Acosta-Herrera, M. et al. (2019) Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. Annals of the Rheumatic Diseases, 78(3), pp. 311-319.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/197306/

Deposited on: 23 September 2019

# Genome-wide meta-analysis reveals shared new *loci* in systemic seropositive rheumatic diseases

Marialbert Acosta-Herrera,[1†]* Martin Kerick,[1†] David González-Serna,[1] Myositis Genetics Consortium,[ɸ] Scleroderma Genetics Consortium,[ɸ] Cisca Wijmenga,[2] Andre Franke,[3] Peter K. Gregersen,[4] Leonid Padyukov,[5] Jane Worthington,[6] Timothy Vyse,[7-8] Marta E Alarcón-Riquelme,[9] Maureen D. Mayes,[10] Javier Martin.[1]*

[1] Institute of Parasitology and Biomedicine López-Neyra, IPBLN-CSIC, PTS Granada, Granada, Spain; [2]Department of Genetics, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands; [3]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; [4]Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Manhasset, NY, USA; [5]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; [6]Manchester NIHR Biomedical Research Centre, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [7]Division of Genetics and Molecular Medicine, King's College London, London, UK; [8]Division of Immunology, Infection and Inflammatory Disease, King's College London, London, UK; [9]Centre for Genomics and Oncological Research (GENYO), Pfizer-University of Granada-Andalusian Regional Government, Granada, Spain; [10]The University of Texas Health Science Center–Houston, Houston, Texas, USA; [ɸ]Members of the Myositis Genetics Consortium (MYOGEN) and Scleroderma Genetics Consortium are listed in the supplemental note.

[†]These authors contributed equally to this work.

**\*Corresponding authors**

Marialbert Acosta-Herrera: Institute of Parasitology and Biomedicine López-Neyra, CSIC, Av. del Conocimiento 17. Armilla (Granada), Spain 18016. Phone: (+34) 958181621; Fax: (+34) 958181632; email: m.acostaherrera@ipb.csic.es

Javier Martín: Institute of Parasitology and Biomedicine López-Neyra, CSIC, Av. del Conocimiento 17. Armilla (Granada), Spain 18016. Phone: (+34) 958181669; Fax: (+34) 958181632; email: javiermartin@ipb.csic.es

## Abstract

**Objective:** Immune-mediated inflammatory diseases (IMIDs) are heterogeneous and complex conditions with overlapping clinical symptoms and elevated familial aggregation, which suggests the existence of a shared genetic component. In order to identify this genetic background in a systematic fashion, we performed the first cross-disease genome-wide meta-analysis in systemic seropositive rheumatic diseases, namely: systemic sclerosis, systemic lupus erythematosus, rheumatoid arthritis and idiopathic inflammatory myopathies.

**Methods:** We meta-analyzed ~6.5 million single nucleotide polymorphisms (SNPs) in 11,678 cases and 19,704 non-affected controls of European descent populations. The functional roles of the associated variants were interrogated using publicly available databases.

**Results:** Our analysis revealed five shared genome-wide significant independent *loci* that had not been previously associated with these diseases: *NAB1*, *KPNA4-ARL14*, *DGQK*, *LIMK1*, and *PRR12*. All of these *loci* are related with immune processes such as interferon and epidermal growth factor signaling, response to methotrexate, cytoskeleton dynamics, and coagulation cascade. Remarkably, several of the associated *loci* are known key players in autoimmunity, which supports the validity of our results. All the associated variants showed significant functional enrichment in DNase hypersensitivity sites, chromatin states and histone marks in relevant immune cells, including shared expression quantitative trait *loci*. Additionally, our results were significantly enriched in drugs that are being tested for the treatment of the diseases under study.

**Conclusions:** We have identified shared new risk *loci* with functional value across diseases and pinpoint new potential candidate *loci* that could be further investigated. Our results highlight the potential of drug repositioning among related systemic seropositive rheumatic IMIDs.

## Introduction

Autoimmunity occurs when the mechanisms related to immune self-tolerance fail, leading to an inappropriate destruction of normal tissue by the immune system. Genetic factors play an important role in the development of more than 80 immune-mediated inflammatory diseases (IMIDs) identified so far.[1] Comorbidity of these diseases, increased familial clustering, and shared risk variants have been widely documented.[2] However, to date, these shared *loci* have been identified by simple comparison between studies, and just recently they have been determined by rigorous and systematic analysis.[3] In this sense, combining genome-wide association studies (GWAS) across several diseases has proven to be a very useful tool for the identification of new genetic risk variants simultaneously associated with several IMIDs, and to expose shared pathways involved in the pathophysiology of these conditions.[4-7] To date, two large studies combining several diseases were recently published following this strategy. One of them was a meta-GWAS across 10 pediatric autoimmune diseases with shared population-based controls that revealed new candidate *loci* with immunoregulatory functions.[8] In the other study, the authors identified new shared associations by combining immunochip data across five chronic inflammatory diseases.[9]

Systemic seropositive rheumatologic IMIDs, such as systemic sclerosis (SSc), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and idiopathic inflammatory myopathies (IIM), are heterogeneous diseases of the connective tissue that share clinical and epidemiological manifestations, as well as life-threatening complications.[10] The common genetic component of these conditions has not been previously assessed systematically, although the overlap of associated genes is elevated when performing a pairwise comparison.[8] Autoantibody production is the main

96  feature of these diseases, comprising additionally a broad deregulation of the innate and

97  adaptive immune response. However, the low prevalence of most of these diseases

98  hinders the collection of large datasets that makes possible to attain sufficient statistical

99  power. Therefore, our study aimed to combine previously published GWAS datasets –

100 all from European descent populations– to identify shared genetic etiologies among

101 systemic seropositive rheumatologic IMIDs in a systematic fashion.

102

103 **Subjects and Methods**

104 **Study population**

105     A total of 12,132 affected subjects with four systemic seropositive rheumatic

106 IMIDs (SSc, SLE, IIM, and RA) and 23,260 controls were included in this study from

107 previously published GWAS [11-16] (Table S1).

108

109 **Data quality control and imputation**

110     Unified quality control (QC) of the 18 case-control collections was conducted

111 separately, based on stringent criteria using PLINK v.1.07.[17] Given that related and/or

112 duplicated subjects may have been recruited for different studies, genome-wide

113 relatedness was assessed and one individual from each pair was removed. Samples with

114 <95% of successfully called genotypes were removed.

115     Further, single nucleotide polymorphisms (SNPs) with genotyping call rate

116 <98%, minor allele frequencies (MAF) <1% and deviating from Hardy-Weinberg

117 equilibrium (HWE) with a *p*-value <0.001 in the control group were removed. To

118 control for possible population stratification, we performed principal component (PC)

119 analysis using GCTA64 and R-base software under GNU Public license v.2.

120    Imputation of autosomal SNPs was conducted in the Michigan Imputation

121    Server using Minimac3.[18] The software SHAPEIT[19] was used for haplotype

122    reconstruction and the Haplotype Reference Consortium r1.1 was used as the reference

123    population.[20]

124

125    **Statistical analyses**

126    *Disease-specific association testing:* Association testing for allele dosages was

127    performed by logistic Wald test using EPACTS software,[21] adjusting by the first two

128    or five PCs as appropriate to control for the genomic inflation factor in European

129    population ($\lambda$<1.05) (Table S1). SNPs with a MAF $\geq$1% and squared correlation (Rsq)

130    $\geq$0.3 were maintained in the analyses as suggested by the imputation software.

131    Additionally, we calculated a concordance rate by comparing imputed and true

132    genotypes.

133    *Cross-phenotype meta-analysis:* to identify shared *loci*, the summary-level statistics

134    were meta-analyzed using METASOFT.[22] A fixed-effects model was applied for

135    those SNPs without evidence of heterogeneity (Cochran's Q test *p*-value Q > 0.05), and

136    random-effects model was applied for SNPs displaying heterogeneity of effects between

137    studies (Q $\leq$ 0.05). Genome-wide significance was established at a *p*-value $\leq 5 \times 10^{-08}$.

138    SNP independence was assessed with the software GCTA-COJO (Table S2).[23, 24] To

139    annotate the independent signals SNPnexus[25] was used to the build37 genomic

140    coordinates.

141    *Model search to identify the diseases contributing to the association:* to identify the

142    diseases most likely contributing to the association signals, we performed an exhaustive

143    disease-subtype model search  with the R statistical package ASSET.[24] The

144     contribution of a disease was considered if at least two independent case-control

145     collections from the same disease were grouped with consistent effects.

146     *Novelty of the variants:* Our independent SNP associations were classified into "known"

147     or "new" associations based on the information retrieved from the NHGRI-EBI GWAS

148     catalog and the Phenopedia and Genopedia from HuGE Navigator.[26]

149     *Functional enrichment analysis:* in order to systematically characterize the functional,

150     cellular and regulatory contribution of the associated variants, a non-parametric

151     enrichment analysis implemented in GARFIELD was performed.[27] Furthermore, the

152     online tools HaploReg v.4.1[28] and the Genotype-Tissue Expression project

153     (GTEx)[29] were queried to determine whether any of the lead associated variants was

154     an expression quantitative trait locus (eQTL). The online tool Capture HiC plotter was

155     used to assess physical interactions between restriction fragments containing the

156     variants and the promoter of genes in the three-dimensional nuclear space.[30]

157     *Drug Target Enrichment Analysis:* the target genes of the eQTLs were used to model a

158     protein-protein interaction (PPI) network using String v10.[31] These protein products

159     were then used to query the OpenTargets Platform[32] for drug targets. Moreover, this

160     platform was used to search for drugs indicated or in different phases of drug

161     development for the treatment of SSc, SLE, IIM and RA. The Fisher's exact test was

162     used to calculate if the results of the meta-analysis were significantly enriched in

163     pharmacologically active drug targets.

164     Additional details of the Methods section are available in the online supplementary

165     methods.

166

167     **Results**

168     *Cross-phenotype meta-analysis and disease contribution*

169        Following sample QC and imputation, a total of 11,678 cases and 19,704 non-

170 overlapping controls were included in the genome-wide meta-analysis of 6,450,125

171 SNPs across the four diseases. The mean concordance rate among imputed and true

172 genotypes was 0.999±0.0003. The final λ showed minimal evidence of population

173 stratification in the meta-abalysis (λ=1.025). Moreover, we calculated λ1,000 with

174 consistent results (λ1,000=1.025). Summary of sample/variant QC and QQ plots are

175 shown in Table S1 and Figure S1, respectively.

176        The global meta-analysis revealed 42 non-hla significantly associated *loci*.

177 Subsequent conditional analyses showed that 27 SNPs were independent (Figure 1 and

178 Figure S2). Sixteen variants were meta-analyzed under a fixed effects model, whereas

179 eleven with random effects based on study heterogeneity.

180        To comprehensively explore the combinations of diseases contributing to the

181 associations we applied a subset-based meta-analysis implemented in ASSET.[24] Our

182 model search yielded 26 SNPs associated with at least two IMIDs (Table 1). All of

183 these variants were imputed in at least one dataset.

184

185 Among these 26 associations we found several key players in autoimmunity;

186 interestingly ten of these associations (38%) have never been reported before for SSc,

187 eight (31%) for SLE and RA, respectively, and 20 (77%) for IIM. Remarkably, five

188 SNPs have not been reported previously for any of the diseases under study and thus

189 constitute new shared risk *loci* in systemic seropositive rheumatic IMIDs (Table 1).

190 Amongst these five new associations we found the SNP rs744600 in the 3' region of the

191 NGFI-A binding protein 1 (*NAB1*) (Odds ratio [OR] for the T allele 0.88, Confidence

192 Interval [CI]=0.85-0.92), *p*-value=7.07x10$^{-11}$), and the intronic SNP rs13101828

193 mapping in the gene Diacylglycerol kinase theta (*DGKQ*) (OR for the G allele 1.11,

194 95%CI: 1.07–1.16, *p*-value=1.32x10[-08]). Of note, both genes have been previously

195 associated with a chronic autoimmune liver disease.[33, 34] The intergenic SNP

196 rs112846137, maps between the genes Karyopherin subunit alpha 4 (*KPNA4*) and the

197 ADP ribosylation factor like GTPase 14 (*ARL14*) (OR for the T allele 1.29, 95%CI:

198 1.07–1.56, *p*-value=1.42x10[-08]). Interestingly, the gene *ARL14* showed a suggestive

199 association in a pharmacogenomic GWAS of response to methotrexate in RA

200 patients.[35] In addition, we observe the associated SNP rs193107685 located in the 3'

201 region of the LIM domain kinase 1 *(LIMK1)* gene (OR for the C allele 1.52, 95%CI:

202 1.27–1.83, *p*-value=3.81x10[-09]). The protein encoded by this gene regulates actin

203 polymerization, a critical process in the activation of T cells.[36] Finally, the SNP

204 rs76246107 is located in an intron of the gene Proline rich 12 (*PRR12*) (OR for the G

205 allele 1.28, 95%CI: 1.14–1.43, *p*-value=3.36x10[-08]), which was associated with

206 fibrinogen concentration,[37] and is an active regulator of the inflammatory

207 response.[38]

Table 1. Twenty-six independent variants associated at a genome-wide significance level ($p < 5 \times 10^{-8}$) in the meta-analysis.

| Chr | Position[a] | SNP | Gene[b] | Functionality[c] | Effect Allele | OR (CI 95%) | Meta-Analysis $p$-value[d] | Cochran's $p$-value | Contributing Disease[e] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 67802371 | rs6659932 | *IL12RB2* | Intronic | C | 0.85 (079-0.91) | $6.08 \times 10^{-11}$ | $1.02 \times 10^{-02}$ | **IIM, SLE**, SSc |
| 1 | 114303808 | rs6679677 | *PHTF1-RSBN1* | Intergenic | A | 1.34 (1.21-1.49) | $2.30 \times 10^{-28}$ | $2.14 \times 10^{-04}$ | **IIM, RA**, SLE |
| 1 | 114377568 | rs2476601 | *PTPN22* | Coding (missense) | G | 0.75 (0.67-0.83) | $1.74 \times 10^{-28}$ | $1.06 \times 10^{-4}$ | IIM, RA, SLE |
| 1 | 114433946 | rs1217393 | *AP4B1* | Intronic | A | 0.89 (0.85-0.92) | $5.21 \times 10^{-09}$ | $4.91 \times 10^{-1}$ | **IIM, RA**, SLE, **SSc** |
| 1 | 173337747 | rs2422345 | *TNFSF4-LOC100506023* | Intronic | A | 1.11 (1.05-1.18) | $2.55 \times 10^{-08}$ | $6.00 \times 10^{-03}$ | **IIM**, SLE, **SSc** |
| 1 | 183532580 | rs17849502 | *NCF2* | Coding (missense) | T | 1.36 (1.16-1.59) | $3.93 \times 10^{-15}$ | $2.84 \times 10^{-04}$ | **IIM**, SLE |
| 2 | 191564757 | rs744600 | *NAB1** | 3'Downstream | T | 0.88 (0.85-0.92) | $7.07 \times 10^{-11}$ | $7.60 \times 10^{-1}$ | **IIM, RA, SLE, SSc** |
| 2 | 191933283 | rs13389408 | *STAT4* | Intronic | C | 1.27 (1.20-1.34) | $3.10 \times 10^{-17}$ | $3.99 \times 10^{-1}$ | **IIM**, SLE, SSc |
| 2 | 191973034 | rs10174238 | *STAT4* | Intronic | A | 0.73 (0.67-0.80) | $2.76 \times 10^{-42}$ | $4.31 \times 10^{-07}$ | **IIM**, SLE, SSc |
| 3 | 58183636 | rs35677470 | *DNASE1L3* | Coding (missense) | A | 1.22 (1.14-1.30) | $4.96 \times 10^{-09}$ | $6.78 \times 10^{-01}$ | **IIM**, SLE, SSc |
| 3 | 160312921 | rs112846137 | *KPNA4-ARL14** | Intergenic | T | 1.27 (1.17-1.37) | $1.42 \times 10^{-08}$ | $9.55 \times 10^{-01}$ | **IIM, RA, SLE, SSc** |
| 4 | 965720 | rs13101828 | *DGKQ** | Intronic | G | 1.11 (1.07-1.16) | $1.32 \times 10^{-08}$ | $2.29 \times 10^{-01}$ | **IIM, RA, SLE, SSc** |
| 5 | 150438477 | rs4958880 | *TNIP1* | Intronic | A | 1.16 (1.10-1.22) | $1.45 \times 10^{-11}$ | $2.61 \times 10^{-01}$ | **IIM, RA**, SLE, SSc |
| 5 | 159887336 | rs2431098 | *PTTG1-MIR3142HG* | Intergenic | G | 1.12 (1.05-1.20) | $4.91 \times 10^{-12}$ | $1.42 \times 10^{-01}$ | SLE, **SSc** |
| 6 | 106569270 | rs802791 | *PRDM1-ATG5* | Intergenic | C | 0.87 (0.83-0.92) | $3.65 \times 10^{-12}$ | $1.13 \times 10^{-01}$ | SLE, SSc |
| 6 | 138243739 | rs58721818 | *TNFAIP3* | 3'Downstream | T | 1.64 (1.46-1.84) | $4.64 \times 10^{-23}$ | $1.65 \times 10^{-01}$ | **IIM**, SLE, SSc |
| 7 | 73537902 | rs193107685 | *LIMK1** | 3'Downstream | C | 1.52 (1.27-1.83) | $3.21 \times 10^{-09}$ | $1.18 \times 10^{-01}$ | **RA, SLE, SSc** |
| 7 | 128589633 | rs10954214 | *IRF5* | 3UTR | T | 1.18 (1.13-1.23) | $6.63 \times 10^{-17}$ | $3.64 \times 10^{-01}$ | **IIM**, RA, SLE, SSc |
| 7 | 128647942 | rs13238352 | *TNPO3* | Intronic | T | 1.44 (1.30-1.60) | $1.47 \times 10^{-38}$ | $2.12 \times 10^{-01}$ | SLE, SSc |
| 8 | 11341880 | rs2736337 | *FAM167A-BLK* | Intergenic | C | 1.23 (1.17-1.30) | $4.86 \times 10^{-22}$ | $1.29 \times 10^{-01}$ | **IIM,** RA, SLE, SSc |
| 11 | 633689 | rs7929541 | *SCT-DRD4* | Intergenic | G | 0.89 (0.83-0.95) | $2.14 \times 10^{-10}$ | $4.98 \times 10^{-04}$ | **IIM, RA**, SLE, **SSc** |

| 12 | 112871372 | rs11066301 | *PTPN11* | Intronic | T | 1.11 (1.07-1.15) | $4.20 \times 10^{-08}$ | $5.86 \times 10^{-01}$ | **IIM, SLE, SSc** |
|----|-----------|------------|----------|----------|---|------------------|-----------------------|-----------------------|-------------------|
| 16 | 85994484 | rs35929052 | *IRF8* | Intergenic | T | 0.83 (0.78-0.88) | $1.71 \times 10^{-09}$ | $4.69 \times 10^{-01}$ | **IIM**, SLE, SSc |
| 19 | 10462513 | rs11085725 | *TYK2* | Intronic | A | 0.88 (0.83-0.92) | $2.65 \times 10^{-10}$ | $1.86 \times 10^{-01}$ | **IIM**, SLE, SSc |
| 19 | 50121274 | rs76246107 | *PRR12** | Intronic | G | 1.28 (1.14-1.43) | $3.36 \times 10^{-08}$ | $1.50 \times 10^{-02}$ | **IIM, SLE, SSc** |
| 22 | 21985094 | rs5754467 | *YDJC* | 5'Upstream | G | 1.20 (1.13-1.27) | $1.24 \times 10^{-13}$ | $8.59 \times 10^{-02}$ | **IIM**, RA, SLE, **SSc** |

[a]According to NCBI build GRCh37/hg19.

[b]Variant localization based on the nearest gene.

[c]Functionality obtained from SNPnexus.[23]

[d]Results of meta-analysis either under a fixed effect if no heterogeneity was found based on Cochran's Q test ($p$-value≥0.05) or under a random effect if heterogeneity was found among studies.

[e]Disease contributing to the association observed by the subset meta-analysis method with ASSET.[25] The diseases for which this locus has never been reported before at genome-wide significance level are shown in boldface.

*Denotes novel *loci* in the study.

Chr: chromosome; OR: odds ratio; CI: confidence interval; IIM: idiopathic inflammatory myopathy; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; SSc: systemic sclerosis.

All the variants in the table were imputed in at least one of the 18 case-control collections.

208

*Associated loci and their functional enrichment on regulatory elements*

210     To assess whether the associated variants lie in coding and non-coding

211 regulatory and cell-type-specific elements of the genome, we performed an enrichment

212 analysis with GARFIELD.[39] The results obtained showed marked enrichment

213 patterns mainly in blood cells and skin cells, with 247 significant enrichments

214 ($p{\leq}5{\times}10^{-05}$) (Figure S3 and Table S3). Table 2 summarizes the main enrichment results.

215 We found that the majority of associated variants were enriched in DNase I

216 hypersensitivity site (DHS) hotspots in blood, as depicted in Figure 2. This functional

217 category included a repertoire of cells from the immune system, such as B-lymphocytes

218 (Fold enrichment (FE)=11.68, empirical *p* (*pemp*)$<1{\times}10^{-05}$), T-lymphocytes (FE=10.42,

219 *p*emp$<1{\times}10^{-05}$), including T helper cells (FE=7.81, *p*emp$<1{\times}10^{-05}$), T CD8+ (FE=7.61,

220 *p*emp$<1{\times}10^{-05}$), natural killer cells (FE=11.36, *p*emp$<1{\times}10^{-05}$), and monocytes

221 (FE=8.99, *p*emp$<1{\times}10^{-05}$). In line with this enrichment, disease-associated SNPs were

222 enriched in enhancers (FE=14.99, *p*emp$<1{\times}10^{-05}$), within TSS (FE=14.87,

223 *p*emp$<1{\times}10^{-05}$), and on transcription factor binding sites (FE=12.20, *p*emp$<1{\times}10^{-05}$) in

224 the B-lymphocyte cell line GM12878. Additionally, the highest enrichment was

225 observed in the histone modification H3K9ac (FE=14.02, *p*emp$<1{\times}10^{-05}$), and

226 H3K27ac (FE=10.81, *p*emp$<1{\times}10^{-05}$) in the B-lymphocyte cell line, which are

227 positively associated with gene activation. Although these modifications are increased

228 in the promoters of active genes, the latter has been shown to be associated with active

229 enhancers.[40] Moreover, enrichment was observed in H3K4me1,2,3 sites, which

230 usually surround TSS and are also positively correlated with gene expression.[40]

Table 2. Summary of the most enriched functional annotations for the SNPs associated in the meta-analysis at a genome-wide significance threshold ($p < 5 \times 10^{-8}$).

| Category[a] | Tissue | Cell types | Type | NAnnotThesh[b] | NAnnot[c] | NThresh[d] | N (LD-pruned variants)[e] | Fold Enrichment | Empirical *p*-value |
|---|---|---|---|---|---|---|---|---|---|
| Chromatin_States | Blood | GM12878 | Enhancer | 13 | 10,944 | 33 | 416,420 | 14.99 | $<1 \times 10^{-5}$ |
| | | GM12878 | TSS | 12 | 10,182 | 33 | 416,420 | 14.87 | $<1 \times 10^{-5}$ |
| Footprints | Blood | GM06990 | Footprints | 8 | 3,153 | 33 | 416,420 | 32.02 | $<1 \times 10^{-5}$ |
| | | GM12878 | H3K9ac | 21 | 18,903 | 33 | 416,420 | 14.02 | $<1 \times 10^{-5}$ |
| | | GM12878 | H3K27ac | 22 | 25,674 | 33 | 416,420 | 10.81 | $<1 \times 10^{-5}$ |
| | | GM12878 | H2AFZ | 22 | 25,824 | 33 | 416,420 | 10.75 | $<1 \times 10^{-5}$ |
| Histone modifications | Blood | GM12878 | H3K4me3 | 17 | 25,365 | 33 | 416,420 | 8.46 | $<1 \times 10^{-5}$ |
| | | GM12878 | H3K4me2 | 23 | 34,807 | 33 | 416,420 | 8.34 | $5 \times 10^{-5}$ |
| | | GM12878 | H3K4me1 | 25 | 39,871 | 33 | 416,420 | 7.91 | $<1 \times 10^{-5}$ |
| | | GM12878 | H3K79me2 | 16 | 25,683 | 33 | 416,420 | 7.86 | $<1 \times 10^{-5}$ |
| Hotspots | Blood | GM06990 | Hotspots | 23 | 24,839 | 33 | 416,420 | 11.68 | $<1 \times 10^{-5}$ |
| | Skin | NHEK | Hotspots | 25 | 54,667 | 33 | 416,420 | 5.77 | $<1 \times 10^{-5}$ |
| Peaks | Blood | GM06990 | Peaks | 13 | 6,433 | 33 | 416,420 | 25.50 | $<1 \times 10^{-5}$ |
| TFBS | Blood | GM12878 | TFBS | 19 | 19,650 | 33 | 416,420 | 12.20 | $<1 \times 10^{-5}$ |

[a]Functional categories from the Encode[28] and Roadmap Epigenomics.[29]

[b]Number of LD-pruned annotated variants passing the meta-analysis threshold.

[c]Number of LD-pruned annotated variants in the reference dataset UK10K project.

[d]Number of LD-pruned variants passing the meta-analysis threshold.

[e]Number of LD-pruned variants in the reference dataset UK10K project.

GM12878: B-Lymphocyte; GM06990: B-lymphocyte, lymphoblastoid; NHEK: Normal Human Epidermal Keratinocytes; LD: Linkage disequilibrium; TSS: Transcription Start Site; TFBS: Transcription Factor Binding Sites.

231 *Expression quantitative trait loci (eQTL) and associated variants*

232      *In silico* analysis of eQTLs revealed the role of 16 of the lead SNPs as eQTLs in

233 whole blood, lymphoblastoid cell lines, transformed lymphocytes, skeletal muscle and

234 transformed fibroblasts derived from European individuals from HaploReg v.4.1[28]

235 (Table 3 and Table S4). Focusing on new associated variants, the SNP rs744600

236 modifies *NAB1* gene expression in lymphoblastoid cell lines ($p$=1.30x10$^{-34}$), whereas

237 the T allele increases *HIBCH* expression in skeletal muscles ($p$=8.09x10$^{-07}$). The G

238 allele of rs13101828 increases *DGKQ* expression in whole blood ($p$=3.29x10$^{-45}$),

239 lymphocytes ($p$=5.23x10$^{-19}$), fibroblasts ($p$=4.44x10$^{-06}$), lung cells ($p$=8.42x10$^{-28}$) and

240 several other tissues. The A allele of rs76246107 can reduce *ALDH16A1* expression in

241 lung cells ($p$=6.45x10$^{-06}$), and the protein encoded by this gene is involved in

242 oxidoreductase activity. Reassuringly, 14 of the 16 (87%) reported eQTLs showed a

243 physical interaction between the SNP and the promoter of 15 of the genes affected by

244 the eQTLs (Table 3), as suggested by Capture HiC (C-HiC) data (Table S5). These

245 independent evidences propose a mechanistic approach to understand the modulation of

246 gene expression.

Table 3. Summary of the eQTL results in European samples for the SNPs independently associated in the meta-analysis.

| SNP | Allele | Source | Gene | Tissue | *p*-value |
|---|---|---|---|---|---|
| rs6659932* | C | GTEx2015_v6 | *IL12RB2* | Whole blood | $3.72\times10^{-11}$ |
| rs6679677* | A | Westra 2013 | *PTPN22* | Whole blood | $4.84\times10^{-10}$ |
| rs2476601* | G | Westra2013 | *PTPN22* | Whole blood | $3.36\times10^{-10}$ |
|  |  | GTEx2015_v6 | *AP4B1* | Skeletal muscle | $5.45\times10^{-07}$ |
|  |  | GTEx2015_v6 | *HIPK1* | Whole blood | $7.71\times10^{-09}$ |
| rs1217393* | A | Westra 2013 | *PHTF1* | Whole blood | $9.56\times10^{-05}$ |
|  |  | Westra 2013 | *PTPN22* | Whole blood | $2.67\times10^{-10}$ |
|  |  | Westra 2013 | *RSBN1* | Whole blood | $1.41\times10^{-10}$ |
| **rs744600*** | T | GTEx2015_v6 | *HIBCH* | Skeletal muscle | $8.09\times10^{-07}$ |
|  |  | Lappalainen2013 | *NAB1* | Lymphoblastoid cell line | $1.30\times10^{-34}$ |
| rs13389408 | C | GTEx2015_v6 | *GLS* | Skeletal muscle | $3.42\times10^{-09}$ |
|  |  | Westra 2013 |  | Whole blood | $2.98\times10^{-07}$ |
| rs35677470* | A | GTEx2015_v6 | *PXK* | Skeletal muscle | $7.08\times10^{-06}$ |
| **rs13101828** | G | GTEx2015_v6 | *DGKQ* | Whole blood | $9.28\times10^{-45}$ |
|  |  |  |  | Transformed lymphocytes | $1.21\times10^{-23}$ |
|  |  |  |  | Transformed fibroblasts | $9.78\times10^{-07}$ |
|  |  |  |  | Lung | $8.42\times10^{-28}$ |
| rs4958880* | A | Westra 2013 | *TNIP1* | Whole blood | $1.09\times10^{-03}$ |
| rs10954214* | T | GTEx2015_v6 | *IRF5* | Whole blood | $2.56\times10^{-16}$ |
|  |  | Lappalainen2013 |  | Lymphoblastoid cell line | $7.54\times10^{-31}$ |
| rs13238352* | T | Lappalainen2013 | *IRF5* | Lymphoblastoid cell line | $2.88\times10^{-13}$ |
| rs2736337* | C | GTEx2015_v6 | *FAM167A* | Whole blood | $2.90\times10^{-26}$ |
|  |  |  | *FAM167A* | Transformed fibroblasts | $1.90\times10^{-18}$ |
|  |  |  | *FAM167A* | Transformed lymphocytes | $2.10\times10^{-15}$ |
|  |  |  | *BLK* | Whole blood | $5.30\times10^{-13}$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| rs2736337* | C | GTEx2015_v6 | *BLK* | Transformed fibroblasts | $1.30 \times 10^{-11}$ |
| | | | *BLK* | Transformed lymphocytes | $3.30 \times 10^{-06}$ |
| rs7929541* | C | GTEx2015_v6 | *TMEM80* | Transformed fibroblasts | $1.22 \times 10^{-11}$ |
| rs11085725* | T | GTEx2015_v6 | *TYK2* | Whole blood | $2.30 \times 10^{-06}$ |
| | | | *TMED1* | Whole blood | $8.80 \times 10^{-06}$ |
| **rs76246107*** | A | GTEx2015_v6 | *ALDH16A1* | Lung | $6.45 \times 10^{-06}$ |
| rs5754467* | G | GTEx2015_v6 | *UBE2L3* | Whole blood | $4.68 \times 10^{-06}$ |

New associated SNPs found in our meta-analysis are shown in boldface: rs744600 and rs13101828 associated with Systemic Sclerosis, Systemic Lupus Erythematosus, Rheumatoid Arthritis and idiopathic inflammatory myopathy; rs76246107 associated with Systemic Sclerosis, Systemic Lupus Erythematosus and idiopathic inflammatory myopathy. *Designates those SNPs where a physical interaction has been observed in Promoter Capture HiC data in relevant immune cells.

*Drug target enrichment analysis*

248    Genetic associations have the potential to improve the rates of success in the

249    development of new therapies.[41] We assessed if the protein-products from disease

250    associated eQTLs and their direct protein-protein interaction (PPI) partners were

251    enriched with pharmacologically active targets (Table S6 and Table S7). We identified

252    as eQTLs and PPIs 608 proteins for SSc, 630 for SLE, 632 for IIM, and 413 for RA,

253    based on data on drugs at any stage of development collected from the Open Targets

254    Platform (Table S8).[32] Using this information, we found for SSc that 23 out of 73

255    (32%) proteins are targeted by drugs being studied for the disease (OR=16.80, *p*-

256    value=1.41x10$^{-18}$). Similarly, 7 out of 25 (28%) proteins related to IIM and 13 out of

257    146 (9%) proteins related to SLE are addressed by drugs in consideration for IIM and

258    SLE (OR=13.40, *p*-value=4.62x10$^{-06}$, OR=3.38, *p*-value=2.85x10$^{-04}$, respectively)

259    (Table S9).

260

## Discussion

262    In the present study we identified five unreported shared *loci* associated with

263    systemic seropositive rheumatic IMIDs. This is the first large-scale meta-analysis,

264    including more than 11,000 cases and 19,000 non-overlapping controls aiming to

265    improve our knowledge regarding the genetic resemblances among these conditions.

266    Our results show that 85% of the associated variants were shared by at least

267    three diseases. Interestingly, for several known RA susceptibility *loci* the contribution

268    of RA was limited. In this case, most of the associated variants were independent to the

269    ones previously reported. Among the new associated SNPs, the signals mapping to

270    *NAB1*, *DGKQ* and *KPNA4-ARL14* were associated to all of the diseases under study.

271    NAB proteins are known to interact with early growth response (EGR) family members

272 and act as corepressors induced by type I interferons (IFN).[42] The 'IFN signature'–

273 has been previously described in these diseases.[43-46] Interestingly, two IFN

274 regulatory factors –*IRF5* and *IRF8*– previously associated to the conditions under study,

275 were associated in the meta-analysis. Additionally, the associated SNP is an eQTL in

276 lymphoblastoid cell line, which evidences its role in disease pathogenesis. The DGKQ

277 protein mediates cell signal transduction and can indirectly enhance the epidermal

278 growth factor receptor (EGFR) signaling activity.[47] This pathway regulates cell

279 proliferation and migration, and its expression is augmented in the vasculature of SSc

280 patients with pulmonary involvement.[48] Moreover, the risk allele was associated with

281 an increased expression of the gene in lymphocytes, fibroblasts and lung. In the same

282 line, this gene was associated with Sjögren's syndrome, a related connective tissue

283 disease.[49] The protein encoded by the gene *ARL14* is a GTPase involved in the

284 recruitment of MHC class II containing vesicles and control the movement of dendritic

285 cells (DCs) along the actin cytoskeleton.[50] The protein LIMK1 regulates many actin-

286 dependent processes, including the assembly of the immune synapse between T cells

287 and antigen presenting cells, an expected biological process involved in seropositive

288 IMIDs. Remarkably, rs193107685 and rs112846137 interact physically with the

289 promoters of the genes *LIMK1* and *ARL14*, respectively, in DCs (Figure S4). The gene

290 *PRR12* has been previously associated with fibrinogen concentrations.[37] Fibrinogen is

291 considered a high-risk marker for vascular inflammatory diseases and is considered an

292 accurate predictor of cardiovascular diseases.[38, 51] Moreover, this molecule is an

293 active player in the coagulation cascade, responsible for the spontaneous formation of

294 fibrin fibrils. Cardiovascular events and fibrosis are the most life-threatening

295 complications described in SSc, IIM, and SLE.[52-54]

296  The associated SNPs are highly enriched in functional categories in B and T

297  cells, natural killer and monocytes, highlighting the relevance of these cells in systemic

298  seropositive rheumatic IMIDs. Beyond whole blood, the skin is the other tissue with

299  significant functional categories, which is not surprising given the nature of these

300  connective tissue diseases. Moreover, epithelial cells could transdifferentiate into

301  mesenchymal cells and eventually contribute in fibrotic processes.[55] Moreover, SSc

302  patients are usually stratified according to the extent of skin involvement.[43] On the

303  other hand, the histone modifications observed are consistent with the ones reported in

304  previous studies, where histone hyperacetilation have been described in synovial tissues

305  in RA, in B cells in SSc, and in CD4+ T cells in SLE.[40] Finally, the independent

306  associated SNPs have significant eQTLs in relevant tissues (Table 3) and *in silico* data

307  from promoter capture HiC experiments showed the potential mechanisms in which

308  most eQTLs modulate gene expression. Interestingly, all new associated SNPs interact

309  with the promoters of surrounding genes, suggesting them as putative candidates with a

310  role in the pathophysiology of these conditions (Figure S4 and Table S5).

311  The prevalence of SSc, SLE, and IIM is low and there are no specific treatments

312  for these diseases in comparison with RA; therefore, given our current knowledge on

313  the use of genetic findings in drug target validation and drug repurposing, we evaluated

314  if drugs currently indicated for RA had the potential to be used in any of the other

315  IMIDs under study. Our meta-analysis revealed that ten *loci* overlap with known RA

316  risk genes. For instance, the gene-product of *TYK2* is targeted directly by Tofacitinib,

317  which inhibits janus kinases (https://www.drugbank.ca/drugs/DB08895) or indirectly

318  through the interleukin 6 (IL-6) family signaling pathway by targeting the IL6 receptor

319  with Tocilizumab (https://www.drugbank.ca/drugs/DB06273). Both drugs are currently

320  indicated for moderate to severe RA patients who respond poorly to disease-modifying

321     anti-rheumatic drugs. As *TYK2* is associated with SSc, SLE and IIM, it is a good

322     candidate for therapy repositioning in these diseases. As a proof of concept, Tofacitinib

323     is currently on trial for SLE (clinical trial identifier NCT02535689), SSc

324     (NCT03274076) and Dermatomyositis (NCT03002649). Overall, we found that five of

325     the *loci* identified in our meta-analysis interact with 17 genes that are considered drug

326     targets, six of which are used for the treatment of these diseases (Table 4). Another

327     interesting candidate for drug repurposing is Imatinib, a kinase inhibitor that targets

328     ABL1, which interacts with the gene product of BLK, a known locus associated with

329     SSc and RA (Table 4). Imatinib is currently being tested for SSc (NCT00555581) and

330     RA (NCT00154336).

Table 4. Summary of the plausible target gene products with drug indications in systemic IMIDs.

| Associated SNP | Gene product | Association results[a] | Drugs[b] | Targets | Disease indication[c] |
|---|---|---|---|---|---|
| rs6659932 | IL12RB2 | **IIM, SLE**, SSc | Canakinumab | IL1B | RA |
| | | | Anakinra | IL1R1 | RA |
| | | | Tofacitinib | JAK kinases | RA |
| rs13389408 | GLS | **IIM**, SLE, SSc | Azathioprine | PPAT | RA, SLE |
| rs13101828 | DGKQ | **IIM, RA, SLE, SSc** | Orlistat | LIPF | -- |
| | | | Nintedanib | PDGFRB | SSc |
| | | | Dasatinib | BLK | -- |
| | | | Imatinib | ABL1 | -- |
| rs2736337 | FAM167A-BLK | **IIM**, RA, SLE, SSc | Osimertinib | EGFR | -- |
| | | | Vandetanib | EPHA1 | -- |
| | | | Fingolimod | S1PR1 | -- |
| | | | Bosutinib | SRC | -- |
| | | | Tofacitinib | JAK kinases | RA |
| | | | Tocilizumab | IL6R | RA |
| rs11085725 | TYK2 | **IIM**, SLE, SSc | Interferon Apha-2B | IFNAR1 | -- |
| | | | Idelalisib | PIK3CD | -- |
| | | | Ruxolitinib | JAK1 | -- |

[a]Based on our meta-analysis, diseases contributing to the observed association. The diseases where the association of this variant has never been reported before at genome-wide significance level are shown in boldface.
[b]Drugs from the OpenTarget platform with their corresponding target.
[c]Current indication of the reported drug. Non-immune mediated diseases were omitted.
SSc: Systemic sclerosis; IIM: Idiopathic inflammatory myopathy; SLE: Systemic lupus erythematosus; RA: Rheumatoid arthritis.

331     As compared to previous cross-phenotype studies of autoimmune diseases, our

332     study has the strength of analyzing systemic seropositive rheumatic diseases, which is a

333     consistent clinical phenotype than in the diseases investigated previously, where mixed

334     seropositive and seronegative diseases were analyzed, and combining systemic and

335     organ-specific diseases.[8, 9] The study of a more homogenous phenotype allowed us to

336     determine that the type I IFN signaling pathway and its regulation play a more

337     prominent role in these conditions than in others, based on the associations observed in

338     *NAB1*, *TYK2*, *PTPN11*, *IRF5*, and *IRF8*. Additionally, we performed a genome-wide

339     scan to identify shared genetic etiologies, as opposed to the study performed by

340     Ellinghaus *et al.* whose analyses were limited to the 186 autoimmune disease-associated

341     *loci* implemented in the Immunochip platform. The study performed by Li *et al.* –which

342     was also a meta-analysis of GWAS data– was focused on pediatric autoimmune

343     diseases, whereas our study was on a new combination of diseases in adult population.

344     In summary, this is the first study to investigate shared common genetic

345     variation in four systemic seropositive rheumatic IMIDs in adults. We identified 26

346     genome-wide significant independent *loci* associated with at least two diseases, of

347     which five *loci* had not been reported before. The shared risk variants and their likely

348     target genes are functionally enriched in relevant immune cells and significantly

349     enriched in drug targets, indicating that it may assist drug repositioning among

350     genetically related diseases based on genomics data.

351

## Competing interests

The authors declare no competing interests.

## Contributorship

Data providers: F.W.M., W.C., T.P.O, R.G.C., J.V., L.G.R., K.D., L.R.W., I.E.L., L. M.P., A.M.R., S.R.Y., A.S.O., T.R.R., D.A.I., H.C., W.E.R.O., P.S., B.P, A.L., J.A.L., C.I.A., C.D., D.H.J., P.P., H.V., on behalf of the Myositis Genetics Consortium; O.G., B.R., J.E.M., B.Z.A., R.P.M., M.J.C., M.C.V., A.E.V., A.J.S., J.C.B, P.L.C.M.R., R.S., A.I., R.A.O., G.R., N.H., C.P.S., N.O.C., M.A.G.G, M.G.E., P.A., J.V.L., A.H., R.H., V.S., F.D.K., F.H., M.M.C., R.M., P.S., R.W., A.K., H.K., E.D.B, T.W., L.K., L.B., R.S., B.A.L., A.M.H.V., P.C., J.V., M.H., A.T.L., J.Y., Y.H., S.F.W., F.M.W., L.H., J.L.N., S.K.A, S.A., P.G., F.K.T., B.P.C.K., F.C.A, on behalf of the Scleroderma Genetics Consortium; QC and imputation in the contributing studies: M.A.H., M.K., D.G.S.; Functional and drug enrichment analysis: M.A.H., M.K.; Meta-analysis, tables and figures: M.A.H., M.K., D.G.S.; Drafting and approved version of the manuscript: M.A.H., M.K., D.G.S., C.W., A.F., P.K.G., L.P., J.W., T.V., M.E.A.R., M.D.M., J.M., F.W.M., W.C., T.P.O, R.G.C., J.V., L.G.R., K.D., L.R.W., I.E.L., L. M.P., A.M.R., S.R.Y., A.S.O., T.R.R., D.A.I., H.C., W.E.R.O., P.S., B.P, A.L., J.A.L., C.I.A., C.D., D.H.J., P.P., H.V.; Study design and management: M.A.H., M.K., J.M.

## Acknowledgements

376 would like to thank all the patients and healthy controls for their essential collaboration

377 and the MYOGEN consortium for the provision of data. The results reported in this

378 study were presented in the Genomics of Common Diseases 2017 conference and in

379 EULAR 2018 (Acosta-Herrera M, Kerick M, González-Serna D *et al.* Ann Rheum Dis

380 2018; 77 Supp 2).

381

390

391 **Ethical approval information**

392 This study was conducted using available data included in previously published GWAS

393 (Supplementary references 1-6).

394

395 **Data availability**

396 Summary statistics of the global meta-analysis generated and analyzed in the current

397 study are available from the corresponding author on reasonable request.

398

399 **Key messages**

400 - Systemic sclerosis, systemic lupus erythematosus, rheumatoid arthritis and idiopathic

401 inflammatory myopathies are systemic seropositive rheumatic diseases that share

402 symptoms, progressions, environmental risk factors, high rates of familial aggregation,

403 and susceptibility genes, pointing to a shared genetic architecture.

404 - The assessment of a shared genetic component among these conditions has not been

405 performed before in a systematic fashion.

406 - We have identified five new shared *loci* among systemic seropositive rheumatic

407 immune-mediated inflammatory diseases. The rest of the observed associations

408 constitute firm susceptibility genes in autoimmunity, providing validity to our findings.

409 - The associated variants are enriched in marks related to gene activation in immune

410 cells and constitute shared expression quantitative trait *loci*.

411 - For most of these diseases there are no specific treatments, therefore, therapy

412 repositioning could be possible among genetically related conditions.

413

414

## References

1. Cho JH and Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nat Med* 2015;21:730-738.

2. Zhernakova A, Withoff S, and Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat Rev Endocrinol* 2013;9:646-659.

3. Zhernakova A, van Diemen CC, and Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* 2009;10:43-55.

4. Martin JE, Assassi S, Diaz-Gallo LM, *et al.* A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet* 2013;22:4021-4029.

5. Lopez-Isac E, Martin JE, Assassi S, *et al.* Cross-disease Meta-analysis of Genome-wide Association Studies for Systemic Sclerosis and Rheumatoid Arthritis Reveals IRF4 as a New Common Susceptibility Locus. *Arthritis Rheumatol* 2016;68:2338-2344.

6. Marquez A, Vidal-Bralo L, Rodriguez-Rodriguez L, *et al.* A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis* 2017;76:286-294.

7. Coenen MJ, Trynka G, Heskamp S, *et al.* Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet* 2009;18:4195-4203.

8. Li YR, Li J, Zhao SD, *et al.* Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 2015;21:1018-1027.

440  9.   Ellinghaus D, Jostins L, Spain SL, *et al.* Analysis of five chronic inflammatory

441       diseases identifies 27 new associations and highlights disease-specific patterns at

442       shared loci. *Nat Genet* 2016;48:510-518.

443  10.  Wallace B, Vummidi D, and Khanna D. Management of connective tissue

444       diseases associated interstitial lung disease: a review of the published literature.

445       *Curr Opin Rheumatol* 2016;28:236-245.

446  11.  Stahl EA, Raychaudhuri S, Remmers EF, *et al.* Genome-wide association study

447       meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*

448       2010;42:508-514.

449  12.  Bentham J, Morris DL, Graham DSC, *et al.* Genetic association analyses

450       implicate aberrant regulation of innate and adaptive immunity genes in the

451       pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;47:1457-1464.

452  13.  International Consortium for Systemic Lupus Erythematosus G, Harley JB,

453       Alarcon-Riquelme ME, *et al.* Genome-wide association scan in women with

454       systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK,

455       KIAA1542 and other loci. *Nat Genet* 2008;40:204-210.

456  14.  Dubois PC, Trynka G, Franke L, *et al.* Multiple common variants for celiac

457       disease influencing immune gene expression. *Nat Genet* 2010;42:295-302.

458  15.  Radstake TR, Gorlova O, Rueda B, *et al.* Genome-wide association study of

459       systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet*

460       2010;42:426-429.

461  16.  Miller FW, Chen W, O'Hanlon TP, *et al.* Genome-wide association study

462       identifies HLA 8.1 ancestral haplotype alleles as major genetic risk factors for

463       myositis phenotypes. *Genes Immun* 2015;16:470-480.

464　17.　Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome

465　　　　association and population-based linkage analyses. *Am J Hum Genet*

466　　　　2007;81:559-575.

467　18.　Das S, Forer L, Schonherr S, *et al.* Next-generation genotype imputation service

468　　　　and methods. *Nat Genet* 2016;48:1284-1287.

469　19.　Delaneau O, Coulonges C, and Zagury JF. Shape-IT: new rapid and accurate

470　　　　algorithm for haplotype inference. *BMC Bioinformatics* 2008;9:540.

471　20.　McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976

472　　　　haplotypes for genotype imputation. *Nat Genet* 2016;48:1279-1283.

473　21.　Kang HM. Efficient and parallelizable association container toolbox (EPACTS).

474　　　　http://genome.sph.umich.edu/wiki/EPACTS

475　22.　Han B and Eskin E. Random-effects model aimed at discovering associations in

476　　　　meta-analysis of genome-wide association studies. *Am J Hum Genet*

477　　　　2011;88:586-598.

478　23.　Yang J, Ferreira T, Morris AP, *et al.* Conditional and joint multiple-SNP

479　　　　analysis of GWAS summary statistics identifies additional variants influencing

480　　　　complex traits. *Nat Genet* 2012;44:369-375, S1-3.

481　24.　Yang J, Lee SH, Goddard ME, *et al.* GCTA: a tool for genome-wide complex

482　　　　trait analysis. *Am J Hum Genet* 2011;88:76-82.

483　25.　Chelala C, Khan A, and Lemoine NR. SNPnexus: a web database for functional

484　　　　annotation of newly discovered and public domain single nucleotide

485　　　　polymorphisms. *Bioinformatics* 2009;25:655-661.

486　26.　Yu W, Clyne M, Khoury MJ, *et al.* Phenopedia and Genopedia: disease-centered

487　　　　and gene-centered views of the evolving knowledge of human genetic

488　　　　associations. *Bioinformatics* 2010;26:145-146.

489     27.     Iotchkova V, Graham RSR, Geihs M, *et al*. GARFIELD - GWAS Analysis of

490           Regulatory or Functional Information Enrichment with LD correction. 2016;

491           doi: https://doi.org/10.1101/085738.

492     28.     Ward LD and Kellis M. HaploReg v4: systematic mining of putative causal

493           variants, cell types, regulators and target genes for human complex traits and

494           disease. *Nucleic Acids Res* 2016;44:D877-881.

495     29.     GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx)

496           pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648-

497           660.

498     30.     Schofield EC, Carver T, Achuthan P, *et al*. CHiCP: a web-based tool for the

499           integrative and interactive visualization of promoter capture Hi-C datasets.

500           *Bioinformatics* 2016;32:2511-2513.

501     31.     Szklarczyk D, Franceschini A, Wyder S, *et al*. STRING v10: protein-protein

502           interaction networks, integrated over the tree of life. *Nucleic Acids Res*

503           2015;43:D447-452.

504     32.     Koscielny G, An P, Carvalho-Silva D, *et al*. Open Targets: a platform for

505           therapeutic target identification and validation. *Nucleic Acids Res*

506           2017;45:D985-D994.

507     33.     Mells GF, Floyd JA, Morley KI, *et al*. Genome-wide association study identifies

508           12 new susceptibility loci for primary biliary cirrhosis. *Nat Genet* 2011;43:329-

509           332.

510     34.     Cordell HJ, Han Y, Mells GF, *et al*. International genome-wide meta-analysis

511           identifies new primary biliary cirrhosis risk loci and targetable pathogenic

512           pathways. *Nat Commun* 2015;6:8019.

513    35.    Senapati S, Singh S, Das M, *et al.* Genome-wide analysis of methotrexate

514           pharmacogenomics in rheumatoid arthritis shows multiple novel risk variants

515           and leads for TYMS regulation. *Pharmacogenet Genomics* 2014;24:211-219.

516    36.    Dustin ML and Cooper JA. The immunological synapse and the actin

517           cytoskeleton: molecular hardware for T cell signaling. *Nat Immunol* 2000;1:23-

518           29.

519    37.    de Vries PS, Chasman DI, Sabater-Lleal M, *et al.* A meta-analysis of 120 246

520           individuals identifies 18 new loci for fibrinogen concentration. *Hum Mol Genet*

521           2016;25:358-370.

522    38.    Davalos D and Akassoglou K. Fibrinogen as a key regulator of inflammation in

523           disease. *Semin Immunopathol* 2012;34:43-62.

524    39.    Iotchkova V, Huang J, Morris JA, *et al.* Discovery and refinement of genetic

525           loci associated with cardiometabolic risk using dense imputation maps. *Nat*

526           *Genet* 2016;48:1303-1312.

527    40.    Araki Y and Mimura T. The Histone Modification Code in the Pathogenesis of

528           Autoimmune Diseases. *Mediators Inflamm* 2017;2017:2608605.

529    41.    Nelson MR, Tipney H, Painter JL, *et al.* The support of human genetic evidence

530           for approved drug indications. *Nat Genet* 2015;47:856-860.

531    42.    Kearney SJ, Delgado C, Eshleman EM, *et al.* Type I IFNs downregulate

532           myeloid cell IFN-gamma receptor by inducing recruitment of an early growth

533           response 3/NGFI-A binding protein 1 complex that silences ifngr1 transcription.

534           *J Immunol* 2013;191:3384-3392.

535    43.    Bossini-Castillo L, Lopez-Isac E, and Martin J. Immunogenetics of systemic

536           sclerosis: Defining heritability, functional variants and shared-autoimmunity

537           pathways. *J Autoimmun* 2015;64:53-65.

538    44.    Bezalel S, Guri KM, Elbirt D, *et al.* Type I interferon signature in systemic
539          lupus erythematosus. *Isr Med Assoc J* 2014;16:246-249.

540    45.    Wright HL, Thomas HB, Moots RJ, *et al.* Interferon gene expression signature
541          in rheumatoid arthritis neutrophils correlates with a good response to TNFi
542          therapy. *Rheumatology (Oxford)* 2015;54:188-193.

543    46.    Lundberg IE and Helmers SB. The type I interferon system in idiopathic
544          inflammatory myopathies. *Autoimmunity* 2010;43:239-243.

545    47.    van Baal J, de Widt J, Divecha N, *et al.* Diacylglycerol kinase theta counteracts
546          protein kinase C-mediated inactivation of the EGF receptor. *Int J Biochem Cell*
547          *Biol* 2012;44:1791-1799.

548    48.    Overbeek MJ, Boonstra A, Voskuyl AE, *et al.* Platelet-derived growth factor
549          receptor-beta and epidermal growth factor receptor in pulmonary vasculature of
550          systemic sclerosis-associated pulmonary arterial hypertension versus idiopathic
551          pulmonary arterial hypertension and pulmonary veno-occlusive disease: a case-
552          control study. *Arthritis Res Ther* 2011;13:R61.

553    49.    Lessard CJ, Li H, Adrianto I, *et al.* Variants at multiple loci implicated in both
554          innate and adaptive immune responses are associated with Sjogren's syndrome.
555          *Nat Genet* 2013;45:1284-1292.

556    50.    Paul P, van den Hoorn T, Jongsma ML, *et al.* A Genome-wide multidimensional
557          RNAi screen reveals pathways controlling MHC class II antigen presentation.
558          *Cell* 2011;145:268-283.

559    51.    Emerging Risk Factors Collaboration, Kaptoge S, Di Angelantonio E, *et al.* C-
560          reactive protein, fibrinogen, and cardiovascular disease prediction. *N Engl J Med*
561          2012;367:1310-1320.

562    52.    Krieg T, Abraham D, and Lafyatis R. Fibrosis in connective tissue disease: the

563           role of the myofibroblast and fibroblast-epithelial cell interactions. *Arthritis Res*

564           *Ther* 2007;9 Suppl 2:S4.

565    53.    Lopez-Mejias R, Castaneda S, Gonzalez-Juanatey C, *et al.* Cardiovascular risk

566           assessment in patients with rheumatoid arthritis: The relevance of clinical,

567           genetic and serological markers. *Autoimmun Rev* 2016;15:1013-1030.

568    54.    Tselios K and Urowitz MB. Cardiovascular and Pulmonary Manifestations of

569           Systemic Lupus Erythematosus. *Curr Rheumatol Rev* 2017;13:206-218.

570    55.    Lamouille S, Xu J, and Derynck R. Molecular mechanisms of epithelial-

571           mesenchymal transition. *Nat Rev Mol Cell Biol* 2014;15:178-196.

572

**Figure Titles and Legends**

**Figure 1. Meta-analysis results for the four systemic immune-mediated inflammatory diseases (IMIDs).** The Manhattan plot displays the -log10 transformed *p*-values (*y-axis*) by position on each chromosome (*x-axis*). The red line depicts the genome-wide significance threshold (*p*-value=$5\times10^{-8}$). A total of 26 SNPs were independently associated with at least two systemic IMIDs. Most of the signals map to known susceptibility *loci* in autoimmunity (e.g. *PTPN22, STAT4, TNPO3, FAM167A-BLK*) and five *loci* have never been reported before.

**Figure 2. GARFIELD functional enrichment analyses in DHS hotspots**. The wheel plot shows functional enrichment in systemic IMIDs within DHS hotspot regions in ENCODE and Roadmap Epigenomics. The radial axis depicts the fold enrichment (FE) calculated at different meta-analysis *p*-value thresholds. The font size is proportional to the number of cell types from the tissue, mainly enriched in blood cell types including a repertoire of immune cell lines.

**Supplementary Figures**

**Figure S1**. Distribution of the observed and expected association *p*-values in each individual study that contributed to the meta-analysis. Quantile-Quantile (QQ) plots from: A) Systemic sclerosis case-control collections. B) Systemic Lupus Erythematosus. C) Rheumatoid Arthritis and D) Idiopathic Inflammatory Myopathies.

**Figure S2.** Non-Conditioned and conditioned analysis on the top associated variants from the meta-analysis. In panels where significant variants remained after conditioning, there are several independent variants in the region. In panels E, P, and Y the remaining independent variants were not significant in the meta-analysis.

**Figure S3.** Wheel plots from the functional enrichment analysis with GARFIELD at different thresholds of p-values from the meta-analysis. Functional categories from the ENCODE project and Roadmap Epigenomics.

**Figure S4.** Circular view of the interactions from the new shared risk SNPs with genes nearby obtained from Promoter Capture HiC data in relevant immune cell types. Interactions are displayed as connecting lines depending on the confidence of the interaction. Grey lines are below threshold in the tissue. Only genes with maximum interaction score are reported.