**Integrative Genomic Analysis Identifies the Soluble Receptor for Advanced Glycation End Products as Putatively Causal for Rheumatoid Arthritis**

**Supplementary Figures and Tables**

Figure S1. Mendelian Randomization Overview

Table S1. Protein Quantitative Trait Locus Variants that Coincide with Genetic Variants for Rheumatoid Arthritis

Table S2. Mendelian Randomization Results for 34 Proteins in Relation to Rheumatoid Arthritis

Table S3. sRAGE Levels in Framingham Heart Study Participants

Figure S2. Sensitivity Analysis of sRAGE in Relation to Rheumatoid Arthritis

Figure S3. Sensitivity Analysis of sRAGE in Relation to Rheumatoid Factor Levels



**Figure S1. Mendelian Randomization Overview**

Overview of Mendelian randomization, which can be used to infer causal relationship between an exposure (e.g. protein level) and an outcome (e.g. RA and RF). MR assumes that instrumental variables (IVs) are associated with the exposure, influences the outcome via the exposure only, and are not associated with confounders[10].

**Table S1. Protein Quantitative Trait Locus Variants that Coincide with Genetic Variants for Rheumatoid Arthritis**

 

\*Based on GWAS of protein level in 6,861 FHS participants[6]

\*\*Based on GWAS of RA from the MRC IEU[7, 10]

Only Bonferroni significant *cis*-pQTLs (P < 1.25E-07) and *trans*-pQTLs (P < 7.04E-10) were considered for overlap analysis, in conjunction with genome-wide significant variants (P < 5.13E-09; 0.05/9,739,304) from RA GWAS (P< 3.69E-09; 0.05/13,538,539). Beta for the pQTL GWAS is in terms of protein level differences (pg/mL) per risk allele, and OR for the RA GWAS is the odds ratio per risk allele.

**Table S2. Mendelian Randomization Results for 34 Proteins in Relation to Rheumatoid Arthritis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Protein | nsnp | OR | 95% CI | P Value |
| sRAGE | 2 | 0.482 | 0.374-0.622 | 1.85E-08\* |
| sICAM1 | 1 | 1.200 | 1.030-1.398 | 0.02 |
| NTproBNP | 1 | 1.122 | 0.980-1.284 | 0.096 |
| MMP9 | 1 | 1.124 | 0.970-1.302 | 0.119 |
| LPA | 10 | 0.968 | 0.925-1.012 | 0.154 |
| EFEMP1 | 2 | 0.914 | 0.803-1.040 | 0.17 |
| GMP140 | 3 | 1.049 | 0.968-1.137 | 0.242 |
| CD40L | 1 | 1.254 | 0.854-1.841 | 0.249 |
| MMP8 | 1 | 0.862 | 0.666-1.117 | 0.262 |
| ADM | 2 | 0.909 | 0.759-1.089 | 0.299 |
| BCHE | 1 | 0.959 | 0.881-1.043 | 0.323 |
| CLEC3B | 1 | 0.911 | 0.758-1.096 | 0.323 |
| GDF15 | 1 | 1.040 | 0.961-1.125 | 0.332 |
| DPP4 | 1 | 1.102 | 0.826-1.470 | 0.512 |
| COL18A1 | 1 | 0.942 | 0.788-1.126 | 0.512 |
| CRP | 1 | 0.935 | 0.766-1.142 | 0.512 |
| MCAM | 1 | 1.073 | 0.870-1.323 | 0.516 |
| CD5L | 1 | 0.966 | 0.869-1.073 | 0.516 |
| FGG | 1 | 1.061 | 0.886-1.270 | 0.516 |
| GP5 | 1 | 1.060 | 0.866-1.297 | 0.572 |
| Cystatin C | 1 | 1.025 | 0.939-1.120 | 0.576 |
| UCMGP | 1 | 1.039 | 0.909-1.187 | 0.579 |
| Resistin | 1 | 0.945 | 0.752-1.186 | 0.623 |
| KLKB1 | 2 | 1.028 | 0.905-1.168 | 0.668 |
| MPO | 2 | 1.024 | 0.895-1.173 | 0.732 |
| CXCL16 | 2 | 0.970 | 0.801-1.176 | 0.758 |
| A1M | 1 | 1.035 | 0.761-1.407 | 0.827 |
| AGP1 | 1 | 0.971 | 0.750-1.258 | 0.827 |
| SERPINA10 | 4 | 1.007 | 0.911-1.113 | 0.896 |
| REG1A | 2 | 1.007 | 0.880-1.153 | 0.924 |
| C2 | 2 | 0.948 | 0.114-7.891 | 0.961 |
| sGP130 | 1 | 1.000 | 0.835-1.198 | 1 |
| CNTN1 | 1 | 1.000 | 0.865-1.156 | 1 |
| GRN | 1 | 1.000 | 0.874-1.145 | 1 |

\* denotes Bonferroni-corrected significance at P<0.00147 (0.05/34).

MR results for all 34 proteins that were tested in MR analysis in relation to RA. 6 proteins (Ceruloplasmin, HPX, SAA1, CD14, PON1, and NCAM) out of the 40 total proteins with overlapping *cis-*pQTLs with the outcome GWAS were removed by the MR algorithm as the pQTL variants contained ambiguous palindromes or contained null values in the outcome GWAS. sRAGE and sICAM1 were the only significant proteins, with sRAGE passing the Bonferroni-corrected significance threshold for multiple testing (p<0.05/34) and sICAM1 being only nominally significant (p<0.05). The OR for the MR results is the odds ratio per 1 SD increment in inverse rank-normalized protein levels.

**Table S3. sRAGE Levels in Framingham Heart Study Participants**

|  |  |  |
| --- | --- | --- |
|  |  | sRAGE Levels (pg/mL) |
| **rs2070600** | **N** | **Mean** | **Std** | **Minimum** | **25th** | **Median** | **75th** | **Maximum** |
| C/C | 6260 | 3633.920 | 1164 | 269 | 2810 | 3500 | 4310 | 12600 |
| C/T | 558 | 2895.400 | 948 | 615 | 2230 | 2810 | 3390 | 6380 |
| T/T | 13 | 1661.540 | 432 | 850 | 1450 | 1610 | 2000 | 2470 |

A distribution of plasma sRAGE level by the genotype of rs2070600 was generated. There were a total of 6,831 FHS participants with information on rs2070600 genotype from the 1000G imputation and plasma protein measurements. Among these 6,831 FHS participants, a total of 558 individuals carried one minor T allele, 13 were homozygotes of the T allele, and 6,260 carried the major (C) allele.

****

**Figure S2. Sensitivity Analysis of sRAGE in Relation to Rheumatoid Arthritis**

Panel (a): Forest plot depicting the causal effect calculated for individual sRAGE pQTL variants on RA. 2 *cis-*pQTLs were available for the MR analysis. The inverse-variance weighted average effect of the SNPs were highly influenced by rs2070600 compared to rs9266529. Panel (b): Scatter plot relating the SNP-sRAGE and SNP-RA effect sizes. Consistent with the observation in the forest plot, the scatter plot indicates that rs2070600 is the driver of the causality between sRAGE and RA. While rs2070600 is negatively associated with sRAGE and positively associated with RA, the scatter plot does not seem to reflect that as the plotting algorithm makes the effect sizes of all data points positive.



**Figure S3. Sensitivity Analysis of sRAGE in Relation to Rheumatoid Factor Levels**

Panel (a): Forest plot depicting the causal effect calculated for individual sRAGE pQTL variants on serum RF levels. For this analysis, three SNPs were available. Similar to MR of our primary MR, rs2070600 had the narrowest confidence interval and contributed the most to the inverse variance weighted average effect of sRAGE on the outcome. Panel (b): Leave-one-out analysis showed that leaving rs2070600 out affected the confidence interval of the effect size to a greater extent than other SNPs. Leave-one-out analysis was not available for the primary MR sensitivity analyses as there were only 2 SNPs available for analysis. Panel (c): Scatter plot also showed a trend similar to the primary MR, where rs2070600 was the driver of the causal relation between sRAGE and RF. The directionality of effect is consistent with the primary RA MR. Both inverse variance weighted and MR Egger tests are shown since the number of SNPs was greater than 2, unlike for MR of RA. The two lines are nearly identical, indicating that there is no horizontal pleiotropy as MR Egger is a sensitivity analysis that relaxes an adjustment for potential pleiotropy.