Polygenic Risk Scores have high diagnostic capacity in ankylosing spondylitis


ABSTRACT

Objective We sought to test the hypothesis that Polygenic Risk Scores (PRSs) have strong capacity to discriminate cases of ankylosing spondylitis (AS) from healthy controls and individuals in the community with chronic back pain.

Methods PRSs were developed and validated in individuals of European and East Asian ethnicity, using data from genome-wide association studies in 15,585 AS cases and 20,452 controls. The discriminatory values of PRSs in these populations were compared with other widely used diagnostic tests, including C-reactive protein (CRP), HLA-B27 and sacroiliac MRI.

Results In people of European descent, PRS had high discriminatory capacity with area under the curve (AUC) in receiver operator characteristic analysis of 0.924. This was significantly better than for HLA-B27 testing alone (AUC=0.869, MRI (AUC=0.885) or C-reactive protein (AUC=0.700). PRS developed and validated in individuals of East Asian descent performed similarly (AUC=0.948). Assuming a prior probability of AS of 10% such as in patients with chronic back pain under 45 years of age, compared with HLA-B27 testing alone, PRS provides higher positive values for 35% of patients and negative predictive values for 67.5% of patients. For PRS, in people of European descent, the maximum positive predictive value was 78.2% and negative predictive value was 100%, whereas for HLA-B27, these values were 51.9% and 97.9%, respectively.

Conclusions PRS have higher discriminatory capacity for AS than CRP, sacroiliac MRI or HLA-B27 status alone. For optimal performance, PRS should be developed for use in the specific ethnic groups to which they are to be applied.

INTRODUCTION

Ankylosing spondylitis (AS) affects approximately 0.2%–0.6% of individuals of European descent and Chinese. Early treatment with biologic therapies in those with more severe forms of the disease achieves more effective clinical responses and probably reduces the rate joint fusion in the long term. However, other causes of chronic back pain are common in the community, and AS is responsible for only a minority of these cases. It can be difficult to distinguish AS from other causes of back pain, particularly early in the disease with the consequence that the diagnosis of AS is often significantly delayed; many surveys undertaken in a variety of different health systems suggest an average delay of 6–10 years. A recent North American survey reported that fewer than half of patients with AS reported that they were correctly diagnosed within 1 year of seeking medical attention, and 32.8% waited more than a decade to receive the diagnosis. Population surveys suggest that as many as 80% of cases in the community remain
undiagnosed\(^{19}\) and therefore may not receive appropriate effective treatment. There is thus a great need for improved testing to improve early accurate diagnosis.

Currently, the most widely used tests for AS in those with chronic back pain are measurements of acute phase reactants, such as erythrocyte sedimentation rate and C-reactive protein (CRP), genetic testing for HLA-B27 and imaging—either plain radiographs or MRI of the sacroiliac joints.\(^{9,13}\) However, each of these tests has limitations. In brief, acute phase reactants and MRI are only positive after disease develops and are therefore not useful for predicting disease risk. Acute phase reactants have only moderate sensitivity and specificity, particularly in early disease. MRI is expensive and is not universally available. Genetic factors are the major determinants of the risk of developing AS, with heritability assessed in twins of $>90\%$.\(^{10,11}\) Although HLA-B27 alone contributes $20\%$ of the variation in disease risk,\(^{12}\) the remainder of the genetic risk is determined by thousands of common genetic variants, each of which has only a very small effect. Polygenic Risk Scores (PRS) use combinations of hundreds to thousands of genetic variants to quantify an individual’s genetic risk of disease. Unlike HLA-B27 testing which is categorical or dichotomous in outcome, PRS are continuous measures. They are of particularly strong predictive value for low-frequency diseases with high heritability,\(^{13}\) such as AS. Here, we describe the development and validation of PRS for AS in two different ethnic groups and compare its performance to standard screening or diagnostic tests.

METHODS

Study population

AS was defined according to the modified New York criteria.\(^{14}\) Following genotyping quality control, there were 8244 cases and 14 274 controls of western European descent; 6001 cases and 4493 controls of East Asian (Chinese) descent; and 1340 cases and 1685 controls of Turkish and Iranian origin, respectively. Written informed consent was obtained from all cases, with approval from the relevant research ethics authorities at each participating centre. Cohort details are provided in online supplemental table S1.

Genetic data

Samples were genotyped using the Illumina Core-Exome SNP genotyping microarray, according to the manufacturer’s recommendations (chip versions used per cohort are provided in online supplemental table S1). Bead intensity data were processed and normalised for each sample, and genotypes called, using Genome Studio V2.0 software (GenomeStudio Software Downloads (illumina.com)). Standard quality control measures as outlined in the Supplementary Methods were applied including identification and exclusion of cryptic-related samples, exclusion of samples with an outlying heterozygosity rate (3 SD from the mean in each cohort) or excess missingness ($>5\%$). Single nucleotide polymorphisms (SNPs) with genotyping missing rate $>2\%$, p value of Hardy-Weinberg equilibrium test $<1\times10^{-6}$, or with allele frequency $<1\%$ were removed. Population stratification was accessed using Shellfish (http://www.stats.ox.ac.uk/~davison/software/shellfish/shellfish.php). PRS analyses were performed with and without inclusion of principal components and gender as covariates. Results including principal components and gender as covariates are reported in online supplemental table S2 and are very similar to the results not including these covariates.

HLA-B27 imputation was performed using SNP2HLA, using a deep sequencing Chinese reference panel (n=10 689)\(^{15}\) for East Asian samples and Type 1 Diabetes Genetics Consortium (n=5 225) panel of combined HLA types and MHC SNP genotypes for all other subjects.\(^{16}\)

PRS were calculated for each individual using the adaptive MultiBLUP algorithm (implemented in the software LDAK V5.0).\(^{17,18}\) LDAK first divides the genetic data into chunks of size 75 000 bp and then performs association test for all the chunks and thinned out SNPs in strong linkage disequilibrium. The significant chunks with p value $<1\times10^{-5}$ and all adjacent chunks with p value $<0.01$ are merged into regions. Then the variance components and effect size of SNPs are estimated, and the effect size of the SNPs used to calculate the PRS. A 10-fold cross-validation analysis was performed as internal validation; a separate external validation was performed in the British and North American subjects, as well as through comparison of performance of PRS trained in either European descent or East Asian subjects, then validated in a separate ethnic group. In regard to cross-validation studies, the case–control cohort being studied is divided into 10 equal folds randomly with same case–control ratio. Nine folds of samples were used as a training set and the remaining fold of samples was retained as the validation data for testing the model generated by the training set. The process was repeated 10 times, with each of the 10-folds used only once as the validation data. The out-of-fold predictions based on the effect sizes of the selected SNPs were obtained for the test fold. All the predictions of 10 test folds were merged, after which statistical analysis was performed using all out-of-fold test set predictions to maximise sample size for internal testing. The resulting weighted predictors were then applied to the test cohort to obtain per sample scores from which the area under the curve (AUC) was obtained using receiver operator characteristic (ROC) analysis. R package pROC was used to calculate the 95% CI of the AUC and also compare AUCs from two models.\(^{18}\) Positive (PPV) and negative predictive values (NPV) were then calculated for PRS centiles, assuming different prior probabilities of AS. The continuous net reclassification improvement (NRI),\(^{19}\) a statistic that aims to quantify differences in classification performance of different models, was calculated using the R package PredictABEL\(^{20}\) and used to compare accuracy of diagnostic assignment by HLA-B27 testing and PRS.

RESULTS

ROC analyses of test discriminatory capacity are summarised in table 1. In 10-fold cross-validation in this case–control cohort, the PRS had AUC of 0.924 (95% CI 0.920 to 0.928) (figure 1). The AUC of HLA-B27 testing alone was 0.869 (95% CI 0.865 to 0.874), which was statistically significantly less discriminatory than the PRS (p=$2.2\times10^{-16}$). Additionally, the NRI was positive (0.717, 95%CI 0.692 to 0.743), confirming that the PRS is an improvement on HLA-B27 alone. A PRS including only non-MHC SNPs performed less well (AUC 0.782), as did a PRS including only 103 (genotyped or imputed) loci previously reported to have achieved genome-wide significance in AS (AUC=0.639).\(^{21}\) MRI has a reported sensitivity of 85% and specificity of 92% in AS,\(^{22}\) which correlates with an AUC of 0.885. CRP has a reported sensitivity of 50% and specificity of 80% for the disease (AUC=0.7).\(^{23}\)

To test the performance of the PRS using external validation, the European descent cases were divided into British and North American cohorts, and controls divided in the same proportion as the two case cohorts. PRS was then
Spondyloarthritis

Table 1  ROC analysis findings (AUC) of genetic risk scores in different populations

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Population tested in</th>
<th>European</th>
<th>East Asian</th>
<th>Iranian</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B27 alone</td>
<td>0.869 (0.865–0.874)</td>
<td>0.901 (0.895–0.906)</td>
<td>0.831 (0.807–0.854)</td>
<td>0.821 (0.804–0.838)</td>
<td></td>
</tr>
<tr>
<td>European non-MHC PRS</td>
<td>0.782 (0.776–0.788)*</td>
<td>0.594 (0.539–0.560)</td>
<td>0.534 (0.500–0.569)</td>
<td>0.568 (0.542–0.595)</td>
<td></td>
</tr>
<tr>
<td>European overall PRS</td>
<td>0.924 (0.920–0.928)*</td>
<td>0.788 (0.779–0.796)</td>
<td>0.852 (0.826–0.879)</td>
<td>0.854 (0.836–0.872)</td>
<td></td>
</tr>
<tr>
<td>East Asian non-MHC PRS</td>
<td>0.555 (0.547–0.563)</td>
<td>0.731 (0.722–0.741)*</td>
<td>0.565 (0.531–0.598)</td>
<td>0.554 (0.528–0.581)</td>
<td></td>
</tr>
<tr>
<td>East Asian overall PRS</td>
<td>0.880 (0.875–0.887)</td>
<td>0.948 (0.943–0.952)*</td>
<td>0.872 (0.848–0.895)</td>
<td>0.840 (0.821–0.860)</td>
<td></td>
</tr>
<tr>
<td>MRI EUR</td>
<td>0.885</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI CH41</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*10-fold cross-validation. All other PRS AUC values are external validation statistics.

AUC, area under the curve; CRP, C-reactive protein; PRS, Polygenic Risk Score; ROC, receiver operator characteristic.

Figure 1  Receiver operating characteristic curve plot of performance of Polygenic Risk Scores (PRS) (purple dashes, area under the curve (AUC)=0.924), HLA-B27 (aqua dashes, AUC=0.869), PRS less major histocompatibility complex (MHC) (green line, AUC=0.782) and genome-wide significant loci only (red line, AUC=0.659).
an HLA-B27 test will be positive in 31% of those tested with a PPV of 80.6%, and in the 69% of those with a negative test, the NPV is 92.4%. Using the PRS, the PPV is >80.6% for top 35% of those screened, and achieves a higher maximum value (93.3%) than does HLA-B27 (80.6%) (figure 2). The PRS NPV will be >92.4% for 65% of those screened, and also achieves a higher maximum value (99.6%) than does HLA-B27 (92.4%). Considering the situation where only 10% of screened patients have AS, then HLA-B27 will be positive in 16% of those tested. In this group, HLA-B27 positivity has a PPV of 51.9%, and a negative result (seen in 84% of screened patients) has an NPV of 97.9%. Using the PRS for the top 35% of patients and achieves a slightly higher maximum value than HLA-B27 testing (100% vs 97.9%).

Considering general population screening, at least 8% of the European population carry HLA-B27, yet only 5% of carriers of this allele will develop AS; as such, no higher PPV can be achieved using HLA-B27 testing alone. In contrast, for the PRS, the PPV for the top 8% of the population is three times higher (15.1%), and it is higher than 5% for the top 35% of the population. The NPV for HLA-B27-negative status is 99.9%, which is exceeded by the PRS for 62.5% of the population.

**DISCUSSION**

Distinguishing AS from other causes of chronic back pain remains an important issue in rheumatology. HLA-B27 testing can have a valuable PPV for AS, particularly in clinical settings where the pretest probability of the disease is relatively high compared with the general population. It is therefore included in the Assessment of Spondyloarthritis International Study Group (ASAS) axial spondyloarthritis (axSpA) classification criteria and is an essential criterion for those with no available imaging evidence of disease. HLA-B27 testing has also been recommended for screening patients with chronic back pain to identify those at higher risk of AS or the related group of diseases axSpA, for referral to specialist services. However, HLA-B27 only contributes ~20% of the overall heritability of AS, which is estimated to be ≥90% overall, indicating a substantial non-MHC component. This suggests that PRS, which capture the common-variant component of heritability, are likely to be much more informative than HLA-B27 tests alone. Our study confirms this, with the PRS performing better than HLA-B27 testing in both AUC and continuous NRI analyses, irrespective of the prevalence of AS among those being tested. We confirm these findings both by internal cross-validation and by external validation. For 35% of the population, the PPV is higher for the PRS than for HLA-B27 testing, and the NPV is higher for >65%. In particular, the peak PPV is substantially higher for the PRS than for HLA-B27 and is informative for a far higher proportion of patients, as it is a continuous variable whereas HLA-B27 is dichotomous. PRS testing also has higher discriminatory capacity for AS than MRI, and far higher than CRP. Accurate interpretation of MRI scans is known to be dependent on training and experience, and particularly in inexperienced, untrained hands may perform worse than the average reported performance, in which setting PRS may be particularly valuable.

Chronic back pain of >3 months’ duration has previously been shown to have very low heritability attributable to common genetic variants (minor allele frequency >0.01) such as those included in our AS PRS (common variant heritability=6.43%30–7.6%31) and not to be genetically correlated with AS. Therefore, it is unlikely that the AS
PRS will prove less discriminatory in practice in the clinical setting of patients presenting with chronic back pain than the estimates presented here. A limitation of this study is that the performance of the PRS has not been formally tested in this setting, where it will require further evaluation.

axSpA refers to a spectrum of diseases. Patients with radiographic sacroiliitis are classified as having AS, whereas those without X-ray changes are classified as having non-radiographic (nr)-axSpA. The current PRS may have prognostic value in distinguishing the 16%–24% of nr-axSpA cases that are likely to go on to develop AS. 32 33 Whether the PRS we report here will prove more informative than HLA-B27 testing alone in patients with nr-axSpA itself is unknown. The ASAS have previously demonstrated that patients meeting the ASAS classification criteria for axSpA who do not yet have AS have a much lower average genetic risk score than patients with AS, using only genome-wide significant AS loci. 34 Whether this is because nr-axSpA is actually genetically distinct from AS, or reflects the generally higher and clinically more osteopathic heterogeneity of nr-axSpA, 35 will require further study.

As with the use of PRS in the screening of individuals with chronic back pain, its performance in nr-axSpA will also require further study. Similarly, the performance of the PRS in males compared with females, in subjects with environmental risk factors for the disease such as cigarette smoking, 36 and in subsets of patients such as those with extraskeletal manifestations of AS requires further study. In that regard, the excellent performance of a PRS in patients with acute anterior uveitis complicating AS (AUC=0.96; 95% CI 0.95 to 0.96) suggests that at least in some AS subsets the performance of the PRS will be even better than reported here. 37

PRS testing can be performed using data from any dense SNP microarray. Indeed, the performance of the PRS reported here was high despite our use of a relatively low density SNP microarray—the Illumina Core-Exome chip (>520,000 variants, including many rare and non-polymorphic variants that do not contribute to the PRS). The performance of PRS testing would be likely to improve further with use of microarrays with better SNP coverage, or with whole genome sequencing. It has been estimated that up to 12 million Americans have had SNP microarray testing performed by commercial services such as 23andMe and Ancestry. 38 At little additional cost, these data would probably prove suitable for the calculation of the AS PRS we report, as well as enabling PRS for many other diseases in which they have been shown to be informative. The cost-effectiveness of the PRS we report here needs to be confirmed in further studies. As the genetic profile of AS becomes better understood, the discriminatory capacity of these tests is also likely to increase. For example, it is likely that many of the SNPs included in the PRS at present are not truly associated with AS, but just add noise to the test.

As there is no preventive therapy yet for AS, general population screening to identify patients at high risk of the disease is not recommended except, perhaps, for those at increased risk, such as the relatives of those with AS (given the high sibling recurrence risk of 8.2%). 39 PRS performs significantly better than HLA-B27 testing alone in the general population, with the PPV of the ~8% of the general population who carry HLA-B27 being 5%, compared with the peak PPV of the PRS of 15.1%. Similarly, the NPV for the PRS exceeds that of HLA-B27 testing for most of the population. Although the PPV for PRS testing for general population screening is modest, the test performs well compared with other widely used screening tests. For example, the PPVs for 10-year risk of coronary heart disease of a high total cholesterol (≥240 mg/dL) —a threshold above which many patients will be prescribed cholesterol-lowering therapy—are 10.3% in women and 18.6% in men, 40 similar to the top 20% of PPVs of PRS for AS in general population screening. Among those who have already had SNP microarray testing performed, knowledge of a high AS-PRS even in the absence of symptoms may heighten clinician awareness of the possible diagnosis, reduce delay and assist with earlier appropriate and effective treatment, given the current long diagnostic delays.

Our study shows that the performance of the PRS varies between ethnic groups, although it remains moderately high even when a PRS developed in subjects of (western) European descent is tested in eastern European/west Asian subjects such as Turks and Iranians. The PRS developed specifically for East Asians performed far better in that population than did the European PRS, indicating that at least for populations that are remotely related, ethnic-specific PRSs are preferable.

We conclude that PRS testing for AS has greater discriminatory capacity than HLA-B27 testing, MRI scanning or CRP testing, either alone or in combination. PRS could be used to screen patients with chronic back pain to identify subjects at increased risk of the disease for referral to secondary care and to assist in diagnosing the condition.
We would like to thank all participating subjects with ankylosing spondylitis and healthy individuals who provided the DNA and clinical information necessary for this study. The TASC study was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) grants R01-052915 and R01-AR046208. Funding was also received from the University of Texas Health Science Center at Houston CTSA UL1RR02411, Cedars-Sinai CRD and Grant M01-RR-00042, Intramural Research Program, NIAMS/NH and Rebecca Cooper Foundation (Australia).

**Collaborators**

TCRI AS Group Jian Yin1, Lei Jiang1, Lin Zhou1, Ting Li1, Qingwen Wang1, Tianwen Gao2, Guomin Gao3, Shengqian Xu4, Weiguo Xiao5, Hui Shen5, and Kunxuan Huang6. The second set of authors are those of the author(s) and not necessarily those of the NHS, the NIHR or the University of Health. French sample collection was performed by the Groupe Francais d’Etude Genetique des Spondyloarthrites, coordinated by Professor Maxime Bierbaum and funded by the Agence Nationale de Recherche GEMISA grant reference ANR-10- MEDI-001. We acknowledge and thank the TCRI AS Group for their support in recruiting patients for the study (see below). The authors acknowledge the sharing of data and samples by the BSRBR-AS Register in Auckland. Chief Investigator, Professor Gary MacFarlane and Dr Gareth Jones, Deputy Chief Investigator created the BSRBR-AS study which was commissioned by the British Society for Rheumatology, funded in part by AbbVie, Pfizer and UCB. We are grateful to every patient, past and present staff of the BSRBR-AS Register and all clinical staff who recruited patients, followed them up and entered data—details here: www.abbv.de/uk/ais/research/epidemiology/spondyloarthritis.php?panel1011. The QIMR control samples were from parents of adolescent twins collected in the context of the Brisbane Longitudinal Twin Study 1992–2016, supported by grants from NHMRC (NGM) and ARC (MIW). We thank Anjali Henders, Lisa Bowdler, Tabatha Gonzales for blood collection and serological analyses. Kerrie McClayney and Scott Gordon for curating samples for this study. MAB is funded by a National Health and Medical Research Council (Australia) Senior Principal Research Fellowship (1024879), and support for this study was received from a National Health and Medical Research Council (Australia) program grant (566938) and project grant (569829), and from the Australian Cancer Research Foundation and Rebecca Cooper Medical Research Foundation. We are also very grateful for the invaluable support, received from the National Ankylosing Spondylitis Society (UK) and Spondyloarthritis Association of America in case recruitment. Additional financial and technical support for patient recruitment was provided by the National Institute for Health Research Oxford Musculoskeletal Biomedical Research Unit and NIHR Thames Valley Comprehensive Local Research Ethics Committee (approval reference HREC/05/QPAH/221). The overall programme was reviewed and approved by Metro South Hospital Ethics appr...
Supplemental material

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Zhuix Li http://orcid.org/0000-0002-2924-9120
Nurulah Akkoc http://orcid.org/0000-0002-3718-171X
Maxime Breban http://orcid.org/0000-0002-6992-9395
Gary J Macfarlane http://orcid.org/0000-0003-2322-3314
Mahdi Mahmoudi http://orcid.org/0000-0002-8164-8831
Helena Marzo-Ortega http://orcid.org/0000-0002-9683-3407
Andrew A Harrison http://orcid.org/0000-0003-4372-3252
Simon Stebbings http://orcid.org/0000-0002-2842-4440
Gareth T James-Cheng-Chung Wei http://orcid.org/0000-0003-0310-7269
Minlai Lee http://orcid.org/0000-0002-4329-506X
Xiaobing Wang http://orcid.org/0000-0002-4302-2213
Zi-Bing Jin http://orcid.org/0000-0003-0515-698X
Michael M Ward http://orcid.org/0000-0003-1857-9367
Tae-Hwan Kim http://orcid.org/0000-0003-3542-2276
John D Revellie http://orcid.org/0000-0001-5950-0913
Bryan Paul Worthword http://orcid.org/0000-0001-7512-3468
Huji Xu http://orcid.org/0000-0002-8588-118X
Matthew A Brown http://orcid.org/0000-0003-0538-8211

REFERENCES

Table S1  Case-control cohorts studied. AS=ankylosing spondylitis. Gender is as imputed from SNP array data.

<table>
<thead>
<tr>
<th>Type</th>
<th>Cohort</th>
<th>Males</th>
<th>Females</th>
<th>Number of samples</th>
<th>Version of CoreExome chip</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>United Kingdom AS Cohort ¹</td>
<td>4568</td>
<td>1928</td>
<td>6499*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Australo-Anglo-American Spondyloarthritis Consortium (TASC) ²</td>
<td>751</td>
<td>377</td>
<td>1128*</td>
<td>24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Australian AS Cohort ³</td>
<td>333</td>
<td>92</td>
<td>425*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Groupe Française d’Etude Génétique des Spondylarthrites (GFEGS) ³</td>
<td>123</td>
<td>69</td>
<td>192*</td>
<td></td>
</tr>
<tr>
<td>AS</td>
<td>Chinese AS cohort ³</td>
<td>4786</td>
<td>1215</td>
<td>6001*</td>
<td>24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Turkish AS cases ⁴</td>
<td>668</td>
<td>239</td>
<td>910</td>
<td>24v1-0</td>
</tr>
<tr>
<td>AS</td>
<td>Iranian AS cases ⁴</td>
<td>336</td>
<td>92</td>
<td>430</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>The UK Household Longitudinal Study ¹</td>
<td>4178</td>
<td>5283</td>
<td>9469*</td>
<td>12v1</td>
</tr>
<tr>
<td>Controls</td>
<td>Queensland Institute of Medical Research Berghofer Medical Research Institute Twins Cohort ²</td>
<td>1006</td>
<td>1655</td>
<td>2664*</td>
<td>12v1</td>
</tr>
<tr>
<td>Controls</td>
<td>Advancing Exercise &amp; Sports Science Collaborative Research</td>
<td>765</td>
<td>734</td>
<td>1504*</td>
<td>24v1-1</td>
</tr>
<tr>
<td>Network $^3$</td>
<td>Controls</td>
<td>Oregon Metagenomics Controls $^5$</td>
<td>38</td>
<td>44</td>
<td>82*</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------------------------------</td>
<td>----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Controls</td>
<td>Australian general population healthy controls $^3$</td>
<td>291</td>
<td>264</td>
<td>555*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>Controls</td>
<td>Chinese general population controls $^3$</td>
<td>3159</td>
<td>1784</td>
<td>4943*</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>Turkish general population controls $^4$</td>
<td>661</td>
<td>258</td>
<td>924</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>Iranian general population controls $^4$</td>
<td>590</td>
<td>164</td>
<td>761</td>
<td>24v1-0, 24v1-1</td>
</tr>
</tbody>
</table>

* Included in European-descent cohort.

* Included in East Asian cohort

Please note that as for some sample’s imputation did not assign gender definitively, the total sample number may differ from the sum of the numbers of males and females.

1. The UK AS Cohort was recruited from patients attending hospital-based rheumatology services across the United Kingdom. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

2. Australo-Anglo-American Spondyloarthritis Consortium cohort was recruited from patients attending hospital-based rheumatology services in Australia and the United States. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were centrally reported by expert readers.
3. The Australian AS cohort was recruited from patients attending rheumatology services (both hospital based and from community rheumatology practices) in Australia. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were centrally reported by expert readers.

4. The Groupe Française d’Étude Génétique des Spondylarthrites (GFEGS) was recruited from hospital-based rheumatology services in France. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

5. The Chinese AS Cohort was recruited from self-reporting Han Chinese ethnicity patients attending hospital-based rheumatology services in China and Taiwan. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

6. The Turkish and Iranian AS Cohorts were recruited from patients attending hospital-based rheumatology services in Turkey and Iran respectively. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

**Supplementary Method**

**Genotyping Chips and Quality Control**

The cohorts were all typed on Illumina Core-Exome chip. The chip versions used are 24v1-1, 24v1-0, 12v1 and 12v1. The number of overlapping SNPs among these chips is greater than 95%. All the data were called separately based on the Core-Exome chip versions from GenomeStudio 2.0. Ethnicity-specific data was strand fixed based on the chip version. The ethnicity-specific datasets were then merged into a European-descent cohort and an East Asian cohort using the common SNPs between datasets. Then normal GWAS quality control was applied to all the cohorts in this study. Related samples (PI_HAT >0.185) were excluded. Samples were further excluded if their missingness rate
was >5%, or heterozygosity > 3 standard deviations from mean, or were outliers in the principal component analysis (by SHELLFISH). Variants were excluded if their missingness rate was >2%, or Hardy-Weinberg equilibrium test P-value < $1 \times 10^{-6}$, or there were significant different missingness rate in cases vs controls ($P < 1 \times 10^{-7}$).

Confidence interval of AUC

The confidence interval of AUC was calculated by R package pROC.
Table S2 Performance of HLA-B27 and AS PRS with controlling PCA and sex.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>European</th>
<th>East Asian</th>
<th>Iranian</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B27</td>
<td>0.869</td>
<td>0.901</td>
<td>0.831</td>
<td>0.821</td>
</tr>
<tr>
<td></td>
<td>(0.865-0.874)</td>
<td>(0.895-0.906)</td>
<td>(0.807-0.854)</td>
<td>(0.804-0.838)</td>
</tr>
<tr>
<td>European non-MHC PRS (PCA+SEX)</td>
<td>0.776</td>
<td>0.54</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>(0.770-0.782)</td>
<td>(0.534-0.556)</td>
<td>(0.500-0.569)</td>
<td>(0.553-0.605)</td>
</tr>
<tr>
<td>European overall PRS (PCA+SEX)</td>
<td>0.923</td>
<td>0.761</td>
<td>0.853</td>
<td>0.849</td>
</tr>
<tr>
<td></td>
<td>(0.919-0.927)</td>
<td>(0.752-0.770)</td>
<td>(0.827-0.880)</td>
<td>(0.830-0.867)</td>
</tr>
<tr>
<td>East Asian non-MHC PRS (PCA+SEX)</td>
<td>0.59</td>
<td>0.639</td>
<td>0.59</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>(0.578-0.593)</td>
<td>(0.629-0.650)</td>
<td>(0.553-0.620)</td>
<td>(0.581-0.633)</td>
</tr>
<tr>
<td>East Asian overall PRS (PCA+SEX)</td>
<td>0.88</td>
<td>0.942</td>
<td>0.87</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>(0.879-0.889)</td>
<td>(0.937-0.947)</td>
<td>(0.850-0.896)</td>
<td>(0.820-0.859)</td>
</tr>
<tr>
<td>PCA alone – European</td>
<td>0.630</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCA alone – East Asian</td>
<td>-</td>
<td>0.750</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCA alone – Iranian</td>
<td>-</td>
<td>-</td>
<td>0.536</td>
<td></td>
</tr>
<tr>
<td>PCA alone – Turkish</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.520</td>
</tr>
</tbody>
</table>
Supplementary References


Figure S1: Receiver operating characteristic curve plot of performance of HLA-B27 in the North American cohort (red line, AUC=0.895), PRS trained in the British cohort and externally validated in the North American cohort (aqua line, AUC=0.928), and European PRS cross-validated (purple line, AUC=0.924).
Figure S2: Receiver operating characteristic curve plot of performance of PRS in European-descent cohort (purple line, AUC=0.924 based on 10 folds cross-validation) and East-Asian cohort (red dotted, AUC=0.948 based on 10 folds cross-validation). The validation performance of the European PRS in different cohort are shown in solid line (Iranian (green, AUC=0.852), Turkish (blue line, AUC=0.854), East Asian (red, AUC=0.788)) whilst the validation performance of the East-Asian PRS in other population are in dotted lines (Iranian (green, AUC=0.872), Turkish (blue, AUC=0.840), European (purple, AUC=0.880)).
Figure S3: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 0.55% (the population prevalence of AS)\(^1\).
Figure S4: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 10%.
Figure S5: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 20%.
Figure S6: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 0.55% (the population prevalence of AS).
Figure S7: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 10%.
Figure S8: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 20%.
Figure S9: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 30%.
Figure S10. Scheme of analysis.  A. Analysis flow-chart. B. Scheme of 10 folds cross-validation in EUR or EAS cohort. C. Scheme of EUR PRS external validation. D. Scheme of cross-ethnicity testing.

A

Raw GWAS cohort

HLA imputation

Post-QC GWAS cohort

Sex imputation

PRS

Cross-Validation

External Validation

Cross-Ethnicity test

B

EUR/EAS cohort – 10 folds cross-validation

Training set

90% of cohort samples

Testing set

rest 10% samples

Repeat 10 times to have result for 10 test folds

C

EUR cohort – external validation

Training set

British cases + pro rate EUR controls

Testing set

North American cases + pro rate EUR controls

D

Cross-Ethnicity testing

Training set

Full EUR cohort / Full EAS cohort

Testing set

Cohort in other populations
Figure S11. Principal coordinates analysis plot of PC1 vs PC2 in European-descent cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Figure S12. Principal coordinates analysis plot of PC1 vs PC2 in East Asian-descent cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.

hapmap3 + East Asian GWAS

PC2

PC1

-0.8 -0.6 -0.4 -0.2 0.0

0.0 0.2 0.4

AFR AMR case control EAS EUR SAS
Figure S13: Principal coordinates analysis plot of PC1 vs PC2 in Iranian cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Figure S14: Principal coordinates analysis plot of PC1 vs PC2 in Turkish cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Table S1  Case-control cohorts studied. AS=ankylosing spondylitis. Gender is as imputed from SNP array data.

<table>
<thead>
<tr>
<th>Type</th>
<th>Cohort</th>
<th>Males</th>
<th>Females</th>
<th>Number of samples</th>
<th>Version of CoreExome chip</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>United Kingdom AS Cohort (^1)</td>
<td>4568</td>
<td>1928</td>
<td>6499*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Australo-Anglo-American Spondyloarthritis Consortium (TASC) (^2)</td>
<td>751</td>
<td>377</td>
<td>1128*</td>
<td>24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Australian AS Cohort (^3)</td>
<td>333</td>
<td>92</td>
<td>425*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Groupe Française d’Etude Génétique des Spondylarthrites (GFEGS) (^3)</td>
<td>123</td>
<td>69</td>
<td>192*</td>
<td></td>
</tr>
<tr>
<td>AS</td>
<td>Chinese AS cohort (^4)</td>
<td>4786</td>
<td>1215</td>
<td>6001(^{\circ})</td>
<td>24v1-1</td>
</tr>
<tr>
<td>AS</td>
<td>Turkish AS cases (^5)</td>
<td>668</td>
<td>239</td>
<td>910</td>
<td>24v1-0</td>
</tr>
<tr>
<td>AS</td>
<td>Iranian AS cases (^5)</td>
<td>336</td>
<td>92</td>
<td>430</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>The UK Household Longitudinal Study (^1)</td>
<td>4178</td>
<td>5283</td>
<td>9469*</td>
<td>12v1</td>
</tr>
<tr>
<td>Controls</td>
<td>Queensland Institute of Medical Research Berghofer Medical Research Institute Twins Cohort (^2)</td>
<td>1006</td>
<td>1655</td>
<td>2664*</td>
<td>12v1</td>
</tr>
<tr>
<td>Controls</td>
<td>Advancing Exercise &amp; Sports Science Collaborative Research</td>
<td>765</td>
<td>734</td>
<td>1504*</td>
<td>24v1-1</td>
</tr>
<tr>
<td>Network 3</td>
<td>Controls</td>
<td>Oregon Metagenomics Controls</td>
<td>38</td>
<td>44</td>
<td>82*</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-------------------------------</td>
<td>----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Controls</td>
<td>Australian general population healthy controls 3</td>
<td>291</td>
<td>264</td>
<td>555*</td>
<td>24v1-0, 24v1-1</td>
</tr>
<tr>
<td>Controls</td>
<td>Chinese general population controls 3</td>
<td>3159</td>
<td>1784</td>
<td>4943◦</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>Turkish general population controls 4</td>
<td>661</td>
<td>258</td>
<td>924</td>
<td>24v1-0</td>
</tr>
<tr>
<td>Controls</td>
<td>Iranian general population controls 4</td>
<td>590</td>
<td>164</td>
<td>761</td>
<td>24v1-0, 24v1-1</td>
</tr>
</tbody>
</table>

* Included in European-descent cohort.
◦ Included in East Asian cohort

Please note that as for some sample’s imputation did not assign gender definitively, the total sample number may differ from the sum of the numbers of males and females.

1. The UK AS Cohort was recruited from patients attending hospital-based rheumatology services across the United Kingdom. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

2. Australo-Anglo-American Spondyloarthritis Consortium cohort was recruited from patients attending hospital-based rheumatology services in Australia and the United States. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were centrally reported by expert readers.
3. The Australian AS cohort was recruited from patients attending rheumatology services (both hospital based and from community rheumatology practices) in Australia. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were centrally reported by expert readers.

4. The Groupe Française d’Etude Génétique des Spondylarthrites (GFEGS) was recruited from hospital-based rheumatology services in France. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

5. The Chinese AS Cohort was recruited from self-reporting Han Chinese ethnicity patients attending hospital-based rheumatology services in China and Taiwan. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

6. The Turkish and Iranian AS Cohorts were recruited from patients attending hospital-based rheumatology services in Turkey and Iran respectively. Rheumatology specialists caring for the patients confirmed the diagnosis of AS according to the modified New York Criteria for AS. Pelvic radiographs were reported by local expert readers.

**Supplementary Method**

*Genotyping Chips and Quality Control*

The cohorts were all typed on Illumina Core-Exome chip. The chip versions used are 24v1-1, 24v1-0, 12v1 and 12v1. The number of overlapping SNPs among these chips is greater than 95%. All the data were called separately based on the Core-Exome chip versions from GenomeStudio 2.0. Ethnicity-specific data was strand fixed based on the chip version. The ethnicity-specific datasets were then merged into a European-descent cohort and an East Asian cohort using the common SNPs between datasets. Then normal GWAS quality control was applied to all the cohorts in this study. Related samples (PI_HAT >0.185) were excluded. Samples were further excluded if their missingness rate...
was >5%, or heterozygosity > 3 standard deviations from mean, or were outliers in the principal component analysis (by SHELLFISH). Variants were excluded if their missingness rate was >2%, or Hardy-Weinberg equilibrium test P-value < $1 \times 10^{-6}$, or they have significant different missingness rate in cases vs controls ($P < 1 \times 10^{-7}$).

Confidence interval of AUC

The confidence interval of AUC was calculated by R package pROC.
<table>
<thead>
<tr>
<th>Predictors</th>
<th>European</th>
<th>East Asian</th>
<th>Iranian</th>
<th>Turkish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.869 (0.865-0.874)</td>
<td>0.901 (0.895-0.906)</td>
<td>0.831 (0.807-0.854)</td>
<td>0.821 (0.804-0.838)</td>
</tr>
<tr>
<td>HLA-B27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.776 (0.770-0.782)</td>
<td>0.54 (0.534-0.556)</td>
<td>0.54 (0.500-0.569)</td>
<td>0.58 (0.553-0.605)</td>
</tr>
<tr>
<td>European non-MHC PRS (PCA+SEX)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.923 (0.919-0.927)</td>
<td>0.761 (0.752-0.770)</td>
<td>0.853 (0.827-0.880)</td>
<td>0.849 (0.830-0.867)</td>
</tr>
<tr>
<td>European overall PRS (PCA+SEX)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.59 (0.578-0.593)</td>
<td>0.639 (0.629-0.650)</td>
<td>0.59 (0.553-0.620)</td>
<td>0.61 (0.581-0.633)</td>
</tr>
<tr>
<td>East Asian non-MHC PRS (PCA+SEX)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.88 (0.879-0.889)</td>
<td>0.942 (0.937-0.947)</td>
<td>0.87 (0.850-0.896)</td>
<td>0.84 (0.820-0.859)</td>
</tr>
<tr>
<td>East Asian overall PRS (PCA+SEX)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.630 - - - -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA alone – European</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 0.750 - -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA alone – Iranian</td>
<td></td>
<td></td>
<td>0.536</td>
<td></td>
</tr>
<tr>
<td>PCA alone – Turkish</td>
<td></td>
<td></td>
<td></td>
<td>0.520</td>
</tr>
</tbody>
</table>
**Supplementary References**


Figure S1: Receiver operating characteristic curve plot of performance of HLA-B27 in the North American cohort (red line, AUC=0.895), PRS trained in the British cohort and externally validated in the North American cohort (aqua line, AUC=0.928), and European PRS cross-validated (purple line, AUC=0.924).
Figure S2: Receiver operating characteristic curve plot of performance of PRS in European-descent cohort (purple line, AUC=0.924 based on 10 folds cross-validation) and East-Asian cohort (red dotted, AUC=0.948 based on 10 folds cross-validation). The validation performance of the European PRS in different cohort are shown in solid line (Iranian (green, AUC=0.852), Turkish (blue line, AUC=0.854), East Asian (red, AUC=0.788)) whilst the validation performance of the East-Asian PRS in other population are in dotted lines (Iranian (green, AUC=0.872), Turkish (blue, AUC=0.840), European (purple, AUC=0.880)).
Figure S3: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 0.55% (the population prevalence of AS)\textsuperscript{1}.
Figure S4: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 10%.
Figure S5: Positive and negative predictive values of PRS for AS in Europeans, assuming prior probability of AS of 20%.
Figure S6: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 0.55% (the population prevalence of AS).
Figure S7: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 10%.
Figure S8: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 20%.
Figure S9: Positive and negative predictive values of PRS for AS in East Asian cohort, assuming prior probability of AS of 30%.
Figure S10. Scheme of analysis. A. Analysis flow-chart. B. Scheme of 10 folds cross-validation in EUR or EAS cohort. C. Scheme of EUR PRS external validation. D. Scheme of cross-ethnicity testing.
Figure S11. Principal coordinates analysis plot of PC1 vs PC2 in European-descent cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Figure S12. Principal coordinates analysis plot of PC1 vs PC2 in East Asian-descent cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Figure S13: Principal coordinates analysis plot of PC1 vs PC2 in Iranian cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.
Figure S14: Principal coordinates analysis plot of PC1 vs PC2 in Turkish cases and controls, compared with ancestry-known subjects from HapMap3. AFR=African, AMR=Admixed American (Mexican ancestry from Los Angeles, California), EAS=East Asian, EUR=European, SAS=South Asian.