12. **Remmers EF**, Pflege RM, Lee AT, *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007;**357**:977-86.
13. **Pflege RM**, Seielstad M, Padyukov L, *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N Engl J Med* 2007;**357**:1199-209.
14. **Pflege RM**, Cotsapas C, Davies L, *et al*. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;**39**:1477-82.
15. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661-78.
16. **Thomson W**, Barton A, Ke X, *et al*; Wellcome Trust Case Control Consortium; YEAR Consortium. Rheumatoid arthritis association at 6q23. *Nat Genet* 2007;**39**:1431-3.
17. **Kurreeman FA**, Padyukov L, Marques RB, *et al*. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med* 2007;**4**:e278.
18. **Barton A**, Thomson W, Ke X, *et al*; Wellcome Trust Case Control Consortium; YEAR Consortium; BIRAC Consortium. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 2008;**40**:1156-9.
19. **Suzuki A**, Yamada R, Kochi Y, *et al*. Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. *Nat Genet* 2008;**40**:1224-9.
20. **Raychaudhuri S**, Remmers EF, Lee AT, *et al*. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008;**40**:1216-23.
21. **Burke W**, Psaty BM. Personalized medicine in the era of genomics. *JAMA* 2007;**298**:1682-4.
22. **Ioannidis JP**. Personalized genetic prediction: too limited, too expensive, or too soon? *Ann Intern Med* 2009;**150**:139-41.
23. **Jakobsdottir J**, Gorin MB, Conley YP, *et al*. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet* 2009;**5**:e1000337.
24. **Kraft P**, Wacholder S, Cornelis MC, *et al*. Beyond odds ratios-communicating disease risk based on genetic profiles. *Nat Rev Genet* 2009;**10**:264-9.
25. **Liang MH**, Meenan RF, Cathcart ES, *et al*. A screening strategy for population studies in systemic lupus erythematosus. Series design. *Arthritis Rheum* 1980;**23**:153-7.
26. **Arnett FC**, Edworthy SM, Bloch DA, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315-24.
27. **Padyukov L**, Silva C, Stolt P, *et al*. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum* 2004;**50**:3085-92.
28. **Costenbader KH**, Chang SC, De Vivo I, *et al*. Genetic polymorphisms in PTPN22, PADI-4, and CTLA-4 and risk for rheumatoid arthritis in two longitudinal cohort studies: evidence of gene-environment interactions with heavy cigarette smoking. *Arthritis Res Ther* 2008;**10**:R52.

29. **Fernando MM**, Stevens CR, Walsh EC, *et al*. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 2008;**4**:e1000024.
30. **Lohmueller KE**, Pearce CL, Pike M, *et al*. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177–82.
31. **Maller J**, George S, Purcell S, *et al*. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006;**38**:1055–9.
32. **Zheng SL**, Sun J, Wiklund F, *et al*. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008;**358**:910–19.
33. **Thomas G**, Jacobs KB, Yeager M, *et al*. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 2008;**40**:310–15.
34. **Kathiresan S**, Melander O, Anevski D, *et al*. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008;**358**:1240–9.
35. **Kathiresan S**, Willer CJ, Peloso GM, *et al*. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009;**41**:56–65.
36. **Aulchenko YS**, Ripatti S, Lindqvist I, *et al*; ENGAGE Consortium. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009;**41**:47–55.
37. **Kathiresan S**, Voight BF, Purcell S, *et al*; Myocardial Infarction Genetics Consortium; Wellcome Trust Case Control Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009;**41**:334–41.
38. **Lyssenko V**, Jonsson A, Almgren P, *et al*. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 2008;**359**:2220–32.
39. **Meigs JB**, Shrader P, Sullivan LM, *et al*. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;**359**:2208–19.
40. **Cornelis MC**, Qi L, Zhang C, *et al*. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Intern Med* 2009;**150**:541–50.
41. **DeLong ER**, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–45.
42. **Rebbbeck TR**, Ambrosone CB, Shields PG, eds. *Molecular epidemiology: applications in cancer and other human diseases*. New York: Informa Healthcare, 2008.
43. **Alamanos Y**, Voulgari PV, Drosos AA. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. *Semin Arthritis Rheum* 2006;**36**:182–8.
44. **Kraft P**, Hunter DJ. Genetic risk prediction—are we there yet? *N Engl J Med* 2009;**360**:1701–3.
45. **Weedon MN**, Lango H, Lindgren CM, *et al*. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;**40**:575–83.
46. **Lettre G**, Jackson AU, Gieger C, *et al*. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008;**40**:584–91.
47. **van der Helm-van Mil AH**, le Cessie S, van Dongen H, *et al*. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. *Arthritis Rheum* 2007;**56**:433–40.
48. **van der Helm-van Mil AH**, Detert J, le Cessie S, *et al*. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: moving toward individualized treatment decision-making. *Arthritis Rheum* 2008;**58**:2241–7.
49. **Hemminki K**, Li X, Sundquist J, *et al*. Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. *Arthritis Rheum* 2009;**60**:661–8.
50. **Gabriel SE**, Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955–1985. *Arthritis Rheum* 1999;**42**:415–20.