scientific reports

OPEN

Integrative epigenomics in Sjögren´s syndrome reveals novel pathways and a strong interaction between the HLA, autoantibodies and the interferon signature

María Teruel[1], Guillermo Barturen[1], Manuel Martínez-Bueno[1], Olivia Castellini-Pérez[1], Miguel Barroso-Gil[1], Elena Povedano[1], Martin Kerick[2], Francesc Català-Moll[3,4], Zuzanna Makowska[5], Anne Buttgereit[5], PRECISESADS Clinical Consortium*, PRECISESADS Flow Cytometry Study Group*, Jacques-Olivier Pers[6], Concepción Marañón[1], Esteban Ballestar[3,4], Javier Martin[2], Elena Carnero-Montoro[1,42]✉ & Marta E. Alarcón-Riquelme[1,7,42]✉

Primary Sjögren's syndrome (SS) is a systemic autoimmune disease characterized by lymphocytic infiltration and damage of exocrine salivary and lacrimal glands. The etiology of SS is complex with environmental triggers and genetic factors involved. By conducting an integrated multi-omics study, we confirmed a vast coordinated hypomethylation and overexpression effects in IFN-related genes, what is known as the IFN signature. Stratified and conditional analyses suggest a strong interaction between SS-associated *HLA* genetic variation and the presence of Anti-Ro/SSA autoantibodies in driving the IFN epigenetic signature and determining SS. We report a novel epigenetic signature characterized by increased DNA methylation levels in a large number of genes enriched in pathways such as collagen metabolism and extracellular matrix organization. We identified potential new genetic variants associated with SS that might mediate their risk by altering DNA methylation or gene expression patterns, as well as disease-interacting genetic variants that exhibit regulatory function only in the SS population. Our study sheds new light on the interaction between genetics, autoantibody profiles, DNA methylation and gene expression in SS, and contributes to elucidate the genetic architecture of gene regulation in an autoimmune population.

Primary Sjögren's syndrome (SS) [MIM 270150] is a systemic autoimmune disease characterized by lymphoid infiltration and tissue damage of the exocrine glands, mainly the salivary and lacrimal glands[1]. The prevalence of SS is about 1% of the World population being the main risk group middle-aged women. SS patients show great clinical heterogeneity that is manifested at the serological, genetic, and cellular level and in their capacity to respond to treatment. SS patients with anti-Ro/SSA or anti-La/SSB autoantibodies usually develop a more severe disease with systemic manifestations and the appearance of lymphomas in a percentage of them[2].

[1]GENYO, Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, 18016 Granada, Spain. [2]IPBLN-CSIC, Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, 18016 Granada, Spain. [3]Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916 Badalona, Barcelona, Spain. [4]IDIBELL, Bellvitge Biomedical Research Institute 08907 L'Hospitalet de Llobregat, Barcelona, Spain. [5]Pharmaceuticals Division, Bayer Pharma Aktiengesellschaft, Berlin, Germany. [6]Université de Brest, INSERM, Labex IGO, CHU de Brest, Brest, France. [7]Institute for Environmental Medicine, Karolinska Institutet, 171 67 Solna, Sweden. [42]These authors contributed equally: Elena Carnero-Montoro and Marta E. Alarcón-Riquelme. *List of authors and their affiliations appears at the end of the paper. ✉email: elena.carnero@genyo.es; marta.alarcon@genyo.es

The etiology of SS is complex and not completely understood; environmental, genetic and epigenetic factors are known to be involved in its development[1,3–5]. Genome-wide association studies (GWAS) have signified an important advance for understanding the mechanisms involved in the pathogenesis of SS. There are currently around 10 well-known risk *loci* implicated in SS susceptibility that have been revealed by GWAS studies[3–5]. The *loci* identified suggest the importance of the adaptive and innate immune responses in SS pathology, especially the interferon (IFN) signaling pathway. Other important pathways include B cells signaling and autoantibody production, the NF-kB signaling pathway and the T cell activation through the major histocompatibility complex[6]. Many of the variants detected in GWAS studies are located in non-coding intron or intergenic regions, suggesting that they have a regulatory role that is fairly unexplored[7,8]. Moreover, the identified SS-associated genetic variants can only explain a small proportion of the heritability observed in SS[9], suggesting the possible contribution of many gene variants at the low-frequency spectrum, with lower effect sizes, and/or the implication of gene-by-environment interactions.

Functional approaches based on genome-wide data of the epigenome or the transcriptome are allowing to dive depth into the genetics of transcriptional regulation by detecting variants with potential regulatory effect, i.e., expression quantitative trait loci (eQTLs)[8] or methylation quantitative trait loci (meQTLs) discovery[10,11]. An increasing number of studies are showing widespread regulatory effects of disease associated genetic variants and are helping to mechanistically explain the genetic risk of disease. Functional genomics approaches have been applied very successfully in cancer research, and to some extent to autoimmune diseases as well, especially in rheumatoid arthritis (RA)[12,13] and systemic lupus erythematosus (SLE)[14,15], but only recently in SS[16].

Epigenetic alterations can be integrators of the complex interaction between genes and the environment and are known to play a relevant role in autoimmunity by altering gene expression profiles in response to genetic factors, changing environment and immunological conditions[17,18]. Recently, a few epigenome-wide association studies (EWAS) have been performed in different cell types and have consistently found a pervasive hypomethylation in CpG sites within genes related with the type I IFN signaling, confirming this pathway as key in SS pathology[19–21]. Likewise, gene expression profiling of minor salivary glands or peripheral blood showed a consistent upregulation of IFN-inducible genes associated with SS, which was found more pronounced in the subset of cases serologically defined by increased titers of anti-Ro/SSA and anti-La/SSB autoantibodies[22,23].

Despite an increasing number of studies interrogating different layers of molecular information, the existing integrative approach for SS mainly focuses on deciphering the regulatory roles of previously known disease associated genetic variants discovered by GWAS[16,24]. Therefore, these studies have not contributed to the identification of new risk variants associated with SS with regulatory function that could explain further the missing heritability of this complex disease[20]. Furthermore, the vast majority of large-scale QTL studies have been performed in healthy populations, and do not directly compare the genetic regulatory effects between cases and controls, which could very much be different as the pathological condition imposes a different environmental and cellular context in affected individuals, especially in immune-related conditions, as we have recently described in scleroderma patients[25]. In this regard, several experimental studies have shown that eQTL and meQTL effects are highly context-specific, as for example cell-type specific effects have been found when studying different blood cell types[26–28], and experimental studies of ex-vivo activation of immune cells have shown that eQTLs profiles are altered upon different stimuli[29,30].

This work represents the first epigenome- and transcriptional-wide integrative study that combines high throughput data on genetics, DNA methylation and gene expression coupled with serological and clinical profiles in a large group of SS patients and healthy subjects. It contributes, on the one hand, to identify new molecular pathways involved in SS. On the other hand, this work identifies putative novel genetic variants implicated in the disease through changes in DNA methylation or gene expression variants with regulatory effects exhibiting disease-specificity, therefore contributing to elucidate the genetic architecture of gene regulation in an autoimmune population. Finally, this work dissects the relationships between HLA variation, autoantibody profiles and the interferon signature in SS.

Results

Differentially methylated positions associated with SS.

We explored the DNA methylation patterns associated with SS comparing the whole blood DNA methylation level between 189 SS patients and 220 healthy subjects (Supplementary Table 1) using the Infinium MethylationEPIC BeadChip with which we could interrogate 776,284 autosomic CpG sites. In total, we observed 118 differential methylated positions (DMPs) associated with SS at a Bonferroni-corrected threshold of $P < 6.4 \times 10^{-08}$ (Supplementary Table 2). The majority of SS-associated DMPs exhibited decreased methylation levels in SS patients compared with controls (91.5%), supporting the overall hypomethylation previously described in SS patients (Fig. 1a)[19,20]. The 118 SS-associated DMPs were annotated to 52 unique genes and 7 intergenic regions. The majority of them fell within promoters (49.2%) and gene bodies (40.7%), and only 4.2% were located in the 3'UTR. SS-associated DMPs were mainly distributed in open sea and shore CpGs and only 5 DMPs fell in CpG islands (Supplementary Fig. 1). We included an independent cohort formed by 60 SS patients and 89 healthy individuals (Supplementary Table 1) from whom we had DNA methylation data available from the 450 K array which contained half of the SS-associated DMPs detected in our discovery cohort. At a $P < 0.05$ we could successfully replicate 84.7% (Supplementary Table 2). The top 10 SS-associated DMPs with an average methylation difference $|\Delta\beta| > 0.1$ were located within *IFI44L*, *MX1*, *PARP9-DTX3L*, *NLRC5*, *IFIT1*, *IFIT3*, *IFITM1*, *PLSCR1*, *PDE7A* and *DDX60* genes, all of them known to be regulated by type I IFN. The most significant DMPs was the cg05696877 probe in the *IFI44L* gene for which an average methylation difference $|\Delta\beta| = 0.36$ was observed ($P = 3.5 \times 10^{-27}$; FDR $= 2.7 \times 10^{-21}$) between SS and controls. Functional analyses for genes underlying SS-DMPs confirmed an enrichment of GO terms related with IFN signaling (GO:0060337; $P = 3.5 \times 10^{-31}$; GO:0034340; $P = 31.37 \times 10^{-30}$) as well as with defense response
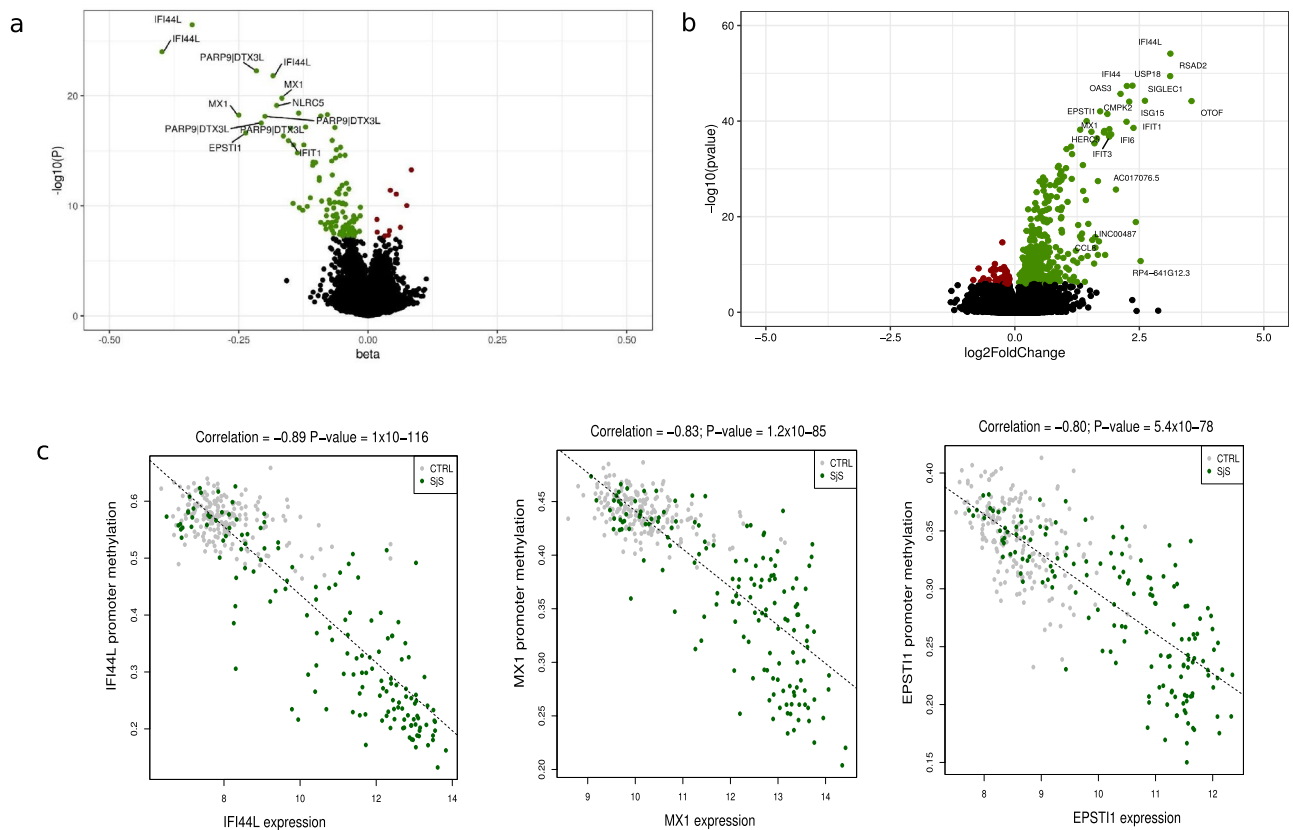
Figure 1. DNA methylation and gene expression patterns associated with SS. Volcano plot for the differential DNA methylation association study in the discovery cohort. *P*-values are represented on the –log10 scale in the y-axis. The effect size and direction obtained for each CpG site is depicted in the x-axis. Green dots represent significant associations with negative sign (hypomethylation). Red dots represent significant associations with positive signs. The top associations are labeled with gene names. (**b**) Volcano plot for the differential expression analysis in the discovery cohort. The effect size and direction obtained for each gene is depicted in the x-axis. Green dots represent significant associations with positive sign (overexpression). Red dots represent significant associations with negative signs. The top associations are labeled with gene names. (**c**) Plots showing high correlation between an average of DNA methylation quantified as β-values at the promoters of the most significant SS-associated DMRs and gene expression at the logarithmic scale. R software[73] and Adobe Illustrator (https://www.adobe.com/) was used to create figures.

to virus (GO:0051607; $P = 2.3 \times 10^{-23}$) and cytokine-mediated signaling (GO:0019221; $P = 1.63 \times 10^{-19}$) (Supplementary Table 4).

We searched for CpGs that exhibit differences in DNA methylation variability between SS patients and controls and named these variable methylated positions (VMPs) (Supplementary Table 2). We observed that 80% of SS-associated DMPs also exhibited evidence of increased DNA methylation variability at a significance level of $P < 0.05$ in SS patients compared with healthy controls. In our replication cohort, we could replicate 83.3% of the VMPs observed. The increased DNA methylation variability associated with SS is especially pronounced for *MX1 and PARP9-DXT3L* genes for which we observed the largest variability differences. Increases in variabiltity might reflect epigenetic plasticity in immune-related cells and/or reflect that DNA methylation is regulated by diverse transcription factors involved in different inflammatory and immune signaling during disease progression[31–35].

DNA methylation is an epigenetic mark that changes with environmental triggers as for example drugs. We investigated whether our SS-associated signatures were driven by the most common treatments in our SS patients by adjusting the linear model for treatments such as antimalarial, steroids, and immunosuppressive therapy (see details of the treatment applied to the individuals in each cohort in Supplementary Table 1). Our results show that most of the SS-DMPs (80.5%) remained significantly associated at our threshold $P < 6.4 \times 10^{-08}$ while the reminder showed suggestive associations, of $P < 1 \times 10^{-04}$, indicating that therapy applied to the SS patients does influence the DNA methylation patterns associated with SS only in a minority of sites, with modest effects (Supplementary Table 3). In fact, a comparison of the effect size of the 118 significant SS-DMPs obtained in these 2 analyses, when including or not the therapy as covariate in the linear regression, showed high correlation (Pearson's correlation R = 0.99, $P = 2.2 \times 10^{-6}$). Similar results were also observed in the replication cohort, for which we observed a high replication rate (81.0%) (Supplementary Table 3). Differences in DNA methylation levels and variability of CpGs located in the X-chromosome were also evaluated in females of the discovery cohort, we could not find any significant differences in females at a significance level corrected for multiple testing (data not shown).

The IFN epigenetic signature in SS is shaped by Anti-La/SSA and associated with classII HLA genetic variation.   Hypomethylation at IFN-regulated genes has been strongly associated with the presence of anti-Ro/SSA and/or anti-La/SSB autoantibodies previously [36]. Stratified analyses based on anti-La/SSB and anti-Ro/SSA autoantibodies positivity show in our data that the epigenetic signature is observed only when positive anti-Ro/SSA patients were compared with healthy subjects (Supplementary Table 3, Fig. 2a-b). Indeed, all associations became far from genome-wide significant ($P > 0.006$) when the model that includes positive and negative SS patients was adjusted by anti-Ro/SSA presence or when only patients negative for SSA were contrasted with controls ($P > 0.05$). Hierarchal clustering shows that DNA methylation profiles in anti-Ro/SSA negative patients resembles that observed in healthy controls rather than to anti-Ro/SSA positive patients (Fig. 2b). While stratifying analyses based anti-La/SSB patients also yielded stronger associations in the positive group, all associations found could be explained by the presence of anti-Ro/SSA autoantibodies and nothing remained significant after adjustment for anti-Ro/SSA (Supplementary Table 3, Fig. 2b). Patients positive for both anti-Ro/SSA and anti-La/SSA exhibited the strongest DNA methylation differences. See for example, in Fig. 2c, the DNA methylation difference at *IFI44L* gene (cg13452062) in all SS patients ($|\Delta\beta| = 0.40$) goes up to $|\Delta\beta| = 0.53$ when only SSA + patients are analysed, and is increased up to $|\Delta\beta| = 0.60$ in the SSA + /SSB + group, and goes down to $|\Delta\beta| = 0.03$ in the negative-group. This trend is conserved across all SS-associated CpGs (Fig. 2b, d, e).

Previous studies have recently shown that genetic association between HLA and SS is dependent on autoantibody profiles[36]. In order to further explore the possible link between HLA and the epigenetic IFN signature, we imputed classical HLA alleles and performed a series of logistic and linear regression models based on stratification and conditional analyses. At a significance level corrected for multiple testing we identified 2 class II HLA alleles (HLA-DQB1-0201 and HLA-DRB1-0301) associated with increased risk for SS ($P < 3 \times 10^{-07}$) which also are associated with the Anti-La/SSA presence ($P < 2 \times 10^{-04}$) and with the IFN epigenetic signature ($P < 1 \times 10^{-04}$) measured as the decreased in DNAm at IFN-related genes (Table 1). Conditional analyses revealed that the relationship between HLA-DQB1-0201 and HLA-DRB1-0301 and the epigenetic IFN signature is lost when the presence of Anti-Ro/SSA is included in the model. We also categorized SS patients as positive or negative for the epigenetic IFN signature (see method) and observed that the association between HLA and SS is only seen in positive IFN signature SS patients ($P < 8 \times 10^{-04}$), and not present in negative IFN signature patients ($P > 0.4$).

In the absence of autoantibodies, we cannot observe the HLA genetic associations with SS, and all the epigenetic signals associated with SS disappear, at least when analyzing single CpG-sites (Supplementary Table 3). Altogether, our results suggest that there is a complex and a strong interaction between genetic variation in the HLA region, the presence of anti-Ro/SSA antibodies and the IFN-epigenetic signature observed in SS.

Differentially methylated regions associated with SS.   We sought to identify differentially methylated regions (DMRs) associated with SS with the purpose of finding subtle, but consistent, methylation changes within a region that could not be detected when analyzing CpG sites by themselves. At FDR 5% threshold, we identified DMRs within 135 gene bodies, 335 promoters and 219 in CpG islands (CGI) (Supplementary Table 5). We found that many of the SS-associated DMPs (55.6%) were SS-associated DMRs, as for examples differentially methylated promoters for *IFI44L, OAS2, RSAD2, BST2, PARP9-DTX3L, IFITM1, MX1, EPSTI1* or *LY6E,* and differentially methylated gene bodies of *AGRN, IRF7, B2M, HERC5, ADAR* and *SP100,* among others.

Importantly, we identified 442 novel differentially methylated genes exhibiting DMRs but for which we did not find DMPs. We could not perform a strict replication of these signatures because DNA methylation was measured in a different array (450 K) in the replication sample, and this implies having lower coverage in the pre-defined regions analyzed. However, we found out that up to 45% novel DMRs, lying within 199 genes, showed robust signals in the independent sample ($P < 0.05$) (Supplementary Table 5). Among the novel genes implicated in SS we found new interferon-regulated genes, such as *SAMHD1, ISG15* and *XAF1*; and other proteins related with the immune system such the Tumor Necrosis Factor, *TNF,* the TNF-receptor *CD27,* the chemokine receptor like protein 2, *CCRL2,* and the tyrosine kinase *LCK*. We also discovered a group of genes belonging to the HOX family, such as *HOXB2, HOXD8, HOXA9, HOXA10* or *HOXA4,* that is implicated in transcriptional processes, as well as other transcription factors such *CEBPD* and *GATA2*. Interestingly, we discovered a large group of genes with differential methylated regions implicated in collagen metabolism such as *COL11A2, COL18A1, COL27A1, COL13A1, COL23A1* and *COL5A1* that were not previously identified. In addition, many DMRs, especially those that fell in CGI, were located within long non-coding RNAs (LncRNAs), as for example *RP11-723C11.2, RP11-326C3.7, RP11-480D4.2* or *RP11-89K21.1* (see Supplementary Table 5). We performed functional enrichment analyses for the set of genes showing lower DNA methylation in the SS group (64%) and separately those with increased DNA methylation (34%). On one hand, DMRs with negative signs showed an enrichment in well known pathways implicated in SS such as interferon and cytokine signaling (R-HSA-913531 and R-HSA-1280215, respectively, from Reactome database), as well as in NOD-like receptor signaling pathway (KEGG path:hsa04621), viral carcinogenesis (KEGG path:hsa05203), transcriptional misregulation in cancer (path:hsa05202), necroptosis (KEEF path:hsa04217). On the other hand, those genes showing DMRs with positive effects were found to be enriched in pathways such as Collagen biosynthesis and modifying enzymes (Reactome, R-HSA-1650814) and extracellular matrix organization (Reactome R-HSA-1474244) (Fig. 2, Supplementary Table 6), which represent novel molecular pathways implicated in SS, and might indicate that the hypermethylation of these genes can be implicated in SS pathogenesis by down-regulating these molecules.

Differential expression around SS-associated epigenetic signals.   In order to detect coordinated epigenetic and transcriptional changes associated with SS, we explored the possibility that the DNA methylation observed at the majority of SS-associated DMRs correlates with the expression of genes in the proximity. We identified 422 differentially expressed genes (DEGs) associated with SS when comparing RNA-seq gene
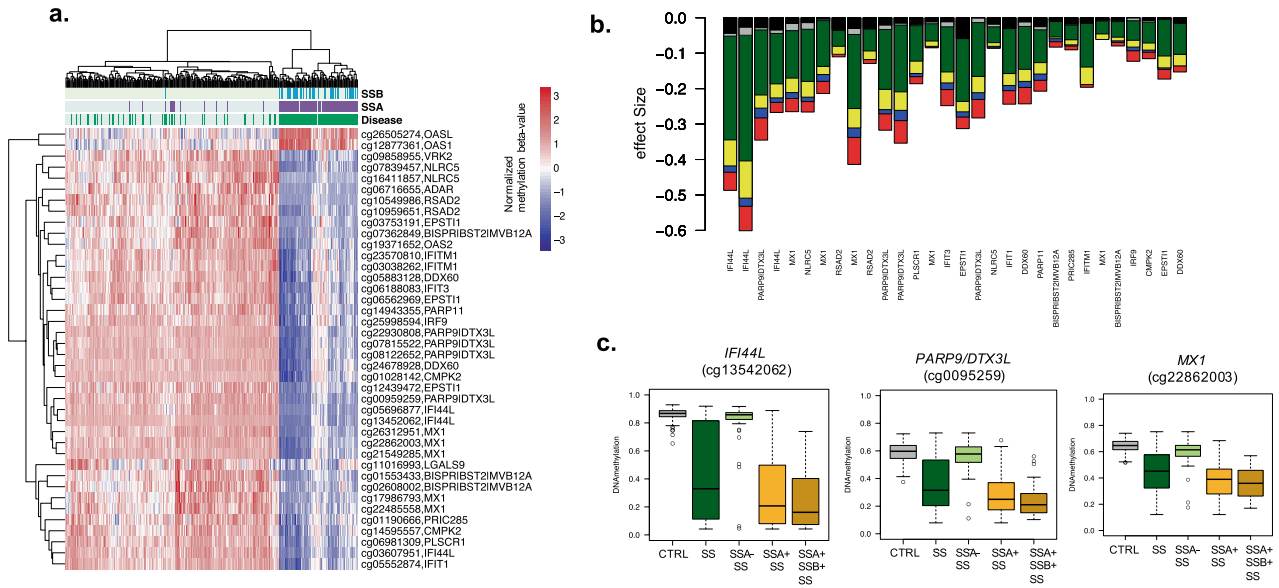
Figure 2. Effect of Autoantibody profile in SS-associated epigenetic signals. (**a**) Hierarchical clustering representation from SS patients and healthy individuals (in columns) accordingly to DNA methylation levels (in rows) at the top SS-associated CpG sites. Subjects are classified according to disease status (green) and the presence of Anti-Ro/SSA (purple) and Anti-La/SSB (light blue) autoantibodies. (**b**) Barplot representing the effect sizes obtained in different models where DNA methylation was contrasting between different SS patients (according to autoantibody profiles). Black bar is a model that contrasted SSA- SS patients with CTRLs. Grey bar is a model that included all SS patients and was adjusted by SSA. Green bar is a model that included all SS patients and was unadjusted by SSA. Yellow line is a model that compared SSA + SS patients and CTRL. Blue line is a model that contrasted SSA + patients with CTRLs adjusted by SSB. Red bar is a model that contrasted SSA + SSB + patients and CTRLs. (**c**) Boxplots representing DNA methylation differences across different groups in three selected genes. R software[73] and Adobe Illustrator (https://www.adobe.com/) was used to create figures.

HLA Alleles	SS ~ HLA		SSA ~ HLA		epigIFN ~ HLA		epigIFN ~ HLA + SSA		SS ~ HLA (positive epigIFN)		SS ~ HLA (negative epigIFN)	
	beta.HLA	P.HLA	beta.HLA	P.HLA	beta.HLA	P.HLA	beta.HLA	P.HLA	beta.HLA	P.HLA	beta.HLA	P.HLA
DRB1_0301	1.14	2.07E-07	0.27	1.02×10^{-04}	-0.17	9.16×10^{-05}	-0.06	0.142	1.40	7.14×10^{-09}	-0.30	0.470
DQB1_0201	1.16	1.39E-07	0.29	2.03×10^{-05}	-0.18	6.55×10^{-05}	-0.05	0.210	1.44	2.58×10^{-09}	-0.29	0.493

Table 1. Genetic associations between HLA variation, presence of Anti-La/SSA autoantibodies and epigenetic IFN signature. β reflects the additive effect of allele dosage for different HLA alleles and P is the associated significance level. The association between HLA genetic variation, SS and SSA was determined by means of logistic regression adjusted by sex and age. The association between HLA genetic variation and epigIFN was determined by linear regression models adjusted by sex, age, cell proportions and batch effects. epigIFN refers to the epigenetic IFN signature. DNA methylation at *IFI44L* gene (cg13452062) was used as a proxy for epigIFN. Patients exhibiting DNAm > 0.8 were classified as negative epigIFN. Patients exhibiting DNAm < 0.8 were classified as positive epigIFN.

expression data available for 174 SS patients and 135 healthy subjects in our discovery cohort ($P < 1.6 \times 10^{-6}$), the majority of them show increased expression in SS patients and were successfully replicated in our independent sample (85.5% and 88.2%, respectively) (Fig. 1b and Supplementary Table 7). *IFI44L* was observed as the most significant differentially expressed gene ($Log_2FC = 3.12$; $P = 7.6 \times 10^{-55}$; FDR = 3.3×10^{-50}). Not surprisingly, many overexpressed genes were related with the IFN signaling (65/422), such as *EPSTI1, RSDA2, USP18 or CMPK2* (Table 2), but a large proportion of SS-DEGs belong to other functional categories, being enriched in pathways such as response to external biotic stimulus (GO:0043207), response to stress (GO:0006950), interspecies interaction between organisms (GO:0044419), protein binding (GO:0005515), RNA binding (GO:0003723) or negative regulation of biological processes (GO:0048519) (Supplementary Table 8).

Next, we investigated whether DNA methylation at DMRs and gene expression at DEGs are correlated and if this correlation is associated with SS status and could, therefore, represent coordinated effects on SS, by expression
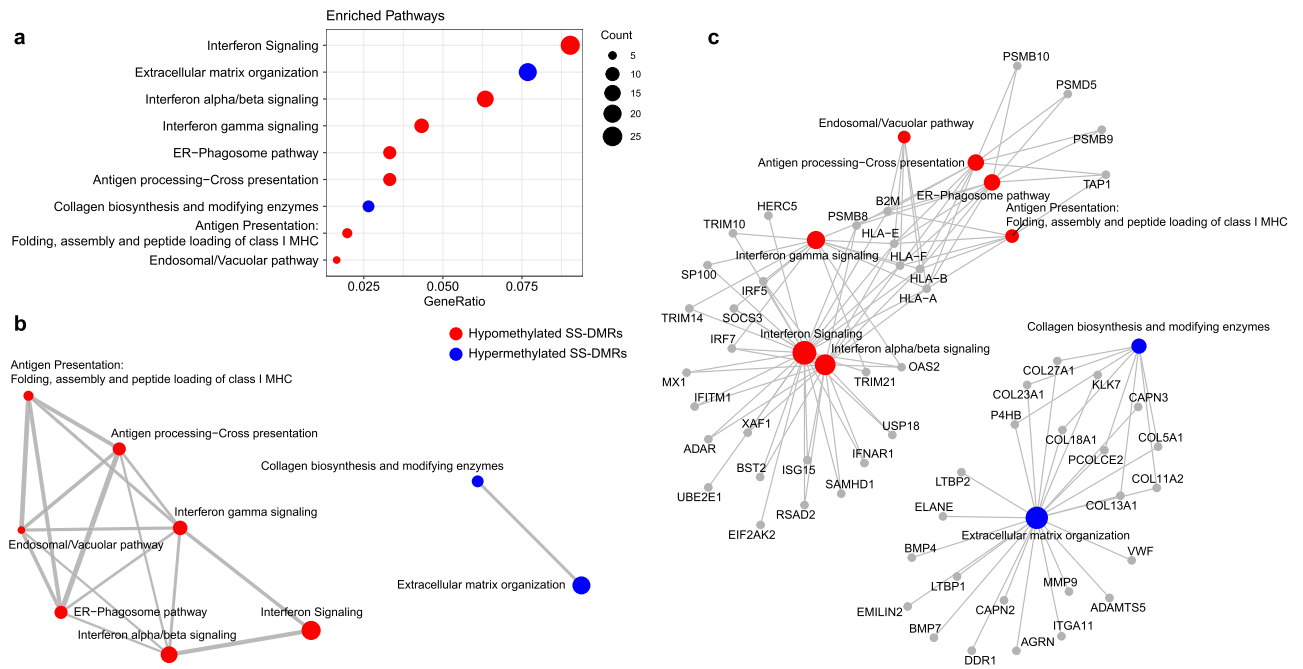
Figure 3. Functional Enrichment Results for differentially methylated regions based in Reactome database. (**a**) Dotplot representing Reactome functional pathways that are enriched in differentially methylated regions. Only significant pathways (adjusted $P < 0.05$) that are represented by more than 5 genes are illustrated. (**b**) Enrichment map that organizes significant enriched terms into a network with edges connecting overlapping gene sets. (c) Genes that are involved in significant terms are connected by linkages. Pathways in red are enriched in hypomethylated genes while pathways in blue are enriched in hypermethylated genes. Dot size represents number of genes with DMRs per pathways. R software[73] and Adobe Illustrator (https://www.adobe.com/) was used to create figures.

quantitative trait methylation (eQTM) analyses. We observed a total of 48 eQTMs or genes showing a strong correlation between DNA methylation levels and gene expression (Pearson's coefficient > 0.60, $P < 1.6 \times 10^{-34}$) (Supplementary Table 9). Many of the SS-eQTMs fell within the promoters of the IFN-related genes, such as *IFI44L* (Fig. 1c), *EPSTI1*, *MX1*, *DTX3L*, *PAPR9*, *LY6E*, *IFITM3*, *DDX60*, *RSAD2*, *PLSCR1* and *ADAR*. For all of them we observed that decreased DNA methylation strongly correlated with increased expression levels in SS patients, while no correlation was apparent in the healthy population. Interestingly, we observed strong coordinated effects at other non IFN-regulated genes implicated in the immune system such as *CD3D*, *FGR*, *PILRA* or *NLRC4*, and in genes with unrecognized function in SS, such as *MDGA1*, *PM20D1* and *SIRPB2* (Pearson's coefficient > 0.80). Genes showing correlated patterns of DNA methylation and gene expression were enriched mainly in immune-related gene ontologies being the most significant, immune response (GO:0006955), defense response (GO:0006952), activation of the immune response (GO:0002253) and response to stress (GO:0006950) (Supplementary Table 10). Our results suggest that SS-associated changes in DNA methylation can impact the transcriptional landscape in SS patients, and this could ultimately lead to alteration in the cellular function and immune response.

Genetic drivers of SS-associated differential methylation and expression.

To obtain insights into the extent to which differential methylation and expression associated with SS is genetically controlled, we performed *cis*-meQTL and *cis*-eQTL analyses (see Methods). By means of linear regression models that adjust for disease conditions, we found evidence for genetic control in 52% of SS-associated DMPs and 39% of SS-associated DEGs (gene variants no farther than 1 Mb to CpG or to transcription start site—TSS). Specifically, at an FDR of 5% we found a total of 4,305 significant meQTLs that included 61 SS-DMPs and 3,508 single-nucleotide polymorphisms (SNPs). We were able to assess the replication of 1475 meQTLs that involved 31 CpGs included in the 450 K array and replicated results for 20 CpGs (64%) that were involved in 754 meQTLs ($P < 0.05$) (Supplementary Table 11). The most significant meQTLs regulate DNA methylation at the *EPSTI1* gene, at an intergenic region in chromosome 3 where CCR cluster is located, and at *VRK2*, *ADAR*, *IRF7* and *MX1* genes. On the other hand, we identified a total of 11,399 significant eQTLs, that included 172 SS-DEGs and 10,620 SNPs, from which we could replicate 6414 eQTLs (56%) at a significance threshold of $P < 0.05$ and with consistent direction of effect, implicating 81 genes (Supplementary Table 12). Genetic variants located close to the TSS of *GBP3* show the strongest association with *GBP3* gene expression, followed by *cis*-genetic variation regulating gene expression of *C3AR1* or *MASTL*, *ETV7* and *IFITM3*.

CpG	Gene	rsID	SNP Position (hg38 Chr:bp)	A1/A2	β_{meQTL}	P_{meQTL}	β_{EWAS}	P_{EWAS}	OR	P	PRECISESADSSADs
A. Genetic effects on SS mediated by changes in DNA methylation											
cg08099136	PSMB8	rs9275569	6:32,710,259	C/T	-0.018	1.92×10^{-04}	-0.062	2.43×10^{-09}	2.03	1.57×10^{-08}	SLE (1×10^{-04})
cg14880222	IFI44	rs1051047*	1:78,663,909	A/G	0.027	5.94×10^{-08}	-0.036	3.61×10^{-09}	0.54	0.0127	SLE (0.042)
cg03879629	intergenic (CCR cluster)	rs9838739*	3:46,095,597	T/C	-0.046	5.21×10^{-05}	-0.071	7.81×10^{-09}	1.71	0.0242	SLE (0.002), MCTD (0.008), UCTD (0.004)
cg12013713	PARP12	rs6962291*#	7:139,971,218	T/A	-0.017	2.96×10^{-05}	-0.048	8.54×10^{-09}	1.32	0.0270	

Transcript	Gene	rsID	SNP Position (hg38 Chr:bp)	A1/A2	β_{eQTL}	P_{eQTL}	$\beta_{diffexp}$	$P_{diffexp}$	OR	P	PRECISESADS
B. Genetic effects on SS mediated by changes in gene expression											
ENSG00000166278	C2	rs1054684$#	6:32,809,044	T/C	0.300	3.57×10^{-04}	0.911	1.46×10^{-13}	4.04	2.16×10^{-13}	SLE(1×10^{-05}), MCTD (0.03), UCTD (0.048)
ENSG00000204267	TAP2	rs4947258	6:32,779,625	A/G	0.168	3.51×10^{-04}	0.478	2.31×10^{-17}	3.53	2.90×10^{-12}	SLE(1×10^{-05})
ENSG00000240065	PSMB9	rs3117106	6:32,375,492	T/C	0.139	2.19×10^{-04}	0.287	1.12×10^{-08}	2.22	1.23×10^{-07}	SLE (0.001)
ENSG00000204713	TRIM27	rs2523425*	6:29,526,364	C/T	-0.069	9.24×10^{-06}	-0.136	1.06×10^{-06}	1.51	7.32×10^{-04}	SADSnoSS (0.03)
ENSG00000128604	IRF5	rs4731531$*	7:128,922,493	G/A	0.084	3.45×10^{-04}	0.192	6.48×10^{-07}	1.46	0.0014	SLE (0.005)
ENSG00000174444	RPL4	rs4482223	15:65,828,392	A/G	0.101	3.18×10^{-04}	-0.195	1.01×10^{-09}	0.53	0.0019	RA (0.004)
ENSG00000124508	BTN2A2	rs3799378*#	6:26,404,046	A/G	0.096	2.74×10^{-04}	0.228	4.61×10^{-08}	1.54	0.0025	
ENSG00000234127	TRIM26	rs9261518	6:30,149,108	G/A	-0.113	2.66×10^{-05}	0.154	3.65×10^{-09}	0.60	0.0029	UCTD (0.01)
ENSG00000175970	UNC119B	rs10849822*	12:120,939,591	G/C	-0.066	3.28×10^{-04}	-0.161	2.88×10^{-09}	1.48	0.0075	SLE (0.02)
ENSG00000162614	NEXN	rs6660288	1:78,927,603	G/A	0.143	2.73×10^{-04}	0.685	3.02×10^{-14}	1.36	0.0156	MCTD (0.04), UCTD (0.02)
ENSG00000002549	LAP3	rs57570581	4:17,705,638	G/A	-0.216	2.76×10^{-05}	0.867	2.61×10^{-22}	0.75	0.0187	UCTD (0.02)
ENSG00000204516	MICB	rs9348883#	6:32,390,672	T/A	-0.174	3.12×10^{-04}	0.242	1.19×10^{-09}	0.46	0.0276	
ENSG00000154451	GBP5	rs76397273*	1:89,333,021	A/G	-0.288	9.70×10^{-05}	0.701	2.54×10^{-13}	0.67	0.0301	SSc (0.007), UCTD (0.03)
ENSG00000184898	RBM43	rs289953	2:151,143,385	T/C	0.183	8.55×10^{-09}	0.378	5.03×10^{-12}	1.29	0.0398	SLE (0.03), SSc (0.01)
ENSG00000166750	SLFN5	rs11080327*	17:35,244,427	G/A	0.222	4.68×10^{-14}	0.255	1.11×10^{-06}	1.28	0.0401	PAPs (0.04)

Table 2. Genetic association of SS-meQTLs and SS-eQTLs with SS and other related SADs mediated by DNA methylation or gene expression changes. Alleles represent major allele first, and then the minor allele, which is the allele tested in each analysis. β_{meQTL} represents the DNA methylation change with the increased in dosage of the minor allele. P_{meQTL} corresponds to the *P* value from the linear regression model that regresses out the number of minor alleles for a given SNP to DNA methylation levels adjusting by age, sex, batch effects, estimated cell proportions, disease status and first genetic component. β_{EWAS} represents the DNA methylation difference between SS and healthy controls from the epigenome-wide association study together obtained by linear regression model in which DNA methylation levels are regressed out by SS status and adjusted by age, sex, batch effects and estimated cell proportions. OR represents the Odd Ratio obtained from genetic association testing based on logistic regression modeling which the SS statuts is regressed out by number of minor alleles for a given SNP adjusted by age, sex, batch effects, estimated cell proportions and first genetic component and its corresponding. P represents the *P*-value obtained in the genetic associations. SAD (OR, P) represents the odd ratio and *P* value obtained in genetic testing for other diseases. RA = Rheumatoid Arthritis, SLE = Systemic Lupus Erythemathosus, UCTD = Undifferentiated Connective Tissue Disease, SSc = Systemic Scleroderma, PAPs = Primary anti-phospholipid syndrome. SADSnoSS = All SADs patients excluding SS. Genomic positions are based on the hg19 human reference sequence build (GRCh37). * eQTL reported in GTEx project (https://www.gtexportal.org/home/), in the case of eQTL the same SNP-gene is reported. # SNP associated with related disease phenotype in GWAS catalog (https://www.ebi.ac.uk/gwas/) or Open Target Genetics Portal (https://genetics.opentargets.org/).

Intermediary role of DNA methylation and gene expression in genetic risk of SS. Then, we investigated the possibility that changes in DNA methylation and gene expression mediate genetic risk in SS, as it has been shown to occur for other autoimmune diseases[13,33]. For that, we interrogated whether SNPs involved in meQTLs and eQTLs could be linked to the disease and showed allele frequency differences in a set of 391 SS patients and 549 healthy controls in a direction that is consistent with a mediation role of DNA methylation or gene expression (Supplementary Tables 13, 14, Methods). At a Bonferroni-corrected significance of P < 0.0008, we detected risk variants in the HLA region of chromosome 6 that regulate DNA methylation at gene *PSMB8* (Table 3). In this case, the minor T-allele of SNP rs7769693, at the *HLA-DRB9* pseudogene, is associated with increased SS risk and with decreased methylation levels, suggesting that the variant might exert its risk by hypo-methylating the *PSMB8* gene. Regarding eQTL results, at a Bonferroni-corrected significant level of P < 0.0002 we detected genetic variants associated with gene expression and SS at three differentially expressed genes *C2, TAP2* and *PSMB9* (Table 2).

CpG	Gene	SNP	SNP Position (hg38 Chr:bp)	Discovery cohort						Replication cohort	
				β_{INT}	P_{INT}	$\beta_{SS.meQTL}$	$P_{SS.meQTL}$	$\beta_{CTRL.meQTL}$	$P_{CTRL.meQTL}$	β_{INT}	P_{INT}
A. SS-dependent meQTLs											
cg08818207	TAP1	rs113547322	6:32,238,742	0.058	1.5×10^{-05}	0.048	0.0004	NA	>0.05	0.059	0.0091
cg14392283	LY6E	rs13273708	8:143,002,764	−0.018	1.6×10^{-04}	−0.016	0.0006	NA	>0.05	−0.020	0.0061
cg01309328	PSMB8	rs3134951	6:32,147,308	−0.040	4.0×10^{-04}	−0.024	0.0124	NA	>0.05	−0.040	0.0023
cg14951497	STAT1	rs4853645	2:191,839,218	−0.029	5.5×10^{-04}	−0.024	0.0012	NA	>0.05	−0.038	0.0418
cg12906975	Intergenic (LY6E)	rs55937049	8:143,033,566	−0.013	0.0018	−0.011	0.0087	NA	>0.05	−0.017	0.0410
cg23387863	SGK269	rs1079396	15:78,137,720	−0.024	0.0022	−0.017	0.0119	NA	>0.05	−0.028	0.0141
cg10734665	ATP10A	rs7169481	15:25,712,123	−0.022	0.0031	0.018	0.0047	NA	>0.05	0.030	0.0138
cg08099136	PSMB8	rs3129943	6:32,370,868	−0.030	0.0038	−0.027	0.0050	NA	>0.05	−0.035	0.0108

Gene ID	Gene	SNP	SNP Position (hg38 Chr:bp)	Discovery cohort						Replication cohort	
				β_{INT}	P_{INT}	$\beta_{SS.eQTL}$	$P_{SS.eQTL}$	$\beta_{CTRL.eQTL}$	$P_{CTRL.eQTL}$	β_{INT}	P_{INT}
B. SS-dependent eQTLs											
ENSG00000183486	MX2	rs9305702	21:40,755,922	−0.300	1.8×10^{-04}	−0.302	1.8×10^{-04}	NA	>0.05	−0.315	0.0172
ENSG00000198785	GRIN3A	rs2417310	9:101,964,703	0.251	4.5×10^{-04}	0.235	7.0×10^{-04}	NA	>0.05	0.262	0.0092
ENSG00000185885	IFITM1	rs12364973	11:1,116,140	0.386	6.9×10^{-04}	0.280	0.0130	NA	>0.05	0.430	0.0476
ENSG00000013374	NUB1	rs77466830	7:151,831,985	−0.143	7.7×10^{-04}	−0.151	3.5×10^{-04}	NA	>0.05	−0.158	0.0345
ENSG00000133106	EPSTI1	rs9525846	13:43,816,621	−0.607	8.8×10^{-04}	−0.577	2.4×10^{-03}	NA	>0.05	−0.651	0.0199
ENSG00000188313	PLSCR1	rs56077428	3:146,981,084	0.471	0.0010	0.357	0.0208	NA	>0.05	0.600	0.0185
ENSG00000157601	MX1	rs9305702	21:40,755,922	−0.602	0.0012	−0.496	0.0083	NA	>0.05	−0.779	0.0085
ENSG00000247317	LY6E − DT	rs902834	8:142,111,410	0.167	0.0038	0.129	0.0149	NA	>0.05	0.149	0.0469
ENSG00000108691	CCL2	rs1490922	17:33,725,536	0.351	0.0044	0.307	0.0159	NA	>0.05	0.367	0.0252
ENSG00000187210	GCNT1	rs2377425	9:75,947,610	0.132	0.0047	0.083	0.0452	NA	>0.05	0.210	0.0068

Table 3. Most significant meQTLs and eQTLs exhibiting disease-dependent genetic effects. β_{INT} represents the interaction effect between a given SNP and SS status in DNA methylation level. P_{INT} represents the P-value obtained for the β_{INT} in a linear regression model that adjusts for SNP, SS status, age, sex, batch effects, estimated cell proportions and the first principal genetic component. $\beta_{SS.meQTL}$ represents the DNA methylation change with the increased in dosage of the minor allele in SS population for a given SNP. $P_{SS.meQTL}$ corresponds to the P value from the linear regression model that regresses out the number of minor alleles for a given SNP to DNA methylation levels adjusting by age, sex, batch effects, estimated cell proportions, disease status and first genetic component in SS population. $\beta_{CTRL.meQTL}$ represents the DNA methylation change with the increased in dosage of the minor allele in the healthy control population for a given SNP. $P_{CTRL.meQTL}$ corresponds to the P value from the linear regression model that regresses out the number of minor alleles for a given SNP to DNA methylation levels adjusting by age, sex, batch effects, estimated cell proportions, disease status and first genetic component in the healthy control population. NA represents effects that are non significant ($P > 0.05$).

To give statistical robustness and further support to the functional and genetic link between meQTL and eQTLs with SS and with autoimmune processes, we first interrogated whether or not the novel identified SS-risk variants that reached a suggestive significance level of $P < 0.05$, show evidence of association with other SADs for which we had access to genotypic data: systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), systemic sclerosis (SSc), mixed connective tissue disease (MCTD), undifferentiated connective tissue disease (UCTD) and primary antiphospholipid syndrome (PAPs) (Table 2, Supplementary Table 15). Moreover, we searched for additional evidence of their functional role in public databases of eQTL and GWAS studies. Importantly, genetic variants regulating DNA methylation at the intergenic region of the CCR cluster in chr3 (Fig. 4a) and DNA methylation at the downstream region of *IFI44* gene (Fig. 4b) show convincing evidence of their link with other SADS and have been described as eQTL for the same genes (Supplementary Table 16). Likewise, genetic variants regulating gene expression of the genes *TRIM27* (Fig. 4c), *BTN2A2*, *UNC119B*, *GBP5* (Fig. 4d) and *SLFN5* show significant associations with other SADS and a link with gene expression has been previously observed in eQTL studies for these genes (Supplementary Table 16). Altogether, these results support the scenario in which differences in DNA methylation and gene expression are intermediates of genetic risk for SS. This study provides a framework for the discovery of new risk *loci* associated with SS via alteration of regulatory landscapes.

Disease-dependent genetic effects on SS-associated differential methylation and gene expression. Finally, we hypothesized that there might exist genetic variants whose effect on molecular phenotypes depends on the specific environment of altered immune activation originated during the disease process. In order to identify such disease-dependent genetic effects or *disease-interacting QTLs*, we performed a gene-by-environment interaction meQTLs and eQTLs analyses in which we looked for gene variants that
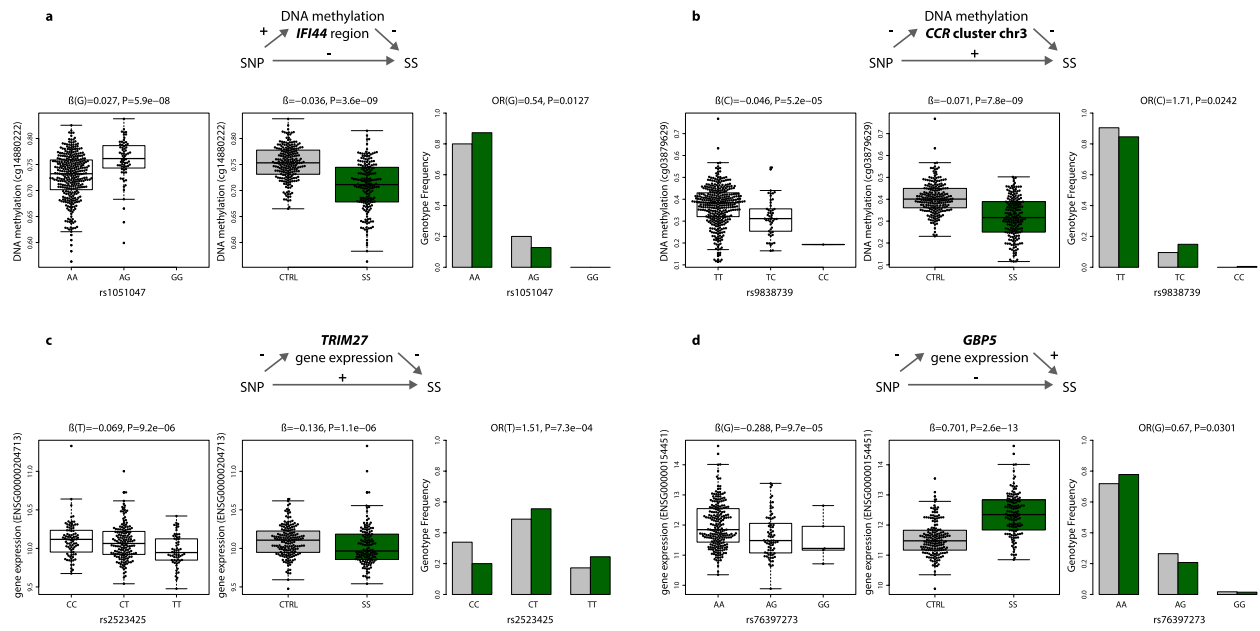
Figure 4. Intermediary role of DNA methylation and gene expression in SS genetic risk. (**a**) The minor G-allele of SNP rs1051047 exerts a protective role on SS susceptibility by increasing DNA methylation levels at the upstream region of gene *IFI44* (cg1488022). (**b**) The minor C-allele of SNP rs9838739 exerts risk on SS susceptibility by decreasing DNA methylation levels at the intergenic region within the *CCR* cluster in chromosome 3 (cg03879629). (**c**) The minor T-allele of SNP rs2523425 exerts risk on SS by decreasing *TRIM27* gene expression. (**d**) The minor G-allele of SNP rs76397273 exerts a protective effect on SS by decreasing *GBP5* gene expression. Green boxplots and barplots represent SS population, while grey plots represent the healthy control population. DNA methylation is quantified with β-values, gene expression is at the logarithmic scale. R software[73] and Adobe Illustrator (https://www.adobe.com/) was used to create figures.

interact with disease status on shaping DNA methylation and gene expression levels at SS-associated DMPs and DEGs. We considered relevant and robust disease interacting QTLs those that accomplish the following strict conditions: *i)* an interaction effect that passes a significance threshold of $P_{INTER} < 0.005$ in the discovery cohort and a significance threshold of $P_{INTER} < 0.05$ in the replication cohort, with consistent direction of the effect, *ii)* genetic variant associated with DNA methylation or gene expression only in the SS group ($P_{SS.QTL} < 0.05$) and without evidence of genetic associations in the healthy population ($P_{CTRL.QTL} > 0.05$), *iii)* genetic variants that show a minor allele frequency higher than 0.10 in the SS group.

For DNA methylation, we observed convincing evidence for disease-interacting meQTLs at 8 SS-associated DMPs within genes *TAP1, LY6E, PSMB8, STAT1, SGK269* and *ATP10A* (Table 3 and Supplementary Table 17). The most significant gene-environment interaction ($\beta_{INTER} = -0.018$, $P_{INTER} = 2.0 \times 10^{-04}$) that fell outside the HLA region involved the genetic variant rs13273708 at chromosome 8 and DNA methylation at *LY6E* (cg14392283) (Fig. 4a, Table 3). For this variant, the minor C-allele is associated with a decrease in *LY6E*-DNA methylation only in SS patients, but this association is not observed in the healthy population (Table 4). A similar scenario is observed for the *STAT1* gene for which we found the second most significant disease-dependent meQTLs effect ($\beta_{INTER} = -0.029$, $P_{INTER} = 5.47 \times 10^{-04}$) outside the HLA region (Fig. 5b). Another convincing example implicates the interaction between SS status and rs1079396 in shaping DNA methylation levels at *SGK269* gene ($\beta_{INTER} = -0.024$, $P_{INTER} = 0.003$, Fig. 5c), and the interaction with rs7169481 in shaping DNA methylation at *ATP10A* (cg10734665, $\beta_{INTER} = 0.022$, $P_{INTER} = 0.003$, Fig. 5d). For gene expression, some IFN-inducible genes such as *MX2, IFITM1, EPST1, MX1* and *LY6E-DT* show convincing evidence that their expression is genetically regulated in a disease-specific manner (Table 3 and Supplementary Table 18). The most significant disease-dependent effect was observed for *MX1* gene expression ($\beta_{INTER} = -0.300$, $P_{INTER} = 1.8 \times 10^{-0}$, Fig. 5e), which is overexpressed in SS patients, and only genetically regulated within the disease population. In this case, the minor T-allele of rs9305702 is associated with decreased *MX1* gene expression in SS patients, without evidence of association in the healthy population (Fig. 5e). Other genes such as *IFITM1* (Fig. 5f), *CCL2* (Fig. 4g), *NUB1* (Fig. 5h), *PLSCR1, GRIN3A* and *GCNT1* also show disease-specific eQTLs that regulate their SS-associated differential expression (Table 3).

For all those disease-interacting meQTLs and eQTLs, the same trend is observed in the replication sample (Supplementary Table 17 and 18) and we found no evidence for them to be genetic regulators of DNA methylation and/or gene expression in a model that corrects for disease status and includes the whole population (Supplementary Table 17 and 18), or in other studies that have interrogated healthy populations (Supplementary Table 19)[8,10,37]. Our findings reveal a differential genetic architecture of gene regulation between SS patients and the healthy population, and suggest that the specific immunological and pathological conditions in autoimmunity
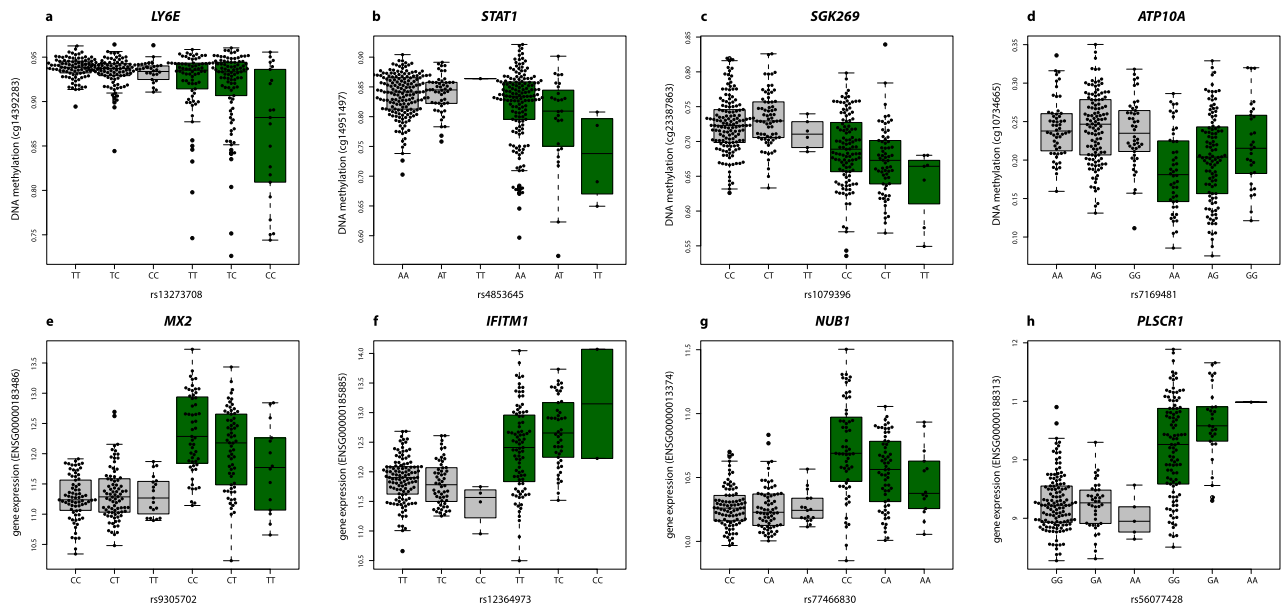
Figure 5. Disease interacting QTLs (**a**) The minor C-allele of SNP rs13273708 is associated with a decrease in DNA methylation levels at *LY6E* gene only in SS patients ($\beta_{SS.meQTL} = -0.018$, $P_{SS.meQTL} = 6 \times 10^{-04}$), but not in the healthy population ($P_{SS.meQTL} > 0.05$). (**b**) The minor A-allele of SNP rs902834 decreases the DNA methylation level at *STAT1* only in SS patients ($\beta_{SS.meQTL} = -0.024$, $P_{SS.meQTL} = 0.0012$), and not the healthy population ($P_{CTRL.meQTL} > 0.05$). (**c**) The minor T-allele of rs1079396 is associated with *SGK269*-methylation in SS patients ($\beta_{SS.meQTL} = -0.017$, $P_{SS.meQTL} = 0.0119$), but not in the healthy population ($P_{CTRL.meQTL} > 0.05$). (**d**) The minor G-allele of rs7169481 in *ATP10A* is associated with increased DNA methylation at *ATP10A* in SS patients ($\beta_{SS.meQTL} = 0.018$, $P_{SS.meQTL} = 0.0047$). However, in the healthy population this allele has no significant effect ($P_{CTRL.meQTL} > 0.05$). (**e**) The minor T-allele of the rs9305702 genetic variant is associated with a decreased *MX2* gene expression in SS patients ($\beta_{SS.eQTL} = -0.302$, $P_{SS.eQTL} = 1.8 \times 10^{-04}$), and shows no evidence of association in the healthy population ($P_{CTRL.eQTL} > 0.05$). (**f**) The minor C-allele of rs12364973 is associated with an increased *IFITM1* gene expression in SS patients ($\beta_{SS.eQTL} = 0.28$, $P_{SS.eQTL} = 0.012$) and shows no evidence of association in the healthy population ($P_{CTRL.eQTL} > 0.05$). (**g**) In SS patients, *NUBI* expression decreases with the dose of the minor A-allele of rs77466830 ($\beta_{SS.eQTL} = -0.15$, $P_{SS.eQTL} = 3.5 \times 10^{-04}$); however, in the healthy population it remains stable ($P_{CTRL.eQTL} > 0.05$). (**h**) In SS patients, *PLSCR1* expression increased with the dose of the minor A-allele of rs56077428 ($\beta_{SS.eQTL} = 0.357$, $P_{SS.eQTL} = 0.0208$); however, in the healthy population it remains stable ($P_{CTRL.eQTL} > 0.05$). Green boxplots represent SS population, while grey boxplots represent the healthy control population. DNA methylation is quantified with β-values, gene expression is at the logarithmic scale. R software[73] and Adobe Illustrator (https://www.adobe.com/) was used to create figures.

modify how genes regulate the transcriptional landscapes of immune cells, which likely has an impact in the function and fate of cells.

Discussion

Here, we present a comprehensive integrative large-scale analysis that serves to discovery new *loci* and pathways involved in SS, to recognize the importance of hypermethylation events in SS pathogenesis, and to unravel the genetic architecture of blood gene regulation in a systemic autoimmune disease.

We could confirm the activation of the IFN system in SS[38] by recognizing blood hypomethylation and over-expression of a large number of genes involved in IFN signaling or specially of type I IFN-inducible genes. Type I IFNs are key immune mediators involve in viral response and in activation of immune responses[39]. Previous large-scale studies have extensively characterized the interferon signature in SS with large scale gene expression data and recently, with DNA methylation as well, in different tissues and cell types[19,20,22,23,40–43]. Our stratified analyses reveal that the IFN epigenetic signature is restricted to only those SS patients that exhibit Anti-Ro/SSA autoantibodies positivity and enhanced in those that also exhibit Anti-La/SSB. It has been reported that patients positive for these autoantibodies[44] and that exhibit the IFN signature[45] are at higher risk for worse diagnosis some disease manifestations such as hypergammaglobulinemia and ongoing lymphomas. Several authors have recognized the important contribution to clinical management that stratifying patients based on gene expression based biomarkers of IFN signature[38]. Recently, a study has shown than an epigenetic-based IFN signature could be even a more suitable biomarker for patient classification[45].

The tight relationship between autoantibodies and IFN signaling has been reported by multiple population-based studies and experimental works in in blood, salivary glands and other target cells[22,36,40]. However, the mechanisms behind this close relationship are still unclear. Whether the IFN signaling have an effect on autoantibodies production or vice versa, is unresolved. On one hand, IFN signaling stimulated upon viral infection

could generate antibodies that cross-react with autoantigens, such as Anti-Ro/SSA and Anti-La/SSB, by molecular mimicry[46]. On the other hand, it is known that the deposition of immuno-complexes formed by autoantibodies, RNA-binding proteins and material releases from apoptotic cells stimulate type I IFN production and subsequently activate IFN-regulated genes[47]. Nevertheless, autoantibodies production seems not to be enough to trigger dysregulating IFN-signaling as this is not observed in healthy subjects that exhibit autoantibodies[48]. In our study we could investigate further the genetic drivers of this strong relationship. Genetics studies have revealed that in SS there is a strong relationship between autoantibody production and class II HLA variation. Indeed, some studies have clearly show that class II HLA is only a risk factor for patients that are positive for Anti-La and/or Anti-Ro[49,50].In a recent SADs molecular stratification[51], we recognized a strong relationship between class II HLA variation and a molecular cluster characterized by the IFN signature enriched in Anti-La and Anti-Ro positive patients. Beyond this previously gathered knowledge, here we have discovered that, specifically, the SS-associated class II HLA-DRB1-0301 and HLA-DQB1-0201 alleles are strongly associated with the IFN signature and with Anti-Ro/SSA and Anti-La/SSB production. Indeed, our stratified analyses revealed that class II HLA is only associated with SS and with autoantibody production in those patients exhibiting the IFN signature. Our findings point towards a complex interaction between class II HLA variation, IFN signaling and autoantibody production Other autoimmune diseases share a common IFN signature with SjS and common HLA risk variants, such as SLE and MCTD. Whether the same HLA alleles and autoantibodies are important for the IFN signature observed in other related autoimmune diseases remains to be explored. More studies that couple observational records with experimental data are needed to disentangle mechanistic routes behind these relationships and recognize potential new drug targets.

Importantly, we also identified a high number of novel hypermethylation events in genes not related to the IFN pathways when applying a statistical method based on gene set enrichments[52], a trend that was not observed when analyzing individual CpG sites. This method permits to detect associations of smaller magnitude but with consistent epigenetic patterns along pre-defined regions. Functional analyses in the group of genes exhibiting increased methylation in SS patients revealed enrichment in important pathways such as those related to the metabolism of collagen and/or implicated in extracellular matrix organization that have not been previously detected, but could explain the increased degradation of extracellular matrix structures and the significant loss of collagen observed in the lacrimal gland and other tissues in SS, and, therefore, have a key role in its pathogenesis[53]. We anticipate that future efforts in biomarker discovery will successfully recognize that these hypermethylated signals can also be of good utility for SS stratification. .

In this study, we discovered new *loci* associated with SS whose functional mechanisms could be the alteration of epigenetic states or gene expression profiles. The association that we have found between genetics, DNA methylation and SS in the *CCR* gene cluster in chromosome 3 is very interesting and represents a novel genetic risk variant. Our study shows that the minor C-allele of rs9838739, located in an active regulatory region overlapping many transcription factor binding sites and upstream the *CCR* gene cluster, increases risk for SS and several related systemic autoimmune diseases (SLE, MTD and UCTD) and decreases DNA methylation levels at this dense regulatory region. Previous eQTL studies from GTEx project identified that C-allele is associated with lower *CCR5* expression in whole blood[37]. Interestingly, the lack of *CCR5* on dendritic cells of a NOD mouse, an experimental model for SS disease and diabetes, promotes a proinflammatory environment in submandibular glands, a target and affected tissue in SS patients[54]. Furthermore, *CCR5* expression is decreased on circulating monocytes from SS patients and is correlated with increased levels of inflammatory chemokines[55] . Our results, together with previous findings, suggest that genetically determined reduced methylation at *CCR* cluste*r*-methylation, could contribute to SS presumably via enhanced gene expression of inflammatory chemokines. Interestingly, the same region has been strongly associated with COVID19 susceptibility and severity in several international efforts[56,57] indicating that common molecular players may be involved in COVID19 and autoimmunity susceptibility. Other than the well described regulatory role of *IRF5* risk variants[58], we also identified that genetic variants at *TRIM27, TRIM26, RLP4, NEXN, LAP3, GBP5, RBM43* and *SLNF5* genes could be implicated in SS through changes in gene expression. For example, we identified a novel risk variant that we are others have demonstrated that downregulates *TRIM27*[37,59]. TRIM27 is a molecule that inhibits the innate immune response and has been recently shown to negatively regulate NOD2 mediated signaling by physical interaction and degradation[60]. These data suggest that genetically determined reduced *TRIM27* expression in SS patients can lead to enhanced innate responses via abnormally enhanced NOD2 activity.

Importantly, the findings of this work corroborate our hypothesis that genetic variants may only manifest genetic regulatory effects in the specific context of altered immune activity exhibited in SS patients. We identified strong meQTL effects only among SS patients and that we did not observed and have not been previously described in the general population for *STAT1, LY6E* and *ATP10A* genes[11]. STAT1 is a signal transducer and transcriptional activator that is activated in response to several cytokines as IFN-alpha, IFN-gamma or IL-6. Stimulation of monocytes and B-cells in SS patients leads to increased sensitivity of immune cells from SS patients to STAT1-activating signals that might partly explain the IFN signature observed in SS[61] In line with these results, our transcriptional data shows overexpression of *STAT1* in SS patients. On the other hand, *LY6E*, an IFN-regulated gene that encodes for the Lymphocyte Antigen 6E, is hypomethylated and overexpressed in SS patients. Interestingly, a previous study discovered trans-eQTL effects on *LY6E* gene expression that is dependent on immune activation[62], supporting that its genetic regulation is context-specific. An earlier GWAS study on cytokine responses found that genetic variants at *ATP10A* are associated with IFN-gamma production in response to vaccinia virus in subjects who had received the smallpox vaccine[63]. These findings, and our results, support a scenario in which inter-individual genetic variation at this gene impacts IFN-regulated gene expression only upon immune activation and via alteration of epigenetic states.

Likewise, we also identified a group of IFN-related genes which expression is genetically regulated in a disease-specific manner, such as *MX1, MX2, EPST1* and *IFITM1*. We also discovered context-specific eQTLS

for other genes with immune-related function such as *CCL2* and *PSLCR1*. *CCL2* encodes for a chemokine ligand that is secreted during inflammatory processes whose serum circulating levels are increased in SS[64]. Intriguingly, the expression of *CCL2* is genetically regulated upon in vitro stimulation but it shows no genetic regulation in naïve conditions[65]. *PSLCR1*, the Phospholipid scramblase 1 gene, has been found to be differentially methylated and expressed in SS, SLE and MCTD[22,36,66]. It is a DNA-binding transcriptional activator that amplifies the IFN-mediated antiviral response and induces cytokine expression and cell proliferation. Interestingly, a previous study has shown that the genetic regulation of *PSLCR1* expression is stronger upon IFN stimulation[62]. We also found disease-interacting genetic variants whose implication in systemic autoimmunity has not been well described. For example, *NUB1* is an IFN-inducible gene that downregulates NEDD8, an ubiquitin-like protein. *NUB1* overexpression is known to induce inhibition of cell growth in cancer cells[67]. Its implication in autoimmune processes is uncertain but it could be related to lymphoma development.

Our study has some important limitations. First of all, we focused this work on the identification of molecular changes from whole-blood derived samples, which is a mixture of different immune-cell types, many of them directly involved in the autoimmune and inflammatory processes occurring in SS. The statistical approaches that we have applied have allowed us to identify epigenetic and transcriptional signatures that are common or shared across different blood cell types. This is, on one side, an advantage to future evaluation of these signals as blood biomarkers for disease diagnosis. However, our work has not revealed molecular changes occurring in specific cell types. Previous works have recognized cell-specific epigenetic signals on data from sorted cells but on a limited sample size[43,68]. We also could not evaluate specific epigenetic changes related to the development of some important disease phenotypes for SS, as for example lymphomas. Defective DNA methylation in SS patients with lymphomas and other cancers has been previously reported[69] and, recently a relationship between lymphoma and the DNA-methylation based interferon signature has been described[45]. However we did not have enough samples in our dataset to assess the epigenome-wide contribution in SS. Future studies that target specific disease phenotypes, that incorporate higher sample sizes, and add other cell-specific layers of molecular information will add more to the understanding of the genetic regulation and transcriptional consequences of epigenetic variation associated with SS and other systemic autoimmune diseases.

To wrap up, this study adds new pieces of evidences to the relationship between class II HLA variation, autoantibody profiles and the IFN molecular signature in SS, identifies new *loci* and pathways involved in SS, recognizes the importance of hypermethylation events in SS pathogenesis, and it contributes to unravel the genetic regulatory architecture of Sjögren's Syndrome, which hopefully will motivate research for the discovery of new drugs and biomarkers for SS in the near future.

Material and methods

Participant recruitment. Samples included in this study were recruited from the PRECISESADS study, which is a European multi center, non-randomized, and observational clinical study with recruitment performed between December 2014 and December 2018 at 19 institutions in 9 countries. PRECISESADS included patients affected by systemic autoimmune (around 400 by disease or group of diseases: rheumatoid arthritis, scleroderma or systemic sclerosis, primary Sjögren's syndrome, systemic lupus erythematosus, and primary antiphospholipid syndrome, mixed connective tissue disease and undifferentiated connective tissue disease) and 554 healthy controls[70,71].

A total of 558 samples were included in the core analyses of the present study (Supplementary Table 1), that included SS patients and healthy controls for which DNA methylation and genotype data was available. All patients diagnosed with SS ($N = 278$) fulfilled the diagnostic criteria of the American and European community published in 2002[72]. The majority of the SS patients (94.8%) were females with a mean of age of 58.8 ± 12.4 years. Healthy individuals ($N = 280$), i.e., not having any history of autoimmune or infectious diseases, were included as controls, and matched to cases to the extent possible. The 74.4% of the healthy subjects were females with a mean of age of 45.1 ± 13.3 years. A large proportion of SS patients exhibited Anti-Ro/SSA positivity and/or Anti-La/SSB (70.9% and 33.7%, respectively).

Samples were divided in two sets; a discovery cohort formed by 189 SS patients and 220 healthy subjects and a replication cohort that included 60 SS patients and 89 healthy individuals. Epidemiological, clinical and serological features of the samples included in the discovery and the replication cohorts are summarized in Supplementary Table 1. Flow cytometry data and determination of autoantibody presence was generated from serological samples as described in Barturen *et al.*[51].

Ethics declarations. An ethical protocol was prepared, reached consensus across all partners, academic and industrial, translated into all participant's languages and approved by each of the local ethical committees of the clinical recruitment centers, and all experimental protocols were approved by each of the local committees. For a list of local committees and centers involved in PRECISESADS please see Supplementary Note 1. All patients recruited to the study were aged 18 years or older and signed an informed consent form, and all methods were carried out in accordance with relevant guidelines and regulations. The study adhered to the standards set by International Conference on Harmonization and Good Clinical Practice, and to the ethical principles that have their origin in the Declaration of Helsinki (2013). The protection of the confidentiality of records that could identify the included individuals is ensured as defined by the EU Directive 2001/20/EC and the applicable national and international requirements relating to data protection in each participating country.

Genome-wide DNA methylation data and differential analyses. DNA methylation data of the discovery cohort was obtained using the Infinium Methylation EPIC BeadChip (Illumina, San Diego, CA, USA) that covers more than 800,000 CpG sites. For replication cohort, we had also available genome-wide DNA meth-

ylation data profiled obtained by Infinium Methylation 450 K BeadChip (Illumina, San Diego, CA, USA), which covers more than 400,000 CpG sites, most of them included in the EPIC array.

After DNA extraction from whole peripheral blood and bisulfite conversion, the genome for each sample was amplified, fragmented and hybridized to the corresponding Illumina arrays according to the manufacturer's protocol. The quality control of samples and the normalization of the data was performed using the meffil R software[73,74] Samples were excluded based on the detection P criteria > 99%, poor bisulfite conversion based on control dashboard check, and sex mismatches according to failed chromosome X and Y clustering. Probes were filtered out based on detection $P > 0.01$ in > 95% of samples. Additionally, probes located at the X and Y chromosomes were separated in different datasets to avoid gender bias. Probes with genetic variants at their CpG sites were also excluded. After applying these filtering steps we obtained 776,284 and 433,337 autosomic probes in the discovery dataset and in the replication dataset, respectively. A total of 17,530 probes located in the X chromosome passed the QC in the discovery cohort and 9282 in the replication cohort. DNA methylation was measured as a beta value ranging from 0 to 1. Zero represents an unmethylated state (0% molecules methylated at a particular sites) while 1 represents a fully methylated state (100% molecules methylated). After QC, the raw methylation beta values were background corrected and normalized using the functional normalization within the meffil R-package.

Differentially DNA methylated positions (DMPs) associated with SS were identified in our discovery cohort using a linear regression model that regresses out the SS status on DNA methylation levels at each CpG. The model was adjusted by sex, age, the first genetic principal component, the observed blood cell proportions obtained at the time of sample extraction, and batch effects as covariates. The technical variables Sentrix_ID and Sample_Plate corrected the batch effect. Data from blood cell proportions were obtained by flow cytometry using Duraclone tubes (Beckman Coulter) by whole blood flow cytometry, as described in Jamin et al.[75] and Barturen et al.[51] (see Supplementary Note 2 for a list of investigators involved in the PRECISESADs Flow Cytometry Study Group). They were calculated as percentages of total leukocytes. Given the population stratification observed in our data, the first genetic principal component was included in the linear model as covariate. We also searched for variable methylated positions (VMP) that show DNA methylated variance differences between SS patients and healthy samples. For that, we first obtain the residuals of DNA methylation levels after correcting for all covariates in a linear regression model. Then, we searched for variance differences in DNA methylation residuals between cases and controls by applying a Levene's test that accounts for mean differences. A Bonferroni-corrected threshold of $P < 6.4 \times 10^{-08}$ was established to consider DMPs and VMPs as significant. For each significant DMP the mean DNA methylation in cases and controls was calculated. The independent cohort was used to evaluate the robustness of the significant associations. The same steps were followed in the replication set. In this case, the replication significance threshold was established at $P < 0.05$. All statistical analyses were performed using R (v3.4.2) Software [73]. The effect of treatment and autoantibodies on the epigenetic associations found was evaluated. First, treatments and autoantibodies were, independently, included as covariates in the linear model and we compared the results of the treatment-, autoantibody-adjusted and unadjusted models. Specifically, we controlled for the use of antimalarial, steroids and immunosuppressants, since we observed that more than 10 SS patients had these drugs prescribed, and the presence of Anti-Ro/SSA and Anti-La/SSB autoantibodies. Then, stratified analyses according to the presence/absence of these treatments were performed using a linear regression model adjusted by sex, age, the first genetic principal component, the cell proportions, and batch effects. Supplementary Table 1 prevalence of treatments and autoantibodies in SS patients included in this study.

Differential DNA methylation regions (DMRs) associated with SS were estimated using mCSEA (methylated CpGs Set Enrichment Analysis) R package that implements a Gene Set Enrichment Analysis method (GSEA) to identify DMRs[52]. Briefly, the first step of mCSEA consists in ranking all the CpG probes by differential methylation and subsequently evaluates the enrichment of CpG sites belonging to the same region in the top positions of the ranked list by applying the GSEA implementation of the *fgsea* package. *mCSEA* allows to perform analysis based on promoters, gene bodies, and CGIs that are defined based on R annotation packages form Illumina for 450 K and EPIC arrays, respectively. A DMRs was considered as significant with a threshold FDR < 0.05 and with a minimum number of 5 CpGs associated to the region.

We used the mCSEA R package[52] to perform expression quantitative trait methylation (eQTM) analysis to discover significant associations between these DMRs and an expression alteration in the closer genes. In this case, the leading edge CpGs of each region is first defined and averaged for each region in each sample. Then, Pearson's correlation coefficient is calculated between each region's methylation and the proximal genes expression within 1500 base pairs upstream and downstream from the region.

RNA-seq data and differential expression analysis.

We analysed gene expression data available from 179 SS patients and 247 healthy subjects. RNA-seq data was obtained as 50 bp single end reads using an Illumina HiSeq 2500 sequencer and its proprietary base caller. Fastq data was aligned against human genome GRCh19 using STAR aligner[76] with GENCODE v19 [URL: https://www.gencodegenes.org/releases/19.html] annotation as single pass alignment. Read counts were obtained using RSEM[77]. Samples were included only if they had more than 8 million reads, a RIN score > 8 and did not appear as outliers in a principal component analysis. After filtering, a total of 42,878 genes were included in the study.

Differential expression analysis was performed using a linear model that regressed out SS status on read-counts as gene expression levels adjusted by age, sex, the first genetic principal component, the cell proportions and sequencing pool (POOL) and RNA integrity number (RIN) as batch effects. A Bonferroni-adjusted P-value threshold $< 1.16 \times 10^{-6}$ was established as significant to identify the differentially expressed genes (DEGs) in a discovery cohort formed by 135 SS patients and 174 healthy controls. In the replication cohort (44 SS patients

and 73 healthy controls), the same steps were followed while a replication significance threshold was established at $P < 0.05$. The analyses were performed using DESeq2[78].

Genotype profiling and imputation methods. Genomic DNA from whole blood was obtained by standard methods. The samples were genotyped using HumanCore -12-v1-0-B, InfiniumCoreExome-24v1-2 and InfiniumCoreExome-24v1-3, all of them of Illumina (San Diego, CA, USA). Only genetic variants, not indels, present on all three platforms were considered for data cleaning and analysis. Quality controls (QC) were performed using PLINK v1.9[79]. Genetic markers were removed if these had a call rate < 90%, exhibited a significant differential missingness between cases and controls ($P < 1 \times 10^{-4}$), and showed a significant deviation from Hardy–Weinberg equilibrium ($P < 0.01$ in controls and $P < 1 \times 10^{-4}$ in cases). Variants with a minor allele frequency of less than 5% were excluded from the analysis. Samples were excluded of the study if they had a call rate < 95% and also high heterozygosity rate, i.e., they deviated 6 standard deviations from the centroid. Duplicated or related individuals were identified using identity-by-descent criteria with REAP[80]. A total of 218,947 variants passed data filtering. Samples were excluded applying a threshold of kinship coefficient < 0.25.

Inference methods based on linkage disequilibrium structure was used in order to increase the number of genetic markers. Imputation was performed using the Michigan Imputation Server [URL: https://imputation server.sph.umich.edu/index.html][81] and Haplotype Reference Consortium (HRC) as reference panel [URL: http://www.haplotype-reference-consortium.org/][82]. We considered the imputed genotypes with a info-value (Minimac R^2) higher than 0.7, i.e., 70% of reliability. Imputed variants were also filtered according to the protocol described above.

In order to study the population structure and prevent population stratification the individual admixture frequency for each individual was estimated. In addition, population stratification was also analyzed by principal component analysis (PCA). Ancestry per cent for each individual was estimated using FRAPPE[83], and a set of 2,707 independent genetic variants that maximized the differences between populations and clustering by $K = 5$, i. e, the 5 global populations: American, African, South Asian, East Asian and European. European published populations from 1000 Genomes phase 3 [URL: http://www.internationalgenome.org/category/phase-3/] were included as reference panel. PCAs were also calculated using SMARTPCA from the Eigensoft software[84] and the same independent set of markers. In this analysis, only the European population from 1000 Genomes, excluding Finns, was included as reference population of our samples[85] (see supplementary Fig. 1). Six standard deviations from the centroid were used as a threshold to filter individuals that deviate from the main European clustering. Moreover, samples with less that 55% European ancestry were excluded from all analyses.

After genotyping QC and filtering of individuals for European ancestry classical HLA alleles were imputed for the PRECISESADS dataset (cases and controls) in the extended MHC region in chromosome 6[86]. The SNP2HLA software[87] was used for imputation using a reference panel consisting of 5,225 European individuals in the Type 1 Diabetes Genetic Consortium[88] containing data of 8,961 variants across the MHC region, and two and four digit-resolution allelic identities of the HLA class I (HLA-A, HLA-B, and HLA-C) and II genes (HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, and HLA-DRB1). Genotypes were extracted as allele dosages for 298 HLA classical alleles.

Quantitative trait loci analysis and genetic associations. Genetic regulation of DNA methylation and gene expression was explored by the quantitative trait *locus* (QTL) approach using Matrix eQTL[89]. A linear regression model was performed to evaluate the effect of closely located SNPs (no further than 1 Mb to CpG site or TSS) with a MAF > 0.05 on DNA methylation levels of DMPs and on gene expression patterns of DEGs. In this case, the gene expression data were firstly normalized into variance stabilizing transformations (VST) using DESeq2[78]. The linear model was corrected by age, sex, disease status, the first genetic principal component, cell proportions, and the respective batch effect. A P-value threshold of false discovery rate (FDR) < 0.05 was considered as significant.

To investigate the possibility of different behavior of genetic effects between SS patients and healthy subjects in regulating DNA methylation and gene expression, meQTL and eQTL interacting analyses that included an interaction term between the considered molecular phenotype and the disease status was included in the linear regression model. Because of the lower statistical power of this analysis, we investigated only genetic variants reaching a minimum allele frequency of 10% among SS patients. We considered as significant those interaction signals that pass a suggestive $P < 0.005$ in the discovery cohort, that show consistent direction of effects in the replication sample and reached a $P < 0.05$ that could only be detected as meQTL or eQTL among SS patients (FDR < 0.05), but not in healthy subjects ($P > 0.05$).

Case–control genetic association analyses were conducted using PLINK v1.9[79]. Logistic regressions under the allelic additive model were performed to interrogate the association between SS diagnosis and the dosage of the minor allele in SNPs involved in meQTL and eQTLs correcting for the first genetic principal component. We used a $P < 0.05$ to report significant genetic associations. Moreover, we tested the genetic association of SS-associated genetic variants with other SADs comparing the allele frequencies of 261 SS, 314 SLE, 297 SSc, 332 RA, 59 PAPs, 69 MCTD and 111 UCTD patients vs. 457 healthy controls.

Logistic and linear regressions under the additive model were performed to assess the genetic association between classical imputed HLA alleles and (i) risk for SS, (ii) presence of autoantibodies, and (iii) DNA methylation levels at the *IFI44* gene (cg13452062) as a proxy for the IFN epigenetic signature (epigIFN). We considered as significant those associations passing a $P < 1.7 \times 10^{-04}$ that corresponds to a Bonferroni-adjusted significance considering the number of HLA alleles tested. We also categorized SS patients as those exhibiting epigenetic hypomethylated IFN signature (DNA methylation-beta value < 0.8 at cg13452062) and those being negative

for the IFN signature (DNA methylation-beta value > = 0.8) and then performed HLA genetic associations in a stratified-fashion between the epigIFN positive and negative groups.

Bioinformatics tools. Significant DMPs were annotated to genes and gene locations according to annotation files provided by Illumina, from which we obtained a list of unique differentially methylated genes. Enrichment analyses based on gene ontology (GO) terms were performed using the web server ConsensusPathDB-human [URL: http://cpdb.molgen.mpg.de/CPDB][90] that integrates different types of functional interactions from 30 public sources in order to assemble a more complete and a less biased picture of cellular biology. Currently, ConsensusPathDB contains metabolic and signaling reactions, physical protein interactions, genetic interactions, gene regulatory interactions and drug-target interactions in human, mouse, and yeast.. We performed gene-set overexpression analyses using as a background the total list of autosomal genes covered by the Illumina EPIC array, a minimum set-size of 2 next-neighbors, and the following pathway databases: Reactome, KEGG, Wikipathways and Biocarta. For gene ontologies, we selected gene ontologies levels of 2 and 3 categories, and pathways, and both biological and molecular functions. Representations of enriched pathways were performed using cluster profiler R packages[91].

Data availability
The cohort datasets generated and analyzed during the current study are available upon request through ELIXIR platform.

References
1. Fox, R. I. Sjögren's syndrome. *Lancet* **366**, 321–331 (2005).
2. Rischmueller, M., Tieu, J. & Lester, S. Primary Sjögren's syndrome. *Best Pract. Res. Clin. Rheumatol.* **30**, 189–220 (2016).
3. Ice, J. A. *et al.* Genetics of Sjögren's syndrome in the genome-wide association era. *J. Autoimmun.* **39**, 57–63 (2012).
4. Burbelo, P. D., Ambatipudi, K. & Alevizos, I. Genome-wide association studies in Sjögren's syndrome: What do the genes tell us about disease pathogenesis?. *Autoimmun. Rev.* **13**, 756–761 (2014).
5. Teruel, M. & Alarcón-Riquelme, M. E. Genetics of systemic lupus erythematosus and Sjögren's syndrome: An update. *Curr. Opin. Rheumatol.* **28**, 506–514 (2016).
6. Lessard, C. J. *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. *Nat. Genet.* **45**, 1284–1294 (2013).
7. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
8. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
9. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
10. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49**, 131–138 (2017).
11. Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**(9), 1311–1321 (2021).
12. Walsh, A. M. *et al.* Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations. *Genome Biol.* **17**, 79 (2016).
13. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**, 142–147 (2013).
14. Odhams, C. A. *et al.* Mapping eQTLs with RNA-seq reveals novel susceptibility genes, non-coding RNAs and alternative-splicing events in systemic lupus erythematosus. *Hum. Mol. Genet.* **26**, 1003–1017 (2017).
15. Imgenberg-Kreuz, J. *et al.* DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann. Rheum. Dis.* https://doi.org/10.1136/annrheumdis-2017-212379 (2018).
16. Arvaniti, P. *et al.* Linking genetic variation with epigenetic profiles in Sjögren's syndrome. *Clin. Immunol.* **210**, 108314 (2020).
17. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
18. Yet, I., Tsai, P.-C., Castillo-Fernandez, J. E., Carnero-Montoro, E. & Bell, J. T. Genetic and environmental impacts on DNA methylation levels in twins. *Epigenomics* **8**, 105–117 (2016).
19. Altorok, N. *et al.* Genome-wide DNA methylation patterns in naïve CD4+T cells from patients with primary Sjögren's syndrome. *Arthritis Rheumatol.* **66**, 731–739 (2014).
20. Imgenberg-Kreuz, J. *et al.* Genome-wide DNA methylation analysis in multiple tissues in primary Sjögren's syndrome reveals regulatory effects at interferon-induced genes. *Ann. Rheum. Dis.* **75**, 2029–2036 (2016).
21. Carnero-Montoro, E. & Alarcón-Riquelme, M. E. Epigenome-wide association studies for systemic autoimmune diseases: The road behind and the road ahead. *Clin. Immunol.* **196**, 21–33 (2018).
22. Emamian, E. S. *et al.* Peripheral blood gene expression profiling in Sjögren's syndrome. *Genes Immun.* **10**, 285–296 (2009).
23. Devauchelle-Pensec, V. *et al.* Gene expression profile in the salivary glands of primary Sjögren's syndrome patients before and after treatment with rituximab. *Arthritis Rheum.* **62**, 2262–2271 (2010).
24. Miceli-richard, C. *et al.* Overlap between differentially methylated DNA regions in blood B lymphocytes and genetic at-risk loci in primary Sjögren' s syndrome. *Ann. Rheum. Dis.* **75**(5), 933–940. https://doi.org/10.1136/annrheumdis-2014-206998 (2016).
25. Kerick, M. *et al.* Expression quantitative trait locus analysis in systemic sclerosis identifies new candidate genes associated with multiple aspects of disease pathology. *Arthritis Rheumatol.* https://doi.org/10.1002/art.41657 (2021).
26. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLOS Genet.* **11**, e1005223 (2015).
27. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
28. Naranbhai, V. *et al.* Genomic modulators of gene expression in human neutrophils. *Nat. Commun.* **6**, 1–13 (2015).
29. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
30. Kim-Hellmuth, S. *et al.* Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **8**, 266 (2017).
31. Ecker, S., Pancaldi, V., Valencia, A., Beck, S. & Paul, D. S. Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. *BioEssays* **40**, 1–11 (2018).

32. Ho, J. W. K., Stefani, M., Dos Remedios, C. G. & Charleston, M. A. Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* **24**, i390 (2008).

33. Carnero-Montoro, E. *et al.* Epigenome-wide comparative study reveals key differences between mixed connective tissue disease and related systemic autoimmune diseases. *Front. Immunol.* **10**, 1880 (2019).

34. Paul, D. S. *et al.* Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat. Commun.* **7**, 13555 (2016).

35. Webster, A. P. *et al.* Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. *Genome Med.* **10**, 64 (2018).

36. Imgenberg-Kreuz, J. *et al.* Shared and unique patterns of DNA methylation in systemic lupus erythematosus and primary Sjögren's syndrome. *Front. Immunol.* **10**, 1686 (2019).

37. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

38. Li, H., Ice, J. A., Lessard, C. J. & Sivils, K. L. Interferons in Sjögren's syndrome: genes, mechanisms, and effects. *Front. Immunol.* **4**, 290 (2013).

39. González-Navajas, J. M., Lee, J., David, M. & Raz, E. Immunomodulatory functions of type I interferons. *Nat. Rev. Immunol.* **12**, 125–135 (2012).

40. Gottenberg, J. E. *et al.* Activation of IFN pathways and plasmacytoid dendritic cell recruitment in target organs of primary Sjögren's syndrome. *Proc. Natl. Acad. Sci. USA* **103**, 2770–2775 (2006).

41. Tang, Y. *et al.* MicroRNA-146a contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins. *Arthritis Rheum.* **60**, 1065–1075 (2009).

42. Cole, M. B. *et al.* Epigenetic signatures of salivary gland inflammation in Sjögren's syndrome's. *Arthritis Rheumatol.* **68**, 2936–2944 (2016).

43. Charras, A. *et al.* Cell-specific epigenome-wide DNA methylation profile in long-term cultured minor salivary gland epithelial cells from patients with Sjögren's syndrome. *Ann. Rheum. Dis.* **76**, 625–628 (2017).

44. Brito-Zerón, P. *et al.* How immunological profile drives clinical phenotype of primary Sjögren's syndrome at diagnosis: analysis of 10,500 patients (Sjögren Big Data Project). *Clin. Exp. Rheumatol.* **36**(Suppl 112), S102–S111 (2018).

45. Imgenberg-Kreuz, J. *et al.* DNA methylation-based interferon scores associate with sub-phenotypes in primary Sjögren's syndrome. *Front. Immunol.* **12**, 2772 (2021).

46. Igoe, A. & Scofield, R. H. Autoimmunity and infection in Sjögren's syndrome. *Curr. Opin. Rheumatol.* **25**, 480–487 (2013).

47. Båve, U. *et al.* Activation of the type I interferon system in primary Sjögren's syndrome: a possible etiopathogenic mechanism. *Arthritis Rheum.* **52**, 1185–1195 (2005).

48. Tb, N., Tl, R., Jp, B. & Mk, C. Serum type I interferon activity is dependent on maternal diagnosis in anti-SSA/Ro-positive mothers of children with neonatal lupus. *Arthritis Rheum.* **58**, 541–546 (2008).

49. Thorlacius, G. E. *et al.* Genetic and clinical basis for two distinct subtypes of primary Sjögren's syndrome. *Rheumatol. (UK)* **60**, 837–848 (2021).

50. Gottenberg, J. E. *et al.* In primary Sjögren's syndrome, HLA class II is associated exclusively with autoantibody production and spreading of the autoimmune response. *Arthritis Rheum.* **48**, 2240–2245 (2003).

51. Barturen, G. *et al.* Integrative analysis reveals a molecular stratification of systemic autoimmune diseases. *Arthritis Rheumatol.* **73**, 1073–1085 (2021).

52. Martorell-Marugán, J., González-Rumayor, V., Carmona-Sáez, P. & Valencia, A. MCSEA: detecting subtle differentially methylated regions. *Bioinformatics* **35**, 3257–3262 (2019).

53. Schenke-Layland, K. *et al.* Increased degradation of extracellular matrix structures of lacrimal glands implicated in the pathogenesis of Sjögren's syndrome. *Matrix Biol.* **27**, 53–66 (2008).

54. Wildenberg, M. E. *et al.* Lack of CCR5 on dendritic cells promotes a proinflammatory environment in submandibular glands of the NOD mouse. *J. Leukoc. Biol.* **83**, 1194–1200 (2008).

55. Lack of CCR5 on dendritic cells promotes a proinflammatory environment in submandibular glands of the NOD mouse - Wildenberg - 2008 - Journal of Leukocyte Biology - Wiley Online Library. Available at: https://jlb.onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1189/jlb.1107794. (Accessed: 12th August 2020)

56. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).

57. Initiative, T. C.-19 H. G. & Ganna, A. TITLE Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *medRxiv* 2021.03.10.21252820 (2021). doi:https://doi.org/10.1101/2021.03.10.21252820

58. Graham, R. R. *et al.* A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* **38**, 550–555 (2006).

59. Nguyen, C. Q. & Peck, A. B. The interferon-signature of Sjögren's syndrome: How unique biomarkers can identify underlying inflammatory and immunopathological mechanisms of specific diseases. *Front. Immunol.* **4**, 142 (2013).

60. Zurek, B. *et al.* TRIM27 Negatively Regulates NOD2 by Ubiquitination and Proteasomal Degradation. *PLoS ONE* **7**, e41255 (2012).

61. Pertovaara, M., Silvennoinen, O. & Isomäki, P. Cytokine-induced STAT1 activation is increased in patients with primary Sjögren's syndrome. *Clin. Immunol.* **165**, 60–67 (2016).

62. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **7**, 343 (2014).

63. Kennedy, R. B. *et al.* Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. *Hum. Genet.* **131**, 1403–1421 (2012).

64. Szodoray, P., Alex, P., Brun, J. G., Centola, M. & Jonsson, R. Circulating cytokines in primary Sjögren's syndrome determined by a multiplex cytokine array system. *Scand. J. Immunol.* **59**, 592–599 (2004).

65. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949–1246949 (2014).

66. Liu, J. *et al.* An integrative cross-omics analysis of DNA methylation sites of glucose and insulin homeostasis. *Nat. Commun.* **10**, 2581 (2019).

67. Hosono, T., Tanaka, T., Tanji, K., Nakatani, T. & Kamitani, T. NUB1, an interferon-inducible protein, mediates anti-proliferative actions and apoptosis in renal cell carcinoma cells through cell-cycle regulation. *Br. J. Cancer* **102**, 873–882 (2010).

68. Altorok, N. *et al.* Genome-wide DNA methylation patterns in naive CD4+ T cells from patients with primary Sjögren's syndrome. *Arthritis Rheumatol.* **66**, 731–739 (2014).

69. Fragkioudaki, S. *et al.* MTHFR gene variants and non-MALT lymphoma development in primary Sjogren's syndrome. *Sci. Rep.* **7**, 1–8 (2017).

70. Hofmann-Apitius, M., Alarcón-Riquelme, M. E., Chamberlain, C. & McHale, D. Towards the taxonomy of human disease. *Nat. Rev. Drug Discov.* **14**, 75–76 (2015).

71. Barturen, G. *et al.* Integrative analysis reveals a molecular stratification of systemic autoimmune diseases. *Arthritis Rheumatol* **73**(6), 1073–1085 (2020).

72. Vitali, C. *et al.* Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann. Rheum. Dis.* **61**, 554–558 (2002).

73. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. *R Foundation for Statistical Computing, Vienna, Austria.* (2013).

74. Min, J. L., Hemani, G., Smith, G. D., Relton, C. & Suderman, M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989 (2018).
75. Jamin, C. *et al.* Multi-center harmonization of flow cytometers in the context of the European 'PRECISESADS' project. *Autoimmun. Rev.* **15**, 1038–1045 (2016).
76. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
77. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).
78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
79. 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).
80. Thornton, T. *et al.* Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91**, 122–138 (2012).
81. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
82. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
83. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
84. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
85. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
86. De Bakker, P. I. W. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
87. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
88. Brown, W. M. *et al.* Overview of the MHC fine mapping data. *Diabetes Obes. Metab.* **11**, 2–7 (2009).
89. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
90. Kamburov, A. *et al.* ConsensusPathDB: Toward a more complete picture of cell biology. *Nucl. Acids Res.* **39**, D712 (2011).
91. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Nat. Protoc.* https://doi.org/10.1016/j.xinn.2021.100141 (2021).

Acknowledgements

Author contributions

E.C.-M., M.E.A.-R., and M.T. contributed to the conception and design of the study and also drafted the manuscript. The PRECISESADS Clinical Consortium members agreed on the clinical data, recruited the patients, and performed their detailed clinical assessment. PRECISESADS Cytometry Consortium obtained flow cytometry data. M.T., E.C.-M., G.B., M.B., O.C.-P., E.P., M.K., S.Z., A.B. and M.M.-B. contributed to pre-processing and/or analyzing the data. M.E.A.-R., J.M., J.O.-P. and E.B. contributed to data generation and sample recruitment. M.E.A.-R. coordinated the entire PRECISESADS project. All authors read and approved the final manuscript.

Competing interests

Zuzanna Makowska and Anne Buttgereit are employees of BAYER AG. All other authors have no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01324-0.

Correspondence and requests for materials should be addressed to E.C.-M. or M.E.A.-R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

PRECISESADS Clinical Consortium

Lorenzo Beretta[8], Barbara Vigone[8], Jacques-Olivier Pers[9], Alain Saraux[9], Valérie Devauchelle-Pensec[9], Divi Cornec[9], Sandrine Jousse-Joulin[9], Bernard Lauwerys[10], Julie Ducreux[10], Anne-Lise Maudoux[10], Carlos Vasconcelos[11], Ana Tavares[11], Esmeralda Neves[11], Raquel Faria[11], Mariana Brandão[11], Ana Campar[11], António Marinho[11], Fátima Farinha[11], Isabel Almeida[11], Miguel Angel Gonzalez-Gay Mantecón[12], Ricardo Blanco Alonso[12], Alfonso Corrales Martínez[12], Ricard Cervera[13], Ignasi Rodríguez-Pintó[13], Gerard Espinosa[13], Rik Lories[14], Ellen De Langhe[14], Nicolas Hunzelmann[15], Doreen Belz[15], Torsten Witte[16], Niklas Baerlecken[16], Georg Stummvoll[17], Michael Zauner[17], Michaela Lehner[17], Eduardo Collantes[18], Rafaela Ortega-Castro[18], Mª Angeles Aguirre-Zamorano[18], Alejandro Escudero-Contreras[18], Mª Carmen Castro-Villegas[19], Norberto Ortego[19], María Concepción Fernández Roldán[19], Enrique Raya[20], Inmaculada Jiménez Moleón[20], Enrique de Ramon[21], Isabel Díaz Quintero[21], Pier Luigi Meroni[22], Maria Gerosa[22], Tommaso Schioppo[22], Carolina Artusi[22], Carlo Chizzolini[23], Aleksandra Zuber[23], Donatienne Wynar[23], Laszló Kovács[24], Attila Balog[24], Magdolna Deák[24], Márta Bocskai[24], Sonja Dulic[24], Gabriella Kádár[24], Falk Hiepe[25], Velia Gerl[25], Silvia Thiel[25], Manuel Rodriguez Maresca[26], Antonio López-Berrio[26], Rocío Aguilar-Quesada[26] & Héctor Navarro-Linares[26]

[8]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milano, Italy. [9]Centre Hospitalier Universitaire de Brest, Hospital de la Cavale Blanche, Brest, France. [10]Pôle de pathologies rhumatismales systémiques et inflammatoires, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. [11]Centro Hospitalar do Porto, Porto, Portugal. [12]Servicio Cantabro de Salud, Hospital Universitario Marqués de Valdecilla, Santander, Spain. [13]Hospital Clinic I Provicia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. [14]Katholieke Universiteit Leuven, Leuven, Belgium. [15]Klinikum der Universitaet zu Koeln, Cologne, Germany. [16]Medizinische Hochschule Hannover, Hannover, Germany. [17]Medical University Vienna, Vienna, Austria. [18]Servicio Andaluz de Salud, Hospital Universitario Reina Sofía Córdoba, Córdoba, Spain. [19]Servicio Andaluz de Salud, Complejo hospitalario Universitario de Granada (Hospital Universitario San Cecilio), Granada, Spain. [20]Servicio Andaluz de Salud, Complejo hospitalario Universitario de Granada (Hospital Virgen de las Nieves), Granada, Spain. [21]Servicio Andaluz de Salud, Hospital Regional Universitario de Málaga, Málaga, Spain. [22]Università degli studi di Milano, Milan, Italy. [23]Hospitaux Universitaires de Genève, Genève, Switzerland. [24]University of Szeged, Szeged, Hungary. [25]Charite, Berlin, Germany. [26]Andalusian Public Health System Biobank, Granada, Spain.

PRECISESADS Flow Cytometry Study Group

Montserrat Alvarez[27], Damiana Alvarez-Errico[28], Nancy Azevedo[29], Nuria Barbarroja[30,31], Anne Buttgereit[32], Qingyu Cheng[33], Carlo Chizzolini[27], Jonathan Cremer[34], Aurélie De Groof[35], Ellen De Langhe[36], Julie Ducreux[35], Aleksandra Dufour[27], Velia Gerl[33], Maria Hernandez-Fuentes[37], Laleh Khodadadi[33], Katja Kniesch[38], Tianlu Li[33], Chary Lopez-Pedrera[35], Zuzanna Makowska[32], Concepción Marañón[40], Brian Muchmore[40], Esmeralda Neves[29], Bénédicte Rouvière[41], Quentin Simon[41], Elena Trombetta[39], Nieves Varela[38] & Torsten Witte[38]

[27]Immunology and Allergy, University Hospital and School of Medicine, Geneva, Switzerland. [28]Chromatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. [29]Serviço de Imunologia EX-CICAP, Centro Hospitalar e Universitário do Porto, Porto, Portugal. [30]IMIBIC, Reina Sofia Hospital, University of Cordoba, Córdoba, Spain. [31]Bayer AG, Berlin, Germany. [32]Pharmaceuticals Division, Bayer Pharma, Berlin, Germany. [33]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany. [34]Department of Microbiology and Immunology, Laboratory of Clinical Immunology, KU Leuven, Leuven, Belgium. [35]Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, Brussels, Belgium. [36]University Hospitals Leuven and Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium. [37]UCB, Slough, UK. [38]Klinik für Immunologie Und Rheumatologie, Medical University Hannover, Hannover, Germany. [39]Laboratorio di Analisi Chimico Cliniche e Microbiologia - Servizio di Citofluorimetria, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy. [40]GENYO, Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain. [41]NSERM, UMR1227, CHRU Morvan, Lymphocytes B et Autoimmunité, University of Brest, BP 824, Brest, France.