Genomic risk scores for juvenile idiopathic arthritis and its subtypes

Supplementary Material

Results

Figure S1: Outline of the study design.
**Figure S2:** Cross-validation results within the UK dataset. The results are from 10x10 cross-validated AUC (LOESS-smoothed) as a function of the number of SNPs assigned a non-zero weight in the model. The best model was selected at 26 SNPs with predictive measure 0.671 with (0.668, 0.674) 95% CI.
Training the genomic risk scores

Figure S3: A) Genome-wide association study (GWAS) of the UK dataset adjusted by its first ten principal components. The highlighted points correspond to the SNPs with non-zero weight selected by the GRS model. B) Weights associated for each of the SNPs selected for the model.
Extensions to the model

MetaGRS

We explored alternative approach to enhance the predictive power of our current JIA model. Given the strong pleiotropy across autoimmune diseases, we hypothesised that it may be possible to extract more predictive signal from GRSs generated for other autoimmune diseases via a metaGRS approach (cite our JACC paper), particularly given the limited sample sizes available for JIA GWAS.

Figure S4 illustrates the pipeline followed to compute the metaGRS model. Table S1 shows 16 autoimmune disease summary statistic (GWAS) sets that were considered. We generated GRSs by LD-clumping (plink --clump) each summary statistic using the CLARITY dataset. Figure S5 shows the correlations between the GRSs including our JIA GRS model. Next, using CLARITY as a training set, we applied elastic-net regression in ten-fold cross-validation (using the glmnet R package\(^1\)) to combine all the GRSs into a single weighted 'meta' GRS. Figure S6 shows the model performance (binomial deviance) across a range of values of the elastic net penalty, in cross-validation on the training set, where we selected the model with the lowest binomial deviance (the parameter s="lambda.min" in the glmnet package). Finally, the weights for this model were used to construct the JIA metaGRS model.

Figure S7 compares the effect size (OR) of each of the GRSs individually (logistic regression) and their effect size as part of the JIA metaGRS (lasso logistic regression), in CLARITY. Finally, Table S2 presents the OR and AUC obtained by our JIA GRS and metaGRS model GRSs. The results showed improvement of >1% in AUC in compare with JIA GRS, however the uncertainty in estimates was large due to limited sample sizes.
Figure S4: Workflow of computing the MetaGRS.

1. **Preparing the Data**
   - **LD-Clumping (plink–clump)**
   - **Compute GRS (plink–score)**

2. **Combining the GRSs**
   - **GRS 1**
   - **GRS 16**
   - **Our JIA GRS**
   - **Phenotype (CLARITY)**
   - **Deriving the MetaGRS (R glmnet)**

3. **Combining the GRSs**
   - **JIA MetaGRS**

4. **Testing**
   - **CHOP**
   - **AUC OR**
**Table S1:** List of external summary statistics (GWAS) used to compute the JIA MetaGRS. Additionally, this table provides the final number of SNPs with non-zero weight from each GRS.

<table>
<thead>
<tr>
<th>GRS label</th>
<th>Trait name</th>
<th>Number of SNPs after clumping</th>
<th>Study reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLE</td>
<td>Systemic lupus erythematosus</td>
<td>1,022</td>
<td>2</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple Sclerosis</td>
<td>546</td>
<td>3</td>
</tr>
<tr>
<td>MS (ic)</td>
<td>Multiple Sclerosis</td>
<td>964</td>
<td>4</td>
</tr>
<tr>
<td>PSO</td>
<td>Psoriasis</td>
<td>615</td>
<td>5</td>
</tr>
<tr>
<td>NAR</td>
<td>Narcolepsy</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>CEL</td>
<td>Celiac</td>
<td>1,141</td>
<td>7</td>
</tr>
<tr>
<td>T1D (cc)</td>
<td>Type 1 Diabetes</td>
<td>380</td>
<td>8</td>
</tr>
<tr>
<td>T1D (meta)</td>
<td>Type 1 Diabetes</td>
<td>457</td>
<td>8</td>
</tr>
<tr>
<td>RA (Okada)</td>
<td>Rheumatoid Arthritis</td>
<td>1,872</td>
<td>9</td>
</tr>
<tr>
<td>RA (Stahl)</td>
<td>Rheumatoid Arthritis</td>
<td>816</td>
<td>10</td>
</tr>
<tr>
<td>RA (Eyre)</td>
<td>Rheumatoid Arthritis</td>
<td>658</td>
<td>11</td>
</tr>
<tr>
<td>PBC (Liu)</td>
<td>Primary Biliary Cirrhosis</td>
<td>449</td>
<td>12</td>
</tr>
<tr>
<td>PBC (Cordell)</td>
<td>Primary Biliary Cirrhosis</td>
<td>524</td>
<td>13</td>
</tr>
<tr>
<td>AS</td>
<td>Ankylosing Spondylitis</td>
<td>378</td>
<td>14</td>
</tr>
<tr>
<td>UC</td>
<td>Ulcerative Colitis</td>
<td>827</td>
<td>15</td>
</tr>
<tr>
<td>JIA (Hinks)</td>
<td>Juvenile Idiopathic Arthritis</td>
<td>256</td>
<td>16</td>
</tr>
</tbody>
</table>
**Figure S5**: Correlation matrix between the JIA GRS model and all the external GRSs generated to create the JIA MetaGRS model. All the GRSs were computed using the CLARITY dataset.

**Figure S6**: 10-fold cross-validated binomial deviance from elastic-net regression on JIA phenotype using our JIA GRS model in combination with the 16 autoimmune GRSs.
**Figure S7:** Comparison of the effect size (95% CI) of each of the GRSs in (i) logistic regression (considering each GRS separately) vs (ii) the conditional effect size in elastic-net logistic regression (used to create the metaGRS individually), in the CLARITY dataset. Confidence intervals are not available for the odds ratios estimated via elastic-net.

```
Table S2: Performace of the JIA GRS and metaGRS in external validation on the CHOP dataset. Based on logistic regression, optionally adjusting for sex and top 10 genetic principal components (PCs).

<table>
<thead>
<tr>
<th></th>
<th>AUC (95% CI)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHOP (US)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex + PCs</td>
<td>0.677 (0.654–0.701)</td>
<td>--</td>
</tr>
<tr>
<td>GRS</td>
<td>0.657 (0.631–0.683)</td>
<td>1.831 (1.685–1.991)</td>
</tr>
<tr>
<td>MetaGRS</td>
<td>0.684 (0.659–0.709)</td>
<td>2.051 (1.870–2.252)</td>
</tr>
<tr>
<td>GRS + Sex + PCs</td>
<td>0.735 (0.712–0.758)</td>
<td>1.838 (1.686–2.007)</td>
</tr>
<tr>
<td>MetaGRS + Sex + PCs</td>
<td>0.748 (0.725–0.771)</td>
<td>2.042 (1.857–2.250)</td>
</tr>
</tbody>
</table>
```

**Enthesitis-related JIA HLA-B27 model**

The human leukocyte antigen B27 gene (*HLA-B27*) has a strong association with the enthesitis-related arthritis JIA subtype\(^{14,17–20}\). We created a model to predict enthesitis-related arthritis by using the most commonly tested tag SNPs for *HLA-B27* (rs13202464, rs116488202 and rs4349859) to classify each individual as *HLA-B27* positive or negative. Then we run a logistic regression model between the case/control status and the presence of the *HLA-B27* in the UK dataset.
Methods

Genotype data and quality control

Figure S8: First five principal components plot for the UK dataset.

Figure S9: First five principal components plot for the CHOP dataset.
**Figure S10:** First five principal components plot for the CLARITY dataset.
Figure S11: First five principal components plot for the selected CHOP subset.

Figure S12: First five principal components plot for the selected CLARITY subset.
Reference


