Variants in the *DDX6-CXCR5* autoimmune disease risk locus influence the regulatory network in immune cells and salivary gland

Mandi M. Wiley[1], Bhuwan Khatri[1], Michelle L. Joachims[1,2], Kandice L. Tessneer[1], Anna M. Stolarczyk[1], Astrid Rasmussen[1], Juan-Manuel Anaya[3], Lara A. Aqrawi[4,5], Sang-Cheol Bae[6], Eva Baecklund[7], Albin Björk[8], Johan G. Brun[9,10], Sara Magnusson Bucher[11], Nick Dand[12], Maija-Leena Eloranta[7], Fiona Engelke[13], Helena Forsblad-d'Elia[14], Cecilia Fugmann[7], Stuart B. Glenn[1], Chen Gong[12], Jacques-Eric Gottenberg[15], Daniel Hammenfors[10], Juliana Imgenberg-Kreuz[7], Janicke Liaaen Jensen[5], Svein Joar Auglænd Johnsen[16], Malin V. Jonsson[9], Jennifer A. Kelly[1], Sharmily Khanam[2], Kwangwoo Kim[17], Marika Kvarnström[8], Thomas Mandl[18], Javier Martín[19], David L. Morris[12], Gaetane Nocturne[20,21], Katrine Brække Norheim[16], Peter Olsson[18], Øyvind Palm[22], Jacques-Olivier Pers[23], Nelson L. Rhodus[24], Christopher Sjöwall[25], Kathrine Skarstein[9], Kimberly E. Taylor[26], Phil Tombleson[12], Gudny Ella Thorlacius[8], Swamy Venuturupalli[27], Edward M. Vital[28], Daniel J Wallace[27], Kiely M. Grundahl[1,2], Lida Radfar[29], Michael T. Brennan[30], Judith A. James[2,31], R. Hal Scofield[2,31,32], Patrick M. Gaffney[1,31], Lindsey A. Criswell[26,33], Roland Jonsson[9], Silke Appel[9], Per Eriksson[25], Simon J. Bowman[34], Roald Omdal[9,16], Lars Rönnblom[7], Blake M. Warner[35], Maureen Rischmueller[36], Torsten Witte[13], A. Darise Farris[2,31], Xavier Mariette[20,21], Caroline H. Shiboski[26], Sjögren's International Collaborative Clinical Alliance (SICCA)[#], Marie Wahren-Herlenius[8,9], Marta E. Alarcón-Riquelme[8,37], PRECISESADS Clinical Consortium[#], Wan-Fai Ng[38,39], UK Primary Sjögren's Syndrome Registry[#], Kathy L. Sivils[2†], Joel M. Guthridge[2,31], Indra Adrianto[40,41], Timothy J. Vyse[12], Betty P. Tsao[42], Gunnel Nordmark[7], Christopher J. Lessard[1,31]*

AFFILIATIONS

[1]Genes and Human Disease Research Program, Oklahoma Medical Research Foundation (OMRF), Oklahoma City, Oklahoma, USA;
[2]Arthritis and Clinical Immunology Research Program, OMRF, Oklahoma City, Oklahoma, USA;
[3]Universidad del Rosario, Bogotá, Colombia;
[4]Department of Health Sciences, Kristiania University College, Oslo, Norway;
[5]University of Oslo, Norway;
[6]Hanyang University, Seoul, Republic of Korea**;**
[7]Uppsala University, Uppsala, Sweden;
[8]Karolinska Institutet, Solna, Sweden;
[9]University of Bergen, Bergen, Norway;
[10]Haukeland University Hospital, Bergen, Norway;
[11]Örebro University, Örebro, Sweden;
[12]King's College London, London, United Kingdom;
[13]Hannover Medical School, Hannover, Germany;
[14]University of Gothenburg, Gothenburg, Sweden;
[15]Strasbourg University Hospitals, Strasbourg, France;
[16]Stavanger University Hospital, Stavanger, Norway;
[17]Kyung Hee University, Seoul, Republic of Korea;
[18]Lund University, Malmö, Sweden;

[19]Instituto de Biomedicina y Parasitología López-Neyra, Granada, Spain;
[20]Université Paris-Saclay, Paris, France;
[21]Assistance Publique – Hôpitaux de Paris, Hôpital Bicêtre, Paris, France;
[22]Oslo University Hospital, Oslo, Norway;
[23]LBAI, UMR1227, University of Brest, Inserm, Brest, France;
[24]University of Minnesota Medical School, Minnesota, USA;
[25]Linköping University, Linköping, Sweden;
[26]University of California San Francisco, San Francisco, California, USA;
[27]Cedars-Sinai Medical Center, Los Angeles, California, USA;
[28]University of Leeds, Leeds, United Kingdom;
[29]University of Oklahoma College of Dentistry, Oklahoma City, Oklahoma, USA;
[30]Atrium Health Carolinas Medical Center, Charlotte, North Carolina, USA;
[31]University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA;
[32]US Department of Veteran Affairs Medical Center, Oklahoma City, Oklahoma, USA;
[33]National Human Genome Research Institute, NIH, Bethesda, Maryland, USA;
[34]University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom;
[35]National Institute of Dental and Craniofacial Research, Bethesda, Maryland, USA;
[36]University of Adelaide, Adelaide, South Australia;
[37]Genyo, Center for Genomics and Oncological Research, Pfizer/University of Granada/Andalusian Regional Government, Spain;
[38]NIHR Newcastle Biomedical Research Centre and NIHR Newcastle Clinical Research Facility, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle Upon Tyne, United Kingdom;
[39]Translational and Clinical Research Institute, Newcastle University, Newcastle Upon Tyne, United Kingdom;
[40]Center for Bioinformatics, Department of Public Health Sciences, Henry Ford Health, Detroit, Michigan, USA;
[41]Department of Medicine, College of Human Medicine, Michigan State University, East Lansing, Michigan, USA;
[42]Medical University of South Carolina, Charleston, South Carolina, USA;
[#]A list of consortia investigators and their affiliations appear as a Supplementary Note;
[†]Current Affiliation: Translational Science, The Janssen Pharmaceutical Companies of Johnson & Johnson, Spring House, Pennsylvania, USA;
*Corresponding Author: chris-lessard@omrf.org

ABSTRACT

Fine mapping and bioinformatic analysis of the *DDX6-CXCR5* genetic risk association in Sjögren's Disease (SjD) and Systemic Lupus Erythematosus (SLE) identified five common SNPs with functional evidence in immune cell types: rs4938573, rs57494551, rs4938572, rs4936443, rs7117261. Functional interrogation of nuclear protein binding affinity, enhancer/promoter regulatory activity, and chromatin-chromatin interactions in immune, salivary gland epithelial, and kidney epithelial cells revealed cell type-specific allelic effects for all five SNPs that expanded regulation beyond effects on *DDX6* and *CXCR5* expression. Mapping the local chromatin regulatory network revealed several additional genes of interest, including *lnc-PHLDB1-1*. Collectively, functional characterization implicated the risk alleles of these SNPs as modulators of promoter and/or enhancer activities that regulate cell type-specific expression of *DDX6*, *CXCR5*, and *lnc-PHLDB1-1*, among others. Further, these findings emphasize the importance of exploring the functional significance of SNPs in the context of complex chromatin architecture in disease-relevant cell types and tissues.

Keywords: Sjögren's Disease, systemic lupus erythematosus, SLE, DEAD-Box helicase 6, DDX6, C-X-C motif chemokine receptor 5, CXCR5, chromatin regulatory network, lnc-PHLDB1-1

INTRODUCTION

Sjögren's disease (SjD) and systemic lupus erythematosus (SLE) are distinct yet related autoimmune diseases with several shared clinical features and genetic associations[1,2]. While autoimmune destruction and impaired function of the exocrine glands (e.g., lacrimal and salivary glands) are unique to SjD[3-5], several clinical manifestations are shared, including serological features like the presence of antinuclear antibody (ANA), anti-Ro/SSA antibodies and overexpression of type I interferons[6-8]. SjD has a higher prevalence in populations of European ancestry while SLE is more frequently diagnosed in those of African descent[9-11]. SjD also has a stronger disparity in women relative to men (SjD: 9-20:1 female: male; SLE: 9-10:1 female: male)[12-17]. The complex etiologies of SjD and SLE are driven by environmental exposures in the context of genetic susceptibility[18,19]. Exploring the functional implications of shared genetic risk loci between SjD and SLE may help identify common disease mechanisms.

DDX6-CXCR5 on chromosome 11p23.3 has been associated through candidate gene and genome-wide association studies (GWAS) with SjD and SLE of European and Asian ancestry[1,20-23]. DDX6 (DEAD-box RNA Helicase 6) is important in viral RNA recognition and modulation of type I interferon responses[23-25]. CXCR5 (C-X-C chemokine receptor type 5) is an important modulator of B and T follicular cell trafficking to peripheral lymphoid organs[26]. Loss of *CXCR5* expression has been reported in purified circulating $CD19^+$ B cells isolated from SjD patients homozygous for the risk allele (T) of rs4938573[27]; a proxy in strong linkage disequilibrium (LD; r^2=0.865) with the top associated SNP, rs7119038, of the *DDX6-CXCR5* SjD risk interval[21]. Loss of

CXCR5 was inversely correlated with increased homing of $CXCR5^+$ cells to the salivary gland[27], suggesting a potential SjD-specific role for this risk interval.

In this study, meta-analysis and genetic fine mapping of the *DDX6-CXCR5* risk interval of SjD and SLE were performed to refine the association signals and identify the most likely functional SNPs shared across the two diseases (**Figure 1A,B**). The *DDX6-CXCR5* locus has a complex local chromatin regulatory network including several genes, gene promoters, and regulatory elements with cell type-specific patterns of epigenetic enrichment. Deep bioinformatic analyses prioritized five SNPs for functional validation using electromobility shift assays (EMSAs), luciferase transactivation assays, and chromatin conformation capture with quantitative PCR (3C-qPCR), identifying cell type- and context-specific functional effects (**Figure 1C**). Understanding how functional SNPs located within this locus regulate the expression of *DDX6*, *CXCR5*, and/or other disease-relevant genes through interactions within the chromatin regulatory network will provide new insights into how genetic risk at the *DDX6-CXCR5* locus may modulate molecular mechanisms of autoimmunity.

RESULTS

Meta-analysis and genetic fine mapping of the *DDX6-CXCR5* risk interval shows similar profiles in SjD and SLE.

Meta-analysis was performed using genotype and Immunochip data from 3,851 SjD cases and 23,652 population controls of European ancestry (Dataset (DS1)) or Immunochip data and summary statistics from 11,840 SLE cases and 28,869 population controls of European or Korean ancestry (DS2), after quality control and imputation, to

define the disease-specific associations on the *DDX6-CXCR5* region (**Figure 1A**)[1,20,28]. Similar association signals were observed, but different peak SNP-trait associations (i.e., index SNPs) were identified for each disease (**Figure 2A, B; Table 1; Supplemental Data 1, 2**). The index SNP, rs7481819, in SjD is positioned near the promoter region of *CXCR5* (**Figure 2A**), whereas the index SNP in SLE, rs11217045, is positioned in a shared regulatory region between *DDX6* and *CXCR5* (**Figure 2B**). The SNP, rs11217045, was also identified as the index SNP when the SjD and SLE datasets were meta-analyzed (**Figure 2C; Table 1; Supplemental Data 3**).

To further refine the *DDX6-CXCR5* association signals in SjD and SLE, Bayesian analyses were performed using Trinculo[29], which leverages genotype data rather than summary statistics to calculate posterior probabilities and identify 95% credible sets of potential functional SNPs (**Figure 1A, B**; **Figure 2D**; **Supplemental Figure 1D-F**; **Supplemental Data 4-7**). Analysis of DS1-SjD, the largest publicly available SjD genotype dataset, identified a total of 66 SNPs in the 95% credible set with a maximum posterior probability of 0.087 (rs7123726) (**Figure 2D; Supplemental Data 4**). Use of SLE genomic summary results in the SLE meta-analysis (DS2) or merged meta-analysis (DS1+DS2) prohibited posterior probabilities calculations using Trinculo. Therefore, separate logistical regression analyses were performed using Immunochip data only (1916 SjD cases (DS3), 3,762 SLE cases (DS4), and 6,194 population controls of European ancestry), revealing similar association signals (**Figure 1B**). Analysis of DS3-SjD identified 105 SNPs in the 95% credible set with the top SNP in LD yielding a posterior probability of 0.034 (rs10790268) (**Supplemental Figure 1D; Supplemental Data 5**). Analysis of DS4-SjD identified 54 SNPs in the 95% credible set

with a maximum posterior probability of 0.020 (rs76409436) (**Supplemental Figure 1E; Supplemental Data 6**). Analysis of the merged Immunochip datasets (DS3 and DS4) identified 15 SNPs in the 95% credible set and a maximum posterior probability of 0.059 (rs11217058) (**Supplemental Figure 1F; Supplemental Data 7**).

Fine mapping revealed that SjD and SLE of European or Korean ancestry exhibit similar association signals at the *DDX6-CXCR5* locus with no heterogeneity observed in the meta-analyses among the top variants (**Supplemental Data 1-3**). While there were several shared SNPs in the 95% credible set, the Bayesian analyses were unable to refine the association with sufficient posterior probability (>0.5) to predict the most likely functional SNPs. To further evaluate the co-inheritance of the index SNP and other possible functional variants within the region, the haplotype structure was examined. Co-inheritance in European ancestry ($D' \geq 0.89$) and LD ($r^2 = 0.64\text{-}0.99$) between several shared SNPs were observed (**Supplemental Figure 2A-C**) with the exception of rs480958 ($D' \leq 0.81$; $r^2 \leq 11$), which was previously associated with SLE of East Asian ancestry (**Supplemental Figure 2D-F**)[22]. Strong co-inheritance of SNPs spanning the risk locus may explain why the Bayesian analyses were unable to sufficiently refine the association and predict likely functional SNPs.

Bioinformatic interrogation identifies multiple SNPs with evidence of functionality on the *DDX6-CXCR5* interval.

Bioinformatics data from multiple databases were interrogated to further refine the shared *DDX6-CXCR5* interval and identify which polymorphisms are most likely to have regulatory potential. From the top 100 SNPs in the 95% credible set from the

196 merged Immunochip analysis (**Supplemental Data 7**), 14 SNPs exhibited bioinformatic
197 evidence of functionally activity, and were as identified index SNPs in this study (**Table**
198 **1**) or were previously associated with SjD and/or SLE[20-22,27] (**Supplemental Figure 3**).
199 Index SNPs, rs7481819, rs11217045, and rs7123726 (**Figure 2A-D**), as well as
200 previously implicated SNPs rs10892286[20] and rs7119038[21], received RegulomeDB
201 scores of 5, suggesting low functional potential (**Supplemental Figure 3;**
202 **Supplemental Data 7**). HaploReg was used to assess epigenetic activity and, along
203 with other eQTL databases (eQTL catalogue, single cell RNA-seq eQTLs, DICE, GTEx
204 v08, QTLBase, and NephQTL), evaluate reported eQTLs in disease-relevant immune
205 cell types, as well as salivary gland and kidney tissue. Consistently, minimal epigenetic
206 activity or eQTLs (n<7) were reported for the index SNPs (**Supplemental Figure 3**).
207 Previously reported SNPs, rs480958 (RegulomeDB Score: 4)[22], rs10892294
208 (RegulomeDB Score: 3a)[21], and rs4936441 (RegulomeDB Score: 3a)[22], also had few
209 reported eQTLs (n<8) and exhibited minimal epigenetic evidence of activity (n<56 total
210 epigenetic promoter/enhancer peaks) (**Supplemental Figure 3; Supplemental Data 7**).
211 Variant rs12365699 received a RegulomeDB score of 1f but only had 5 reported eQTLs
212 and exhibited minimal evidence of epigenetic activity (n=80 epigenetic
213 promoter/enhancer peaks) limited primarily to T cells (**Supplemental Figure 3**).
214 In contrast, five SNPs exhibited bioinformatic evidence of epigenetic regulation
215 (including promoter and/or enhancer activity (n>243)), eQTL enrichment (n>36), and
216 protein binding by ChIP-seq (n>2) across several immune cell types and tissues:
217 rs57494551 (RegulomeDB Score: 2b) positioned in the first intron of *DDX6*, and
218 rs4936443 (RegulomeDB Score: 4), rs4938572 (RegulomeDB Score: 1b), rs7117261

(RegulomeDB Score: 2b), and rs4938573 (RegulomeDB Score: 4) clustered in an 978 bp shared regulatory region approximately 79 kb downstream of *DDX6* and 12.5 kb upstream of *CXCR5* (**Figure 3A**; **Supplemental Figure 3; Supplemental Data 7**). Our comprehensive bioinformatic interrogation revealed additional eQTLs (n>36) that were not reported in RegulomeDB for four of the five SNPs (**Supplemental Figure 3**), suggesting that rs57494551, rs4936443, rs7117261, and rs4938573 should have amended RegulomeDB scores of 1c or better.

Analysis of previously reported chromatin-chromatin interactions from the GM12878 EBV B cell line and eQTLs from multiple different immune cell types revealed a complex regulatory network spanning the *DDX6-CXCR5* SjD/SLE risk locus (**Figure 2E**). Coalescence of reported eQTLs and publicly available promoter-capture Hi-C (hereafter simplified as Hi-C) data further revealed that these five SNPs are positioned in regulatory elements that may modulate gene expression beyond that of *DDX6* and *CXCR5* (**Figure 3B-F**). For example, rs57494551 is an eQTL for several transcripts in immune cells, including *DDX6* in monocytes and neutrophils, *CXCR5* in B cells, $CD8^+$ T cells, monocytes, macrophages, and neutrophils, *IL10RA* in T cells and neutrophils, *BCL9L* (B cell CLL/Lymphoma 9) in B cells and whole blood, and *TRAPPC4* (transport protein particle complex subunit 4) in macrophages and minor salivary gland (**Figure 3B**). SNP rs57494551 is positioned in an intronic regulatory element that forms chromatin-chromatin interactions with the promoters of *DDX6, CXCR5,* and *BCL9L*, but not *IL10RA*, in an immune cell type and/or context specific manner, suggesting that rs57494551 may exhibit regulatory effects on *DDX6*, *CXCR5*, and *BCL9L* through modulating interactions between the intronic enhancer and the promoters of these

genes. Unfortunately, publicly available Hi-C data remains unavailable for the minor salivary gland, therefore any potential coalescence between the reported *TRAPPC4* eQTL and a chromatin-chromatin interaction with the rs57494551 regulatory element could not be inferred.

Coalescence of eQTLs (*DDX6*, *CXCR5*, *IL10RA*, *TRAPPC4*) and Hi-C data were also observed for the four SNPs in the shared regulatory region between *DDX6* and *CXCR5* (**Figure 3C-F**). Interestingly, the four SNPs are also reported eQTLs for a long noncoding RNA, *lnc-PHLDB1-1*, in several immune cell types and/or the minor salivary gland (**Figure 3C-F**). Further, Hi-C data indicate cell type-specific chromatin-chromatin interactions between the regulatory region carrying rs4938572 and rs4938573 and the promoter of *lnc-PHLDB1-1* (**Figure 3D, F**). Collectively, the bioinformatic evaluation of shared polymorphisms associated with the *DDX6-CXCR5* SjD/SLE risk locus identified five SNPs with functional potential warranting further characterization.

Prioritized SNPs demonstrate allele-specific nuclear protein complex binding.

EMSAs were performed in EBV B cells, Daudi B cells, Jurkat T cells, THP1 monocytes, and/or A253 cell line originating from a submandibular gland squamous cell carcinoma to assess the cell type- and allelic effects of the five prioritized SNPs on nuclear protein binding (**Supplemental Table 1)**. Binding specificity was confirmed by competition assay with unlabeled probes. All five SNPs demonstrated varying allele- and cell type-specific nuclear protein binding (**Figure 4; Supplemental Figure 4-9**). Banding patterns varied by SNP and cell type, requiring separate analyses of multiple bands.

The risk allele of rs57494551 impaired nuclear protein binding compared to the non-risk allele in EBV B cells (P_{EMSA}<5.8x10^{-3}) and THP1 monocytes (P_{EMSA}<3.9x10^{-2}), but increased binding in A253 cells (P_{EMSA}=1.4x10^{-2}) (**Figure 4A-B, 4E; Supplemental Figure 4A-B**). The risk alleles of rs4938572 (P_{EMSA}<4.2x10^{-2}) (**Figure 4C-E; Supplemental Figure 4C-D**) and rs7117261 (P_{EMSA}<2.0x10^{-3}) (**Figure 4E; Supplemental Figure 5**) increased protein binding compared to the non-risk allele in all tested immune cell lines and A253 cells. The risk allele of rs4936443 decreased nuclear protein binding compared to the non-risk allele in Daudi B cells, Jurkat T cells, THP1 monocytes, and A253 cells (P_{EMSA}<3.9x10^{-2}), but did not change binding in EBV B cells (**Figure 4E; Supplemental Figure 6**). The risk allele of rs4938573 increased protein binding in both EBV B and Daudi B cell lines (P_{EMSA}<2.5x10^{-2}) but decreased binding in the Jurkat T cells (P_{EMSA}<3.7x10^{-3}) (**Figure 4E; Supplemental Figure 7**). Interestingly, the risk allele of rs4938573 exhibited increased or decreased nuclear protein binding compared to the non-risk allele in the THP1 monocytes and A253 cells, respectively, depending on the band analyzed (**Figure 4E; Supplemental Figure 7**).

Despite cell type-specific epigenetic evidence of function in T cells, rs12365699 exhibited no change in allele-specific binding to nuclear proteins from Jurkat T cells, Daudi B, or THP1 monocytes (**Figure 4E; Supplemental Figure 8**). SNP rs10892294 also exhibited no change in allele-specific binding to nuclear proteins from Daudi B, Jurkat T, or THP1 monocytes (**Figure 4E; Supplemental Figure 9**). These results further suggest that the five prioritized SNPs are likely functional in multiple immune cell types and/or A253 cells. Further, allele-specific nuclear protein binding to rs12365699 or rs10892294 showed limited epigenetic evidence of functionality.

Prioritized SNPs demonstrate allele-specific effects on regulatory activity.

To determine if the regulatory elements carrying the prioritized SNPs exhibit allele-specific and/or cell type-specific promoter or enhancer activity, luciferase expression assays were performed using a promoter-less vector or minimal-promoter vector, respectively. The regulatory activity and allelic effects were first screened using 293T human embryonic kidney cells, which are a model for high transfection efficiency and optimal luciferase expression. Regulatory elements carrying rs57494551, rs4936443, rs4938572, or rs7117261 exhibited significantly increased enhancer and/or promoter activity compared to the empty vector control in 293T cells (**Figure 5A-D,F)**. In contrast, rs10892294 and rs12365699 exhibited minimal enhancer or promoter activity above that of the empty vector in 293T cells and, therefore, were excluded from additional study (**Supplemental Figure 10**). Polymorphism rs4938573 also exhibited minimal enhancer or promoter activity in 293T cells (**Figure 5E**). However, rs4938573 exhibited strong bioinformatic evidence of function and allele-specific nuclear protein binding in EMSAs from multiple immune cell types. Further, the SjD/SLE risk allele of rs4938573 was previously reported to modulate *CXCR5* expression and $CXCR5^{+}$ cell homing to the salivary gland[27]. For these reasons, rs4938573 was selected, along with rs57494551, rs4936443, rs4938572, and rs7117261 for further evaluation.

Consistent with decreased nuclear protein binding in EBV B and THP1 cells, the risk allele of rs57494551 significantly decreased promoter activity compared to the non-risk allele in Jurkat ($P_{luc}<1.0x10^{-4}$), THP1 ($P_{luc}=1.5x10^{-2}$), and A253 cells ($P_{luc}=8.5x10^{-3}$) (**Figure 5A**). However, the risk allele of rs57494551 also significantly increased

enhancer activity compared to the non-risk allele in both B cell lines (EBV B (P_{luc}<1.0x10^{-4}) and Daudi (P_{luc}=4.0x10^{-4})), THP1 (P_{luc}<1.0x10^{-4}), A253 (P_{luc}<1.0x10^{-4}), and 293T (P_{luc}=2.4x10^{-2}) cells. The risk allele of rs4936443 increased promoter and enhancer activity in THP1 (P_{luc}<2.4x10^{-3}) and A253 cells (P_{luc}<3.0x10^{-4}) (**Figure 5B**). No allele specificity was observed in the EBV B, Daudi, Jurkat, or 293T cells (**Figure 5B)**. The risk allele of rs4938572 increased promoter (P_{luc}=1.8x10^{-3}) and enhancer (P_{luc}=3.0x10^{-3}) activity in Jurkat cells, increased promoter activity in A253 cells (P_{luc}=7.0x10^{-4}), and increased enhancer activity in 293T cells (P_{luc}=1.9x10^{-2}), but decreased promoter activity in THP1 and 293T cells (P_{luc}<4.0x10^{-2}) (**Figure 5C**).

The risk allele of rs7117261 increased promoter activity (P_{luc}=7.0x10^{-4}) but decreased enhancer activity (P_{luc}=7.0x10^{-4}) in A253 cells, decreased promoter activity in 293T cells (P_{luc}=1.9x10^{-2}), and decreased enhancer activity in EBV B (P_{luc}=1.8x10^{-2}) and Jurkat cells (P_{luc}=2.5x10^{-2}) (**Figure 5D**). The risk allele of rs4938573 significantly decreased promoter activity in EBV B (P_{luc}=2.4x10^{-2}), Jurkat (P_{luc}<1.0x10^{-4}), and A253 (P_{luc}=3.1x10^{-3}) cells, but increased promoter activity in THP1 cells (P_{luc}<1.7x10^{-3}) (**Figure 5E**). Additionally, the rs4938573 risk allele decreased enhancer activity in EBV B cells (P_{luc}=4.3x10^{-2}), but increased enhancer activity in THP1 (P_{luc}=4.3x10^{-2}) and A253 (P_{luc}<1.0x10^{-4}) cells (**Figure 5E**). In summary, all five prioritized SNPs exhibited cell type- and allele-specific effects on promoter and/or enhancer activity with the A253 cells demonstrating the most allele-specific differences among all tested cell types; THP1 cells demonstrated the most allele-specific effects among tested immune cell lines (**Figure 5F**).

Three of the four SNPs in the regulatory region between *DDX6* and *CXCR5* (rs4936443, rs4938572, rs7117261) span a 292 bp region positioned 5' of a CTCF binding site that separates the SNP cluster from rs4938573 (**Figure 3A**). Further, the three SNPs have similar epigenetic enrichment, reported eQTLs, and chromatin-chromatin interactions, suggesting that they likely modulate the same regulatory element. To test whether the SNPs facilitate concordant or discordant allele-specific effects on regulatory activity, the non-risk or risk alleles of the three SNPs were cloned together on the promoter-less or minimal-promoter luciferase vector. When cloned together, the risk alleles of rs4936443, rs4938572, and rs7117261 significantly increased promoter (P_{luc}<1.5x10^{-2}) and enhancer activity (P_{luc}<2.8x10^{-2}) in EBV B, Jurkat, and 293T cells relative to the non-risk alleles (**Figure 5G**). In A253 cells, the three risk alleles also significantly increased promoter activity (P_{luc}=1.7.0x10^{-3}), but impaired enhancer activity (P_{luc}=9.0x10^{-4}) (**Figure 5G**). In contrast to the allele-specific activities observed when the SNPs were separated, THP1 cells did not exhibit allele-specific differences in promoter or enhancer activity when the alleles were cloned together (**Figure 5G**).

Collectively, these findings suggest that the three clustered SNPs likely function together to modulate the haplotype- and cell type-specific regulatory activity at this locus. Further, in A253 cells, a model of salivary gland epithelial cell function, the risk alleles of rs4936443, rs4938572, and rs7117261 may suppress enhancer activity and, therefore, expression of genes that interact within the chromatin regulatory network. The opposing allelic- and cell type-specific effects observed between rs4938573 and the three SNP cluster suggest that the ~978 bp region between *DDX6* and *CXCR5* likely

has more than one regulatory element, separated by a CTCF site, that are influenced by the *DDX6-CXCR5* SjD/SLE risk haplotype.

Chromosome architecture is similar between immune cells and salivary gland at the *DDX6-CXCR5* risk locus.

Coalescence of reported Hi-C chromatin-chromatin interactions and eQTLs suggest that the five prioritized SNPs are positioned in regulatory elements that likely modulate the expression of *DDX6*, *CXCR5*, and/or other genes within the chromatin regulatory network. In further support, the SNPs are positioned in areas of open chromatin (ATAC-seq peaks), elevated promoter and/or enhancer activity (H3K4me1, H3K4me3, H3K27ac peaks), and active transcription factor binding in B cells, T cells, monocytes, macrophages, and neutrophils (**Supplemental Figures 11-16**), resulting in high IMPACT scores (**Supplemental Figure 17**). Lastly, the cell type- and allele-specific effects observed in the EMSAs and luciferase assays indicate that the risk alleles of these five variants likely influence the regulatory activity across the interval to modulate gene expression in the context of SjD and/or SLE.

To test this and address the limited availability of Hi-C data for the tested cell lines used in the functional assays, 3C-qPCR was performed in four patient-derived EBV B cell lines (two homozygous for the *DDX6-CXCR5* non-risk or risk haplotype; **Supplemental Table 2**), A253, Jurkat, THP1 and 293T cell lines using rs574994551 or rs4938572 to tag the anchor fragments determined by HindIII digestion (**Figure 3A, 6A-B**). 3C-qPCR results were also plotted in the context of cell type-specific RNA-seq, ATAC-seq, epigenetic marks, and/or Hi-C data (**Figure 6C-D; Supplemental Figures**

11-16). Publicly available epigenetic and genomic data on salivary gland tissues are limited and Hi-C or long-range chromatin interaction data are nonexistent. To provide additional insights in the salivary gland, 3C-qPCR results from the A253 cells were plotted in the context of in-house CUT & RUN (H3K27me3 pull-down primer), ATAC-seq, and RNA-seq data (**Figure 6D**).

3C-qPCR designed with the anchor spanning rs57494551 revealed the highest relative interaction frequency (RIF) with the upstream promoter region of the long non-coding RNA, *lnc-PHLDB1-1*, in all tested cell types except THP1 and 293T (**Figure 6A**). Hi-C data also revealed this interaction in all examined cell types except B cells, M2 macrophage, neutrophils, and 293T cells (**Figure 6C**; **Supplemental Figure 11-16**). Consistent with the allele-specific increase in luciferase activity across several cell types (**Figure 5A**), the scaled effect size of reported eQTLs also indicate that the rs57494551 risk allele increases *lnc-PHLDB1-1* expression in the minor salivary gland and whole blood (**Supplemental Figure 18A**).

The rs57494551 regulatory region also interacted with the shared promoter region of *BCL9L* and *UPK2*, albeit with lower RIF, across all cell types (**Figure 6A**). Hi-C consistently showed this interaction in other immune cell types (**Figure 6C; Supplemental Figure 11-16**), as well as an additional cell type-specific (EBV B, B, and $CD8^+$ T cells) interaction with the promoter of *TRAPPC4* further downstream (**Figure 6C; Supplemental Figure 11, 12B**). An interaction with *TRAPPC4* above the RIF of the short-range control was not observed in the 3C-qPCR of EBV B cells (**Figure 6A,B**). Interestingly, the scaled effect sizes of the *TRAPPC4* eQTL (minor salivary gland) and *BCL9L* eQTL (blood and EBV B cells) suggest that the rs57494551 risk allele may

exhibit cell type-specific suppression of the long-range enhancer effect of this region (**Supplemental Figure 18A**). However, a discordant observation of increased luciferase enhancer activity in the context of the rs57494551 risk allele (**Figure 5A**) is indicative of a more complex regulatory mechanism that is likely allele-specific, cell type-specific, and modulated in the context of the chromatin regulatory network.

3C-qPCR designed with the anchor spanning the SNP cluster (tagged by rs4938572) also revealed the highest RIF with the promoter region of *lnc-PHLDB1-1*, in all tested cell types (**Figure 6B**) and Hi-C data from T cells (**Supplemental Figure 12**). Consistent with impaired luciferase enhancer activity in A253 cells in the context of the combined risk alleles (**Figure 5G**), the scaled effect size of *lnc-PHLDB1-1* was reduced in minor salivary gland in the context of the rs4936443, rs4938572, and rs7117261 risk alleles (**Supplemental Figure 18B-D**). The risk alleles of rs4938572 and 4938573 also exhibited decreased effect size for *lnc-PHLDB1-1* in $CD4^+$ and $CD8^+$ T cells and natural killer cells but increased scaled effect size in peripheral $CD14^+$ monocytes (**Supplemental Figures 18C, E**). Consistently, increased luciferase enhancer activity in THP1 cells, but decreased activity in Jurkat cells, was observed in the context of rs4938573 alone (**Figures 5E**). Allele-specific enhancer activity was not observed in THP1 cells in the context of rs4938572 alone or in combination with rs4936443 and rs7117261 (**Figure 5E, G**). Given that the rs4938573 is separated from the three SNP cluster by a CTCF site, the discordant observations in observed luciferase enhancer activity and eQTL effect sizes may indicate that rs4938573 modulates the activity of a separate enhancer that forms independent interactions with the *lnc-PHLDB1-1* promoter.

Chromatin-chromatin interactions were also observed, albeit with lower RIF, between the regulatory region tagged by rs4938572 and the shared promoter region of *BCL9L* and *UPK2* in all tested cell types (**Figure 6B**) and in Hi-C data from EBV B, $CD4^+$ and $CD8^+$ T cells (**Figure 6C; Supplemental Figure 12**). The risk alleles of all four SNPs in this region are associated with increased expression of *BCL9L* in blood, EBV B cells, and $CD16^+$ neutrophils (rs4938572 and rs4938573 only) (**Supplemental Figure 18B-E**). Interestingly, Hi-C data from B and T cells, macrophage, and neutrophils demonstrated interactions between this regulatory region and the rs57494551 regulatory region, but not the *BCL9L-UPK2* promoter (**Supplemental Figure 11-12, 14-15**). The risk alleles of rs4938572, rs4936443, and/or rs4938573 all exhibited reduced scaled effect size for *DDX6* in PBMCs (rs4938572, rs4936443 only), B cells (rs4938572 only), T cell, and monocyte populations (**Supplemental Figure 18B,C,E**). This finding is discordant with the observed increase in luciferase enhancer activity in EBV B and Jurkat cells in the context of three risk alleles together (rs4936443, rs4938572, and rs7117261) and rs4938573 alone (**Figure 5E, G**). Discordance between the observed allele-specific effects in the luciferase assays and the directionality of reported eQTLs may result from the potential regulatory effects of the native chromatin architecture that cannot be recapitulated in the luciferase assay, the influence of additional SNPs that are carried on *DDX6-CXCR5* risk haplotype, and additional regulation by non-coding RNAs, among others.

DISCUSSION

This study performed fine mapping across two unique but related autoimmune diseases, SjD and SLE, to identify the most likely functional SNPs in the shared *DDX6-CXCR5* risk locus, hypothesizing that the identification and functional characterization of common SNPs may identify common disease mechanisms. Although the statistical genetic approach revealed a similar genetic association profile across the two diseases, strong co-inheritance and LD across the interval rendered it unable to identify SNPs with likely functional potential. As an alternative approach, this study leveraged publicly available datasets and databases to perform bioinformatic fine mapping of the top 100 common SNPs, identifying SNPs with cumulative evidence of transcription factor binding enrichment, epigenetic marks of regulatory activity, and reported eQTLs and their position in the chromatin regulatory network. This approach successfully prioritized five SNPs that exhibited both allele- and cell type-specific regulatory activity in subsequent functional assays and predicted the observed minimal functional activity of several additional SNPs, including the index SNPs, providing strong justification for using bioinformatic fine mapping to discern likely functional SNPs from other co-inherited non-functional SNPs carried on a disease risk haplotype. Although the statistical and bioinformatic fine mapping examined an extensive credible SNP set, as well as SNPs previously implicated in either SjD or SLE, the *DDX6-CXCR5* risk interval likely has additional functional SNPs not examined in this study.

Noncoding functional SNPs influence disease mechanisms by altering the expression of key regulatory genes[30-34]. Risk alleles can modulate transcription factor binding, activity of promoters and/or regulatory elements, and organization of the chromatin regulatory network resulting in altered long-range interactions between

promoters and distant regulatory elements. Deep bioinformatic interrogation of the five selected SNPs identified three potential regulatory elements on the *DDX6-CXCR5* locus with evidence of allele- and cell type-specific promoter and/or enhancer activity that likely modulate expression of *DDX6* and *CXCR5*, as well as other genes spanning the interval through local and/or distant regulatory activity. Many of these transcripts have functional implications in established autoimmune disease mechanisms, including *DDX6*, *CXCR5*, and *BCL9L*[30,32,34]. Others, such as *lnc-PHLDB1-1*, have not been previously reported in SjD or SLE.

DDX6 is a central regulator of the viral RNA recognition pathway of the innate immune system and, as such, an important modulator of type I interferon responses[23-25,35]. Genetic ablation of *DDX6* in human cells induced global upregulation of type I interferon-stimulated genes (ISGs)[24]. Chronic upregulation of ISG expression, i.e., the interferon signature, is a hallmark of SjD and SLE[7,24,36-42]. *DDX6* is a reported eQTL of rs4936443, rs4938572, and rs4938573, where the risk alleles exhibited strong negative relative eQTL effect sizes in PBMCs and isolated B and T cell populations (**Supplemental Figure 18B-C,E**). Further, an interaction between the regulatory region carrying rs4936443, rs4938572, and rs4938573 and a region spanning part of the *DDX6* promoter was observed in the immune and A253 cell lines by 3C-qPCR (**Figure 6B**) and primary human B cells and $CD4^+$ and $CD8^+$ T cells by Hi-C (**Supplemental Figure 11-12**). Collectively these data suggest that the expression of *DDX6* may be regulated through a long-range interaction with the three SNP cluster (tagged by rs4938572) and/or rs4938573 enhancer regions.

The three SNP cluster enhancer and the rs4938573 enhancer are separated by a CTCF site and exhibit opposing allele-specific nuclear binding and luciferase activity across several of the cell lines tested but, due to the location of the HindIII sites used in the 3C-qPCR and Hi-C, are positioned in the same loop anchor. Consistent with the reported eQTL effect sizes, the risk allele of rs4938573 significantly impaired promoter and enhancer luciferase activity, as well as nuclear protein binding, in the EBV B and Jurkat T cells (**Figure 4,5**). Despite being a reported eQTL of *DDX6*, the risk allele of rs4938572 increased luciferase activity and nuclear protein binding across several cell types when examined separately and as part of the three SNP cluster (**Figure 4,5**). Since eQTLs are examined in the context of LD and co-inheritance, it is possible that the discordance between the functional studies and the reported eQTL effect sizes for the three SNP cluster are due to the co-inheritance between the three SNPs and rs4938573. In summary, these findings suggest that the rs4938573 risk allele may modulate a long-range interaction between the rs4938573 enhancer and the *DDX6* promoter to impair *DDX6* expression in the context of the *DDX6-CXCR5* SjD/SLE risk haplotype.

CXCR5 is expressed on the surface of B and T cell subsets. CXCR5-CXCL13 signaling activates the recruitment and infiltration of immune cells to lymphoid germinal centers and the germinal center-like structures that form in the salivary gland of SjD patients and kidneys of SLE patients[26,27,43]. SjD patients carrying the *DDX6-CXCR5* risk haplotype (tagged by the risk allele (T) of rs4938573) reportedly exhibit increased infiltration of $CXCR5^+$ immune cells in the salivary gland, but lower *CXCR5* expression in circulating $CXCR5^+$ cells[27]. Aqwari, et al.[27] hypothesized that impaired *CXCR5*

expression in $CD19^+$ B cells may initially delay activation of the CXCR5-CXCR13 signaling axis in genetically predisposed individuals, resulting in early glandular destruction. CXCR5 was a reported eQTL for the SNPs in the three SNP cluster and rs4938573, where the risk alleles exhibited strong negative relative eQTL effect sizes in B cells, but positive effect sizes in neutrophils (**Supplemental Figure 18B-E**). CXCR5 was also a reported eQTL for rs57494551, where the risk allele exhibited increased relative expression in macrophage and neutrophils. Since rs4938573 is positioned ~12.5 kb upstream of the CXCR5 promoter and exhibited allele-specific loss of promoter and enhancer activity in the EBV B, Jurkat, and A253 cell lines, but had increased nuclear protein binding, we hypothesize that the rs4938573 regulatory element may function as a proximal repressor to reduce *CXCR5* expression in the context of the *DDX6-CXCR5* risk haplotype. The location of HindIII restriction cut sites prohibited separation of rs4938573 from the three SNP cluster and, therefore, examination of potential long-range interactions between the three SNP cluster and the *CXCR5* promoter in this study.

Like in *DDX6*, the rs57494551 regulatory element likely modulated *CXCR5* through chromatin-chromatin interactions in immune, salivary gland, and kidney cells. Luciferase and eQTL data suggest that the rs57494551 risk allele increases enhancer activity and expression of *CXCR5* in B cells, but decreases in T cells; opposite effect of the other SNPs. Because activation and modulation of the CXCR5-CXCL13 axis is critical for adaptive immune responses, it is possible that these regulatory mechanisms work synergistically to titrate *CXCR5* expression and that genetic susceptibly from the *DDX6-CXCR5* risk haplotype dysregulates this mechanism, resulting in delayed &

538 exacerbated signaling. SNP rs57494551 exhibited minimal evidence of *DDX6*
539 regulation, despite being positioned within its first intron.

540 *Lnc-PHLDB1-1* is a previously uncharacterized lncRNA located in the
541 promoter/enhancer of *TREH* and downstream of *DDX6.* Mined RNA-seq datasets from
542 the Expression Atlas revealed that although not characterized, lnc-PHLDB1-1 does
543 exhibit downregulated expression in whole blood of individuals with juvenile idiopathic
544 arthritis (JIA), several primary and cultured immune cell types under varying
545 inflammatory conditions, and in multiple types of cancer (**Supplemental Figure 19**)[44].
546 Further, rs4938572 and rs4938573 are strong eQTLs of *lnc-PHLDB1-1* in T cells and
547 macrophages (**Figure 3E, G; Supplemental Figure 18C, E**). This lncRNA region has
548 no CTCF binding sites or H3K27Ac peaks (data not shown) but, because the looping
549 activity is so strong, we propose that *lnc-PHLDB1-1* may be acting as a regulatory
550 mechanism in this risk interval. The functional significance of *lnc-PHLDB1-1* remains
551 unclear; however, we speculate that *lnc-PHLDB1-1* may function as an enhancer
552 lncRNA (elncRNA). In a previous lupus study, it was discovered that a functional SNP
553 regulated allele-specific expression of IRF8 through a mechanism involving the
554 enhancer RNA AC092723.1[45]. Similar to the findings of this study, the prioritized SNPs
555 in this study are located in (rs57494551) or near (rs4938572, rs4936443, rs7117261,
556 and rs4938573) an H3K27ac peak that could modulate the expression of the risk
557 interval through *lnc-PHLDB1-1* by looping the DNA into physical proximity[45].
558 Identification and characterization of lncRNAs and their functions in health and human
559 disease is of growing interest, but continues to be limited by classification standards,

560 inclusion in publicly available bioinformatic databases, and technologies to characterize
561 them.

562 *BCL9-like* (*BCL9L*) gene is a homology gene of BCL9 that functions within the
563 Wnt and B-catenin signaling pathways to drive EMT. In breast cancer, *BCL9L* was also
564 shown to promote T cell infiltration of the lymphoid tissue[46,47]. 3C-qPCR and Hi-C
565 analyses suggest that the rs57494551 enhancer may form a chromatin interaction with
566 the *BCL9L* promoter in cell lines examined. Although rs4938573 cluster also
567 demonstrated interactions in the 3C-qPCR, the interactions were not observed by Hi-C.
568 The eQTL analyses suggest that *BCL9L* expression is upregulated in blood and
569 neutrophils in the context of the risk allele. Further, rs4938573 is also a reported eQTL
570 of follicular lymphoma[48,49]. The evidence provided here suggest that rs4938573 may
571 act as a tag SNP for rs57494551 and that risk alleles of rs57494551 may increase
572 enhancer activity and interactions with the *BCL9L* promoter.

573 TRAPPC4 is a reported eQTL with elevated positive effect size in the salivary
574 gland. Interestingly, Hi-C revealed a chromatin interaction between the rs57494551
575 enhancer and the promoter of *TRAPPC4* in primary B and T cell subsets. This
576 interaction was not observed by 3C-qPCR above the RIF threshold, indicating potential
577 important differences between cell lines and primary cells. *TRAPPC4* encodes an
578 important regulator of autophagy[50,51]. Dysregulated autophagy and subsequent
579 accumulation of potential autoantigens is implicated in the loss of self-tolerance in
580 autoimmune diseases, including SjD and SLE[52]. Collectively, evidence discovered in
581 this study suggest that the *DDX6-CXCR5* risk haplotype may play a role in the

582 autophagic dysregulation in target tissues (salivary gland and kidney) through regulation
583 of *TRAPPC4*.

584 In summary, deep interrogation of publicly available and cell type-specific
585 bioinformatic data to prioritize and gain insights into the functionality of common SNPs
586 spanning the *DDX6-CXCR5* SjD/SLE risk haplotypes. Functional interrogation of
587 nuclear protein binding, enhancer/promoter activity, and interactions with the local
588 chromatin regulatory network revealed important insights into how genetic susceptibly at
589 this locus may influence immune cell dysregulation and loss of self-tolerance in disease
590 target tissues through cell type- and allele-specific expression of *DDX6* and *CXCR5*, as
591 well as other genes in the local network (*BCL9L*, *TRAPPC4*, *lnc-PHLDB1-1*) that have
592 not been previously implicated and/or functionally characterized in autoimmune disease.
593 Further, four of the five SNPs prioritized for functional characterization in this study have
594 also been associated with other autoimmune diseases: rheumatoid arthritis (rs7117261,
595 rs4938572, and rs4936443)[53], primary biliary cholangitis (rs7117261)[54], and follicular
596 lymphoma (rs4938573)[48,49]. SNP rs4938573 is also a reported SLE- and SjD-
597 associated sex-influenced eQTL[55]. The results of this study performed across multiple
598 immune and disease-specific tissue cell types may offer important mechanistic insights
599 into other diseases beyond autoimmunity.

MATERIALS AND METHODS

Study Datasets and Quality Control.

Deidentified genotype, Immunochip, and genomic summary results were obtained with approval from and analyzed in accordance with the Oklahoma Medical Research Foundation Institutional Review Board. Dataset (DS) 1 included a) genotype data from 3,232 SjD case and 17,481 controls of European ancestry and b) Immunochip data from 619 SjD cases and 6,171 controls of European ancestry (**Figure 1A**)[1]. DS2 included a) Immunochip data from 3,762 SLE cases and 6,194 controls of European ancestry[20], b) summary statistics from the meta-analysis of 6,904 SLE cases and 18,429 controls of European ancestry, and c) GWA scan summary results from 1,174 SLE cases and 4,246 controls of Korean ancestry (**Figure 1A**)[28]. DS3 included Immunochip data from an independent cohort of 1916 SjD cases analyzed with the 6,194 population controls from DS2 (**Figure 1B**)[20]. DS4 included Immunochip data from the 3,762 SLE cases and 6,194 population controls from DS2 (**Figure 1B**)[20].

Written informed consent was obtained in accordance with the institutional review board of each respective study[1,20,28]. All SjD cases fulfilled the America-European Consensus Group (AECG) criteria for primary SjD according to clinical evaluations performed within their respective studies[56-58]. All SLE cases fulfilled at least four of the eleven American College of Rheumatology revised criteria for SLE according to clinical evaluations performed within their respective studies[59,60].

All data were subjected to the following quality control (QC) measures as previously described[1]: i) a well-defined cluster scatter plot; ii) MAF>1%; iii) SNP call rate >95%; iv) sample call rate >95%; v) Hardy-Weinberg equilibrium test with p>0.001 and

rate >95% in controls; vi) p>0.001 for differential missingness between cases and controls. Individual data with excessive heterozygosity (>5 s.d. from mean) and relatedness determined by identity-by-descent (IBD) >0.4 using PLINK (v1.9)[61] were excluded. EIGENSTRAT[62] was used to identify population substructure using independent genetic markers (r^2<0.2 between variants).

Imputation and Statistical Analyses

Genotype and Immunochip data were imputed for the *DDX6-CXCR5* region (chr11: 116194547-121194547; hg19). Imputation was performed using the Haplotype Reference Consortium panel version 1.1 in the Michigan Imputation Server[63]. Data were prephased using SHAPEIT and imputed using Minimac3[64,65]. Imputed variants with imputation quality score (INFO) >0.5 and that fulfilled the above QC criteria were included. GWA Scan data was imputed separately using Z-scores of the missing variants in the summary statistic. Meta-analysis summary statistics were imputed using Ssimp v.0.5.6 imputation software and the East Asian population from 1000 Genome Projects as a reference panel for LD calculation[66]. After filtering SNPs with MAF>1%, 66,397 variants were available for a meta-analysis.

Logistic regression models in PLINK (v1.9) were utilized to assess SNP-trait association in the genotype and Immunochip data, adjusting for the first 3 (or 4) principal components. Stepwise logistic regression was used to adjust for the most significant variant with a P<0.0001. Results from the logistic regressions or summary statistics were meta-analyzed using a fixed-effect model in METAL by weighing the SNP effect by sample size. Meta-analyses were performed between/among: 1) data from SjD datasets (DS1), 2) data from SLE datasets (DS2), and 3) data from merged

SjD and SLE datasets (**Figure 1A**). Both Cochran's Q test statistic and I^2 index were used to evaluate meta-analysis heterogeneity. LocusZoom[67] was used to plot meta-analyses results. Linkage disequilibrium (LD) and haplotype reconstruction were calculated utilizing the solid spine of the LD block with minimum r^2 value of 0.8 at HAPLOVIEW (4.2)[68].

Fine mapping and Functional Annotations

Fine mapping was implemented using Trinculo[29] with multinomial logistic regression in the *DDX6-CXCR5* region using genotype and/or Immunochip data (**Figure 1A – DS1; Figure 1B – DS3, DS4, Merged DS3+DS4**). For each SNP tested, population structure was accounted for by including the first four principal components as covariates. The 95% credible sets were derived from the likelihood function of the multinomial logistic model. Posterior probabilities were calculated for each SNP.

Publicly available databases were mined and bioinformatic functional fine mapping performed to further prioritize candidate functional variants in the 95% credible set. RegulomeDB[69,70], HaploReg v.4.1[71], and UCSC Genome Browser[72] were used to screen the 95% credible set for likely functional SNPs (**Figure 1C**). Candidate functional SNPs in the *DDX6-CXCR5* region were further evaluated for reported cell/tissue-specific cis-eQTL (P<0.05) using FUMA[73], the eQTL catalogue[74], single-cell RNA sequencing eQTLs[75], DICE (https://dice-database.org), GTEx v08 (https://gtexportal.org), QTLBase (http://www.mulinlab.org/qtlbase), and NephQTL (https://nephqtl.org) (**Figure 1C**). eQTL significance was further grouped based on p-values ($p<5x10^{-15}$ to $5x10^{-3}$, $p<5x10^{-3}$ to $1x10^{-2}$, and $p<1x10^{-2}$ to 0.05), Promoter-

capture Hi-C data from the 3D Genome Browser at the YUE lab, Northwestern University (http://3dgenome.fsm.northwestern.edu/chic.php), were mined for chromatin-chromatin interaction start and end coordinates within ± 5kb of each SNP (**Figure 1C**). The pyGenomeTracks tool was used to visualize chromatin interactions in genomic space[76]. Reported eQTLs and chromatin interactions were analyzed in the following cell types/tissues: GM12878 EBV B cell line and primary human total B cells, $CD4^+$ and $CD8^+$ T cells, monocytes, macrophages (M0, M1 and M2), neutrophils, salivary gland tissue, kidney tissue, and whole blood. Promoter-capture Hi-C data was not available for kidney and salivary gland.

Tissue Culture and Maintenance

Patient-derived EBV B cells[77], Daudi B, and Jurkat T cells (ATCC #CCL-213 and #TIB-151 respectively) were grown to 80% confluency using RPMI 1640 complete media (Corning #10-040-CM) supplemented with 10% fetal bovine serum (FBS) (Atlanta Biologicals, #S12450H), 1% penicillin/streptomycin, and 1% L-glutamate in 75-cm^2 vented tissue culture flasks at 37°C under 5% CO_2. THP1 monocytes (ATCC, #TIB-202) were grown in the above complete RPMI media supplemented with 0.05mM 2-mercaptoethanol (Sigma, #M3148). 293T human embryonic kidney epithelial cells (ATCC, #CRL-3216) were grown in complete Dulbecco's Modified Eagle's Medium (DMEM) (Thermo Fisher, #11995-065) supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% L-glutamate at 37°C under 5% CO_2. Adherent A253 cells (ATCC, #HTB-41) were grown in McCoy's 5a medium (ATCC, #30-2007) supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% L-glutamate at 37°C under 5% CO_2. Cell viability was assessed using Trypan blue as a live/dead stain and

visualized and counted on the Countess II Automated Cell Counter (Thermo Fisher, #AMQAX1000).

SNP Genotyping of Tissue Cultures (TaqMan Genotyping Assay)

TaqMan assays was used to genotype Jurkat T cells, THP1 monocytes, and A253 cells. Total genomic DNA was isolated from $1x10^6$ cells using the Zymo Quick-DNA Miniprep Plus Kit (ZymoResearch, #D4068) and quantified. Taqman reactions were performed on a QuantStudio 6 qPCR machine (Thermo Fisher, #A43168) following manufacturer instructions using ~25ng of DNA, TaqMan Genotyping Master Mix (ThermoFisher, #4371355), and TaqMan SNP Genotyping probes for rs57494551 (ThermoFisher, #4351379, Assay ID#C_90471390_10) and rs4938572 (ThermoFisher, #4351379, Assay ID#C_28012712_10) (**Supplemental Table 1**). Genotypes for each allele were compared to EBV B cell lines homozygous or heterozygous for the minor or major allele of rs57494551 and rs4938572 (**Supplemental Table 2**). A minimum of three biological replicates were analyzed in triplicate. Genotyping results were analyzed using the Design & Analysis Software package from ThermoFisher.

Electromobility shift assay (EMSA)

Nuclear lysates from EBV B cells, Daudi B, Jurkat T, THP1 monocyte, and A253 cells were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Fisher Scientific, #78833) supplemented with Halt Protease Inhibitor Cocktail 100x (Thermo Fisher, #78425), DTT (Fisher Scientific, #R0861) and PMSF (Sigma Aldrich, #P7626-5G). Complementary pairs of ~60bp non-risk or risk oligonucleotide probes for each SNP were chemically synthesized (IDT) (**Supplemental Table 3**), annealed using 10X

Oligonucleotide Annealing Buffer (100mM Tris-HCl pH 7.5, 1M NaCl, 10mM EDTA pH 7.4) in 95°C for 2 minutes, and end-labeled with (γ-P^{32}) adenosine triphosphate (MP Biomedicals or Perkin Elmer, #NEG002A250U) using T4 polynucleotide kinase (NEB, #M0201S). Labeled oligos were purified using MicroSpin G-25 Columns (VWR, #95017-621), then scintillation counts were performed using Liquid Scintillation Analyzer Tri-Carb 4910 TR (Perkin Elmer). Nuclear protein extracts (5µg) were incubated for 30 minutes at room temperature with labeled probes (50,000cpm/reaction) in binding buffer (1µg poly (dI-dC), 20mM HEPES, 10% Glycerol, 1mM $MgCl_2$, 0.5mM EDTA, pH 8.0, 50mM NaCl and 10mM Tris-HCl, pH 7.4). DNA-protein complexes were resolved on non-denaturing 5% acrylamide gels. Competition assays were performed by adding 100x unlabeled probe and incubating for 15 minutes on ice. Differential binding was assessed by densitometry analysis of photostimulated luminescence (PSL) using Molecular Imager PharosFX System (Bio-Rad, #170-9450) and Quantity One 1-D analysis software (Bio-Rad). Band density was normalized to the background and statistically analyzed by Student's t-test in Prism 9 software.

Luciferase Reporter Assay

A 500 bp gBlock (IDT) carrying the non-risk or risk allele of each SNP (**Supplemental Table 4**) were cloned into the no-promoter firefly luciferase reporter plasmid pGL4.14 [luc2/Hygro] (Promega, #E6691) or minimal promoter firefly luciferase reporter plasmid pGL4.26 [luc2/miniP/Hygro] (Promega, #E8441). Note, each gBlock had 1 SNP with risk allele in the context of the non-risk alleles for all other SNPs in close proximity. Plasmids were double digested with HindIII-HF (NEB, #R3104S) and

738 KpnI-HF (NEB, #R3142S) restriction enzymes. A Thymidine Kinase (TK) promoter-
739 driven Renilla luciferase plasmid pGL4.74 [hRluc/TK] (Promega, #E6921) was used as
740 a transfection efficiency control.

741 Plasmids were transfected using either the 4D Nucleofector (EBV B, Daudi,
742 THP1; Lonza, 4D-Nucleofector Core Unit #AAF-1002B and 4D-Nucleofector X Unit
743 #AAF-1002X), the Neon Transfection System (Jurkats; ThermoFisher, #MPK5000S), or
744 Lipofectamine 3000 Transfection Reagent (HEK 293T, A253 cells; Life Technologies,
745 #L3000015). The SF Cell Line 4D-Nucleofector X Kit S (Lonza, #V4XC-2032) was used
746 to transfect EBV B cells and Daudi cells; the SG Cell Line 4D-Nucleofector X Kit S
747 (Lonza, #V4XC-3032) was used to transfect THP1 monocytes. One million cells were
748 resuspended in 20µL of SF or SG buffer and transfected with either pGL4.14 or
749 pGL4.26 plasmid (1µg/µL) containing the appropriate gBlock or empty vector, and
750 pGL4.74 (125ng/µL) Renilla as a transfection control using program EH-100 (EBV B),
751 CA-137 (Daudi), or FF-100 (THP1). After transfection, cells were rested at room
752 temperature for 15 minutes, then 80µL of pre-warmed complete media was added and
753 cells incubated at 37^{o}C for 1 hour. Transfected cells were then transferred into a 96-
754 well plate containing 100µL of pre-warmed complete media for a final volume of 200µL.
755 Cells were collected, washed, lysed, and luciferase activity measured after 24 hours
756 using the Dual-Luciferase Reporter Assay System (Promega, #E1980) and the Synergy
757 H1 Hybrid Multi-Mode Microplate Reader with dual injectors (BioTek, #8041000) or the
758 GloMax Explorer Multimode Microplate Reader (Promega, #GM3500). Relative
759 luciferase activity (RLA) was reported as a ratio of firefly luciferase to Renilla

transfection control. A minimum of three experimental replicates were plotted and analyzed by Student's t-test.

Jurkat cells were transfected using the Neon Transfection System with the Neon Transfection System 10µL Kit (Thermo Fisher, #MPK1096) according to manufacturer instructions with modification. Briefly, $4x10^5$cells/mL were transfected with either pGL4.14 or pGL4.26 (1µg/µL) containing the appropriate gBlock or empty vector, and pGL4.74 (62.5ng/µL) Renilla transfection control in Buffer R. Cells were pulsed with a voltage of 1600v, pulse width 10, and pulse number 3. Luciferase activity was measured and analyzed after 24 hours as described above.

A253 and 293T cells were transfected with Lipofectamine 3000 Transfection Reagent following manufacturer protocols. Briefly, $8x10^4$ cells were seeded per well of a 24-well plate. After 24 hours, media was replaced with a mixture of Lipofectamine 3000 and pGL4.14 (1µg/µL) or pGL4.26 (1µg/µL) containing the appropriate gBlock or empty vector, and pGL4.74 (10ng/µL) Renilla transfection control diluted in Opti-MEM Reduced Serum Medium (Life Technologies, #31985-070). Cells were lysed and luciferase activity analyzed after 24 hours as described above.

Chromatin Conformation Capture with Quantitative PCR (3C-qPCR)

3C was performed as described with few modifications[78]. Briefly, $1x10^6$ cells from EBV B cell lines homozygous for non-risk or risk genotypes at the SNP of interest, Jurkat T cells, or A253 cells were crosslinked with 2% formaldehyde for 10 minutes. Crosslinking was quenched with 0.125M glycine for 20 minutes then cell pellets were flash frozen in liquid nitrogen and stored at -80°C overnight. Cell pellets were thawed

on ice and lysed with fresh lysis buffer (10mM TrisHCl, 0.2% NP-40, 10mM NaCl, 1x Protease Inhibitor Cocktail (Life Technologies, #78425)) for 30 minutes on ice. Nuclei were pelleted and washed with 1.25x CutSmart Buffer (New England Biolabs (NEB), #B7204S), then incubated with 0.3% SDS at 37°C for one hour and then 2% Triton X-100 at 37°C for one hour. Nuclei were incubated with 3000U of HindIII-HF (NEB, #R3104M) at 37°C overnight while shaking to digest chromatin, then ligation incubation buffer (10mg/mL BSA, 1x T4 DNA Ligase Reaction Buffer (NEB, #B0202S) and 2000U of T4 DNA Ligase (NEB, #M0202S) were added and incubated overnight on ice. 3C ligation mixtures were de-crosslinked with Proteinase-K at 65°C overnight. Samples were treated with RNaseA for 1 hour at 37°C and then purified with phenol:chloroform:isoamyl alcohol 25:24:1, twice.

Primers were designed to target 100 bp upstream of HindIII cut sites positioned in or near a Hi-C anchor; only fragments <5kb were considered (**Supplemental Table 5**). Primer efficiency was validated using three Bacterial Artificial Chromosome (BAC) clones (Thermo Fisher, #RP11-1011H15, #CTC-3245B9, #RP11-45N4) spanning the *DDX6-CXCR5* locus, mixed in equimolar concentrations. All samples were run on a QuantStudio 6 qPCR machine (Thermo Fisher, #A43168). CT values were normalized to standard curves generated from the BAC clones. Relative interaction frequency (RIF) of each 3C primer pair was calculated by comparing the interaction frequency from each 3C primer pair with the interaction frequency from the short-range interaction primer control. Graphs were plotted using Prism V9 and student's t-test was used to calculate standard deviation.

806 **RNA-sequencing**

807 A253 cells grown to 80-85% confluence were washed, trypsinized, and counted.
808 One million cells in triplicate were collected and resuspended in 700µL of Trizol. RNA
809 was isolated using the Direct-zol RNA Microprep kit following manufacturer instructions
810 (Zymo Research, #R2060) and quantified by fluorometric analysis (Thermo Fisher Qubit
811 fluorometer). mRNA was isolated and libraries made using the Swift Biosciences Rapid
812 RNA Library kit (Swift Biosciences, #R2024) according to manufacturer instructions.
813 Libraries were indexed using Swift Biosciences Unique Dual Indexing primers (Swift
814 Biosciences, #X9096). Libraries were quantified, pooled, and sequenced using an
815 Illumina NovaSeq 6000 with PE150 reads. Twenty million total sequenced reads were
816 quality assessed using FastQC tool and cleaned for adapters using trimmomatic
817 software (Bolger et al., 2014). QC reads were aligned to hg19 reference genome using
818 HISAT2 (Kim et al., 2019). SAMtools was used to sort and index the aligned files into
819 bam files (Li et al., 2009). Bam files were split by strandedness to obtain read coverage
820 for plus(+) and minus(-) strands.

821

822 **ATAC-sequencing**

823 In-house chromatin accessibility maps were generated using ATAC-seq data
824 from A253 cells. A253 cells grown to 80-85% confluence were released by
825 trypsinization, washed, and counted. Six replicates of $5x10^4$ cells were pelleted by
826 centrifugation and resuspended in 50 µL of cold lysis buffer (10 mM Tris-HCL, pH 7.4,
827 10 mM NaCl, 3mM MgCL2, 0.1% IGEPAL (NP-40)) for ~20 seconds. Nuclei were
828 collected by centrifugation and resuspended in 50 µL transposase mix (25 µL 2x TD

Buffer, 2.5 µL TDE1 transpose enzyme (Illumina Tagment DNA Enzyme and Buffer Small Kit, #20034197), and 22.5 µL nuclease-free water). After incubation at 37^{o}C for 30 minutes, all samples were cleaned using the Qiagen MinElute Kit (Qiagen, #28006) following manufacturer instructions and eluted in 10 µL EB buffer. DNA was amplified using NEBNext Ultra II Q5 PCR Master Mix (NEB, #M0544S) with Index Primers from Illumina (IDT for Illumina DNA/RNA UD Indexes Set A (Illumina, #20027213)). PCR amplified libraries were cleaned up using AMPure XP beads (Beckman Coulter, #A63880). Libraries were quantified, pooled, and sequenced using an Illumina NovaSeq 6000 with PE150 reads. Fifty million reads from each replicate were analyzed as follows: Paired-end reads (100 bp) were sequenced, quality of the reads was determined using FASTQC tool, adapters were trimmed using trimmomatic and bowtie2 was used to align the reads to hg19 reference genome with parameters --very-sensitive -X 2000 -k 10 --threads 8 –dovetail[79]. Aligned reads were sorted by SAMtools[80], then additional processing was performed to remove duplicate and nonuniquely mapped reads. Finally, peak calling was performed in sorted files using Genrich tool[81].

CUT&RUN

The CUT & RUN protocol (CUTANA ChIC/CUT & RUN Kit, EpiCypher, #14-1048) was performed according to manufacturer protocols with optimizations for A253 cells. For example, A253 cell viability increased when cells were detached with 0.05% trypsin, as opposed to manual scraping. A final concentration of 0.01% digitonin was optimal for 92-100% permeabilization of A253 cells and 2.5 µL of pAG-MNase for 2 hours was optimal sufficient cleavage of 6x10^{5} A253 cells into mononucleosome

fragments of ~300bp. Heterochromatin formation or gene inactivation were detected using antibodies against H3K4me3 (0.5 µg/sample, EpiCypher #13-0041), H2K27me3 (0.5 µg/sample, Invitrogen #MA5-11198), IgG (0.5 µg/sample, EpiCypher #13-0042). Isolated DNA pull-down fragments were quantified, and libraries prepared using NEB Ultra II DNA Library Prep Kit (NEB, #E7645S) and NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs Set 4) (NEB, E6446S) as index primers. Libraries were run on an Agilent TapeStation to assess quality and ensure enrichment of ~300bp fragments (~170bp DNA fragments + library sequencing adapters). After PCR enrichment, libraries were pooled and sequenced using Illumina NovaSeq 6000 with PE150 reads, 20 million reads per sample. Raw reads were quality controlled using FastQC and adapters were trimmed using trimmomatic tools. Clean reads were aligned to hg19 reference genome using Bowtie2. SAMtools was used to convert and sort SAM files into BAM format. Picard was used to mark and remove duplicate reads[82]. SEACR was used for peak calling[83]. Signals from sorted BAM files via bamCoverage were normalized by deeptools then computeMatrix was used to calculate overall signal distribution around the peaks center called by SEACR[84].

IMPACT Regulatory Annotation

To better understand binding patterns of transcription factors (TFs) around putative functional SNPs, the 707-cell type-specific IMPACT regulatory annotations resource was used[85]. IMPACT annotations for *DDX6-CXCR5* region were extracted and the total number of transcription factors with their probabilities was used to annotate

874 SNPs in the region. pyGenomeTracks tool was then used to visualize the number of
875 active TFs and their IMPACT scores in different cell types[76].
876
877
878 **Epigenetic colocalization**
879 Regulatory functions of putative SNPs were investigated by alignment with
880 publicly available and in-house epigenetic data. Chromatin segmentation maps of
881 GM12878, B, $CD4^+$ T, $CD8^+$ T cells, monocytes and neutrophils were obtained from the
882 Roadmap Epigenomics Consortium
883 [https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmMod
884 els/coreMarks/jointModel/final/]. RNA-seq and histone marks data of B, CD4+ T, CD8+
885 T cells, macrophages, monocytes, and neutrophils were obtained from BluePrint
886 Epigenome [https://epigenomesportal.ca/tracks/Blueprint/hg19/]. Histone marks data of
887 GM12878 and 293T, and RNA-seq data of GM12878 cell lines were obtained from
888 ENCODE database. The RNA-seq data of 293T was from NCBI short read archive
889 (GSM2258993).
890 The putative functional SNPs were colocalized in different cell types using a
891 SNP-trait association, fine mapping, mRNA expressions, histone marks, capture Hi-C,
892 3C, chromatin segmentation marks, and impact annotations along with hg19 reference
893 genome. The results were visualized using pyGenome tracks[76].
894
895 **Data availability**

SjD genome-wide association summary statistics and individual-level data used in Dataset 1 of this study were previously reported in[1] and are available through the Databases of Genotypes and Phenotypes (dbGaP) under accession number: phs002723.v1.p1 [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002723.v1.p1], phs000672.v1.p1 (n=735) [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000672.v1.p1], phs000428.v2.p2 (n=8,519) [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000428.v2.p2], phs000196.v3.p1 (n=995) [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000196.v3.p1], phs000187.v1.p1 (n=602) [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000187.v1.p1]. SLE Immunochip and genome-wide association (GWA) scan data used in Dataset 2 of this study were previously reported in[20,28] and are available from the corresponding authors of the original manuscript in accordance with respective Institutional Review Board approval and subject consent. All remaining data were generated by coauthors in this study and will be made available upon request to the corresponding author (chris-lessard@omrf.org) in accordance with an established material transfer agreement. All other data presented in this study were previously published and can be accessed by URLs provided in the methods section.

ACKNOWLEDGEMENTS

We thank all the research and clinical staff, consortium investigators, study participants, and funding agencies who made this study possible.

We also thank the investigators and funding mechanisms for the following dbGaP studies:

Phs002723.v1.p1: Genotype data from the international Sjogren's Genetics Network (SGENE) was reported in Khatri, et al. *Nat Commun*, 2022 and obtained through dbGAP accession number phs002723.v1.p1. The study was supported by the National Institutes of Health (NIH): R01AR073855 (Lessard); R01AR065953 (Lessard); R01AR074310 (Farris); P50AR060804 (Sivils), N01DE32636 (SICCA), HHSN26S201300057C (SICCA), U01DE028891 (SICCA), R03DE029800 (SICCA), U01HG004446 (SICCA-GWAS), P30AR070155 (SICCA-GWAS), R61AR076803 (Adrianto); NIDCR Sjögren's Syndrome Clinic and Salivary Disorders Unit were supported by funds awarded to BMW (Z01-DE000704) by the NIDCR Division of Intramural Research at the National Institutes of Health (Warner); Birmingham NIHR Biomedical Research Centre (Bowman); Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2155 – Projektnummer 390874280 (Witte); Research Council of Norway (Oslo, Norway) – Grant 240421 (Reksten), Grant 316120 (Wahren-Herlenius); Western Norway Regional Health Authority (Helse Vest) – 911807, 912043 (Omdal); Swedish Research Council for Medicine and Health (Rönnblom, Nordmark, Wahren-Herlenius); Swedish Rheumatism Association (Rönnblom, Nordmark, Wahren-Herlenius); King Gustav V's 80-year Foundation (Nordmark); Swedish Society of Medicine (Rönnblom, Wahren-Herlenius); Swedish Cancer Society (Baecklund); The Stockholm County Council (Wahren-Herlenius); The Swedish Twin Registry is managed through the Swedish Research Council under the grant 2017-000641. The SNP&SEQ Technology Platform was supported by Science for Life Laboratory, Uppsala University, the Knut and Alice Wallenberg Foundation and the Swedish Research Council (Nordmark).

Phs000428.v2.p2: This study used control data from the Health and Retirement Study in dbGaP (phs000428.v2.p2) submitted by David Weir, PhD at the University of Michigan and funded by the National Institute of Aging RC2 AG036495 and RC4 AG039029.

Phs000672.v1.p1: Genotype data from the Sjögren's International Collaborative Clinical Alliance (SICCA) Registry was obtained through dbGAP accession number phs000672.v1.p1. This study was supported by the National Institute of Dental and Craniofacial Research (NIDCR), the National Eye Institute, and the Office of Research on Women's Health through contract number N01-DE-32636. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health (NIH) to the Johns Hopkins University (contract numbers HHSN268200782096C, HHSN268201100011I, HHSN268201200008I). Funds for genotyping were provided by the NIDCR through

CIDR's NIH contract. Assistance with data cleaning and imputation was provided by the University of Washington. We thank investigators from the following studies that provided DNA samples for genotyping: the Genetic Architecture of Smoking and Smoking Cessation, Collaborative Genetic Study of Nicotine Dependence (phs000404.v1.p1); Age-Related Eye Disease Study (AREDS) - Genetic Variation in Refractive Error Substudy (phs000429.v1.p1); and National Institute of Mental Health's Human Genetics Initiative (phs000021.v3.p2, phs000167.v1.p1). We thank the many clinical collaborators and research participants who contributed to this research.

Phs000196.v3.p1: Investigators and Parkinson Disease patients that contributed to this Genome-wide Association Study of Parkinson Disease.

Phs000187.v1.p1: Research support to collect data and develop an application to support the High Density SNP Association Analysis of Melanoma project was provided by 3P50CA093459, 5P50CA097007, 5R01ES011740, and 5R01CA133996.

FUNDING

The content of this publication is solely the responsibility of the authors and does not represent the official views of the funding agencies. Research reported in this publication was supported by the National Institutes of Health (NIH): P3AR0073750 (James), UM1AI144292 (James), R01AR074310 (Farris), P50AR060804 (Farris), R33AR076803 (Adrianto), R21AR079089 (Adrianto), R01AR071410 (Tsao), R01AR071947 (Tsao); R01AR073855 (Lessard*), R01AR065953 (Lessard*); NIDCR Division of Intramural Research at the NIH: Z01-DE000704 (Warner); Sjögren's Foundation (Lessard*); Presbyterian Health Foundation (Lessard*); Assistance Publique-Hôpitaux de Paris (Ministry of Health) : PHRC 2006 P060228 (French ASSESS); Birmingham NIHR Biomedical Research Centre (Bowman); Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Germany's Excellence Strategy EXC2155: 390874280 (Witte); European Innovative Medicines Initiative Joint-Undertaking (IMI-JU): #115565 (Alarcón-Riquelme, PRECISESADS); NECCESSITY IMI-JU: #806975 (Alarcón-Riquelme); French Society of Rheumatology (French ASSESS) ; FOREUM Foundation for Research in Rheumatology (Jonsson, Appel, Wahren-Herlenius); King Gustav V's 80-year Foundation (Alarcón-Riquelme, Nordmark); National Research Foundation of Korea – Basic Science Research Program: NRF-2021R1A6A1A03038899 (Bae); NIHR Newcastle Biomedical Research Centre (UK Primary Sjögren's Syndrome Registry); NIHR Newcastle Clinical Research Facility (UK Primary Sjögren's Syndrome Registry); Research Council of Norway: 316120 (Wahren-Herlenius); Stockholm County Council (Wahren-Herlenius); Swedish Cancer Society (Baecklund); Swedish Heart-Lung Foundation (Wahren-Herlenius); Swedish Research Council for Medicine and Health (Rönnblom, Wahren-Herlenius, Nordmark); Swedish Research Council – Twin Registry: 2017-000641; Swedish Rheumatism Association (Rönnblom, Wahren-Herlenius, Nordmark); Swedish Society of Medicine (Rönnblom); Torsten Söderberg Foundation (Wahren-Herlenius); UK Medical Research Council: G080062 (Ng, UK Primary Sjögren's Syndrome Registry);

Western Norway Regional Health Authority (Helse Vest): 911807 (Omdal), 912043 (Omdal).

Author Contributions

C.J.L.* supervised the research and serves as senior author (*); M.M.W., B.K., M.L.J., A.M.S. conceived and designed experiments under the supervision of C.J.L.*; C.J.L.*, B.K., A.R., J-M.A., L.A.A., S-C.B., E.B., A.B., J.G.B., S.M.B., N.D., M-L.E., F.E., H.F-d'E., C.F., C.G., J-E.G., D.H., J.I-K., J.L.J., S.J.A.J., M.V.J., J.A.K., S.K., K.K., M.K., T.M., J.M., D.L.M., G.Nocturne, K.B.N., P.O., Ø.P., J-O.P., N.L.R., C.S., K.S., K.E.T., P.T., G.E.T., S.V., E.M.V., D.J.W., K.M.G., L.R., M.T.B., J.A.J., R.H.S., P.M.G., L.A.C., R.J., S.A., P.E., S.J.B., R.O., L.R., B.M.W., M.R., T.W., A.D.F., X.M., C.H.S., M.W-H., M.E.A-R., W-F.N., K.L.A., J.M.G., I.A., T.J.V., B.P.T., G. Nordmark collected, characterized, and/or genotyped the Sjögren's and/or SLE cases and population controls used in this study; S.B.G., J.A.K., K.M.G., J.M.G. manages the technology and/or data required for processing and analysis of large-scale sequencing approaches used; M.M.W., B.K., M.L.J., A.M.S. performed experiments; M.M.W., B.K., M.L.J., A.M.S. performed statistical analyses; M.M.W., B.K., M.L.J., K.L.T., A.M.S., B.P.T., G. Nordmark analyzed the data; C.J.L.*, M.M.W., B.K., K.L.T. wrote the paper; all authors critically reviewed the paper and accepted authorship responsibilities.

Competing Interests

C.J.L.* and **A.D.F.** have an active collaborative research agreement with Janssen. **E.B.** has an active research collaboration with Pfizer. **T.M.** is employed as medical solutions lead in rheumatology at UCB. **R.H.S.** is a consultant for Jansen Pharmaceuticals. **S.J.B.** provided consultancy services for Abbvie, BMS, Galapagos, Iqvia, J&J, Kiniksa, and Novartis in 2020-2021. **L.R.** provided consultancy services for AstraZeneca. **B.M.W.** has active collaborative research agreements with Astellas Bio and Pfizer, Inc. **M.R.** received grants from Amgen, AstraZeneca, Bristol Myers-Squibb, Novartis, and Servier for clinical trials in Sjögren's Syndrome and SLE. All other authors have reported that they have no competing interests to report.

REFERENCES

1 Khatri, B. *et al.* Genome-wide association study identifies Sjogren's risk loci with functional implications in immune and glandular cells. *Nat Commun* **13**, 4287 (2022). https://doi.org:10.1038/s41467-022-30773-y

2 Thorlacius, G. E., Bjork, A. & Wahren-Herlenius, M. Genetics and epigenetics of primary Sjogren syndrome: implications for future therapies. *Nat Rev Rheumatol* **19**, 288-306 (2023). https://doi.org:10.1038/s41584-023-00932-6

3 Baer, A. N. & Hammitt, K. M. Sjogren's Disease, Not Syndrome. *Arthritis Rheumatol* **73**, 1347-1348 (2021). https://doi.org:10.1002/art.41676

4 Mariette, X. & Criswell, L. A. Primary Sjogren's Syndrome. *N Engl J Med* **379**, 97 (2018). https://doi.org:10.1056/NEJMc1804598

5 Nocturne, G. & Mariette, X. Advances in understanding the pathogenesis of primary Sjogren's syndrome. *Nat Rev Rheumatol* **9**, 544-556 (2013). https://doi.org:10.1038/nrrheum.2013.110

6 Pasoto, S. G., Adriano de Oliveira Martins, V. & Bonfa, E. Sjogren's syndrome and systemic lupus erythematosus: links and risks. *Open Access Rheumatol* **11**, 33-45 (2019). https://doi.org:10.2147/OARRR.S167783

7 Crow, M. K., Olferiev, M. & Kirou, K. A. Type I Interferons in Autoimmune Disease. *Annu Rev Pathol* **14**, 369-393 (2019). https://doi.org:10.1146/annurev-pathol-020117-043952

8 Emamian, E. S. *et al.* Peripheral blood gene expression profiling in Sjogren's syndrome. *Genes Immun* **10**, 285-296 (2009). https://doi.org:10.1038/gene.2009.20

9 Danchenko, N., Satia, J. A. & Anthony, M. S. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* **15**, 308-318 (2006). https://doi.org:10.1191/0961203306lu2305xx

10 Kvarnstrom, M., Ottosson, V., Nordmark, B. & Wahren-Herlenius, M. Incident cases of primary Sjogren's syndrome during a 5-year period in Stockholm County: a descriptive study of the patients and their characteristics. *Scand J Rheumatol* **44**, 135-142 (2015). https://doi.org:10.3109/03009742.2014.931457

11 Patel, R. & Shahane, A. The epidemiology of Sjogren's syndrome. *Clin Epidemiol* **6**, 247-255 (2014). https://doi.org:10.2147/Clep.S47399

12 Petri, M. Epidemiology of systemic lupus erythematosus. *Best Pract Res Clin Rheumatol* **16**, 847-858 (2002). https://doi.org:10.1053/berh.2002.0259

13 Chakravarty, E. F., Bush, T. M., Manzi, S., Clarke, A. E. & Ward, M. M. Prevalence of adult systemic lupus erythematosus in California and Pennsylvania in 2000: estimates obtained using hospitalization data. *Arthritis Rheum* **56**, 2092-2094 (2007). https://doi.org:10.1002/art.22641

14 Cervera, R. *et al.* Systemic lupus erythematosus: clinical and immunologic patterns of disease expression in a cohort of 1,000 patients. The European Working Party on Systemic Lupus Erythematosus. *Medicine (Baltimore)* **72**, 113-124 (1993).

15 Helmick, C. G. *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum* **58**, 15-25 (2008). https://doi.org:10.1002/art.23177

16 Goransson, L. G. *et al.* The point prevalence of clinically relevant primary Sjogren's syndrome in two Norwegian counties. *Scand J Rheumatol* **40**, 221-224 (2011). https://doi.org:10.3109/03009742.2010.536164

17 Goules, A. V., Kapsogeorgou, E. K. & Tzioufas, A. G. Insight into pathogenesis of Sjogren's syndrome: Dissection on autoimmune infiltrates and epithelial cells. *Clin Immunol* **182**, 30-40 (2017). https://doi.org:10.1016/j.clim.2017.03.007

18 Bjork, A., Mofors, J. & Wahren-Herlenius, M. Environmental factors in the pathogenesis of primary Sjogren's syndrome. *J Intern Med* **287**, 475-492 (2020). https://doi.org:10.1111/joim.13032

19 Imgenberg-Kreuz, J., Rasmussen, A., Sivils, K. & Nordmark, G. Genetics and epigenetics in primary Sjogren's syndrome. *Rheumatology (Oxford)* **60**, 2085-2098 (2021). https://doi.org:10.1093/rheumatology/key330

20 Langefeld, C. D. *et al.* Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat Commun* **8**, 16021 (2017). https://doi.org:10.1038/ncomms16021

21 Lessard, C. J. *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat Genet* **45**, 1284-1292 (2013). https://doi.org:10.1038/ng.2792

22 Yin, X. *et al.* Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann Rheum Dis* **80**, 632-640 (2021). https://doi.org:10.1136/annrheumdis-2020-219209

23 Zhang, J. *et al.* Three SNPs in chromosome 11q23.3 are independently associated with systemic lupus erythematosus in Asians. *Hum Mol Genet* **23**, 524-533 (2014). https://doi.org:10.1093/hmg/ddt424

24 Lumb, J. H. *et al.* DDX6 Represses Aberrant Activation of Interferon-Stimulated Genes. *Cell Rep* **20**, 819-831 (2017). https://doi.org:10.1016/j.celrep.2017.06.085

25 Nunez, R. D. *et al.* The RNA Helicase DDX6 Associates with RIG-I to Augment Induction of Antiviral Signaling. *Int J Mol Sci* **19** (2018). https://doi.org:10.3390/ijms19071877

26 Wiener, A. *et al.* CXCR5 is critically involved in progression of lupus through regulation of B cell and double-negative T cell trafficking. *Clin Exp Immunol* **185**, 22-32 (2016). https://doi.org:10.1111/cei.12791

27 Aqrawi, L. A. *et al.* Diminished CXCR5 expression in peripheral blood of patients with Sjogren's syndrome may relate to both genotype and salivary gland homing. *Clin Exp Immunol* **192**, 259-270 (2018). https://doi.org:10.1111/cei.13118

28 Lessard, C. J. *et al.* Identification of a Systemic Lupus Erythematosus Risk Locus Spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. *Arthritis Rheumatol* **68**, 1197-1209 (2016). https://doi.org:10.1002/art.39548

29 Jostins, L. & McVean, G. Trinculo: Bayesian and frequentist multinomial logistic regression for genome-wide association studies of multi-category phenotypes. *Bioinformatics* **32**, 1898-1900 (2016). https://doi.org:10.1093/bioinformatics/btw075

30 Suzuki, A., Guerrini, M. M. & Yamamoto, K. Functional genomics of autoimmune diseases. *Ann Rheum Dis* **80**, 689-697 (2021). https://doi.org:10.1136/annrheumdis-2019-216794

31 Fu, Y., Tessneer, K. L., Li, C. & Gaffney, P. M. From association to mechanism in complex disease genetics: the role of the 3D genome. *Arthritis Res Ther* **20**, 216 (2018). https://doi.org:10.1186/s13075-018-1721-x

32 Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, 424 (2020). https://doi.org:10.3389/fgene.2020.00424

33 Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015). https://doi.org:10.1038/nature13835

34 Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179-189 (2020). https://doi.org:10.1038/s41586-019-1879-7

35 Huang, J. H., Ku, W. C., Chen, Y. C., Chang, Y. L. & Chu, C. Y. Dual mechanisms regulate the nucleocytoplasmic localization of human DDX6. *Sci Rep* **7**, 42853 (2017). https://doi.org:10.1038/srep42853

36 Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A* **100**, 2610-2615 (2003). https://doi.org:10.1073/pnas.0337679100

37 Niewold, T. B. *et al.* Association of the IRF5 risk haplotype with high serum interferon-alpha activity in systemic lupus erythematosus patients. *Arthritis Rheum* **58**, 2481-2487 (2008). https://doi.org:10.1002/art.23613

38 Ronnblom, L. & Eloranta, M. L. The interferon signature in autoimmune diseases. *Curr Opin Rheumatol* **25**, 248-253 (2013). https://doi.org:10.1097/BOR.0b013e32835c7e32

39 Bodewes, I. L. A. *et al.* Systemic interferon type I and type II signatures in primary Sjogren's syndrome reveal differences in biological disease activity. *Rheumatology (Oxford)* **57**, 921-930 (2018). https://doi.org:10.1093/rheumatology/kex490

40 Imgenberg-Kreuz, J. *et al.* Transcription profiling of peripheral B cells in antibody-positive primary Sjogren's syndrome reveals upregulated expression of CX3CR1 and a type I and type II interferon signature. *Scand J Immunol* **87**, e12662 (2018). https://doi.org:10.1111/sji.12662

41 Joachims, M. L. *et al.* Dysregulated long non-coding RNA in Sjogren's disease impacts both interferon and adaptive immune responses. *RMD Open* **8** (2022). https://doi.org:10.1136/rmdopen-2022-002672

42 Hjelmervik, T. O., Petersen, K., Jonassen, I., Jonsson, R. & Bolstad, A. I. Gene expression profiling of minor salivary glands clearly distinguishes primary Sjogren's syndrome patients from healthy control subjects. *Arthritis Rheum* **52**, 1534-1544 (2005). https://doi.org:10.1002/art.21006

43 Salomonsson, S. *et al.* Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjogren's syndrome. *Arthritis Rheum* **48**, 3187-3201 (2003). https://doi.org:10.1002/art.11311

44 Papatheodorou, I. *et al.* Expression Atlas update: from tissues to single cells. *Nucleic Acids Res* **48**, D77-D83 (2020). https://doi.org:10.1093/nar/gkz947

45 Zhou, T. *et al.* Lupus enhancer risk variant causes dysregulation of IRF8 through cooperative lncRNA and DNA methylation machinery. *Nat Commun* **13**, 1855 (2022). https://doi.org:10.1038/s41467-022-29514-y

46 Wang, X. *et al.* BCL9/BCL9L promotes tumorigenicity through immune-dependent and independent mechanisms in triple negative breast cancer. *Oncogene* **40**, 2982-2997 (2021). https://doi.org:10.1038/s41388-021-01756-y

47 Vafaizadeh, V. *et al.* The interactions of Bcl9/Bcl9L with beta-catenin and Pygopus promote breast cancer growth, invasion, and metastasis. *Oncogene* **40**, 6195-6209 (2021). https://doi.org:10.1038/s41388-021-02016-9

48 Skibola, C. F. *et al.* Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. *Am J Hum Genet* **95**, 462-471 (2014). https://doi.org:10.1016/j.ajhg.2014.09.004

49 Zhong, C. *et al.* Follicular lymphoma polygenic risk score is associated with increased disease risk but improved overall survival among women in a population based case-control in Los Angeles County California. *Cancer Epidemiol* **65**, 101688 (2020). https://doi.org:10.1016/j.canep.2020.101688

50 Lamb, C. A. *et al.* TBC1D14 regulates autophagy via the TRAPP complex and ATG9 traffic. *EMBO J* **35**, 281-301 (2016). https://doi.org:10.15252/embj.201592695

51 Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466**, 68-76 (2010). https://doi.org:10.1038/nature09204

52 Keller, C. W., Adamopoulos, I. E. & Lunemann, J. D. Autophagy pathways in autoimmune diseases. *J Autoimmun* **136**, 103030 (2023). https://doi.org:10.1016/j.jaut.2023.103030

53 Freudenberg, J., Gregersen, P. & Li, W. Enrichment of Genetic Variants for Rheumatoid Arthritis within T-Cell and NK-Cell Enhancer Regions. *Mol Med* **21**, 180-184 (2015). https://doi.org:10.2119/molmed.2014.00252

54 Juran, B. D. *et al.* Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum Mol Genet* **21**, 5209-5221 (2012). https://doi.org:10.1093/hmg/dds359

55 Linden, M. *et al.* Sex influences eQTL effects of SLE and Sjogren's syndrome-associated genetic polymorphisms. *Biol Sex Differ* **8**, 34 (2017). https://doi.org:10.1186/s13293-017-0153-7

56 Rasmussen, A. *et al.* Comparison of the American-European Consensus Group Sjogren's syndrome classification criteria to newly proposed American College of Rheumatology criteria in a large, carefully characterised sicca cohort. *Ann Rheum Dis* **73**, 31-38 (2014). https://doi.org:10.1136/annrheumdis-2013-203845

57 Vitali, C. *et al.* Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* **61**, 554-558 (2002). https://doi.org:10.1136/ard.61.6.554

58 Vitali, C. *et al.* Classification criteria for Sjogren's syndrome: we actually need to definitively resolve the long debate on the issue. *Ann Rheum Dis* **72**, 476-478 (2013). https://doi.org:10.1136/annrheumdis-2012-202565

59 Hochberg, M. C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* **40**, 1725 (1997). https://doi.org:10.1002/art.1780400928

60 Tan, E. M. *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* **25**, 1271-1277 (1982). https://doi.org:10.1002/art.1780251101

61 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007). https://doi.org:10.1086/519795

62 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006). https://doi.org:10.1038/ng1847

63 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283 (2016). https://doi.org:10.1038/ng.3643

64 Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-784 (2015). https://doi.org:10.1093/bioinformatics/btu704

65 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181 (2011). https://doi.org:10.1038/nmeth.1785

66 Rueger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet* **14**, e1007371 (2018). https://doi.org:10.1371/journal.pgen.1007371

67 Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010). https://doi.org:10.1093/bioinformatics/btq419

68 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005). https://doi.org:10.1093/bioinformatics/bth457

69 Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-1797 (2012). https://doi.org:10.1101/gr.137323.112

70 Dong, S. & Boyle, A. P. Predicting functional variants in enhancer and promoter elements using RegulomeDB. *Hum Mutat* **40**, 1292-1298 (2019). https://doi.org:10.1002/humu.23791

71 Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-881 (2016). https://doi.org:10.1093/nar/gkv1340

72 Lee, B. T. *et al.* The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* **50**, D1115-D1122 (2022). https://doi.org:10.1093/nar/gkab959

73 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017). https://doi.org:10.1038/s41467-017-01261-5

74 Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* **53**, 1290-1299 (2021). https://doi.org:10.1038/s41588-021-00924-w

75 van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet* **50**, 493-497 (2018). https://doi.org:10.1038/s41588-018-0089-9

76 Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422-423 (2021). https://doi.org:10.1093/bioinformatics/btaa692

77 Rasmussen, A. *et al.* The lupus family registry and repository. *Rheumatology (Oxford)* **50**, 47-59 (2011). https://doi.org:10.1093/rheumatology/keq302

78 Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722-1733 (2007). https://doi.org:10.1038/nprot.2007.243

79 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012). https://doi.org:10.1038/nmeth.1923

80 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352

81 Gaspar, J. M. *Genrich: Detecting sites of genomic enrichment*, <https://github.com/jsh58/Genrich> (2018).

82 Institute, B. *Picard Toolkit*, <https://github.com/broadinstitute/picard> (2019).

83 Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019). https://doi.org:10.1186/s13072-019-0287-4

84 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160-W165 (2016). https://doi.org:10.1093/nar/gkw257

85 Amariuta, T. *et al.* Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat Genet* **52**, 1346-1354 (2020). https://doi.org:10.1038/s41588-020-00740-8

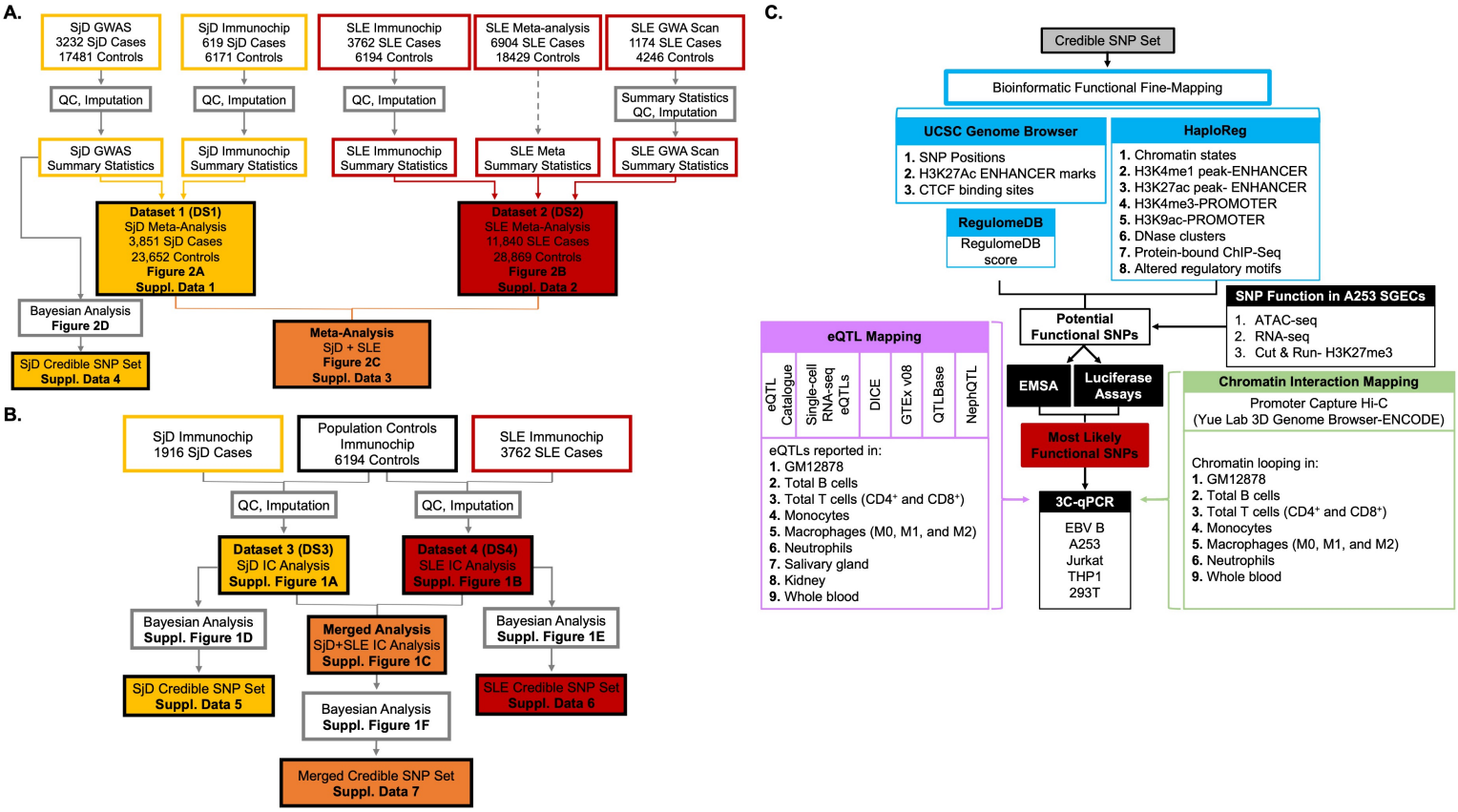


Figure 1. Schematic of the meta-analysis, fine mapping, and bioinformatic workflow of the shared *DDX6-CXCR5* interval associated with Sjögren's disease (SjD) and systemic lupus erythematosus (SLE). **(A)** Composition and workflow of three datasets (DS) used in the meta-analysis of SjD (DS1; yellow), SLE (DS2; red), and SjD and SLE merged (DS1+DS2; orange). DS1 was also used to perform SNP-SjD single marker trait analysis and identify a SjD credible SNP set. **(B)** Composition and workflow of the SjD (DS3; yellow) and SLE (DS4; Red) Immunochip datasets used for SNP-single marker trait analysis and identification of credible SNP sets for SjD, SLE, and SjD and SLE merged. **(C)** Workflow of the bioinformatic fine mapping applied to the 95% credible SNP sets from the *DDX6-CXCR5* risk locus to identify, prioritize, and functionally characterize SNPs common between the SjD and SLE risk associations.

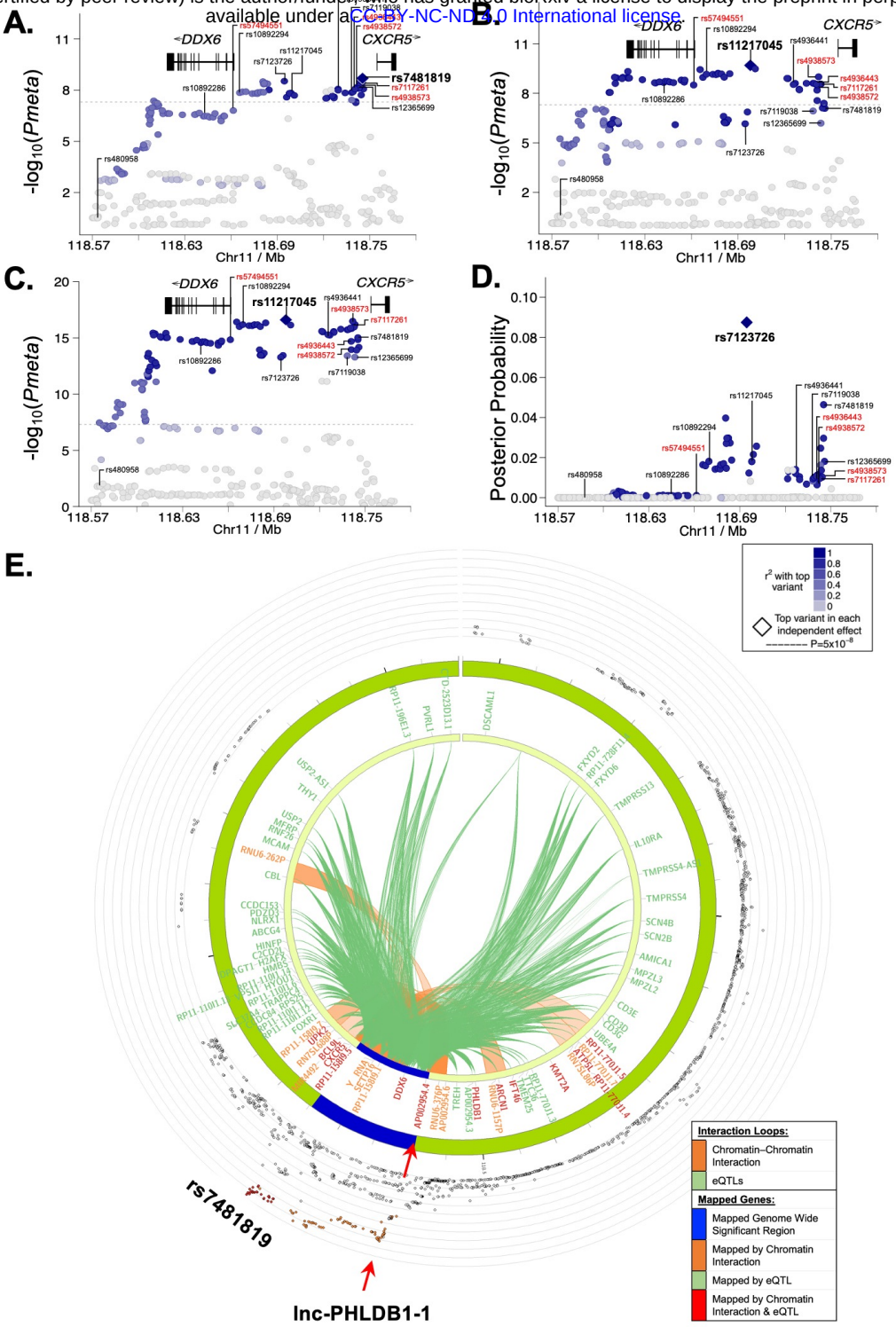


Figure 2. Fine mapping of the *DDX6-CXCR5* region in Sjögren's disease (SjD) and systemic lupus erythematosus (SLE) after imputation and meta-analysis. (A-C) Logistic regression analysis was performed on **(A)** Dataset 1 (DS1)-SjD (3851 SjD cases; 23652 controls), **(B)** DS2-SLE (11840 SLE cases; 28869), and **(C)** DS1+DS2 (merged SjD and SLE) after quality control and imputation, identifying the top SNPs (e.g., index SNPs indicated in bold) of the *DDX6-CXCR5* region. **(D)** Posterior probability distribution of SNPs in the *DDX6-CXCR5* region of DS1-SjD, identifying rs7123726 as having the highest posterior probability. For **A-D,** SNPs prioritized for bioinformatic screening are indicated in black; five SNPs prioritized for functional characterization are labeled in red. **(E)** Circos plot showing chromatin-chromatin interactions in GM12878 EBV B cell line (orange) and reported eQTLs from different cells/tissues (green). The outer ring shows the logistic regression analysis of DS1-SjD with index SNP rs7481819 indicated.

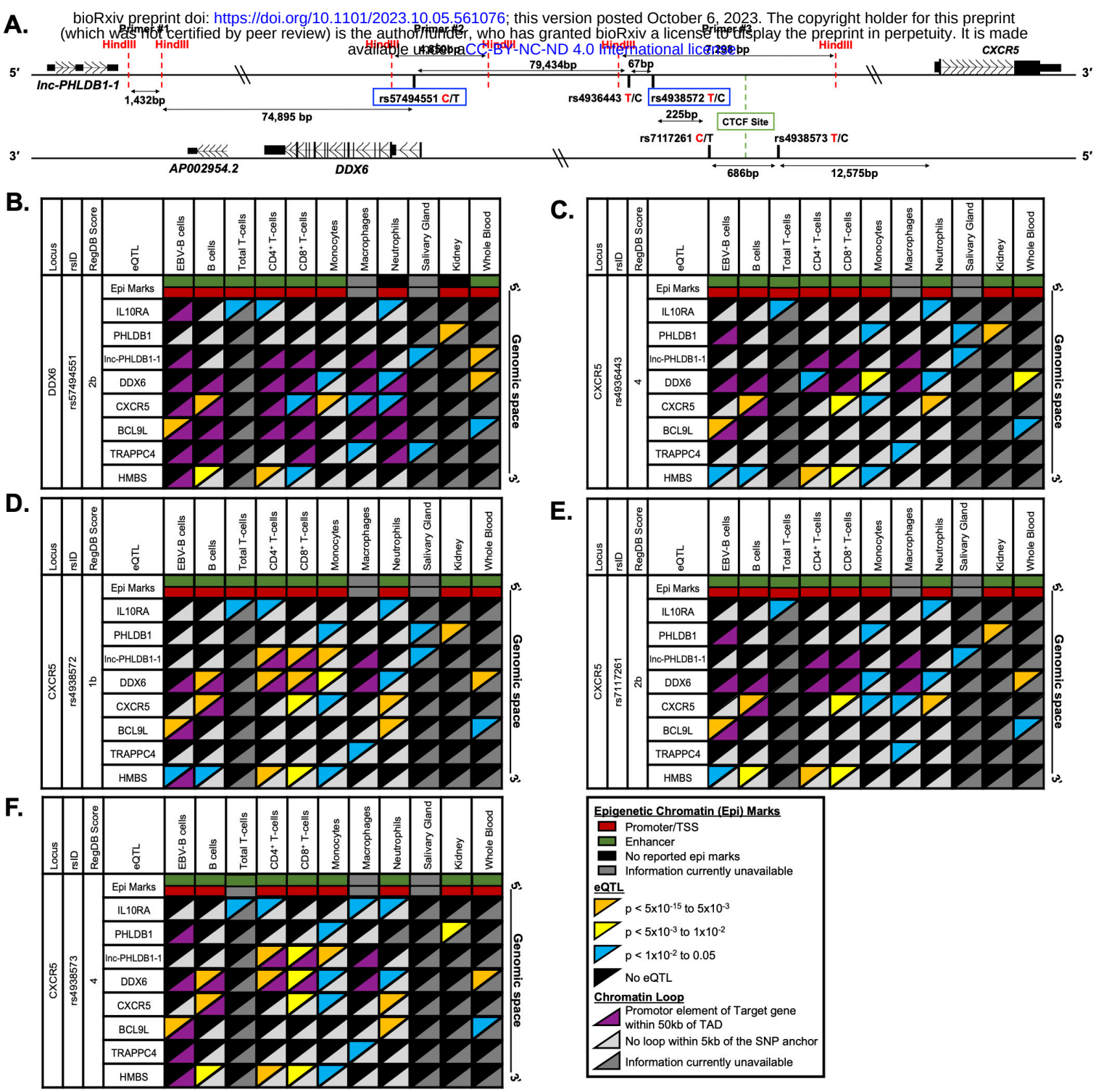
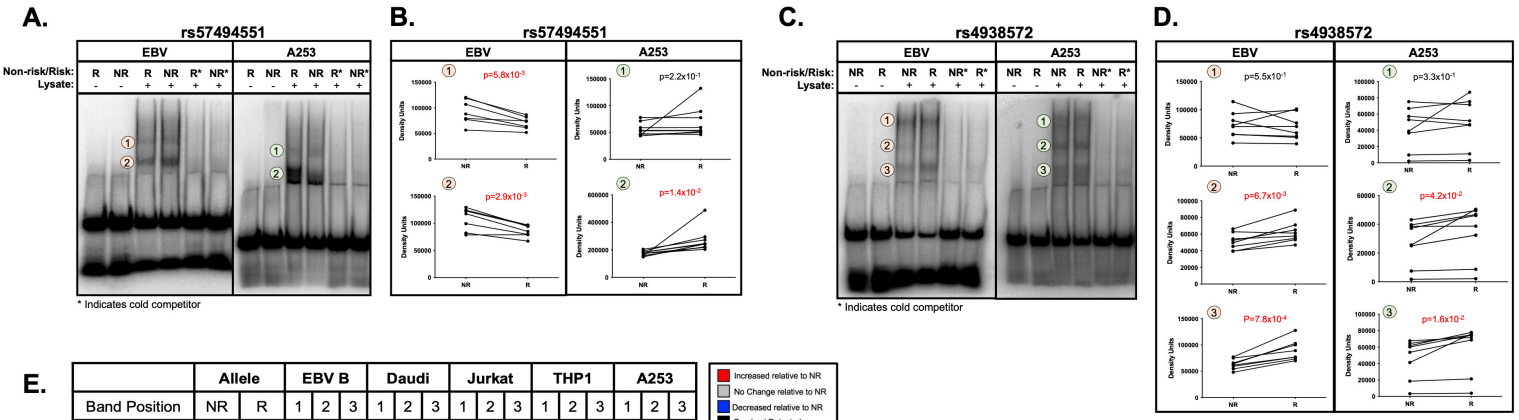


Figure 3. Reported expression quantitative trait loci (eQTLs) and chromatin-chromatin interactions in immune and disease-relevant tissues for the prioritized SNPs on the *DDX6-CXCR5* risk interval. **(A)** Schematic of the five prioritized SNPs arranged in genomic space. SNP rs57494551 is in the first intron of *DDX6*. The other four SNPs span 978 bp in a shared promoter/enhancer region between *DDX6* and *CXCR5*. Risk alleles are indicated by red font; non-risk alleles by black. SNPs rs57494551 and rs4938572 (blue boxes) are representative SNPs. *HindIII* restriction sites (red dotted line), CTCF site (green box), and positions of 3C-qPCR primers #1-3 are also indicated. **(B-F)** Publicly reported cell type-specific functional annotations (horizontal rectangles), select eQTLs (top triangles), and chromatin-chromatin interactions (bottom triangles) are shown for **(B)** rs57494551, **(C)** rs4936443, **(D)** rs4938572, **(E)** rs7117261, and **(F)** rs4938573 across 10 different immune cell types or disease-specific tissues (GM12878 EBV B cells and primary human B cells, CD4+ T cells, CD8+ T cells, monocytes, macrophages, neutrophils, salivary gland tissue, kidney tissue, and whole blood).

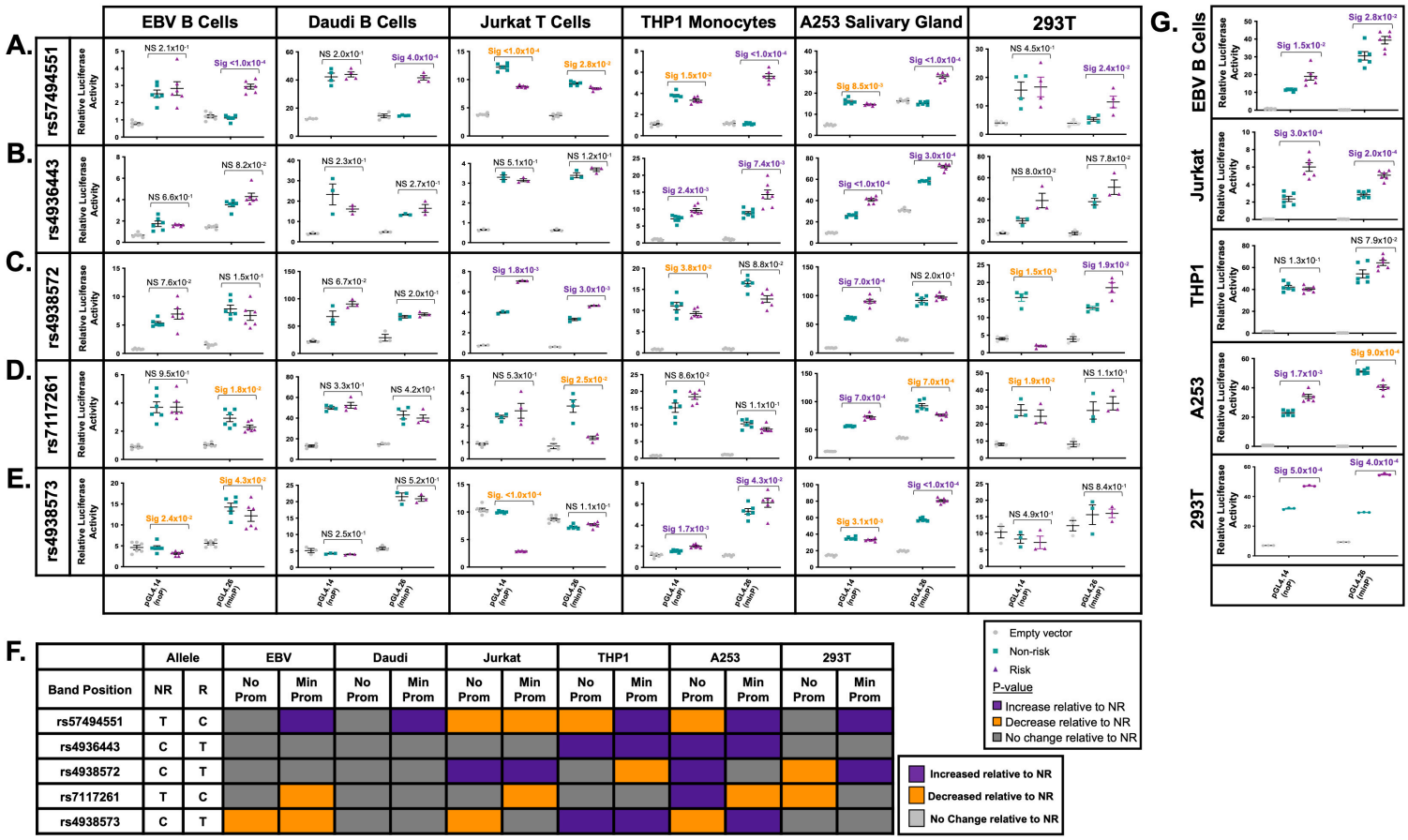


E.

	Allele		EBV B			Daudi			Jurkat			THP1			A253		
Band Position	NR	R	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
rs57494551	T	C															
rs10892294	C	G															
rs4936443	C	T															
rs4938572	C	T															
rs7117261	T	C															
rs4938573	C	T															
rs12365699	G	A															

Increased relative to NR
No Change relative to NR
Decreased relative to NR
Band not Detected
Data Not Available

Figure 4. Allele- and cell-type specific differential nuclear protein affinities of the prioritized SNPs rs57494551 and rs4938572 on the shared *DDX6-CXCR5* risk region. **(A-D)** Radiolabeled electromobility shift assays (EMSA) were performed to assess the binding affinity of ribonucleoproteins isolated from EBV B or A253 cells to oligonucleotides containing the non-risk (NR) or risk (R) allele of **(A-B)** rs57494551 or **(C-D)** rs4938572. Probes incubated in the absence of nuclear lysate were used as negative control (Lanes 1, 2). Cold competitors were used to assess non-specific binding (Lanes 5, 6). Images shown in (A) and (C) are representative of n>6 biological replicates. **(B, D)** Bands indicated in (A, C) by the orange or green circles were quantified by densitometry and analyzed using paired t-test (n>6); p-values indicated. **(E)** Summary analysis of the allele-specific nuclear protein affinities of the five prioritized SNPs in EBV B, Daudi, Jurkat, THP1, and A253 cells shown in A-D and Supplemental Figures 4-9. Increases in binding relative to NR are shown in red; decreases relative to NR in blue; no change relative to NR in grey; no detected band in black; data not available in white.



F.

	Allele		EBV		Daudi		Jurkat		THP1		A253		293T	
Band Position	NR	R	No Prom	Min Prom	No Prom	Min Prom	No Prom	Min Prom	No Prom	Min Prom	No Prom	Min Prom	No Prom	Min Prom
rs57494551	T	C												
rs4936443	C	T												
rs4938572	C	T												
rs7117261	T	C												
rs4938573	C	T												

Increased relative to NR
Decreased relative to NR
No Change relative to NR

Figure 5. Allele- and cell-type specific promoter and enhancer activity of the prioritized SNPs on the shared *DDX6-CXCR5* risk region. **(A-E)** gBlocks carrying the non-risk or risk alleles of **(A)** rs57494551, **(B)** rs4936443, **(C)** rs4938572, **(D)** rs7117261, or **(E)** rs4938573 were cloned into a promoter-less (pGL4.14; noP) or minimal promoter (pGL4.26; minP) luciferase vector. Plasmids were transfected into EBV B, Daudi, Jurkat, THP1, A253, or 293T cells. Luciferase activity was measured after 24 hours and normalized to the Renilla transfection control and then the vector-only control; reported as Relative Luciferase Activity. Statistical comparisons were performed using a paired t-test (n>3); p-values are indicated. **(F)** Summary analysis of the allele-specific luciferase activity of the five prioritized SNPs in EBV B, Daudi, Jurkat, THP1, A253 and 293Tcells. Increases in luciferase activity relative to non-risk are shown in purple; decreases relative to non-risk in orange; no change relative to non-risk in grey. **(G)** gBlocks carrying all non-risk or all risk alleles of rs4936443, rs4938572, and rs7117261 were cloned into the promoter-less or minimal promoter above, transfected into EBV B, Jurkat, THP1, A253, or 293T cells, and luciferase activity tested as described above. Statistical comparisons were performed using a paired t-test (n>3); p-values are indicated.

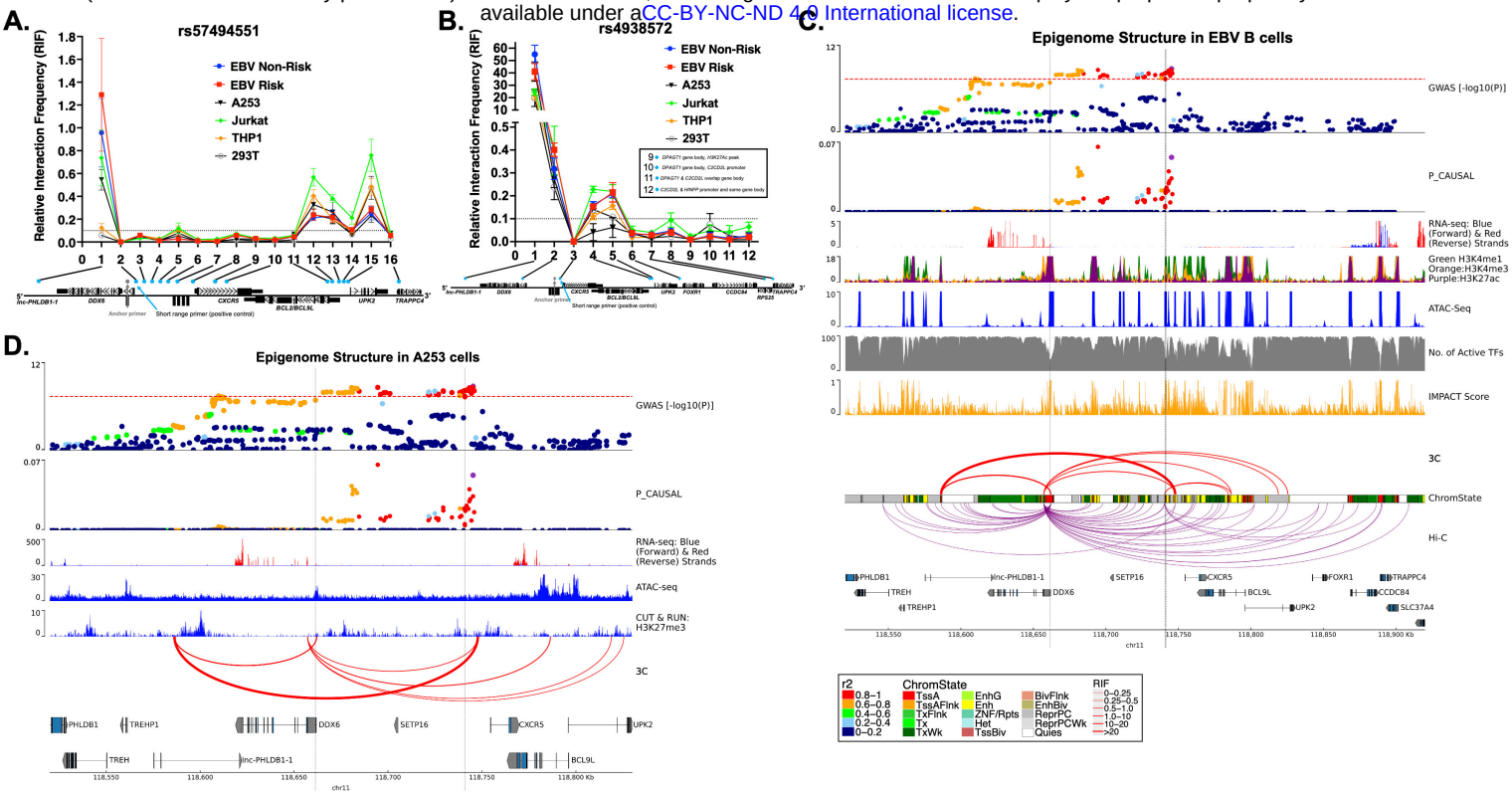


Figure 6. Complex chromatin architecture revealed across the *DDX6-CXCR5* region in immune cells, salivary gland, and kidney by 3C-qPCR. (A, B) Chromatin conformation capture with quantitative PCR (3C-qPCR) across the *DDX6-CXCR5* region where **(A)** rs57494551 or **(B)** rs4938572 is the anchor SNP (grey dot). Relative interaction frequency (RIF) is plotted relative to the primer number in 5'-3' genomic orientation (blue dots; see Figure 3A for additional detail). Primers 9-12 are shown in a text box for simplicity in (B). **(C)** SjD GWAS association (top panel) and publicly available epigenomic enrichment across the *DDX6-CXCR5* region in GM12878 EBV B cells. Vertical grey lines indicate the locations of rs57494551 or rs4938572, respectively. Promoter-capture Hi-C looping (purple lines) contrasts the summary 3C-qPCR results (red lines); line thickness indicates relative interaction frequency (RIF) of the 3C data. **(D)** SjD GWAS association (top panel) across the *DDX6-CXCR5* region in A253 cells. In house ATAC-seq, CUT & RUN: H3K27me3 (Epicypher), and RNA-seq data from A253 cells are shown because of limited publicly available epigenetic data on salivary gland. Vertical grey lines indicate the locations of rs57494551 or rs4938572, respectively. Summary 3C-qPCR results are shown (red lines); line thickness indicates relative interaction frequency (RIF).

Table 1: Summary Results of the Fine Mapping of the *DDX6-CXCR5* Risk Locus in SjD and SLE of European Ancestry

SNP rsID	Chr	Base pair	Minor	Major	SjD Meta P value	SLE Meta P value	SjD+SLE Meta P value
rs480958	11	118577990	G	A	8.11×10^{-1}	1.77×10^{-1}	4.19×10^{-1}
rs10892286	11	118642085	C	A	3.39×10^{-7}	5.91×10^{-4}	2.93×10^{-9}
rs57494551	11	118661398	T	C	1.57×10^{-7}	3.16×10^{-9}	1.45×10^{-15}
rs10892294	11	118667357	C	G	1.26×10^{-8}	1.09×10^{-9}	9.87×10^{-17}
rs7123726	11	118694297	C	T	2.98×10^{-9}	7.02×10^{-7}	5.31×10^{-14}
rs11217045	11	118697865	A	C	1.54×10^{-8}	2.02×10^{-10}	2.42×10^{-17}
rs4936441	11	118725660	C	G	2.23×10^{-8}	2.09×10^{-1}	7.42×10^{-6}
rs7119038	11	118738281	G	A	3.77×10^{-8}	2.46×10^{-2}	8.54×10^{-2}
rs4936443	11	118740864	C	T	4.57×10^{-8}	4.58×10^{-2}	5.79×10^{-2}
rs4938572	11	118740931	C	T	5.48×10^{-8}	9.47×10^{-2}	3.33×10^{-2}
rs7117261	11	118741157	T	C	8.01×10^{-8}	2.23×10^{-1}	3.75×10^{-6}
rs4938573	11	118741842	C	T	6.69×10^{-8}	3.92×10^{-2}	3.50×10^{-2}
rs12365699	11	118743286	A	G	7.01×10^{-8}	6.49×10^{-7}	5.17×10^{-14}
rs7481819	11	118745278	C	T	2.02×10^{-8}	1.53×10^{-1}	7.47×10^{-3}