Meta-analyses uncover the genetic architecture of Idiopathic Inflammatory Myopathies

Catherine Zhu M.S.[1], Younghun Han Ph.D.[1], Jinyoung Byun Ph.D. [1], Xiangjun Xiao Ph.D. [1], Simon Rothwell Ph.D. [2], Frederick W. Miller M.D., Ph.D.[3], Ingrid E. Lundberg M.D., Ph.D.[4], Peter K. Gregersen M.D.[5], Jiri Vencovsky M.D., DSc[6], Vikram R. Shaw M.S.[1], Neil McHugh M.D.[7], Vidya Limaye Ph.D.[8], Albert Selva-O'Callaghan M.D., Ph.D.[9], Michael G. Hanna M.D.[10], Pedro M. Machado M.D., Ph.D.[10,11], Lauren M. Pachman M.D.[12], Ann M. Reed M.D.[13], Lisa G. Rider M.D.[3], Øyvind Molberg M.D.[14], Olivier Benveniste M.D.[15], Timothy Radstake M.D., Ph.D.[16], Andrea Doria M.D.[17], Jan L. De Bleecker M.D., Ph.D.[18], Boel De Paepe Ph.D.[18], Britta Maurer M.D.[19], William E. Ollier Ph.D.[20], Leonid Padyukov M.D., Ph.D.[4], Lucy R. Wedderburn Ph.D.[21], Hector Chinoy M.D., Ph.D.[22], Janine A. Lamb Ph.D.[23], Christopher I. Amos Ph.D. [1,24,25]; Myositis Genetics Consortium (MYOGEN)

[1] Institute for Clinical and Translational Research, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 USA

[2] Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

[3] Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health, Bethesda, Maryland.

[4] Division of Rheumatology, Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

[5] The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute, Manhasset, New York.

[6] Institute of Rheumatology and Department of Rheumatology, First Medical Faculty, Charles University, Prague, Czech Republic.

[7] Department of Life Sciences, University of Bath, Bath, UK

[8] Rheumatology Unit, Royal Adelaide Hospital and Discipline of Medicine, University of Adelaide, Adelaide, Australia.

[9] Internal Medicine Department, Vall d'Hebron General Hospital, Universitat Autonoma de Barcelona, Barcelona, Spain.

[10] Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK.

[11] Centre for Rheumatology, UCL Division of Medicine, University College London, London, UK.

[12] Division of Pediatric Rheumatology, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

[13] Department of Pediatrics, Duke University, Durham, North Carolina.

[14] Department of Rheumatology, Oslo University Hospital, Oslo, Norway.

[15] Sorbonne Université, Assistance Publique Hôpitaux de Paris, Myology Research Center UMR974, Pitié-Salpêtrière Hospital, Department of Internal Medicine and Clinical Immunology, Paris, France.

[16] Department of Rheumatology and Clinical Immunology, University Medical Center, Utrecht, the Netherlands.

[17] Rheumatology Unit, Department of Medicine, University of Padova, Padova, Italy.

[18] Department of Neurology, Ghent University, Ghent, Belgium.

[19] Department of Rheumatology and Immunology, University Hospital, Bern, Switzerland.

[20] Manchester Metropolitan University, School of Healthcare Sciences, Manchester, UK.

[21] NIHR Biomedical Research Centre at Great Ormond Street Hospital, Centre for Adolescent Rheumatology Versus Arthritis, and UCL Great Ormond Street Institute of Child Health, University College London, London, UK.

[22] National Institute for Health Research Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, The University of Manchester, Manchester, UK, and Department of Rheumatology, Salford Royal Hospital, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, Salford, UK, and Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.

[23] Epidemiology and Public Health Group, Division of Population Health, Health Services Research & Primary Care, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.

[24] Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

[25] Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX USA

Funding acknowledgment: This project was partially supported by a TPEHS Training Program fellowship T32 ES027801. This research was supported in part by the Intramural Research Program of the National Institute of Environmental Health Sciences (NIH). This research was supported by the NIHR Manchester Biomedical Research Centre (NIHR203308), the NIHR GOSH Biomedical Research Centre, Myositis UK, Great Ormond Street Children's Charity, Versus Arthritis (18474, 20380, 20164, 21593), and Medical Research Council (MR/n003322/1). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Dr. Ingrid E. Lundberg is supported by the Swedish Research Council (2020-01378), Region Stockholm (ALF project), the Swedish Rheumatism Association, King Gustaf V 80 Year Foundation. Dr. Jiri Vencovsky receives support from the Czech Ministry of Health for Conceptual Development of Research Organization 00023728 (Institute of Rheumatology). Dr. Younghun Han, Dr. Jinyoung Byun, Dr. Christopher I. Amos, and Dr. Lucy R Wedderburn are supported by The Cure JM Foundation.

Abstract

Objective: Idiopathic inflammatory myopathies (myositis, IIMs) are rare, systemic autoimmune disorders that lead to muscle inflammation, weakness, and extra-muscular manifestations, with a strong genetic component influencing disease development and progression. Previous genome-wide association studies identified loci associated with IIMs. In this study, we imputed data from two prior genome-wide myositis studies and analyzed the largest myositis dataset to date to identify novel risk loci and susceptibility genes associated with IIMs and its clinical subtypes.

Methods: We performed association analyses on 14,903 individuals (3,206 cases and 11,697 controls) with genotypes and imputed data from the Trans-Omics for Precision Medicine (TOPMed) reference panel. Fine-mapping and expression quantitative trait locus co-localization analyses in myositis-relevant tissues indicated potential causal variants. Functional annotation and network analyses using the random walk with restart (RWR) algorithm explored underlying genetic networks and drug repurposing opportunities.

Results: Our analyses identified novel risk loci and susceptibility genes, such as *FCRLA, NFKB1*, *IRF4*, *DCAKD*, and *ATXN2* in overall IIMs; *NEMP2* in polymyositis; *ACBC11* in dermatomyositis; and *PSD3* in myositis with anti-histidyl-tRNA synthetase autoantibodies (anti-Jo1). We also characterized effects of HLA region variants and the role of *C4*. Colocalization analyses suggested putative causal variants in *DCAKD* in skin and muscle, *HCP5* in lung, and *IRF4* in EBV-transformed lymphocytes, lung, and whole blood. RWR further prioritized additional candidate genes, including *APP, CD74*, *CIITA, NR1H4,* and *TXNIP*, for future investigation.

Conclusion: Our study uncovers novel genetic regions contributing to IIMs, advancing our understanding of myositis pathogenesis and offering new insights for future research.

Introduction

Idiopathic inflammatory myopathies (IIMs), characterized by chronic muscle weakness and muscle inflammation, are rare, heterogeneous autoimmune diseases[1,2]. The underlying pathogenesis of myositis involves genetic components, which increase susceptibility to environmental insults and can confer an elevated risk of the disease[3].

Genome-wide association studies (GWAS) have emerged as powerful tools for elucidating the genetic basis of complex conditions, including myositis. Recent GWAS investigations have highlighted the involvement of genetic variants within the human leukocyte antigen (HLA) region, contributing to the risk of myositis[4]. Several alleles included in the 8.1 extended haplotype, which encompasses loci in class I, II, and III regions, have been associated not only with increased susceptibility to IIMs but also with other autoimmune disorders such as rheumatoid arthritis, systemic lupus erythematosus, and Hashimoto's thyroiditis in populations of European ancestry. Furthermore, variants in non-HLA loci, such as *PTPN22*, *STAT4*[5]*, SDK2*, *LINC00924*, *NAB1*[6], and *C4A* deficiency[7,8] have been implicated in the pathogenesis of myositis; however, the lower number of cases and controls available in previous studies has limited novel discoveries

and our ability to gain a comprehensive understanding of the genetic architecture of myositis.

To address these limitations and enhance our understanding of the underlying pathogenesis of myositis, we conducted meta-analyses, which combined two studies providing a total of 14,903 individuals of European descent, including 3,206 myositis cases and 11,697 healthy controls. By integrating data from multiple cohorts, we confirmed known signals and discovered ten novel associations for myositis and its clinical subtypes. Additional analysis with the complement *C4* system provided insights into the relative impacts of *HLA* and *C4* genes to disease susceptibility. Fine-mapping and expression quantitative trait locus (eQTL) co-localization suggested specific disease-associated genetic variants within the identified risk loci, highlighting their role in gene expression modulation in myositis-related tissues. We also constructed multiplex networks based on biological pathways of susceptibility markers to explore potential risk candidate genes by using a random walk with restart (RWR) algorithm, which provided insights for hypothesis generation in future myositis research and offered potential avenues for targeted therapies and precision medicine approaches.

Materials and Methods

Study Populations

Samples from the Immunochip dataset[5] (7,486 control, 2,688 cases) and an earlier GWAS dataset[4] (4,712 controls, 1,710 cases) were obtained from the Myositis Genetics Consortium (MYOGEN)[4–6]. Subtypes of cases included in the overall analyses of the IIM group (total IIMs) contained cases with polymyositis (PM), dermatomyositis (DM), juvenile

polymyositis (JPM), juvenile dermatomyositis (JDM), anti-synthetase syndrome (ASyS), inclusion body myositis (IBM), and necrotizing myopathy (NM). Cases were selected based on the classification criteria in the previous studies[4–8]. Clinical subtype analyses were performed on PM, DM, JDM, and myositis with anti-histidyl-tRNA synthetase autoantibodies (anti-Jo1), that comprised the largest numbers of subsets of cases participants.

Genotyping, Quality Control, and Imputation

For each dataset, samples without clinical information or misdiagnoses were excluded. The genotyping data were converted to GRCh38 positions using UCSC's liftOver tool. Chromosome X was excluded from the analyses due to its absence in the GWAS dataset. Single-nucleotide polymorphisms (SNPs) with a call rate < 95% were removed. Individuals with > 5% missing genotypes were excluded. Variants deviating from Hardy-Weinberg equilibrium (HWE) in controls ($p<10^{-6}$) and in cases ($p<10^{-10}$) were removed before imputation. Relatedness was checked using the KING-robust kinship estimator[9]. Duplicates or related individuals (kinship>0.12) within each dataset were excluded to reduce correlations among participants. Principal components analysis (PCA) in the GWAS dataset was performed using PLINK 1.9[10]. In the Immunochip dataset, PCA was calculated using Ancestry Inference using Principal component analysis and Spatial analysis (AIPS)[11] based on the HapMap III reference panel (Supplementary Information). Outliers identified based on the PCA were removed from further analyses (Figure S1, S2). In the meta-analyses, for closely related individuals between datasets (kinship>0.15) and any duplicates between the two datasets, the data from the Immunochip dataset was retained (Figure S1).

To expand the coverage of our investigation, genotypes from the arrays were imputed separately against the Trans-Omics for Precision Medicine (TOPMed) reference panel[12]. SNPs located within 28-34Mb on chromosome 6 were selected for HLA-imputation against the Multi-ethnic HLA reference panel (version 2.0 2022)[13]. Variants with imputation quality $r^2<0.6$, minor allele frequency (MAF) < 0.005 , and post-imputation HWE $p<10^{-6}$ in controls were excluded from further analyses. C4 alleles were imputed

against the European population from HapMap III, according to the imputec4 protocol[14]. Supplementary Information provides further detail on data processing.

Statistical Analyses

Association analyses were conducted for each dataset on the genotyped and imputed data from the total myositis samples (n=3,206 cases), DM (n=1,131 cases), JDM (n=645 cases), PM (n=1,094 cases) subtypes, and myositis with anti-Jo1 (n=388 cases) using SNPTEST 2.5.6[15] (-method expected), with adjustments for population variation based on the PCA analyses as described in the Supplementary Information. Inverse-variance fixed-effects meta-analyses and sex-stratified analyses for the total IIMs and subtypes were then performed using METASOFT[16]. To identify independent and secondary signals, we implemented stepwise conditional and joint association analyses (COJO) using GCTA v1.94[17]. Interaction analyses were performed to assess potential differences by sex and across subtypes, including PM, DM, and JDM. In the HLA region, meta-analyses were carried out on the HLA-imputed data. Additionally, C4 association analysis, along with conditional and joint analyses of both C4 and HLA imputed data, were performed using R. Model details and parameters are provided in the Supplementary Information.

Fine-mapping and Co-localization Analyses

susieR is a Bayesian approach that evaluates multiple causal signals in a region simultaneously based on the Sum of Single Effects (SuSiE) regression framework. We identified 95% credible sets within the non-HLA region and 99% credible sets for the HLA region, and subsequently computed the posterior inclusion probability (SNP.PIP) of causality. The co-localization analyses of meta-analyzed data and eQTL studies were

performed using the coloc 5.2.1 R package[18]. We selected eQTL studies of five myositis-associated tissue types from the Genotype-Tissue Expression (GTEx) v8[19], including EBV-transformed lymphocytes (n=147), skin without sun exposure (n=517), whole blood (n=670), lung (n = 515), and skeletal muscle (n=706) for 32 myositis associated SNP signals. Specifically, we studied 60 analyses for the total IIM group, 25 for the PM group, 35 for the DM group, 15 for the JDM group, and 25 for the anti-Jo1 group. Co-localization in specific tissue was considered when the posterior probability of shared causal variants in meta-analyzed myositis data and tissue-specific eQTL (PP.H4) exceeded 80%, with the identified variants residing in 90% credible sets. Details of the fine-mapping and co-localization analyses are described in the Supplementary Information.

Annotation and Enrichment Analyses

Significantly associated variants were annotated using ANNOVAR[20] based on the GRCh38 UCSC refGene and FAVOR[21]. For intergenic variants, their corresponding nearest up and/or downstream genes were reported. The RegulomeDB 2.2[22,23] was utilized to annotate variants and assess potential regulatory impact. This database integrates information such as histone-sequencing, DNase hypersensitivity footprints, transcription factor ChIP-seq, chromatin accessibility, and position weight matrix information (motif), to assign probability scores and rank the variants based on their likelihood of regulatory significance. The Gene Set Analysis Toolkit (WebGestaltR)[24] was used to perform functional enrichment. We report the results with FDR<0.05.

Random Walk with Restart on Multiplex Networks

Random Walk with Restart (RWR) Algorithm

Network propagation has been widely applied in genetic analyses to prioritize potential genes based on biological networks of risk markers[25]. RWR is a network propagation algorithm, in which an iterative stimulate starts from seed nodes and transverse the network with a probability to move to adjacent nodes or back to the seed nodes. Upon reaching convergence, nodes are ranked based on their level of connection to the seed nodes (rank scores). Nodes with higher rank scores indicate stronger connections, suggesting their importance in the disease-related networks. In our study, genes represent nodes, and biological interactions denote edges. We adapted the RWR algorithm on multiplex networks[26] with the modifications detailed in the Supplementary Information. Disease genes reached suggestive significant threshold ($p=1\times10^{-5}$) from the COJO analyses served as seed nodes, with *-log10*(p) as the initial score. We used the default settings for the transition probability δ, the restart probability 1-α, and the restart probability of a given layer t. Multiplex network patterns were evaluated by permutation test. We reported the top 50 candidates with $p<0.05$.

Construction of Multiplex Network

For the total IIMs and each clinical subgroup, we built their corresponding multiplex networks. Each layer of the multiplex network was composed of genes or proteins from the protein-protein interaction network (PPI), pathways from the KEGG databases, and co-expression RNA seq data. The multiplex network details are available in the Supplementary Information.

Data and Code availability

The data, housed at Baylor College of Medicine, are part of the MYOGEN consortium and are available for collaboration through the network. Software, packages, and code utilized for the study can be found in the Supplementary Information.

Results

The heritability of IIMs was estimated at 25% excluding the MHC region and 59% including it, based on meta-analyzed GWAS and Immunochip data. This indicates a significant genetic component for myositis. To identify genetic risk factors, we performed analyses on total myositis and clinical subtypes, as detailed in Figure 1.

Meta-analyses identified novel risk loci in myositis and subtypes

We conducted meta-analyses for the total IIMs and its subtypes to identify genome-wide significant signals. Many significant signals within risk loci are highly correlated due to linkage disequilibrium (LD), which complicates causal inference and functional interpretation. To elucidate the genetic pattern between correlated signals and distinguish independent variants that contribute to myositis, we employed a stepwise model that executed iterative conditional and joint analyses. This approach captured the combined effects of multiple conditionally independent variants within a risk locus and allowed discoveries of additional novel variants (detail in Supplementary Information).

Our analyses showed that the majority of genome-wide significant signals ($p<5\times10^{-8}$) are within the HLA region, confirming that HLA alleles are the strongest genetic risk factors

for overall IIM and its subtypes. In addition, we identified several significant associations outside of the HLA region (Table1, Figure 2, Table S1).

In the total IIMs (Figure 2A), an intergenic variant near *HLA-DRB1*/*HLA-DQA1* (rs535777, $p = 3.78\times10^{-106}$, OR = 2.327) was the most significant signal within the HLA (Figure S3A, Figure S4G). Outside of the HLA region, we discovered novel variants within the total IIMs group and its subtypes and confirmed previously reported associations. *STAT4* (rs4853540, $p = 5.93\times10^{-9}$, OR = 0.808) showed higher significance in our analysis compared to previous studies (Figure 3B, Figure S4B). We also identified a novel intronic locus in *NFKB1* (rs230514, $p = 3.86\times10^{-8}$, OR = 1.185) (Figure 3C, Figure S4C). An independent signal rs12203592 in *IRF4* outside the HLA region on chromosome 6 was significant for total IIMs ($p = 8.41\times10^{-18}$, OR = 1.439) (Figure 3D, Figure S4D). Several novel variants in the non-coding region showed genome-wide significant associations with myositis, including signals near *GJA1* and *HSF2* (rs7754730, $p = 1.47\times10^{-9}$, OR = 1.189), *PINX1* (rs113538396, $p = 1.57\times10^{-10}$, OR = 3.081), *ATXN2* (rs35350651, $p = 3.30\times10^{-9}$, OR = 0.843), and *DCAKD* (rs9898793, $p = 1.65\times10^{-9}$, OR = 1.252) (Figure 3E-H, Figure S4I-L). Near *FCRLA*, a novel intergenic locus (rs6668534, $p = 5.39\times10^{-8}$, OR = 0.826)(Figure 3A, Figure S4A) was suggestively associated with total IIMs.

In analyses stratified by clinical subtypes, we identified some subset-specific and novel findings. In the PM subgroup (Figure 2B), we confirmed a previously identified risk variant of *PTPN22* (rs2476601, $p = 5.70\times10^{-10}$, OR = 1.503) and identified a novel risk locus in *NEMP2* (rs74925618, $p = 4.52\times10^{-9}$, OR = 1.804) (Figure 4A, Figure S5B). In the MHC

region, the most significant signal was an intergenic variant near *MICA*/*LINC01149* (rs3132473, $p = 6.88\times10^{-68}$, OR = 2.772) (Figure S6, Figure S5D). For DM (Figure 2C), the strongest signal came from *HCP5*/*HCG26* (rs3131617, $p = 1.56\times10^{-44}$, OR =2.306) (Figure S7D, S8A). Non-HLA intronic variants in *ABCB11* (rs145940036, $p = 4.91\times10^{-8}$, OR =3.172) and in *PINX1* (rs113538396, $p = 5.22\times10^{-9}$, OR = 3.868) reached the genome-wide significance threshold in DM (Figure 4B,D, Figure S7A,G). Additionally, rs12203592 in *IRF4* was also significantly associated with both DM ($p = 1.69\times10^{-9}$, OR =1.471) and JDM ($p = 9.58\times10^{-9}$, OR =1.578)(Figure 4C, E).In JDM, the strongest signal observed was near *HLA-DRB5*/*HLA-DRB6* (rs1894553, $p = 2.24\times10^{-18}$, OR =1.894) at the HLA region (Figure 2D, Figure S7K, Figure S9A).

In the anti-Jo1 group, in addition to the strongest signal in *HCP5* (rs3132090, $p = 1.75\times10^{-84}$, OR = 5.643) (Figure 2E, Figure S5F, Figure S10A) located within the HLA region, a non-HLA risk locus in *PSD3* (rs6991531, $p = 5.01\times10^{-9}$, OR = 4.229) was significantly associated with anti-Jo1 myositis (Figure 4F, Figure S5J).

Sex-stratified analysis showed similar effect sizes in females and males, except a slightly stronger effect for *IRF4* in males. Formal analysis of sex interactions did not reveal significant differences in the identified signals between males and females, except for *IRF4* in the DM group, where the effect was also marginally stronger in males (Table S2-3). Pairwise interaction analyses among DM, JDM, and PM suggested that many of the identified variants have distinct effect sizes across subtypes and contribute to subtype-specific disease risk ($|Z_{weighted}| > 1.96$) (T able S 4). The vari ant in *PTPN22* was significant in the interaction test across different subtypes, with a larger effect or stronger

association in PM compared to JDM and adult-onset DM, consistent with previous observations[27]. *IRF4* showed significant differences between JDM and other subtypes, with a higher odds ratio in JDM compared to DM and total IIMs in the meta-analysis. This suggested that *IRF4* may play a more pivotal role in JDM, contributing to the significant interaction observed when comparing JDM with other subtypes. Variants in *ABCB11* and *PINX1* in the interaction tests between DM and JDM were not significant. The lack of significance may be due to low MAF of the variants, which could limit the power to detect associations and subtle differences in variant effects between subtypes.

Within the HLA region, an additional variant in non-coding RNA locus in *TSBP1-AS1* ($p = 2.71\times10^{-40}$, $pC = 9.06\times10^{-9}$, OR = 1.487) was jointly significant within the HLA region in total IIMs (Table S5, Figure S3) after conditioning on top signals. Variants that reached the suggestive threshold ($p = 1\times10^{-5}$) in autosomal regions of total IIMs and subtypes were reported (Table S6).

Investigation of HLA variants and C4 revealed their respective roles as risk factors

To investigate the genetic architecture of myositis in the HLA region, we conducted HLA conditional and joint analyses on the imputed HLA data. These analyses confirmed previously reported signals in the MHC class I and class II regions (Figure S11) and identified novel associations, such as *HLA-DRB1*16:01* ($pC = 1.57\times10^{-8}$) in the DM group after conditioning on the previously reported *HLA-B*08:01* allele. Additional HLA risk loci rs1265764 in *TSBP1-AS1* and rs116312062 in *HLA-DRB6* were also genome-wide significant (Table S7). The list of HLA alleles and HLA amino acids that reached the genome-wide significant threshold is shown in Table S8. We then further studied the MHC

class III region, calculating the mean and standard error of imputed dosages of C4 genes (*C4A*, *C4B*, *C4L*, and *C4S*) in both cases and controls. Cases showed lower levels of *C4A* and *C4L* compared to controls (Table S9). *C4A* and *C4L* were strongly correlated, as were C4B and C4S (Figure S12). Significant associations of *C4A* and *C4L* with myositis were found across total IIMs and subtypes (Table S10). Additionally, conditional and joint analyses of HLA variants and C4 genes suggested that *HLA-DRB1*03:01* remained the strongest HLA risk allele in total IIMs, and *HLA-B*08:01* with rs116312062 were significant in PM and anti-Jo1 groups. After accounting for effects from *C4A* or *C4L*, rs126574 and *HLA-DQA1*05:01* were no longer significant, suggesting these associations are due to LD with *C4A* or *C4L* (Table S11).

Fine-mapping and eQTL co-localization analyses pinpointed credible sets of causal variants within the identified risk loci

Discoveries from the conditional and joint analyses provided significant insights into the involvement of novel genetic risk loci in the underlying pathogenesis of myositis. To gain more perspective into the causal variants within the identified risk loci, we employed a Bayesian fine-mapping approach to identify candidates with potential direct effects on disease phenotypes. The fine-mapping analyses yielded multiple 95% credible sets (99% for the HLA region), with each set containing a distinct group of variants with assigned SNP.PIP of causality (Table S12). The SNP.PIP of each variant in a credible set indicates its likelihood of being causal within a specific risk locus. We reported candidate variant with the highest PIP in each credible set from the fine-mapping analyses (Table S13). Most of the highly probable causal variants were the lead variants detected from the COJO analyses. Fine-mapping identified rs535777 near *HLA-DRB1* as a top potential

variant in causal credible sets for total IIMs and anti-Jo1 in the HLA region. This variant confers a strong regulatory impact in the MHC class II region according to the RegulomeDB annotation (rank = 1b, probability = 1).

To assess the likelihood that genetic variants are causally related to disease phenotypes and gene expression in myositis-related tissues, we performed co-localization analyses and identified credible sets of variants (Table S14). The co-localization posterior probability of rs12203592 in the total IIMs, DM, and JDM indicated plausible shared causality in both the meta analyzed myositis data, and the expression level of *IRF4* in EBV-transformed lymphocytes(PP.H4 = 0.990, SNP.PIP = 1.000), lung (PP.H4 = 1.000, SNP.PIP = 1.000), and whole blood cells (PP.H4 = 1.000, SNP.PIP = 1.000) (Figure 5A). rs12950988 in *DCAKD* was the lead variant in the credible sets of total IIMs co-localization analysis in skeletal muscle (PP.H4 = 0.984, SNP.PIP = 0.928) and skin tissue (PP.H4 = 0.970, SNP.PIP = 0.428) (Figure 5B). We also observed colocalization of eQTLs for *HCP5* in lung tissue with both total myositis and DM. The most likely causal variants associated with these colocalization credible sets are rs3132090 for the *HCP5* eQTL with total myositis (PP.H4 = 0.902, SNP.PIP = 0.442) and rs3131618 for the *HCP5* eQTL with DM (PP.H4 = 0.945, SNP.PIP = 0.651) (Figure 5C).

RWR identified candidate genes and potential drug repurposing targets through biological networks

To provide more comprehensive insights into the genetic architecture of myositis and support future research, we performed network propagation with the RWR algorithm. This

approach prioritizes candidate genes based on the functional connectivity of disease-associated genes within related biological pathways. RWR on multiplex networks that integrate various biological datasets (Table S15) enables robust and comprehensive analyses, capturing the global network topology while preserving individual network properties. Additional candidate genes were identified within each clinical group (Table S16, Figure S13) based on their strength of connectivity to disease genes.

Notably, several genes belonging to nuclear transcription factor Y (NFY), including *NFYA*, *NFYB*, *NFYC*, regulatory factor X (*RFX*), and the transmembrane protein family(TMEM) ranked as top candidates in the overall myositis group. Class II MHC transactivator (*CIITA*), the β-amyloid precursor protein (*APP*), and *CD74* were also the leading candidates across multiple groups. Gene set over-representation analyses of candidate and disease gene sets showed enrichment in immune-related processes and signaling pathways, prioritized for future investigation based on statistical significance (Figure S14). To explore the potential for drug repurposing in myositis, we also investigated if any disease (Table S6) and additional candidate genes (Table S16) from RWR were targets of drugs on the market according to the DrugBank database[28]. Potential treatments are listed in Table S17.

Discussion

In our study, we explored the genetic architecture underlying myositis to gain a deeper insight into the genetic basis of IIMs. Our meta-analyses on 15,350 individuals of

European descent identified novel genetic risk loci within and outside the HLA region in overall IIMs and subtypes.

In the total IIMs group, the odds ratio of minor alleles in *FCRLA*, *STAT4*, and *ATXN2* suggested their protective effects against myositis. rs4853540 in *STAT4*, which targets the IRF7 motif, may enhance transcription activity, as indicated by H3K4me1 and H3K27ac, in both T cells and B cells, suggesting its critical role in modulating immune responses (Figure S15A). rs35350651 within the 3'UTR of *ATXN2* potentially regulates the highly conserved DNA-binding motif in AT-Rich Interaction Domain (ARID) subfamily *ARID3A*, which is associated with autoimmune diseases[29,30] and neuromuscular diseases, including diseases with similarities to IBM (e.g. amyotrophic lateral sclerosis (ALS))[31], possibly through histone modification. (Figure S15B). Future studies are needed to further understand mechanisms by which these variants may influence risk for IIMs development.

Through conditional and joint analyses, rs12203592 in *IRF4*, genotyped in the Immunochip dataset, was significant in the association study of Immunochip dataset and the meta-analyses of total, DM, and JDM groups, but not in the GWAS, where it was imputed. *IRF4* is associated with pigmentation and its allele prevalence varies across European populations[32]. Based on fine-mapping and co-localization analyses, rs12203592 might contribute to both the risk of myositis and gene expression levels in EBV-transformed lymphocytes, lung, and whole blood cells, but further studies in more homogeneous populations would be helpful. The novel genetic locus around *DCAKD*, rs9898793, with high posterior probability and in LD with the additional putative causal variant rs12950988, might modify regulatory elements, influencing *DCAKD* expression in

skeletal muscle and skin tissue. rs3132090 near *HCP5* locus was identified as the top variants in the co-localized credible sets for lung tissue in total myositis, and rs3131618 was the top variant in the co-localized credible sets in DM. Variants in the 3'UTR proximal to the *HCP5* loci have been associated with myositis[33]. The novel variant rs3132090, upstream of the transcription start site of *HCP5*, may influence DNA-binding motifs in the transcription factor activator protein 2 (*TFAP2*) members, including *TFAP2A, TFAP2B,* and *TFAP2C*, potentially modulating the expression of multiple targets (Figure S15C). Fine-mapping analyses also identified rs535777, a putative causal variant with regulatory impacts on nuclear receptor *NR1H3* and sine oculis homeobox 1 (*SIX1*) in various immune cell types, such as CD4+ T cells (Figure S15D), which are related to autoimmune diseases and inflammatory processes[34,35].

When conducting stratified analysis based on clinical subtypes, we revealed distinct risk loci specifically associated with each subtype. In the PM group, rs74925618 in *NEMP2* was significant. *NEMP2* encodes a nuclear envelope integral membrane protein, suggesting a potential role in signaling pathways that might influence PM risk. *PINX1*, associated with total IIMs and DM, has also been linked to other autoimmune diseases in European ancestry[36]. In addition to subtype analyses, we investigated myositis with anti-Jo1, the most common myositis-specific autoantibody. Despite the limited sample size in the anti-Jo1 myositis group, rs3132090 near the *HCP5* loci showed the strongest association among the subtype-specific signals. Outside the HLA region, *PSD3* had boarder odds ratio ranges, likely due to the small anti-Jo1 sample size and low allele frequencies. Larger patient cohorts or combining with other anti-tRNA synthetase

autoantibodies could provide more compelling evidence for these novel signals. Certain subgroups, such as IBM, had insufficient sample sizes to support genetic analyses. Future studies with expanded cohorts are necessary to investigate these important but rarer subtypes. Another limitation of the study is the limited density of genotypes available in the arrays that we studied. In particular, three SNPs had limited support from nearby genotyped SNPs. The *PINX1* variant (rs113538396) was imputed in both datasets, and the *IRF4* (rs12203592) and *GJA1* (rs7754730) were genotyped only on the Immunochip. Further studies that directly genotype additional SNPs around these variants will help to validate these findings.

By exploring the impact of C4 located within the MHC class III region, we showed that cases had fewer copies of *C4A* and *C4L* than controls, with reduced copy numbers significantly associated with total IIMs and its subtypes. These observations are consistent with previous studies linking lower copy number variations of *C4A* and *C4L* to increased risks of myositis and other autoimmune diseases[37,38].

The relative contributions of C4 isotypes and HLA variants as risk factors were also investigated, in addition to the analyses on HLA imputed data. Conditional and joint analysis were performed, with copy number of each C4 isotype included as predictors to assess their impact on HLA associations. The strongest HLA alleles, *HLA-DRB1*03:01* in total IIMs, and *HLA-B*08:01* in PM and anti-Jo1, retained statistical significance, indicating their independence from C4 as risk factors in myositis. rs1265764 in *TSBP1-BTNL2* Antisense RNA 1 (*TSBP1-AS1*), which showed joint significance with other HLA

variants in nearly all subtypes, was not significant after conditioning on C4, suggesting its dependence on C4.

To expand our understanding of the genetic network of myositis, we prioritized additional candidate genes based on the biological networks of risk markers using RWR. Several *RFX* genes, such as *RFX5*, ranked highly across groups. *RFX5* is evolutionary conservative, and its mutation can disrupt HLA expression, leading to immunodeficiency[39]. *CIITA* was also among the top candidates, playing an essential role in regulating transcriptional activity of the HLA class II promoter, and together with *NFY* and *RFX* genes, affecting the function of the immune system[40–42]. *CD74* ranked at the top across multiple groups and has been previously suggested in animal models as a key regulator, binding to migration inhibitory factor, and a potential therapeutic target in alphavirus-induced myositis, in addition to its intracellular role in MHC class II antigen-presenting[43]. Another candidate, *APP*, is associated with neurodegenerative and neuroinflammatory conditions similar to IBM, such as ALS and multiple sclerosis[44].

NR1H4 (*FXR*) was one of the leading candidates in DM. Experiments have shown that overexpression of *NR1H4* inhibits expression of proinflammatory cytokine in inflammatory bowel diseases[45], which are associated with DM and possibly sharing similar immunopathogenesis[46]. *TXNIP*, a top candidate in JDM, interacts with the nucleotide-binding oligomerization domain (NOD)-like receptor protein 3 (*NLRP3*) inflammasome, which has been suggested to associate with myositis[47]. The discoveries from RWR suggest new avenues for exploration and provide a foundation for future research

hypothesis generation. Further investigations are helpful to understand the contribution and underlying pathogenicity of these candidate genes in myositis.

Due to the absence of standardized therapeutic guidelines for the treatment of IIM, the current therapeutic interventions for myositis are mainly guided by expert experience, case reports, and small clinical trials. Medications for myositis primarily consist of drugs that have received approval for other conditions and are utilized "off-label" in myositis[48]. To explore the potential for drug repurposing in myositis, we investigated if any disease and additional candidate genes from RWR were targets of drugs on the market. Current medications for myositis, such as human immunoglobulin and rituximab (combined with glucocorticoids)[48], were confirmed in the analyses (Table S17). Antithymocyte immunoglobulin (rabbit) and valproic acid have also been investigated to treat certain autoimmune conditions[49,50]. Other medications targeting B-lymphocyte antigen CD20, serine/threonine-protein kinase, and dipeptidyl peptidase 4 could also be considered for potential relevance to myositis treatment. While our initial findings suggested possible opportunities for drug repurposing in myositis, further investigations and validations are necessary to provide reliable evidence, ultimately improving patient outcomes.

Overall, our study has identified novel genetic associations, providing valuable insights into the genetic architecture underlying myositis and its clinical subtypes. A limitation of our study is the variable spacing of SNP data across the genome. Because data were derived, in part, from the Immunochip, which focused on loci known to influence autoimmune conditions, there are regions where there are highly significant signals but

for which further studies are needed to identify the most likely locus. Future studies using denser genotyping or whole genome sequencing would help to refine signals in these loci. Nevertheless, integrating results from the current GWAS studies not only contributes to our understanding of disease etiology, but also helps guide future investigations and facilitate the development of more efficient and effective interventions and treatments for these complex autoimmune disorders.

Author contributions: CZ and CIA designed the study. CZ performed the analyses. JB and YH provided data for the eQTL association studies. CZ wrote the manuscript. All authors contributed to the interpretations of the results and provided feedback during manuscript revision.

Acknowledgments:

The authors thank Dr. Michael Ombrello, Dr. Elaine Remmers, and Dr. Sandeep Agarwal for useful comments on the manuscript. Jan L. De Bleecker is a member of the European Reference Network for Neuromuscular Diseases (ERN EURO-NMD). Minimal usage of Chat-GPT and Grammarly for checking spelling and grammar. Citation style was generated using Schiwheel.

References

1. Lundberg IE, de Visser M, Werth VP. Classification of myositis. *Nat Rev Rheumatol*. 2018;14(5):269-278. doi:10.1038/nrrheum.2018.41

2. Rider LG, Miller FW. Deciphering the clinical presentations, pathogenesis, and treatment of the idiopathic inflammatory myopathies. *JAMA*. 2011;305(2):183-190. doi:10.1001/jama.2010.1977

3. Lundberg IE, Fujimoto M, Vencovsky J, et al. Idiopathic inflammatory myopathies. *Nat Rev Dis Primers*. 2021;7(1):86. doi:10.1038/s41572-021-00321-x

4. Miller FW, Chen W, O'Hanlon TP, et al. Genome-wide association study identifies HLA 8.1 ancestral haplotype alleles as major genetic risk factors for myositis phenotypes. *Genes Immun*. 2015;16(7):470-480. doi:10.1038/gene.2015.28

5. Rothwell S, Cooper RG, Lundberg IE, et al. Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. *Ann Rheum Dis*. 2016;75(8):1558-1566. doi:10.1136/annrheumdis-2015-208119

6. Rothwell S, Amos CI, Miller FW, et al. Identification of Novel Associations and Localization of Signals in Idiopathic Inflammatory Myopathies Using Genome-Wide Imputation. *Arthritis Rheumatol*. 2023;75(6):1021-1027. doi:10.1002/art.42434

7. Rose MR, ENMC IBM Working Group. 188th ENMC International Workshop: Inclusion Body Myositis, 2-4 December 2011, Naarden, The Netherlands. *Neuromuscul Disord*. 2013;23(12):1044-1055. doi:10.1016/j.nmd.2013.08.007

8. Hilton-Jones D, Miller A, Parton M, Holton J, Sewry C, Hanna MG. Inclusion body myositis: MRC Centre for Neuromuscular Diseases, IBM workshop, London, 13 June 2008. *Neuromuscul Disord*. 2010;20(2):142-147. doi:10.1016/j.nmd.2009.11.003

9. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559

10. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

11. Byun J, Han Y, Gorlov IP, Busam JA, Seldin MF, Amos CI. Ancestry inference using principal component analysis and spatial analysis: a distance-based analysis to account for population substructure. *BMC Genomics*. 2017;18(1):789. doi:10.1186/s12864-017-4166-8

12. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299. doi:10.1038/s41586-021-03205-y

13. Luo Y, Kanai M, Choi W, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat Genet*. 2021;53(10):1504-1516. doi:10.1038/s41588-021-00935-7

14. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530(7589):177-183. doi:10.1038/nature16549

15. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010;11(7):499-511. doi:10.1038/nrg2796

16. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88(5):586-598. doi:10.1016/j.ajhg.2011.04.014

17. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375, S1. doi:10.1038/ng.2213

18. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet*. 2021;17(9):e1009440. doi:10.1371/journal.pgen.1009440

19. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603

21. Zhou H, Arapoglou T, Li X, et al. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. *Nucleic Acids Res*. 2023;51(D1):D1300-D1311. doi:10.1093/nar/gkac966

22. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797. doi:10.1101/gr.137323.112

23. Dong S, Zhao N, Spragins E, et al. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nat Genet*. 2023;55(5):724-726. doi:10.1038/s41588-023-01365-3

24. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47(W1):W199-W205. doi:10.1093/nar/gkz401

25. Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. *Nat Rev Genet*. 2017;18(9):551-562. doi:10.1038/nrg.2017.38

26. Valdeolivas A, Tichit L, Navarro C, et al. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*. 2019;35(3):497-505. doi:10.1093/bioinformatics/bty637

27. Chinoy H, Platt H, Lamb JA, et al. The protein tyrosine phosphatase N22 gene is associated with juvenile and adult idiopathic inflammatory myopathy independent of the HLA 8.1 haplotype in British Caucasian patients. *Arthritis Rheum*. 2008;58(10):3247-3254. doi:10.1002/art.23900

28. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037

29. Li Y, Li Z, Chen R, et al. A regulatory variant at 19p13.3 is associated with primary biliary cholangitis risk and ARID3A expression. *Nat Commun*. 2023;14(1):1732. doi:10.1038/s41467-023-37213-5

30. Garton J, Barron MD, Ratliff ML, Webb CF. New frontiers: arid3a in SLE. *Cells*. 2019;8(10). doi:10.3390/cells8101136

31. Van Daele SH, Moisse M, van Vugt JJFA, et al. Genetic variability in sporadic amyotrophic lateral sclerosis. *Brain*. 2023;146(9):3760-3769. doi:10.1093/brain/awad120

32. Praetorius C, Grill C, Stacey SN, et al. A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell*. 2013;155(5):1022-1033. doi:10.1016/j.cell.2013.10.022

33. Kulski JK. Long noncoding RNA HCP5, a hybrid HLA class I endogenous retroviral gene: structure, expression, and disease associations. *Cells*. 2019;8(5). doi:10.3390/cells8050480

34. Wang Z, Sadovnick AD, Traboulsee AL, et al. Nuclear receptor NR1H3 in familial multiple sclerosis. *Neuron*. 2016;90(5):948-954. doi:10.1016/j.neuron.2016.04.039

35. Zhan H, Chen H, Tang Z, Liu S, Xie K, Wang H. SIX1 attenuates inflammation and rheumatoid arthritis by silencing MyD88-dependent TLR1/2 signaling. *Int Immunopharmacol*. 2022;106:108613. doi:10.1016/j.intimp.2022.108613

36. Gorlova OY, Li Y, Gorlov I, et al. Gene-level association analysis of systemic sclerosis: A comparison of African-Americans and White populations. *PLoS ONE*. 2018;13(1):e0189498. doi:10.1371/journal.pone.0189498

37. Zhou D, King EH, Rothwell S, et al. Low copy numbers of complement C4 and C4A deficiency are risk factors for myositis, its subgroups and autoantibodies. *Ann Rheum Dis*. 2023;82(2):235-245. doi:10.1136/ard-2022-222935

38. Lundtoft C, Pucholt P, Martin M, et al. Complement C4 copy number variation is linked to ssa/ro and ssb/la autoantibodies in systemic inflammatory autoimmune diseases. *Arthritis Rheumatol*. 2022;74(8):1440-1450. doi:10.1002/art.42122

39. Sugiaman-Trapman D, Vitezic M, Jouhilahti E-M, et al. Characterization of the human RFX transcription factor family by regulatory and target gene analysis. *BMC Genomics*. 2018;19(1):181. doi:10.1186/s12864-018-4564-6

40. Reith W, Siegrist CA, Durand B, Barras E, Mach B. Function of major histocompatibility complex class II promoters requires cooperative binding between factors RFX and NF-Y. *Proc Natl Acad Sci USA*. 1994;91(2):554-558. doi:10.1073/pnas.91.2.554

41. Ly LL, Yoshida H, Yamaguchi M. Nuclear transcription factor Y and its roles in cellular processes related to human disease. *Am J Cancer Res*. 2013;3(4):339-346.

42. Sachini N, Papamatheakis J. NF-Y and the immune response: Dissecting the complex regulation of MHC genes. *Biochim Biophys Acta Gene Regul Mech*. 2017;1860(5):537-542. doi:10.1016/j.bbagrm.2016.10.013

43. Herrero LJ, Sheng K-C, Jian P, et al. Macrophage migration inhibitory factor receptor CD74 mediates alphavirus-induced arthritis and myositis in murine models of alphavirus infection. *Arthritis Rheum*. 2013;65(10):2724-2736. doi:10.1002/art.38090

44. Nguyen KV. β-Amyloid precursor protein (APP) and the human diseases. *AIMS Neurosci*. 2019;6(4):273-281. doi:10.3934/Neuroscience.2019.4.273

45. Gadaleta RM, van Erpecum KJ, Oldenburg B, et al. Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease. *Gut*. 2011;60(4):463-472. doi:10.1136/gut.2010.212159

46. Sharif K, Ben-Shabat N, Mahagna M, et al. Inflammatory Bowel Diseases Are Associated with Polymyositis and Dermatomyositis-A Retrospective Cohort Analysis. *Medicina (Kaunas)*. 2022;58(12). doi:10.3390/medicina58121727

47. Dubuisson N, Versele R, Davis-López de Carrizosa MA, Selvais CM, Brichard SM, Abou-Samra M. Walking down Skeletal Muscle Lane: From Inflammasome to Disease. *Cells*. 2021;10(11). doi:10.3390/cells10113023

48. Barsotti S, Lundberg IE. Current treatment for myositis. *Curr Treatm Opt Rheumatol*. 2018;4(4):299-315. doi:10.1007/s40674-018-0106-2

49. Lytton SD, Denton CP, Nutzenberger AM. Treatment of autoimmune disease with rabbit anti-T lymphocyte globulin: clinical efficacy and potential mechanisms of action. *Ann N Y Acad Sci*. 2007;1110:285-296. doi:10.1196/annals.1423.030

50. Seet L-F, Toh LZ, Finger SN, Chu SWL, Wong TT. Valproic acid exerts specific cellular and molecular anti-inflammatory effects in post-operative conjunctiva. *J Mol Med*. 2019;97(1):63-75. doi:10.1007/s00109-018-1722-x

Group	Nearest Gene	LeadSNP	CHR:POS	Function	Minor Allele	MAF	Meta b	Meta se	Meta p	Meta OR	bJ	ORJ	pJ
total IIMs	FCRLA(dist=5024)*	rs6668534	1:161702205	intergenic	G	0.234	-0.192	0.035	5.39×10^{-8}	0.826	-0.192	0.035	5.53×10^{-8}
	STAT4	rs4853540	2:191052591	intronic	T	0.222	-0.213	0.037	5.93×10^{-9}	0.808	-0.213	0.808	6.13×10^{-9}
	NFKB1*	rs230514	4:102550782	intronic	G	0.353	0.169	0.031	$3.86 \times 10_{-8}$	1.185	0.169	1.185	3.96×10^{-8}
	IRF4*	rs12203592	6:396321	intronic	T	0.157	0.364	0.042	8.41×10^{-18}	1.439	0.364	1.439	9.97×10^{-18}
	LINC01149(dist=5050);HCP5(dist=11157)	rs2516457	6:31452023	intergenic	A	0.466	-0.428	0.029	1.05×10^{-47}	0.652	-0.272	0.762	6.15×10^{-19}
	TSBP1-AS1	rs1980496	6:32372293	ncRNA_intronic	T	0.398	0.397	0.030	2.71×10^{-40}	1.487	0.201	1.223	5.36×10^{-10}
	HLA-DRB1(dist=20020);HLA-DQA1(dist=27550)	rs535777	6:32609856	intergenic	C	0.154	0.845	0.039	3.78×10^{-106}	2.327	0.691	1.996	2.89×10^{-59}
	HLA-DRB1(dist=37244);HLA-DQA1(dist=10322)	rs147774179	6:32627079	intergenic	G	0.049	0.313	0.060	1.83×10^{-7}	1.367	0.448	1.565	1.50×10^{-13}
	GJA1(dist=225016);HSF2(dist=724808)	rs7754730	6:121674743	intergenic	C	0.385	0.173	0.030	4.68×10^{-9}	1.189	0.173	1.189	4.83×10^{-9}
	PINX1*	rs113538396	8:10777742	intronic	A	0.007	1.125	0.176	1.57×10^{-10}	3.081	1.125	3.081	1.64×10^{-10}
	ATXN2*	rs35350651	12:111469627	intronic	AC	0.485	-0.171	0.029	3.30×10^{-9}	0.843	-0.171	0.843	3.41×10^{-9}
	DCAKD*	rs9898793	17:45038945	intronic	T	0.241	0.225	0.037	1.65×10^{-9}	1.252	0.225	1.252	1.73×10^{-9}
PM	PTPN22	rs2476601	1:113834946	exonic	A	0.101	0.407	0.066	5.70×10^{-10}	1.503	0.407	1.503	5.90×10^{-10}
	NEMP2*	rs74925618	2:190513057	intronic	C	0.066	0.59	0.101	4.52×10^{-9}	1.804	0.59	1.804	4.72×10^{-9}
	PSORS1C3(dist=1826);HCG27(dist=9617)	rs28360059	6:31188143	intergenic	A	0.163	-0.573	0.072	1.63×10^{-15}	0.564	-0.403	0.668	3.03×10^{-8}
	MICA(dist=25237);LINC01149(dist=1115)	rs3132473	6:31440552	intergenic	A	0.132	1.02	0.059	6.88×10^{-68}	2.772	0.893	2.443	6.79×10^{-49}
	HLA-DQB1(dist=2299);HLA-DQA2(dist=72435)	rs3135000	6:32668956	intergenic	A	0.472	0.496	0.046	1.41×10^{-26}	1.642	0.318	1.374	3.50×10^{-11}
DM	ABCB11*	rs145940036	2:169006750	intronic	A	0.007	1.154	0.212	4.91×10^{-8}	3.172	1.154	3.172	5.00×10^{-8}
	IRF4*	rs12203592	6:396321	intronic	T	0.154	0.386	0.064	1.69×10^{-9}	1.471	0.386	1.471	1.77×10^{-9}
	LINC01149(dist=5050);HCP5(dist=11157)	rs2516457	6:31452023	intergenic	A	0.476	-0.432	0.045	1.62×10^{-21}	0.649	-0.286	0.752	1.86×10^{-9}
	HCP5(dist=3152);HCG26(dist=2268)	rs3131617	6:31468961	intergenic	T	0.128	0.835	0.060	1.56×10^{-44}	2.306	0.556	1.743	2.74×10^{-17}
	HLA-DRB5(dist=19601);HLA-DRB6(dist=2816)	rs371760589	6:32549887	intergenic	G	0.017	0.637	0.127	4.95×10^{-7}	1.892	0.765	2.149	1.91×10^{-9}
	HLA-DQB1	rs9274258	6:32663671	intronic	G	0.41	0.567	0.045	1.99×10^{-36}	1.762	0.451	1.57	2.64×10^{-21}
	PINX1*	rs113538396	8:10777742	intronic	A	0.006	1.353	0.232	5.22×10^{-9}	3.868	1.353	3.868	5.35×10^{-9}
JDM	IRF4*	rs12203592	6:396321	intronic	T	0.155	0.456	0.080	9.58×10^{-9}	1.578	0.456	1.578	9.93×10^{-9}
	HLA-DRA(dist=20244);HLA-DRB5(dist=52063)	rs9268926	6:32465290	intergenic	G	0.191	0.273	0.068	6.72×10^{-5}	1.314	0.403	1.496	7.90×10^{-9}
	HLA-DRB5(dist=4802);HLA-DRB6(dist=17624)	rs1894553	6:32535089	intergenic	A	0.131	0.639	0.073	2.24×10^{-18}	1.894	0.721	2.058	3.84×10^{-22}
anti-Jo1	HCP5(dist=205)	rs3132090	6:31462975	upstream	A	0.128	1.73	0.089	1.75×10^{-84}	5.643	1.017	2.765	1.99×10^{-20}
	HLA-DRA(dist=8250);HLA-DRB5(dist=64057)	rs9268791	6:32453296	intergenic	T	0.379	1.056	0.080	2.10×10^{-39}	2.874	0.916	2.5	6.26×10^{-24}
	HLA-DQA1(dist=14581);HLA-DQB1(dist=1205)	rs4713570	6:32658263	intergenic	T	0.258	0.967	0.076	1.65×10^{-37}	2.630	0.541	1.717	3.90×10^{-10}
	HLA-DQB1(dist=753)	rs9273370	6:32658715	downstream	G	0.408	-0.56	0.080	3.27×10^{-12}	0.571	-0.728	0.483	1.94×10^{-16}
	PSD3*	rs6991531	8:18883152	intronic	C	0.012	1.442	0.247	5.01×10^{-9}	4.229	1.442	4.229	5.11×10^{-9}

Table 1. Joint analyses on conditionally independent variants reaching the genome-wide significance threshold in the total IIMs, IIM subtypes and myositis with anti-Jo1 antibodies. CHR:POS. chromosome and position in hg38; MAF, minor allele frequency in combined cases and controls; Meta b, Meta se, Meta p, and Meta OR indicate effect size of minor alleles, standard error of effect size, p-value, and Odds Ratio from the result of meta-analyses; bJ, ORJ, and pJ indicate effect size, Odds Ratio, and p-value from a joint analysis of all the selected SNP. * novel association outside of HLA.

Figure Legends

Figure 1. Overview of the study. Genotypes from GWAS and Immunochip datasets are imputed after quality control. Meta-analyses are performed. Independent lead variants are then identified using multi-SNP-based conditional and joint analyses (COJO) on the results of meta-analyses. We also investigate credible sets of causal variants and prioritize additional candidate genes via Random Walk with Restart (RWR) on multiplex networks. Each multiplex network integrates three different biological networks, including co-expression, KEGG pathway, and protein-protein interaction network (PPI). Leveraging the results of COJO and RWR, we perform gene set enrichment analyses and explore potential drug repurposing opportunities. Figure is created with BioRender.

Figure 2. Manhattan Plots of the total IIMs and subtypes of Myositis. Signals reaching genome-wide significance level ($P = 5\times10^{-8}$, red line) are highlighted in green. The most significant signal at each risk locus is annotated. Locus at chromosome 6 is truncated at -log10P = 20. The arrow points to the most significant signal with its p-value

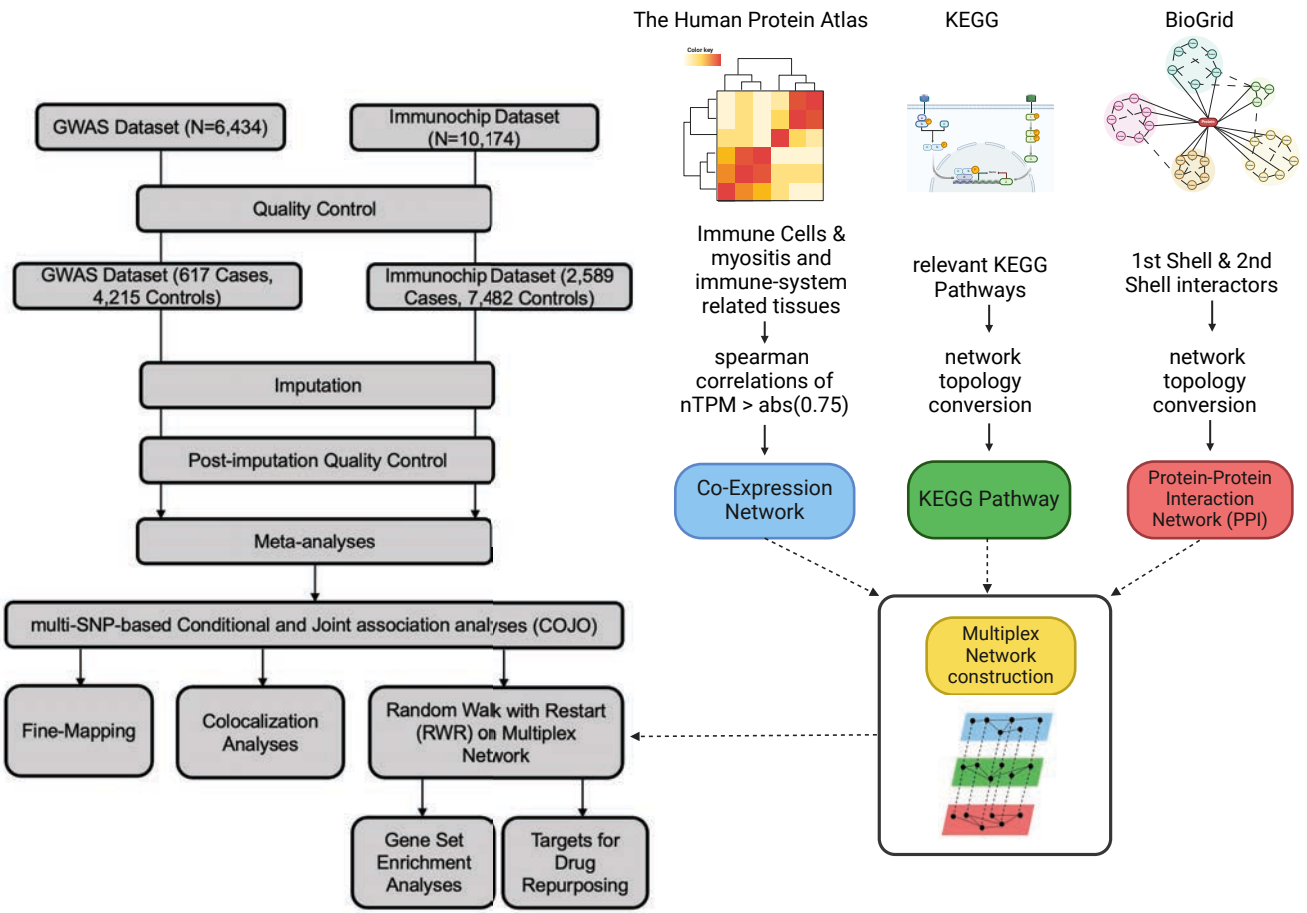
in each plot. **A.** total IIMs: total myositis; **B.** PM, polymyositis; **C.** DM, dermatomyositis; **D.** JDM, juvenile dermatomyositis; **E.** anti-Jo-1 autoantibody-positive myositis.

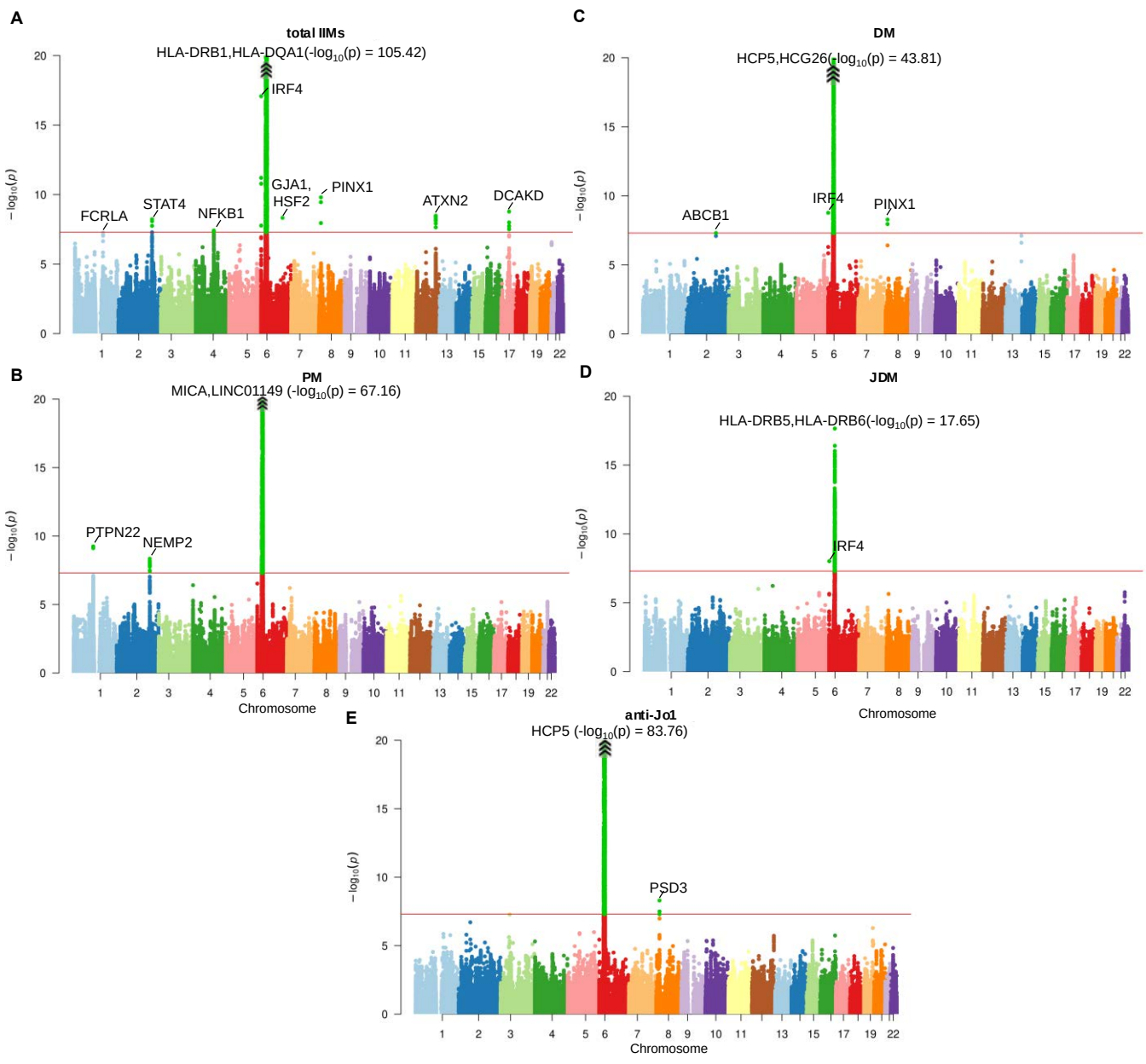
Figure 3. Regional Plots of the most significant novel signals outside the HLA region in the total IIMs. The purple diamond indicates the index SNP. Variants imputed in both studies are represented by downward-pointing triangles, while those genotyped in at least one study are indicated by upward-pointing triangles. The color of each variant indicates the approximate value of the LD squared coefficient of correlation (r2) between the index SNP and the corresponding variant. Index SNPs: **A.** rs6668534 in *FCRLA* at chr1; **B.** rs4853540 in *STAT4* at chr2; **C.** rs230514 in *NFKB1* at chr4; **D.** rs12203592 in *IRF4* at chr6; **E.** rs7754730 in *GJA1*, *HSF2* at chr6; **F.** rs113538396 in *PINX1* at chr8; **G.** rs35350651 in *ATXN2* at chr12; **H.** rs9898793 in *DCAKD* at chr17.

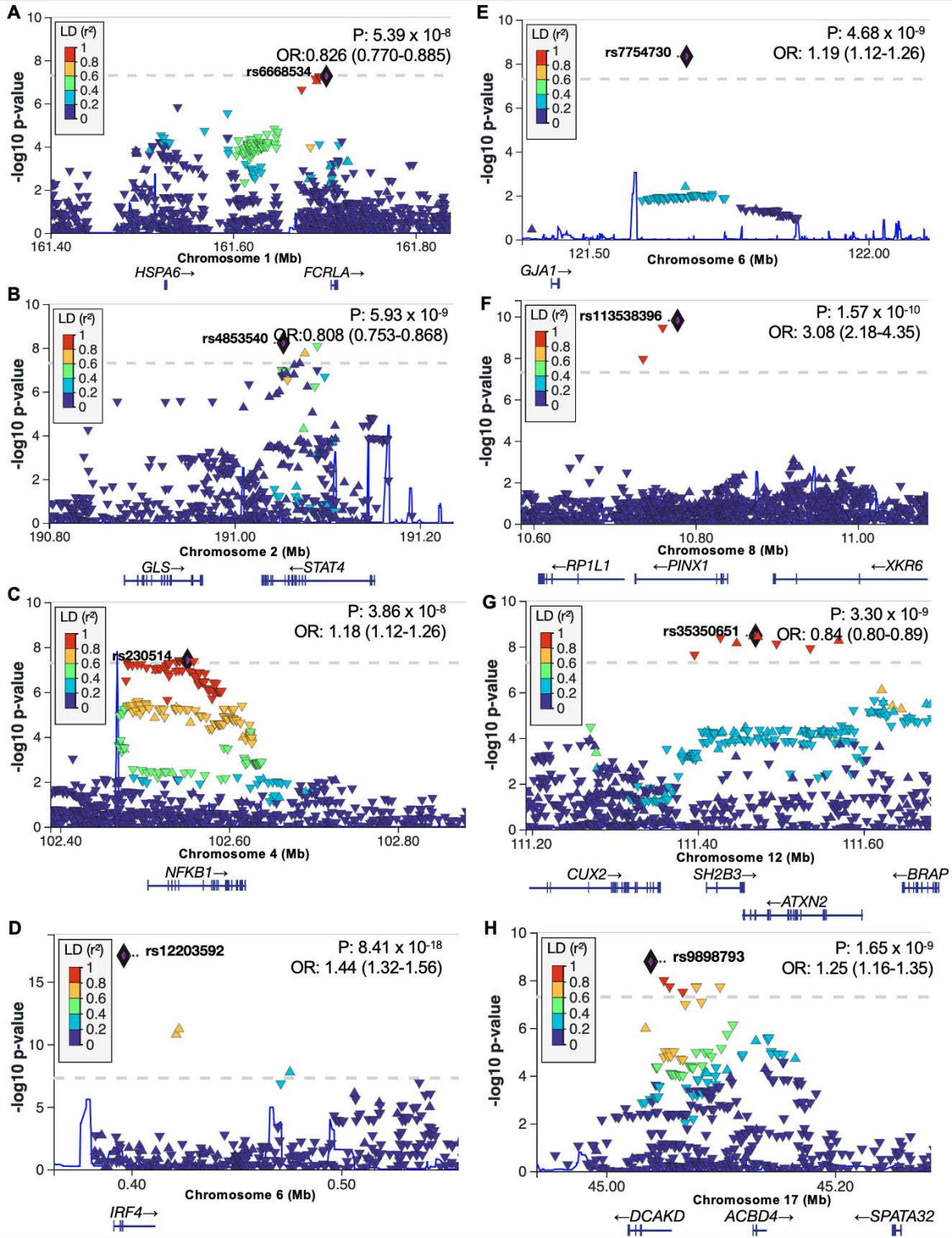
Figure 4. Regional Plots of the most significant novel signals outside of the HLA region in PM, DM, and myositis with anti-Jo1. Purple diamond: the index SNP. Variants imputed in both studies are represented by downward-pointing triangles, while those genotyped in at least one study are indicated by upward-pointing triangles. The color of each variant corresponds to the approximate value of the LD squared coefficient of correlation (r2) between the index SNP and a given variant. Index SNPs are **A.** rs74925618 in *NEMP2* in PM ; **B.** rs145940036 in *ABCB11* in DM; C. rs12203592 in *IRF4* in DM; D. rs113538396 in *PINX1* in DM; E. rs12203592 in *IRF4* in JDM; F. rs6991531 in *PSD3* in anti-Jo1.

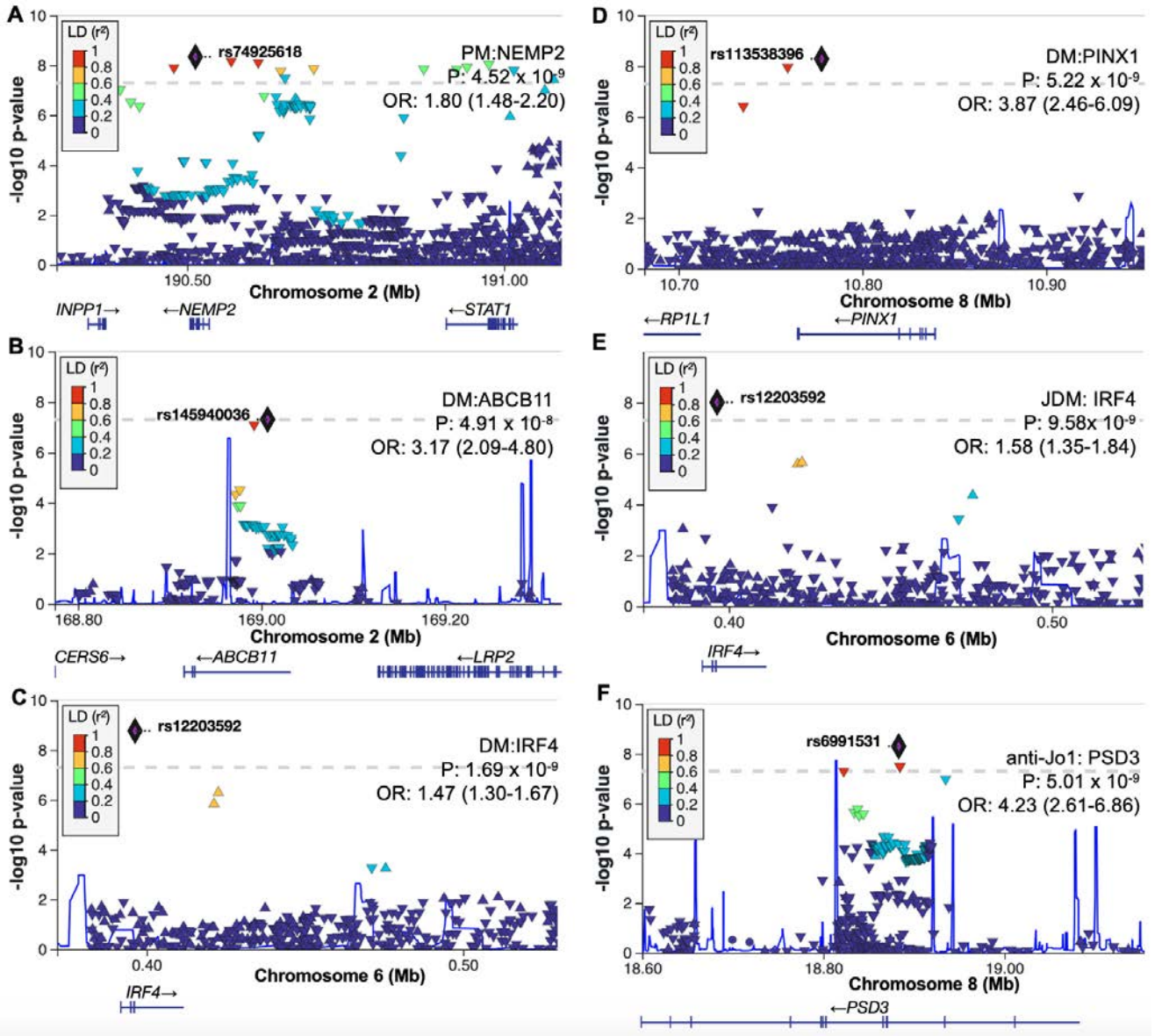
Figure 5. eQTL co-localization analyses within risk loci in significant tissues. Regional plots and plots of Z-score from eQTL studies vs. meta-analyses of **A.** *IRF4* in total IIMs, DM, and JDM. eQTL variants from ENSG00000137265.IRF4. circle (right

panel): top variant in credible sets of meta-analyses, eQTL, and co-localization. **B.** *DCAKD* in total IIMs. eQTL variants from ENSG00000172992.DCAKD. circle (right panel): top variant in credible sets of meta-analyses; triangle (right panel): top variant in credible sets of eQTL and co-localization. **C.** *HCP5* in total IIMs and DM. eQTL variants from ENSG00000206337.HCP5. circle (right panel): top variant in credible sets of meta-analyses and DM co-localization ; triangle (right panel): top variant in credible sets of eQTL. square (right panel): top variant in credible sets of total IIMs co-localization. Purple Diamond: LD index. SNP.PP.H4: the probability of colocalization for the credible set.* Only common variants between meta-analyses and eQTL are analyzed.


The Human Protein Atlas
KEGG
BioGrid
GWAS Dataset (N=6,434)
Immunochip Dataset (N=10,174)
Quality Control
GWAS Dataset (617 Cases, 4,215 Controls)
Immunochip Dataset (2,589 Cases, 7,482 Controls)
Imputation
Post-imputation Quality Control
Meta-analyses
multi-SNP-based Conditional and Joint association analyses (COJO)
Fine-Mapping
Colocalization Analyses
Random Walk with Restart (RWR) on Multiplex Network
Gene Set Enrichment Analyses
Targets for Drug Repurposing
Immune Cells & myositis and immune-system related tissues
spearman correlations of nTPM > abs(0.75)
relevant KEGG Pathways
network topology conversion
1st Shell & 2nd Shell interactors
network topology conversion
Co-Expression Network
KEGG Pathway
Protein-Protein Interaction Network (PPI)
Multiplex Network construction


A
total IIMs
HLA-DRB1,HLA-DQA1(-log10(p) = 105.42)
IRF4
GJA1,
HSF2
PINX1
ATXN2
DCAKD
FCRLA
STAT4
NFKB1
B
PM
MICA,LINC01149 (-log10(p) = 67.16)
PTPN22
NEMP2
C
DM
HCP5,HCG26(-log10(p) = 43.81)
ABCB1
IRF4
PINX1
D
JDM
HLA-DRB5,HLA-DRB6(-log10(p) = 17.65)
IRF4
E
anti-Jo1
HCP5 (-log10(p) = 83.76)
PSD3
−log10(p)
Chromosome


A
P: 5.39 x 10-8
OR:0.826 (0.770-0.885)
rs6668534
LD (r2)
1
0.8
0.6
0.4
0.2
0
-log10 p-value
161.40
161.60
161.80
Chromosome 1 (Mb)
HSPA6→
FCRLA→
B
P: 5.93 x 10-9
rs4853540
OR:0.808 (0.753-0.868)
190.80
191.00
191.20
Chromosome 2 (Mb)
GLS→
←STAT4
C
P: 3.86 x 10-8
OR: 1.18 (1.12-1.26)
rs230514
102.40
102.60
102.80
Chromosome 4 (Mb)
NFKB1→
D
P: 8.41 x 10-18
OR: 1.44 (1.32-1.56)
rs12203592
0.40
0.50
Chromosome 6 (Mb)
IRF4→
E
P: 4.68 x 10-9
OR: 1.19 (1.12-1.26)
rs7754730
121.50
122.00
Chromosome 6 (Mb)
GJA1→
F
P: 1.57 x 10-10
OR: 3.08 (2.18-4.35)
rs113538396
10.60
10.80
11.00
Chromosome 8 (Mb)
←RP1L1
←PINX1
←XKR6
G
P: 3.30 x 10-9
OR: 0.84 (0.80-0.89)
rs35350651
111.20
111.40
111.60
Chromosome 12 (Mb)
CUX2→
SH2B3→
←BRAP
←ATXN2
H
P: 1.65 x 10-9
OR: 1.25 (1.16-1.35)
rs9898793
45.00
45.20
Chromosome 17 (Mb)
←DCAKD
ACBD4→
←SPATA32


A
rs74925618
PM:NEMP2
P: 4.52 x 10-9
OR: 1.80 (1.48-2.20)
LD (r2)
1
0.8
0.6
0.4
0.2
0
-log10 p-value
190.50
191.00
Chromosome 2 (Mb)
INPP1→
←NEMP2
←STAT1
B
rs145940036
DM:ABCB11
P: 4.91 x 10-8
OR: 3.17 (2.09-4.80)
168.80
169.00
169.20
Chromosome 2 (Mb)
CERS6→
←ABCB11
←LRP2
C
rs12203592
DM:IRF4
P: 1.69 x 10-9
OR: 1.47 (1.30-1.67)
0.40
0.50
Chromosome 6 (Mb)
IRF4→
D
rs113538396
DM:PINX1
P: 5.22 x 10-9
OR: 3.87 (2.46-6.09)
10.70
10.80
10.90
Chromosome 8 (Mb)
←RP1L1
←PINX1
E
rs12203592
JDM: IRF4
P: 9.58x 10-9
OR: 1.58 (1.35-1.84)
0.40
0.50
Chromosome 6 (Mb)
IRF4→
F
rs6991531
anti-Jo1: PSD3
P: 5.01 x 10-9
OR: 4.23 (2.61-6.86)
18.60
18.80
19.00
Chromosome 8 (Mb)
←PSD3

A

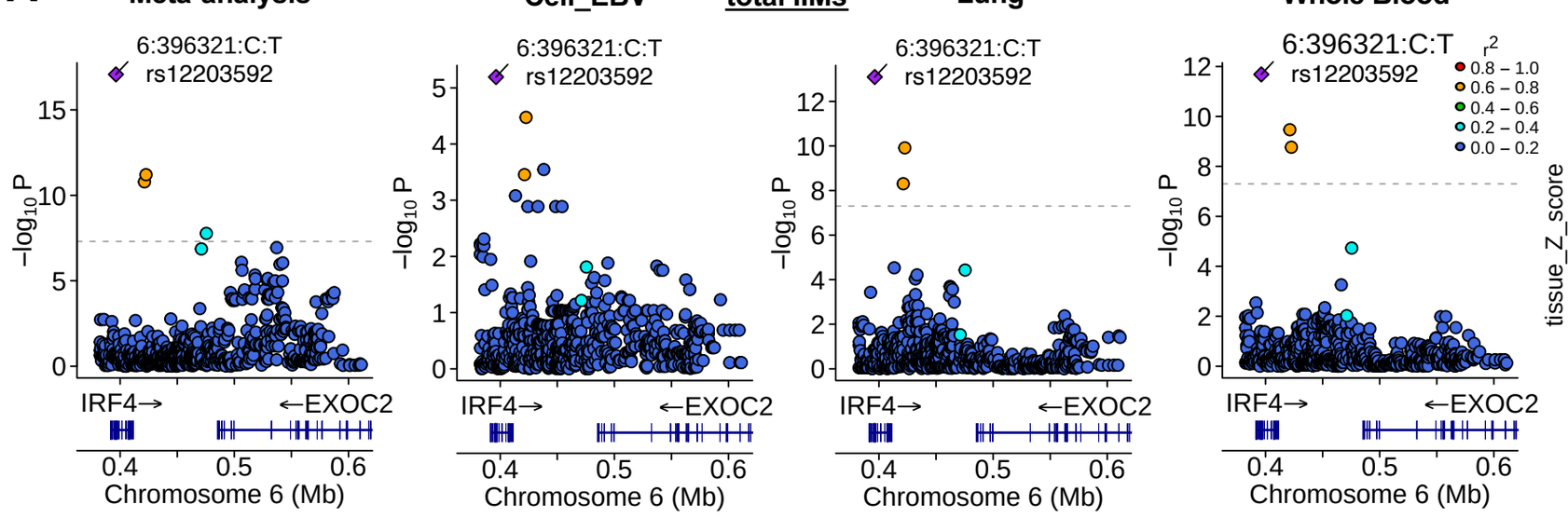
Meta-analysis

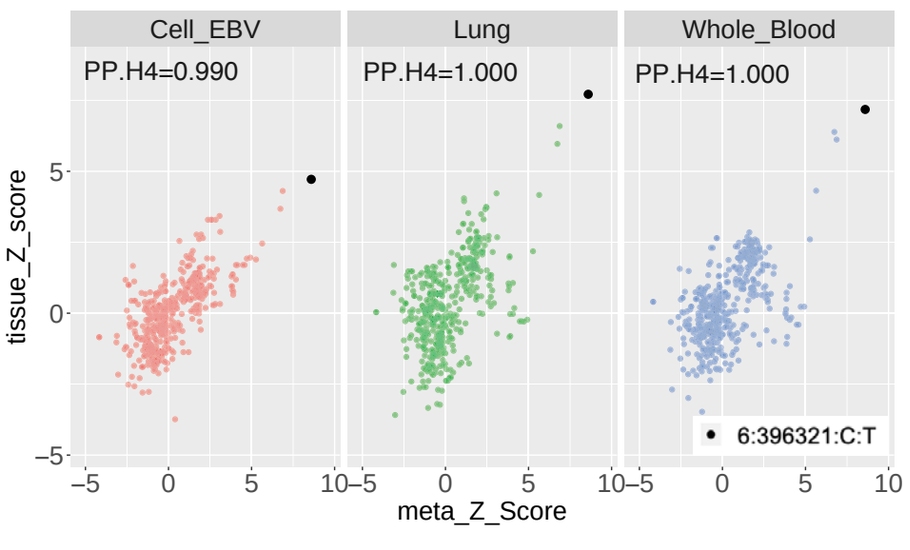
total IIMs

Cell_EBV

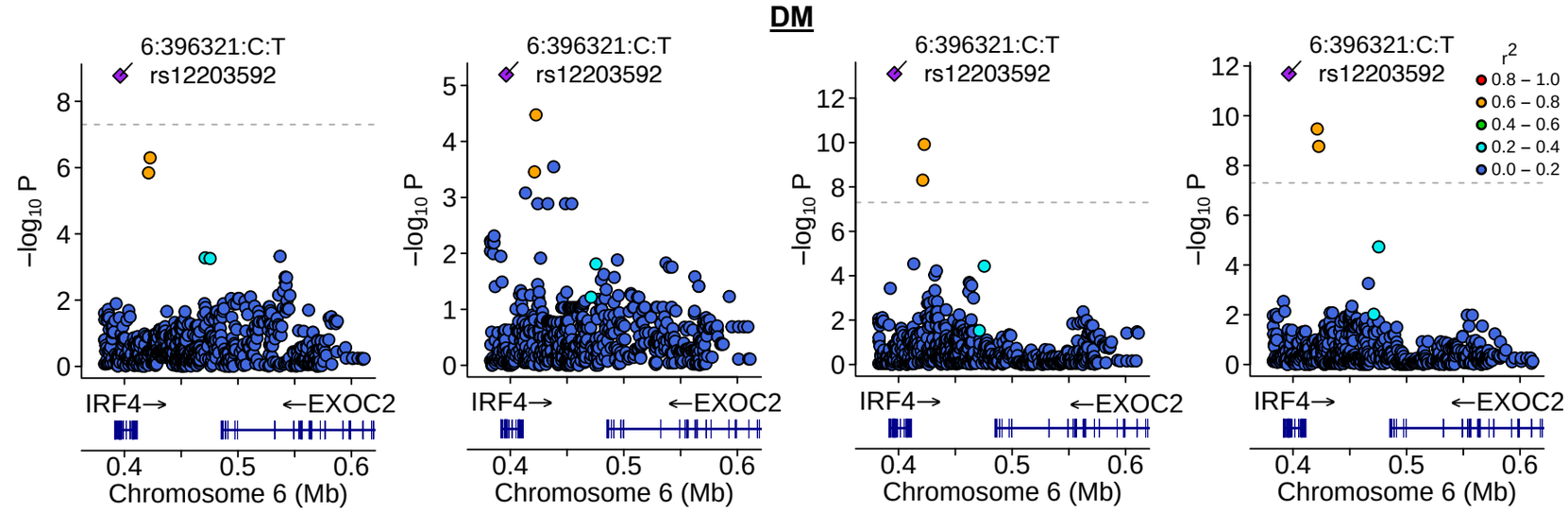
Lung

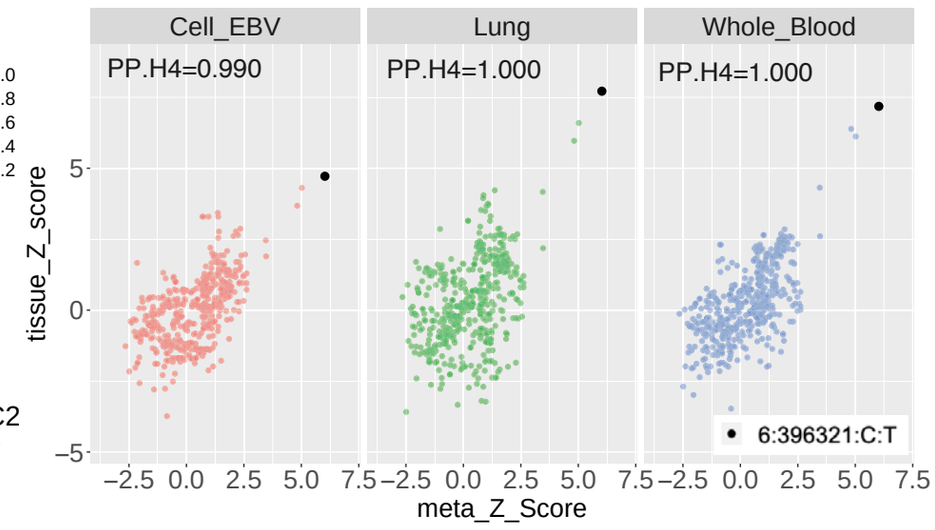
Whole Blood



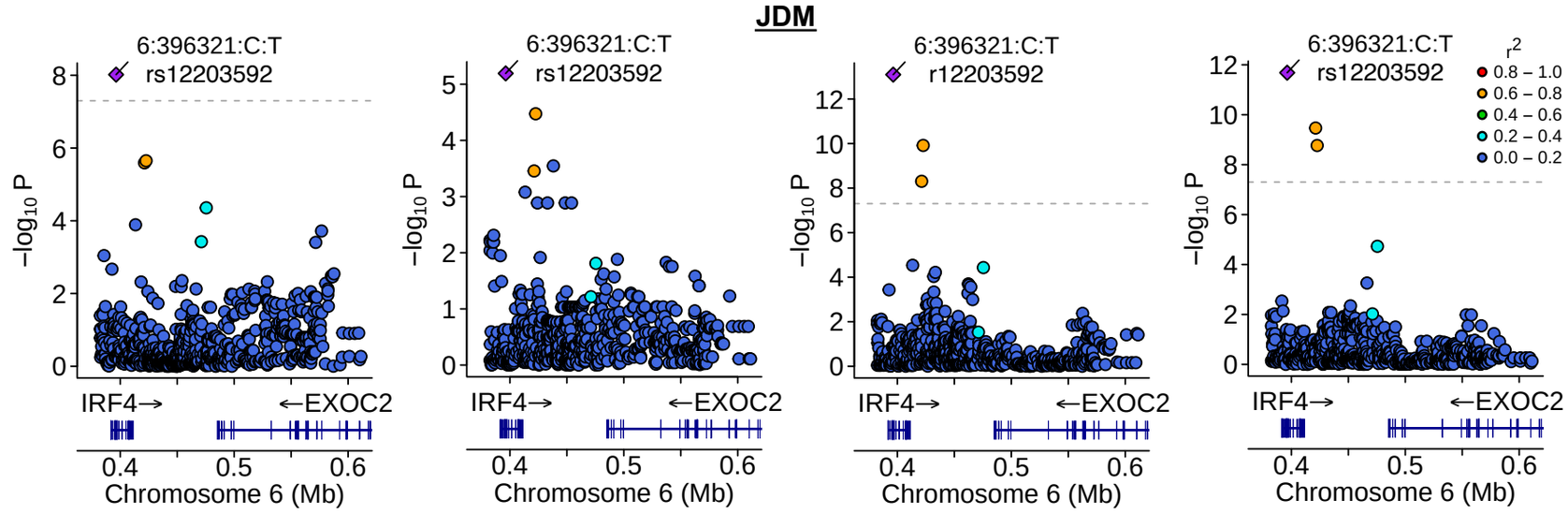


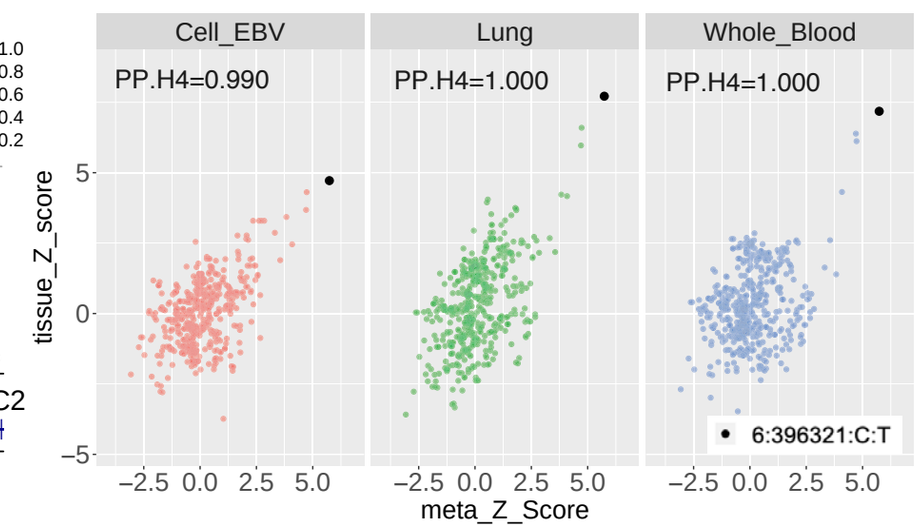
DM





JDM





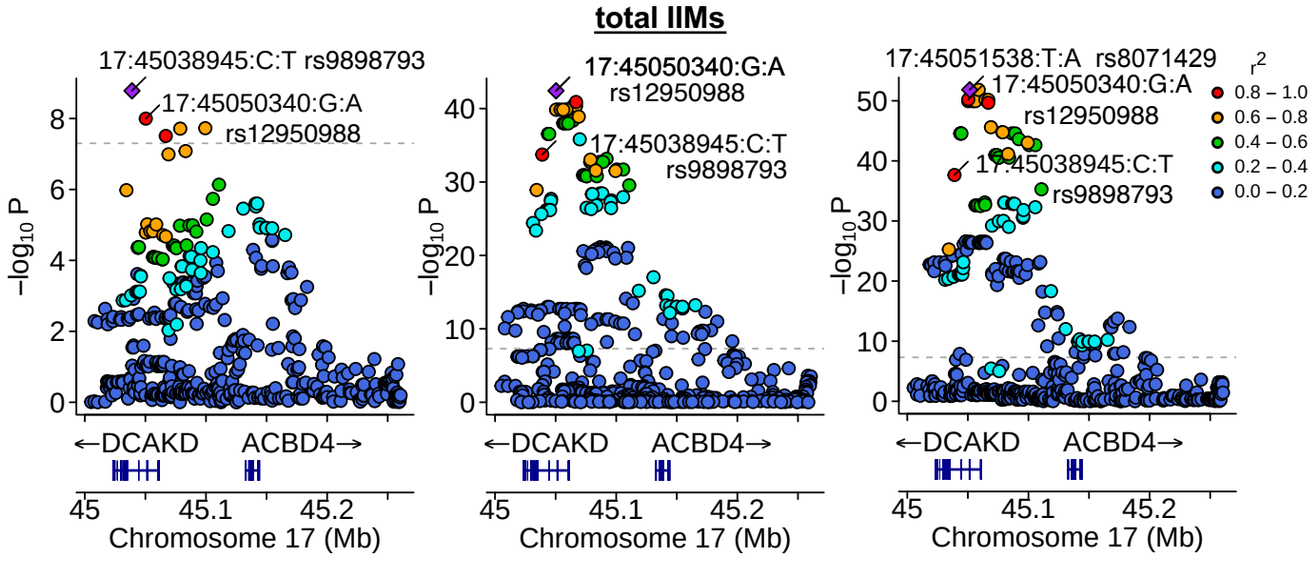
B

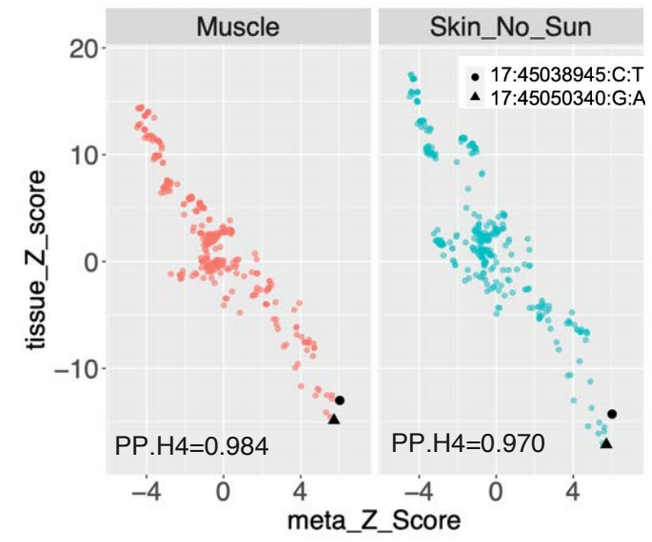
Meta-analysis

Muscle

total IIMs

Skin_No_Sun





C

Meta-analysis

Lung

Meta-analysis

Lung

total IIMs

DM