

<https://doi.org/10.1038/s41525-024-00420-0>

# Molecular subtypes explain lupus epigenomic heterogeneity unveiling new regulatory genetic risk variants

Check for updates

Olivia Castellini-Pérez<sup>1,2</sup>, Elena Povedano<sup>1,3,4</sup>, Guillermo Barturen<sup>1,5</sup>, Manuel Martínez-Bueno<sup>1</sup>, Andrii Iakovliev<sup>6</sup>, Martin Kerick<sup>7</sup>, Raúl López-Domínguez<sup>1</sup>, Concepción Marañón<sup>1</sup>, Javier Martín<sup>7</sup>, Esteban Ballestar<sup>8</sup>, PRECISEADS Clinical Consortium\*, PRECISEADS Flow Cytometry Study Group\*, María Orietta Borghi<sup>9</sup>, Weiliang Qiu<sup>10</sup>, Cheng Zhu<sup>11</sup>, Srinivas Shankara<sup>11</sup>, Athina Spiliopoulou<sup>6</sup>, Emanuele de Rinaldis<sup>11,46</sup>, Elena Carnero-Montoro<sup>1,2,46</sup> ✉ & Marta E. Alarcón-Riquelme<sup>1,12,46</sup> ✉

The heterogeneity of systemic lupus erythematosus (SLE) can be explained by epigenetic alterations that disrupt transcriptional programs mediating environmental and genetic risk. This study evaluated the epigenetic contribution to SLE heterogeneity considering molecular and serological subtypes, genetics and transcriptional status, followed by drug target discovery. We performed a stratified epigenome-wide association studies of whole blood DNA methylation from 213 SLE patients and 221 controls. Methylation quantitative trait loci analyses, cytokine and transcription factor activity - epigenetic associations and methylation-expression correlations were conducted. New drug targets were searched for based on differentially methylated genes. In a stratified approach, a total of 974 differential methylation CpG sites with dependency on molecular subtypes and autoantibody profiles were found. Mediation analyses suggested that SLE-associated SNPs in the HLA region exert their risk through DNA methylation changes. Novel genetic variants regulating DNAm in disease or in specific molecular contexts were identified. The epigenetic landscapes showed strong association with transcription factor activity and cytokine levels, conditioned by the molecular context. Epigenetic signals were enriched in known and novel drug targets for SLE. This study reveals possible genetic drivers and consequences of epigenetic variability on SLE heterogeneity and disentangles the DNAm mediation role on SLE genetic risk and novel disease-specific meQTLs. Finally, novel targets for drug development were discovered.

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease (SAD) caused by the activation of autoreactive T and B cells, the release of inflammatory cytokines and the formation of immune complexes that deposit in tissues, resulting in organ damage. It predominantly appears in young to middle-aged women with a 9:1 female:male bias<sup>1</sup>. Treating and

diagnosing SLE is challenged by the patients' heterogeneity in terms of diversity of symptoms, manifestations<sup>2</sup>, the organs affected, and a diverse array of autoantibody (AAb) specificities. Although AAb production helps in the diagnosis of some autoimmune diseases and is related to several clinical manifestations, the complexity of SLE is such that patients can

<sup>1</sup>GENYO. Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, 18016 Granada, Spain. <sup>2</sup>University of Granada, Granada, Spain. <sup>3</sup>Spanish National Research Council (CSIC), Institute of Economy, Geography and Demography, Madrid (IEGD), Madrid, Spain.

<sup>4</sup>Autonomous University of Madrid, Madrid, Spain. <sup>5</sup>Department of Genetics, Faculty of Sciences, University of Granada, Granada, Spain. <sup>6</sup>Usher Institute of Population Health Sciences and Informatics, University of Edinburgh Medical School, EH8 9YL Edinburgh, UK. <sup>7</sup>IBPLN-CSIC, Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, 18016, Spain. <sup>8</sup>Epigenetics and Immune Disease Group, Josep Carreras Research Institute (JRC), 08916 Badalona, Barcelona, Spain. <sup>9</sup>Università degli Studi di Milano and Istituto Auxologico Italiano, Milan, Italy. <sup>10</sup>Sanofi, Early Development and Research, Cambridge, MA, USA. <sup>11</sup>Sanofi, Precision Medicine & Computational Biology (PMCB), R&D, Cambridge, MA, USA. <sup>12</sup>Institute for Environmental Medicine, Karolinska Institutet, 171 67 Solna, Sweden. <sup>46</sup>These authors contributed equally: Emanuele de Rinaldis, Elena Carnero-Montoro, Marta E. Alarcón-Riquelme. \*Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: [elena.carnero@genyo.es](mailto:elena.carnero@genyo.es); [marta.alarcon@genyo.es](mailto:marta.alarcon@genyo.es)

present a wide range of specificities, being, moreover, not solely identified in SLE or in specific SLE manifestations.

The understanding of genetic, environmental, and molecular mechanisms disrupting immunity and triggering autoantibody production is not well understood. Genome-wide association studies (GWAS) identified several susceptibility genes among which HLA class II locus and genes such as *TNIP*, *BANK1*, and *IRF5* exhibit the strongest risk effects<sup>3,4</sup>. A strong HLA genetic association with the presence of autoantibodies such as anti-SSB, anti-SSA, anti-RNP, anti-SM<sup>5</sup>, and anti-dsDNA production has been recognized since long<sup>5,6</sup>. The most important molecular characteristic of SLE is the overexpression of several interferon-regulated genes (IRG) known as the interferon (IFN) signature, also widely observed at the epigenetic level (epigenIFNsig) in all blood cell types and tissues<sup>7</sup> as well as in other SADS<sup>8,9</sup>. A strong interaction between HLA genetic variation, the production of anti-SSA AAb and the epigenIFNsig has been reported in primary Sjogren's syndrome<sup>9</sup>, however the genetic and autoantibody determinants of the IFN signature in SLE are still not clear. Recently, the study by *Barturen et al.*<sup>8</sup> molecularly reclassified SLE and other six SADs into different molecular subtypes with important clinical implications. SLE and SADs patients could be stratified into an inflammatory subtype, with increased activity of genes related to the function of monocyte and neutrophil; a lymphoid subtype, with genes related to the function of these immune cells; and an IFN subtype, defined by enhanced activity in genes induced by IFN. This molecular classification needs to be taken into account to overcome SLE heterogeneity.

Heritability studies showed that genetics explains only a small fraction, <10% of SLE susceptibility<sup>10-12</sup>, suggesting the important contribution of environmental or non-genetic factors<sup>13</sup> and possibly the role of epigenetics mediating gene-by-environment interactions. Epigenetic modifications allow to retrieve different phenotypes from a unique genotype, and are fundamental for immune cells to exhibit diverse and

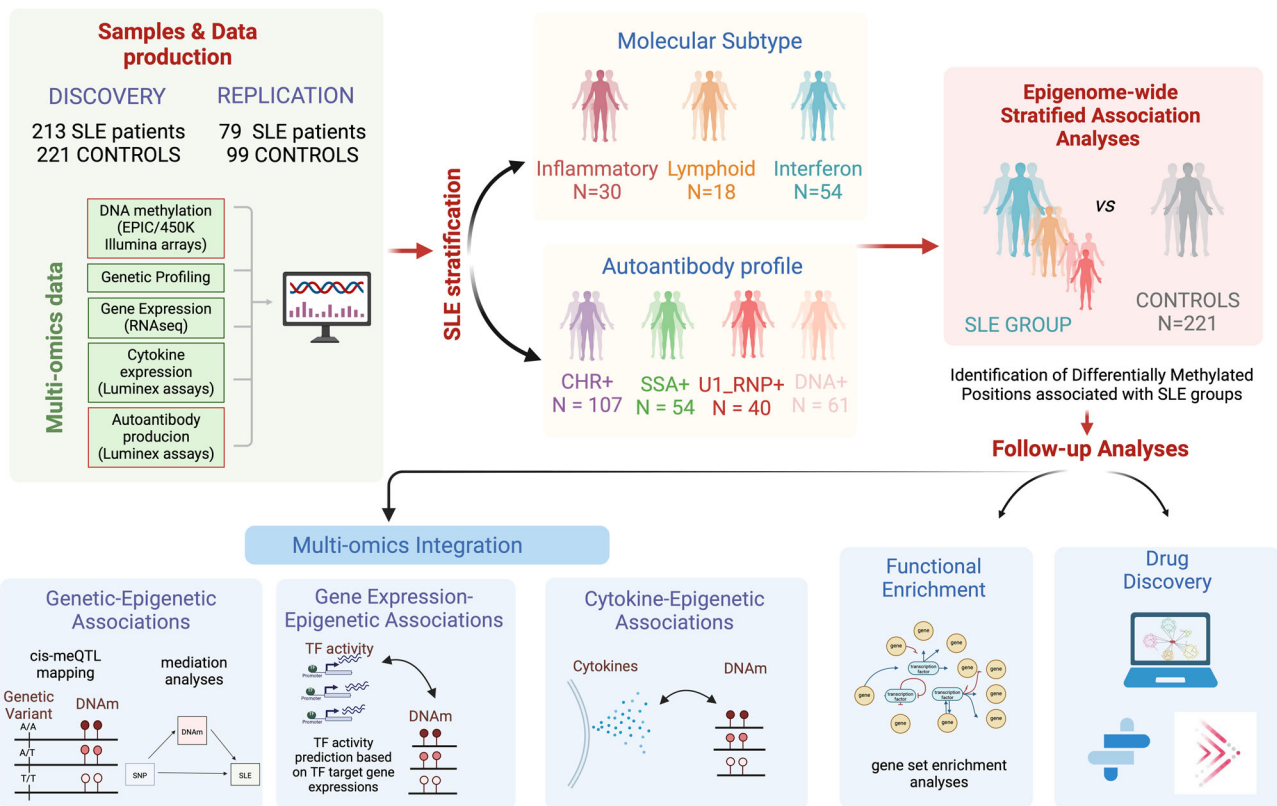
plastic functions responding to evolving environments, stimuli and differentiation processes<sup>14</sup>. Dysregulation at the epigenetic level has been identified in association with SLE and some SLE manifestations<sup>15-17</sup>. The genome of SLE patients is globally hypomethylated -unmethylated state of CpGs in a normally methylated sequence - and a large percentage of SLE patients exhibit the epigenIFNsig but, intriguingly, not all of them. Functional genomics deciphers the regulatory role of many disease-associated non-coding genetic variants<sup>18</sup>, but studies are scarce in SLE<sup>17</sup>. Approaches such as methylation quantitative trait loci (meQTL)<sup>18</sup> allows identifying genetic variants influencing disease by modifying DNA methylation (DNAm)<sup>17</sup>. However, large meQTL studies have been performed in non-disease populations despite the increasing recognition that regulatory genetic effects can be dependent on age, context, and pathological status<sup>9,18,19</sup>.

Despite the increasing number of molecular and epigenetic studies in SLE, previous work has not paid attention to disease heterogeneity nor to the possible role of genetic factors or their consequences at the transcriptional and cytokines levels.

The present work moves a step forward in the understanding of epigenetic landscapes in SLE, and their possible drivers and consequences, as well as in the identification of potential new SLE drug targets. We speculate that stratifying epigenome-wide association studies (EWAS) based on molecular subtypes and autoantibody specificities, and integrating different multi-omics layers, provides greater statistical power to find new and group-dependent epigenetic associations that might have a specific genetic regulation and context-specific epigenetic correlation with transcriptional factors and cytokine expression.

## Results

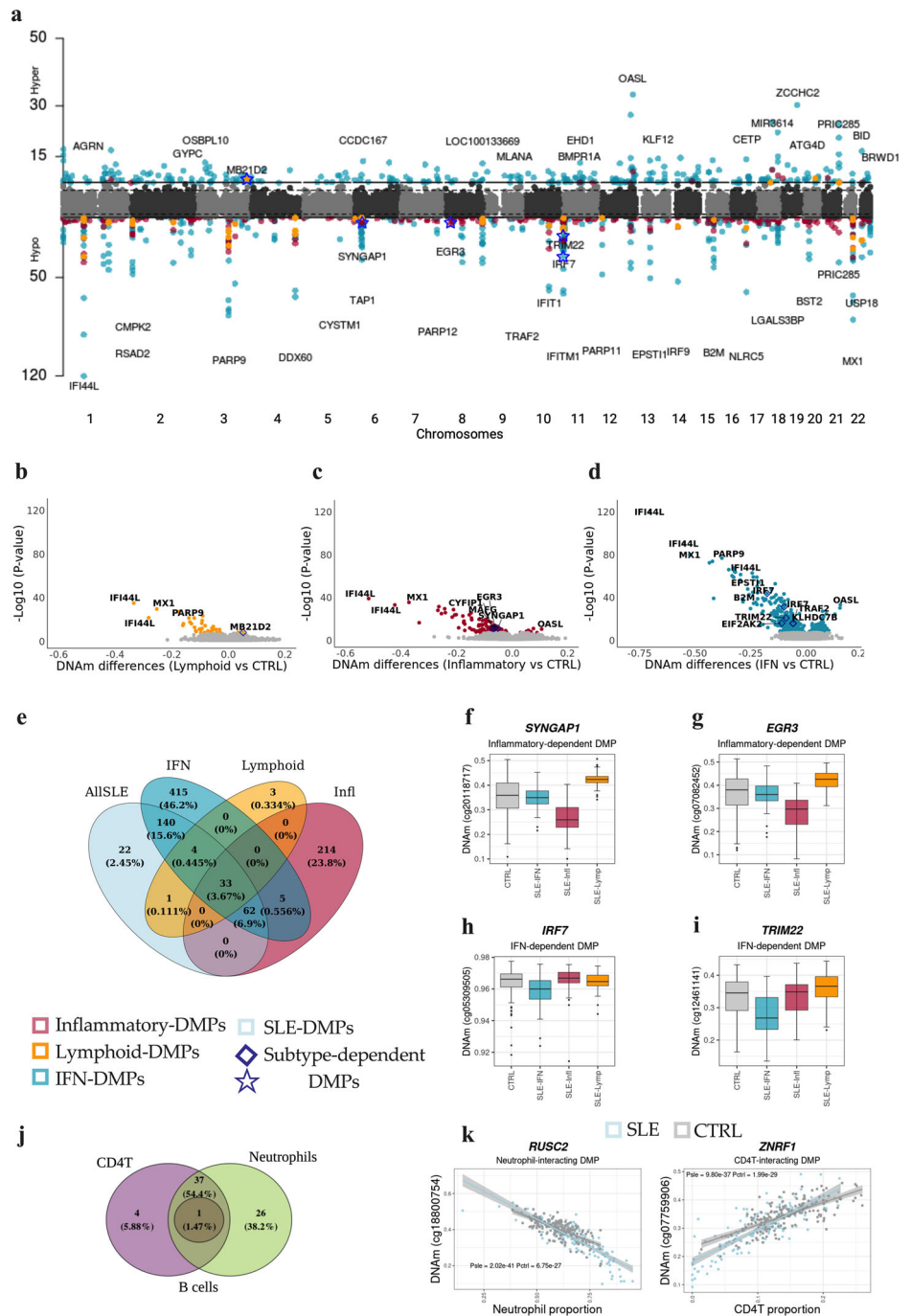
An overview of the study design is depicted and described in Fig. 1 and Supplementary Table 1.



**Fig. 1 | Overview of the study design.** Schematic outline illustrating the data analysis and the use of patients' genetic and molecular information to stratify patients into molecularly homogenous groups and autoantibody-positivity profiles to further

perform EWAS functional enrichment analyses, meQTL analyses, cytokine association, and drug discovery according to molecular and serological subtypes.

**Fig. 2 | Epigenetic signatures of SLE molecular subtypes and cell type interactions.** **a** Manhattan plot illustrating EWAS results for SLE and different molecular groups when compared with controls. X-axis represents the chromosomal locations of CpG sites and Y-axis represents the  $\log_{10}(P)$  obtained in linear regression models. **b–d** Volcano Plots representing EWAS results. The X-axis represents the DNAm differences between each pair of groups tested. **e** Overlap of genome-wide significant results for each EWAS was performed. **f–i** Examples of subtype dependent-DMPs. Colored dots represent significant DMPs after Bonferroni correction of different groups according to the legend. Diamonds and starts dots represents subtypes-dependent DMPs. **j** Venn diagram illustrating the overlap of cell-type-interacting DMPs that remain significant after Bonferroni correction. **k** Scatter plots depicting DNAm changes between SLE and CTRL individuals at a specific CpG site identified as a significant DMP in the cell-type-interacting analysis. Neutrophil proportions are represented on the y-axis in the left plot, while CD4T proportions are depicted on the x-axis in the right plot.



**Genome-wide DNAm patterns associated with different molecular SLE subtypes**

We conducted an EWAS with a sample size of 213 SLE cases and 221 controls. This analysis led to the identification of 262 SLE-DMPs (97% replication rate in an independent 450 K sample), as depicted in Fig. 2a and detailed in Supplementary Data. Of these DMPs, 64% displayed hypomethylation effects, while the remaining 36% exhibited hypermethylation effects. Not unexpectedly, the top SLE-DMPs implicated large reductions of DNAm at IRGs and genes enriched in IFN pathways (Supplementary Table 2), as for example *IFI44L*, *MX1*, *NLR5*, *IFITM1*, *IFIT1*, *IRF9* or *PARP9* but also genes involved in antigen processing and class I presentation (*TAP1* and *B2M*). Large hypermethylation effects were also observed in IRGs such as *OASL*. These results corroborate previous findings and is likely a consequence of the higher IFN levels observed in SLE patients<sup>20</sup>.

Stratifying SLE patients into the molecular subtypes ( $N_{IFN} = 54$ ,  $N_{Infl} = 30$ ,  $N_{Lymph} = 18$ ) allowed us to raise the number of associations despite reducing the sample size (Fig. 2a). We identified 314 inflammatory-DMPs (96.18% hypomethylation), 41 lymphoid-DMPs (90.24% hypomethylation) and 659 IFN-DMPs (54.48% hypomethylation), among which 214, 3, and 420, respectively, had not been previously detected when all SLE patients were pooled together (Fig. 2e). Next, we unveiled the outcomes of the enrichment analysis, emphasizing the concentration of identified genes within specific pathways. The IFN subtype showcased the most pronounced impact on genes associated with IFN pathways. Nevertheless, a persistent epigenetic IFN signature and analogous enrichment were also apparent in patients within the lymphoid and inflammatory subtypes (Fig. 2b–d and Supplementary Table 3). Additionally, synaptic functions exhibited enrichment in the set of genes with inflammatory differential DNA methylation.

We identified 71 inflammatory-dependent DMPs, 1 lymphoid-dependent DMPs and 232 IFN-dependent DMPs (Supplementary Fig. 1a–c and Supplementary Data). For example, we observed a DNAm decrease at *SYNGAP1* and *EGR3* genes in SLE inflammatory patients ( $P < 4.08 \times 10^{-04}$ ) that we did not observe in healthy controls or in patients from other subtypes ( $P < 0.05$ ; Fig. 2f, g). For the IFN subtype, we found large DNAm differences at *IRF7* and *TRIM22* genes, (Fig. 2h, i) both of which are related to type I IFN pathways<sup>21</sup>.

Through EWAS interaction models, we addressed the question of whether DNAm differences at DMPs are modified by specific blood cell types. The adjustment of EWAS models by blood composition guarantees that DMPs are not explained by cell proportions, however, we did find that DNAm differences between controls and SLE patients at some DMPs is modified in relation to cell proportions, and we called this as cell-interacting DMPs. We uncovered 145 cell-interacting DMPs ( $P_{int} < 6 \times 10^{-8}$ ), distributed as follows: 2 CD8T-interacting DMPs, 71 CD4T-interacting DMPs, 123 neutrophils-interacting DMPs, and 3 B cell-interacting DMPs. Notably, 68 of these DMPs were previously identified in our inflammatory stratified analyses, but none in the interferon or in the lymphoid groups (Supplementary Data). For example, DNAm at genes *RUSC2* and *ZNRF1* show greater DNAm differences between SLE and CTRLs in subjects with higher neutrophil and higher CD4T proportions respectively (Fig. 2k).

Intriguingly, we observed a contrast DNAm relationship with cell proportion in these DMPs, with 100% of those interacting CD4T displaying an opposite sign than those interacting with neutrophils. This could be explained by the tight negative correlation between neutrophils and CD4T cell proportions. Nevertheless, some DMPs were also found to be shared across different cell types (Fig. 2j), while 30 cell-type-interacting DMPs exhibited significance in one group but not in the other (Fig. 2j, k). We could also identify 77 novel DMPs showing cell-type-interacting DMPs. For example, DNAm at *FCGRT* gene for neutrophil, *ECE1* for B cell or *TMEM173* for CD4T cell. Due to low number of genes, functional enrichment analysis did not yield any pathway enriched in cell-interacting DMPs (Supplementary Data).

### The relationship of autoantibody profile on the SLE epigenome

We stratified patients based on their positivity for the most prevalent AAbs ( $N_{SSA+} = 54$ ,  $N_{CHR+} = 107$ ,  $N_{DNA} = 61$ ,  $N_{U1-RNP} = 40$ ) to perform the EWAS analysis. We identified 388 anti-SSA-DMPs, 223 anti-chromatin-DMPs, 256 anti-dsDNA-DMPs and 164 anti-U1-RNP-DMPs (Fig. 3a, Supplementary Data) when compared with CTRLs, yielding a total of 466 AAb-DMPs, from which 238 had not been previously detected as SLE-DMPs, and 81 were not detected as molecular subtype-DMPs. The epigenIFNsig, and in general every CpG effect, was stronger in SSA + SLE patients for which we identified 155 DMPs not observed in the other groups (Fig. 3b–f). However, the epigenetic signature was still persistent and dominant in SLE patients positive for other autoantibodies (Fig. 3b–e). A small proportion of 27 anti-chromatin-DMPs and 38 anti-dsDNA-DMPs were exclusively observed in these groups when establishing a genome-wide significance level. However, at a significant threshold of  $P < 0.05$ , we could only identify one example of an anti-SSA-dependent-DMPs in the *NLCR5* gene that reached genome-wide significance level in the SSA group and was not significant in the rest. (Fig. 3g). Functional enrichment analyses on AAb-DMPs identified IFN pathways as well as antigen processing and other pathways regulatory of immune responses (Supplementary Table 4). Furthermore, the examination considered negative antibody spectra. When comparing negative AAbs with DNAm changes in healthy controls, the observed alterations were limited. The analysis predominantly highlighted IFN signals without yielding any novel discoveries (Supplementary Fig. 2a, b). Our results indicate that epigenetic signature is highly shared across SLE patients with different autoantibody profiles, and that there is little specificity on SLE autoantibody-related epigenetic signals.

### Genetic drivers of SLE epigenetic signals and mediation role of DNAm on SLE risk

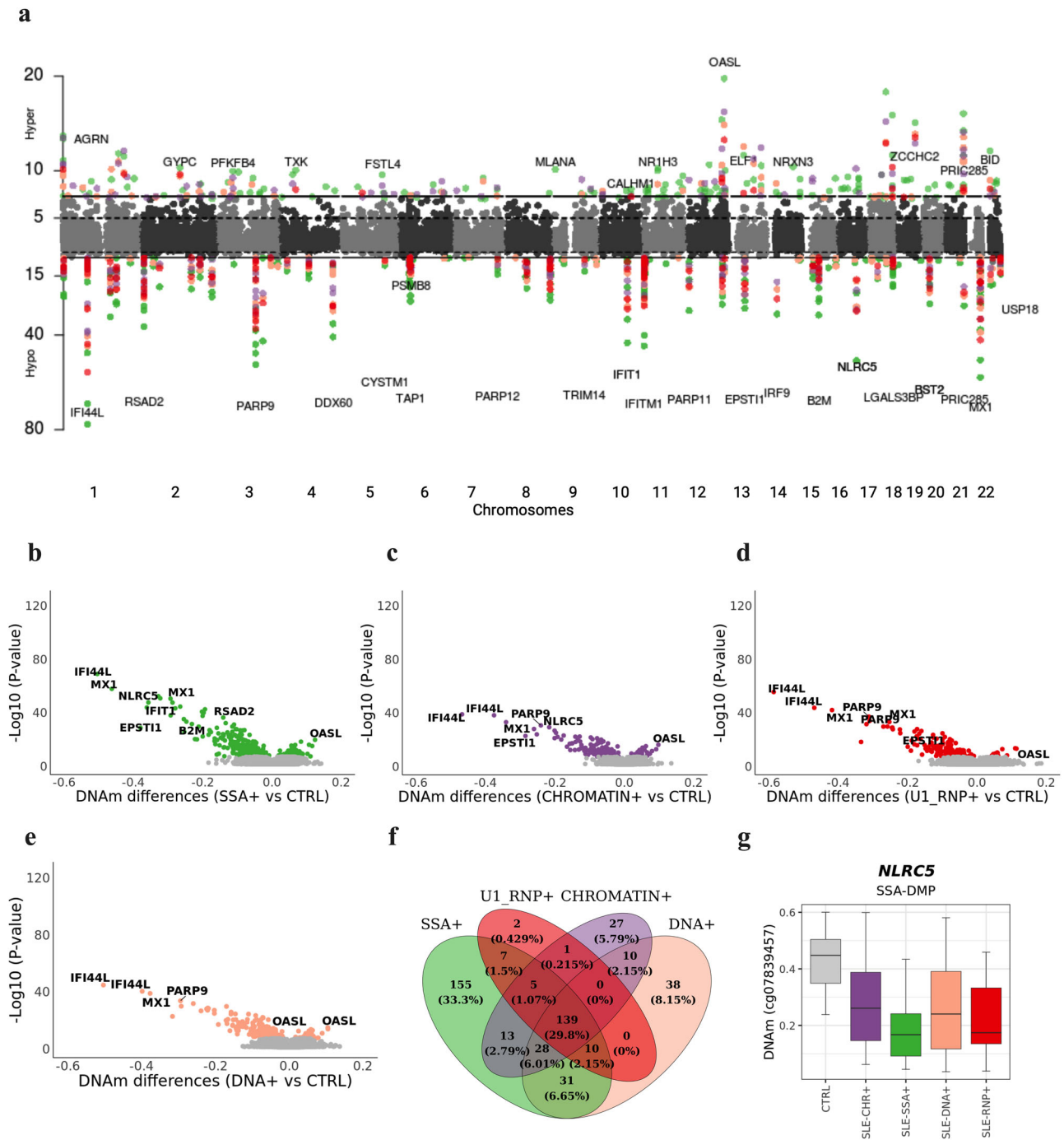
We searched for *cis* genetic variants associated with DNAm at 148 DMPs (*cis*-meQTLs-DMPs, FDR  $< 0.05$ ) (Fig. 4a). Up to 31 loci involved in *cis*-meQTL-DMPs also associate with SLE (SLE-*cis*-meQTLs-DMPs, Bonferroni significance  $P < 8 \times 10^{-05}$ ; Fig. 4a). Mediation analyses on SLE-*cis*-meQTLs-DMPs (Supplementary Data) revealed that SLE genetic risk is significantly reduced when DNAm is incorporated in the model (Fig. 4b), and that a significantly high proportion of SLE genetic risk at the HLA class I region (12–65% proportion mediated) is mediated by DNAm in *HLA-F*, *HLA-A*, *C6orf136*, *HLA-C*, *HLA-B*, *HCP5*, *TAP1* and *PSMB9* genes (Fig. 4c, d). Other SLE-associated SNPs such as those located in *STAT1* or the intergenic region at chromosome 5 nearby microRNA mir-146 also showed a significant indirect effect of SNP on SLE suggesting a mediation role of DNAm, but the proportion mediated did not reach statistical significance ( $P > 0.05$ ), probably due to the low sample size. Interestingly, for SLE-associated genes such as *IRF7* and *ITGAX*, the SLE genetic effect increased when DNAm was included in the model, suggesting a cofounded or more complex relationship between DNAm, genetic variation and SLE (Fig. 4d).

### Context-dependent meQTL regulatory function in SLE

We discovered *cis*-genetic variants associated with DNAm with dependency on disease status or molecular subtype by the identification of meQTL in different groups together with significant interaction effects. We identified SLE-dependent meQTLs for 394 DMPs among which significant disease-dependent effects were observed in *CDHR5* and *MAML2* (Fig. 4e, f). Likewise, we discovered inflammatory- and IFN-dependent meQTLs for 283 and 316 DMPs, respectively (Supplementary Data). For instance, we observed genetic regulation on DNAm in *RSAD2* gene in inflammatory SLE patients, but not in the IFN group or in controls where *RSAD2*-DNAm does not show a relationship with the genotype (Fig. 4g). IFN-dependent meQTL effects were observed for example for *HLA-F* (Fig. 4h). Intriguingly, we also identified meQTLs with strong opposite effects (op-meQTLs) between SLE and CTRLs (119 DMPs) or between the IFN and inflammatory subtypes (35 DMPs) among which *IFI44* gene shows the greatest opposite effect (Fig. 4i). Some genetic variants involved in context-dependent meQTLs were associated with SLE at a Bonferroni-corrected threshold ( $P < 1.2 \times 10^{-04}$ ) (Supplementary Data). The strongest SLE-associated SNPs involved SLE-dependent meQTL that associated with DNAm at *HLA-B* and *HLA-E* genes. Interestingly, we also identified strong genetic-disease interaction ( $P_{int} = 6.5 \times 10^{-09}$ ) in the regulation of DNAm within the *CDHR5* gene at chromosome 11, that involve SNPs strongly associated with SLE ( $P = 1.9 \times 10^{-07}$ ; Fig. 4e).

### SLE-associated epigenetic signals correlate with transcription factor activity and cytokine production

Transcription factor (TF) binding has an important role in shaping DNAm levels and vice versa<sup>22</sup>. Here, we identified up to 61 different TFs whose activity correlated with DNAm at several DMPs ( $P < 0.05$ ) (Supplementary Data). Interestingly, the large interaction between DNAm and TF activity (TFact) for *IRF9*, *IRF1*, *STAT2*, *STAT1*, *STAT3*, *TFDP1*, *FOXM1*, *E2F3*, *E2F2*, *GLI2*, and *RUNX3* was restricted to SLE patients (Fig. 5a, b). While, TFs of IRF and STAT family have a well described role in the activation of IFN pathways and are associated with DNAm at IRGs. The TFact of *TFDP1*, *E2F3*, and *FOXM1* correlated also with DNAm at IRGs such as *IFI44L*, *MX1*, *TRIM22*, and *ISG15*. Interestingly, *RUNX3* activity was associated with DNAm at genes such as *FCGR3B*, *HLX*, *LGALS12*, *BACH*, and *SYNGAP* which are not IFN-regulated. Intriguingly, we observed differential DNAm-TFact associations when comparing between inflammatory and IFN SLE patients. For example, NR2F2 TFact is strongly negatively associated with DNAm at the *LETM1* gene, but such effect is not observed in the IFN subtype (Fig. 5c). Likewise, ZEB1 TFact is associated with DNAm at *LGALS9*, *RSAD2*, *TMEM123*, *HECA*, and *IFI44L* genes only in IFN patients (Fig. 5d).

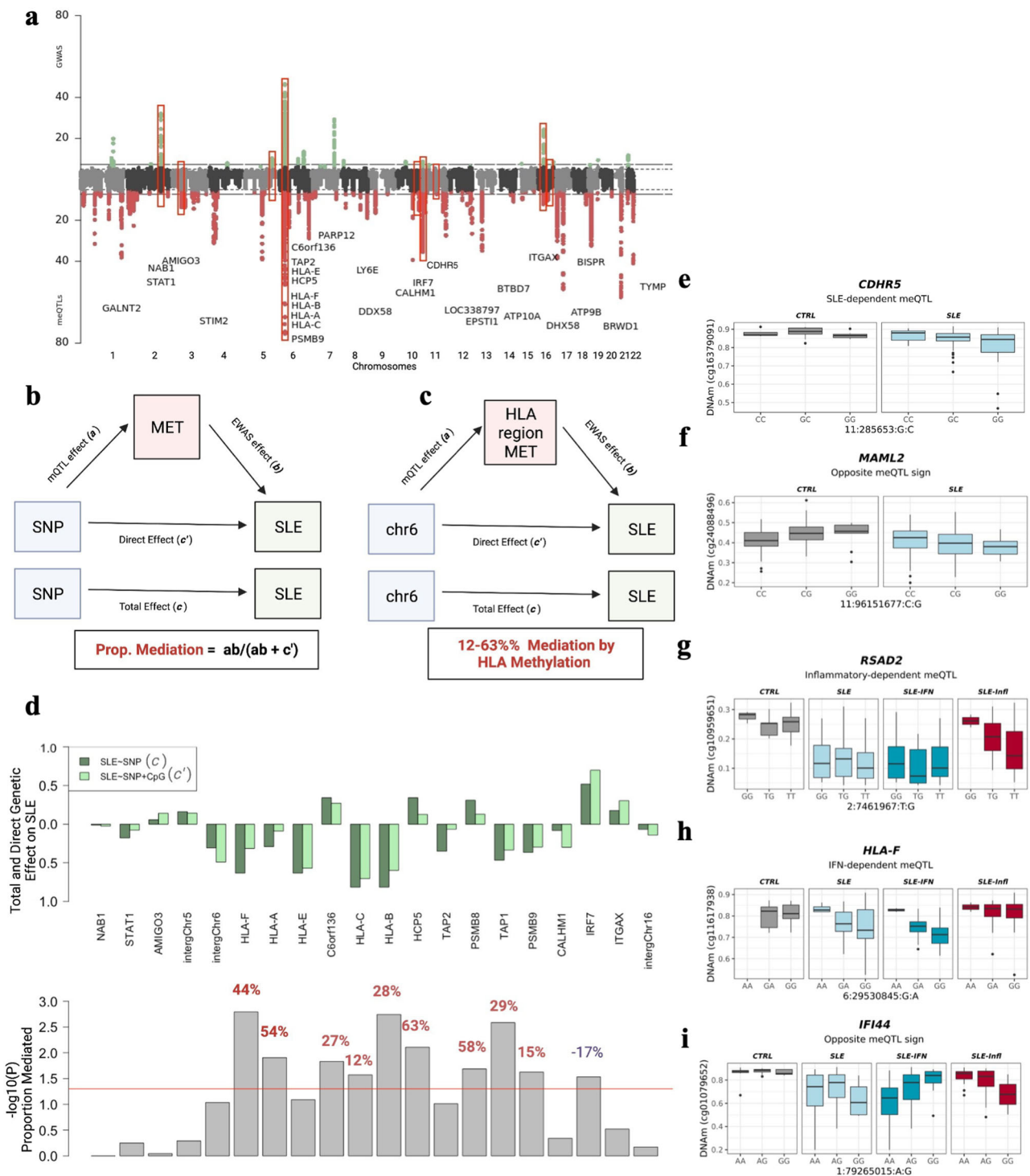


**Fig. 3 | Epigenetic signatures of SLE autoantibody profiles.** **a** Manhattan plot illustrating EWAS results for different groups of SLE patients according to their autoantibody profiles when compared with controls. X-axis represents the chromosomal locations of CpG sites and Y-axis represents the log<sub>10</sub> (*P*) obtained in linear regression models. **b–e** Volcano Plots representing EWAS results. The X-axis

represents the DNAm differences between each pair of groups tested. **f** Overlap of genome-wide significant results for each EWAS performed. **g** Example of DNAm distribution across different autoantibody groups, and healthy subjects. Colored dots represent significant DMPs after Bonferroni correction of different groups according to the legend.

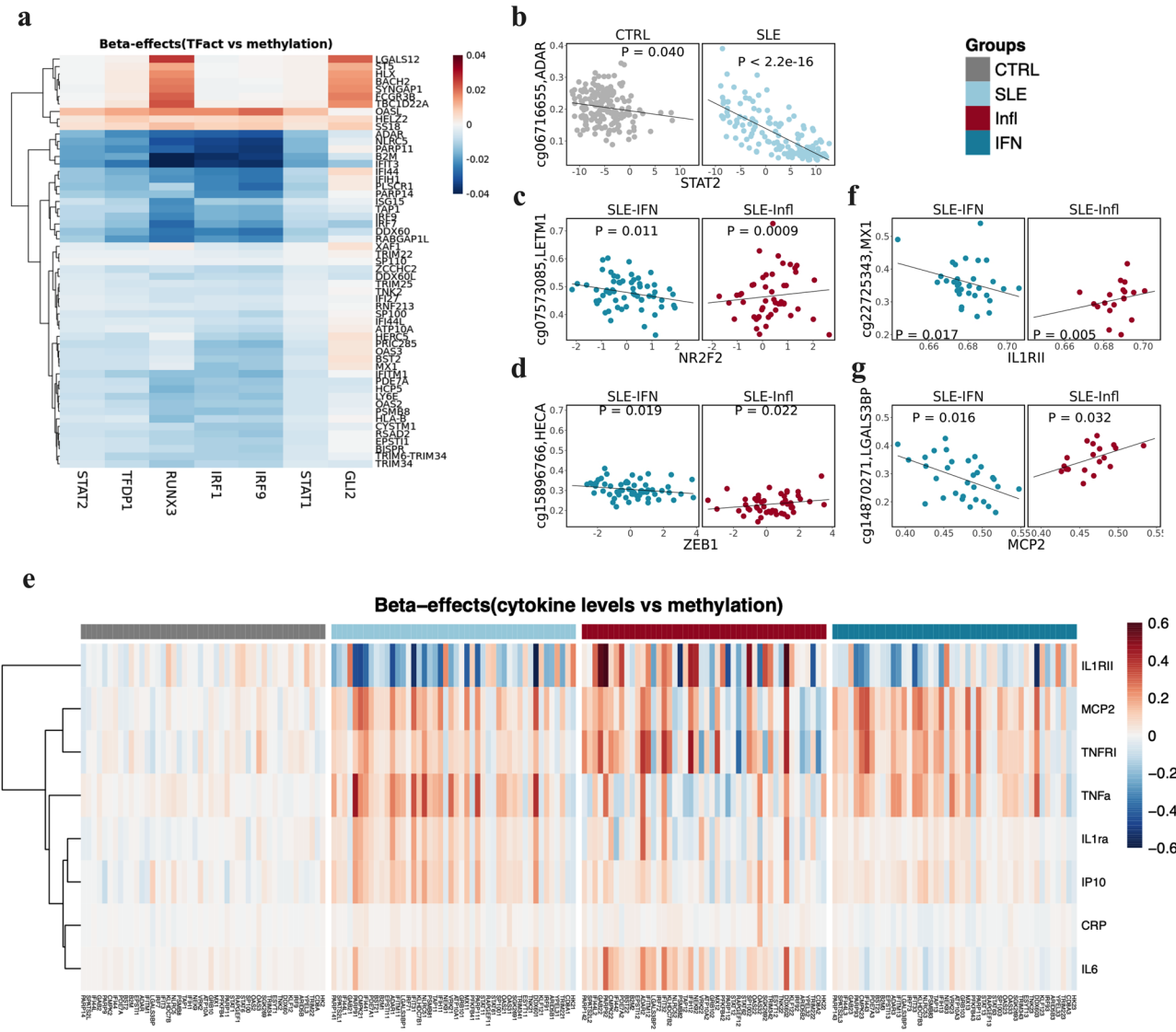
We explored whether DMPs could be associated with cytokine production in each molecular group. In total, 82 DMPs were significantly associated with levels of 8 cytokines ( $P < 5 \times 10^{-05}$ ). The strongest (and negative) associations were found between IL1RII levels and SLE-DMPs at a number of IRGs such as *PARP9* and *IFI44L*. We found SLE CpG-cytokine association for TNFa (*VRK2*, *PARP9*), IL1ra (*BST2*, *ATP10A*), IL1RII (*IFITM1*, *ARID5B*), MCP2 (*IFI44L*, *CPMK2*), and IP10 (*NLRC5*, *B2M*) which were not observed in CTRLs. We also observed inflammatory-dependent association for TNFR1 (*RAPGEF1*) and CRP (*OAS3*) (Supplementary Data). Figure 5e illustrates the strength of the

group-dependent associations. It can be observed that associations differed between groups and that for some CpGs, for cytokines such as IL1RII, these were stronger within the inflammatory as compared to the IFN subtype and stronger than CTRLs. For example, IL1RII was negatively associated with DNAm at *MX1* in IFN subtype but it is positively associated in the inflammatory subtype (Fig. 5f), similarly MCP2 showed opposite effects at the DNAm level for the *LGALS3BP* gene for the inflammatory and IFN subtypes (Fig. 5g). Altogether, our results indicate that the relationship between DNAm, TFact and cytokine production is determined in a subtype-dependent manner.



**Fig. 4 | Genetic drivers of SLE-epigenetic signatures. a** Top Manhattan plot (MP) shows GWAS results contrasting allele frequencies between a group of SLE patients ( $N = 4212$ ) and a group of CTRL ( $N = 4065$ ). Bottom MP illustrates meQTL results for DMPs in the whole sample. X-axis represents the chromosomal locations of CpG sites and the Y-axis represents the  $\log_{10}(P)$  obtained in a logistic regression model or meQTL analyses. Genetic associations above the red line marks the statistical association at a significant threshold of  $P < 1 \times 10^{-06}$  for logistic associations and  $FDR < 0.05$  for meQTLs. Red boxes show overlap in GWAS and meQTL results and represent meQTLs associated with SLE diagnosis. **b** Mediation model in which SLE genetic risk is exerted partly through DNAm changes. **c** Examples of SLE-associated SNPs in chr6 that are mediated by DNAm changes at DMPs in the HLA region. **d** Mediation results for the best SLE-meQTL-DMPs by gene. Upper barplot shows

the Total and the Direct Effect of SLE-associated genetic variants. Bottom barplot show the significance of the proportion mediated via DNA met resulted from mediation models. Percentage of mediation is illustrated in red below each bar only for those significant genes ( $P_{\text{proportion mediated}} < 0.05$ ). **e–i** Group-dependent meQTL effects. **e** A meQTL significant effect is observed among SLE patients ( $FDR < 0.05$ ) but not in the CTRL population ( $P > 0.05$ ). **f** A meQTL significant effect is observed in SLE patients and CTRLs but with different signs. **g** A meQTL significant effect is observed among SLE from inflammatory group ( $FDR < 0.05$ ) but not in the CTRL population or in the IFN group ( $P > 0.05$ ). **h** A meQTL significant effect is observed among SLE from IFN group ( $FDR < 0.05$ ) but not in the CTRL population or in the inflammatory group ( $P > 0.05$ ). **i** A meQTL with opposite direction effects between Inflammatory-IFN subtypes.



**Fig. 5 | Relationship between SLE-epigenetic signatures, transcription factor activity, and cytokine production.** **a** Heatmap representing SLE-dependent associations between DNAm at DMPs and TFact inferred from RNAseq data. **b** SLE dependent example showing the effect between TFact STAT2 and DNAm at ADAR gene vs CTRL. **c**, **d** Subtypes dependent examples showing the effect between TFact

ZEB1 and NR2F2 and DNAm at SPATS2L and LETM1 genes. **e** Heatmap showing the effect distribution of CpGs-genes associated to cytokine levels exhibiting group specificity. Color gradient from blue to red correspond to effect sizes. **f**, **g** Examples for cytokine opposite associations at DNAm levels for inflammatory and IFN subtypes.

**SLE-associated epigenetic signals inform drug discovery**

We observed an enrichment of the list of unique 549 SLE differentially methylated genes (DMG) among known drug targets (Fold enrichment of 1.4,  $P < 0.01$  in phase 1 or above in OpenTargets and Informa databases), and identified a total of 62 DMG being known drug targets, including 8 known SLE drug targets: SYK, JAK3, BCL2, PIK3CD, VDR, BTLA, FGR, GDP2, and NDUFS8 (Supplementary Table 5 and Fig. 6).

To identify and prioritize candidate novel SLE gene targets, we collected different features of each gene and ranked them based on two criteria: (i) previous evidence of genetic association to SLE (ii) “network proximity” (Fig. 6, see “Methods” section for scoring details). We showed strong enrichment of validated targets within the groups of genes with scores equal or higher than one (Supplementary Fig. 3). Using the different annotations available for each gene of the list, we could prioritize genes of interest and focus on potentially novel targets (see **discussion**) for SLE, and for different SLE subtypes. For example, we observed that 4 of the known drug targets (VDR, ALOX5AP, ITGA5, and ECE1) overlapped with genes that are differentially methylated only in the inflammatory SLE patients. Likewise,

17 genes (CLU, CETP, TSHR, ITGA2, ACACA, TNK2, STAT4, PSMB8, SCN8A, INPP5D, BMPRIA, TAPI, TYMP, QPCT, GPD2, PSMB9, and LAMB1) overlapped with interferon specific epigenetic signatures.

**Discussion**

In this study, we present the first integrative EWAS that contrasts genome-wide DNAm data in several stratifications of SLE patients and integrates results with genetic, transcription, clinical, and serological data.

Stratifying SLE patients across homogenous molecular subtypes allowed us to significantly increase statistical power expanding the epigenetic signals and reporting new loci associated with SLE that had not been revealed when analyzing SLE heterogeneous populations. The largest effects were mostly found at interferon-related genes as for example IFI44L, MX1, NLRC5, PARP9/DTX3L, which has been extensively identified in previous SLE-EWAS across most blood cell types and in other SADs<sup>7</sup>. Here we show that the epigenetic IFN signature is present across all SLE molecular subtypes at different intensities, however we could discover some epigenetic signals exclusively in those SLE patients from the IFN subtype, as DMPs at

Gene	CpG(chr:location)	Drug target		SLE drug target		SLE drug target PPI connection		SLE genetic association		Total Score	Score (SLE drug target PPI connection and SLE genetic association)	Group
		Informa	OpenTargets	Informa	OpenTargets	Informa	OpenTargets	DisGeNET	UKBB			
SYK	cg13782919(chr9:93588842)									3	1	SLE
JAK3	cg20959703(chr19:17959783)									3	1	SLE
BCL2	cg01863731(chr18:60810665)									3	1	SLE
PIK3CD	cg04287259(chr1:9730401)									3	1	SLE
VDR	cg06195428(chr12:48295148)									3	1	Infl
STAT4	cg15636519(chr2:191894418)									3	2	IFN
BTLA	cg19281794(chr3:112218761)									3	1	SLE
GPD2	cg02039987(chr2:157291262)									3	1	SLE
NDUFS8	cg12070987(chr11:67804055)									3	1	IFN
FGR	cg09370867(chr1:27961680)									2	1	SLE
ETS1	cg03702171(chr11:128333168)									2	2	SLE
CD8A	cg26730653(chr2:87022189)									2	1	SLE
FCGR3B	cg00974864(chr1:161601053)									2	1	SLE
STAT1	cg00676801(chr2:191876673)									2	1	SLE
CBLB	cg12614898(chr3:105581164)									2	1	SLE
INPP5D	cg01218053(chr2:233940417)									2	1	SLE
INSR	cg19961945(chr19:7192407)									2	1	SLE
RARA	cg16680942(chr17:38500161)									2	1	IFN
HIST1H2BL	cg19832381(chr6:2776524)									2	2	SLE
TXK	cg11132334(chr4:48136298)									2	1	Infl

**Fig. 6 | Representative SLE Epigenomic signature genes.** We list the top 20 (of the 549 genes from the identified 974 CpG sites) based on summary of gene total scores derived from individual criteria (filled box indicates criterion satisfied). Filled boxes

indicate an overlap with the data source described in each column. For full results, see Supplementary Data.

*IRF7* and *TRIM22* genes. *IRF7* has been established as an SLE genetic risk locus that alters IFN type I expression<sup>21</sup> and TRIM molecules have been studied as autoantigens in some autoimmune diseases, especially in Sjögren’s syndrome (SS) where *TRIM22* protein showed little or no immunoreactivity in a sub-population of SS patients<sup>23</sup> suggesting that *TRIM22* was acting as a potential SSA60 regulated gene. Importantly, we also discovered epigenetic signals that are specific of SLE patients with an inflammatory molecular profile, as those in the *EGR3* and *SYNGAP* genes. *EGR3* is a member of a zinc finger transcription factor that plays an important role in regulating immune responses inducing also the expression of anti-inflammatory cytokines such as IL-10 and TGFB1<sup>24</sup>, however the role of *EGR3* in autoimmunity it is not clear yet. *SYNGAP1* encodes a Ras GTPase activating protein that is member of the N-methyl-D-aspartate receptor complex and has been identified as a differentially methylated gene associated to gene expression in systemic sclerosis (SSc)<sup>25</sup>.

By incorporating interaction terms in the EWAS model, we could identify DMP which DNAm differences between SLE and controls is modified by specific cell type proportions, most of them implied neutrophil-interacting DMPs, and implied CpGs discovered in the SLE inflammatory cluster. Our findings are in agreement with previous research by Barturen et al.<sup>8</sup>, which demonstrated specific neutrophil epigenetic signature associated with the inflammatory molecular subtype. On the other hand, the fact that cell-interacting DMS were not that spread among IFN-related CpGs do suggest that these changes are shared in the same magnitude among different blood cell type, as it has been described before.

We also observed a noteworthy opposite effect in cell-interacting DMPs between neutrophils and CD4T cells. This contrasting methylation pattern aligns with the known regulatory role of neutrophils in restricting CD4 + T cell expansion during adaptive immune responses<sup>26</sup>. We also identified unique DMPs that are only observed in one particular cell type. For instance, we noted a CD4T-interacting DMP within the *ZNRF1* gene, encoding an E3 ubiquitin-protein ligase involved in neural-cell differentiation, with known associations to diabetic retinopathy<sup>27</sup>. Additionally, genes such as *RUSC2* and *RBL1*, contributing to the interaction with neutrophils, have previously been linked to lupus nephritis<sup>28,29</sup>. These cell-interacting DMPs results shed light on the role of cell type differential DNAm in immune-related disease processes and demonstrated that new SLE-associated epigenetic signals

can be revealed. Further analyses based on cell-type derived DNAm in larger sample sizes will expand these results.

Our stratification approach also included an exploration of the AAb relationship on the SLE epigenome. The DMPs with largest effects were found within SSA positive SLE patients. Our previous work showed that the epigenIFNsig in Sjogren’s syndrome is restricted to SSA positive patients and driven by HLA genetic variation<sup>9</sup>. However, in SLE many of the signals observed are shared across SLE patients with different AAb profiles, which suggests that in SLE, the epigenetic signature is not AAb-specific or AAb specificities are highly correlated.

One of the most important findings of this work is that SLE-associated genetic variation might exhibit its risk through DNAm changes, this is especially true for SLE-associated SNPs within the HLA region. Our findings strongly suggest that beyond the impact on antigen presentation, genetic risk at the HLA region is also mediated through the epigenetic and transcriptional alterations of many genes residing within the HLA such as *HLA-F*, *HLA-A*, *HLA-C*, *HCP5*, *PSMB8*, *TAP1*, and *PSMB9*. These results are in concordance with previous work showing an overexpression of these genes in SLE patients<sup>7</sup>. Building upon these findings another study found that genetic variation in the HLA region can influence the transcriptional regulation of genes in memory CD4 + T cells during activation<sup>30</sup>. Outside the HLA region, our mediation results also support an indirect effect of SLE-associated genetic variation via DNAm within *STAT1* and *microRNA146* gene, previously genetically associated with SLE. However, given the small effects of these SNPs on SLE, and the sample size analyzed here, confirmatory studies are needed to provide conclusive results, and larger sample sizes are needed to explore differential mediation effects between different SLE molecular subgroups. In the same line, analyzing public repositories of genetic associations by means of Mendelian Randomization approaches would help to resolve puzzling relationships as those observed for genetic variation, DNAm and SLE risk at the *IRF7* gene.

Our findings show that there exists a clear influence of disease and molecular status on the genetic architecture of DNAm, given the fact that DNAm at a large proportion of DMPs is regulated by meQTLs but exclusively in the SLE population context or in certain molecular states. For example, this is the case for *CDHR5* gene which DNAm is regulated by genetics only in the IFN and SLE subtypes, respectively. Interestingly, our results show that genetic variants involved in *CDHR5* meQTLs are also



associated with SLE risk. The SLE-associated SNPs represent novel disease variants as they have not been previously identified in GWAS and are likely to mediate their effects through DNAm changes. Previous studies have identified *CDHR5*-SNPs located close to *IRF7*<sup>31,32</sup> and associated with systemic sclerosis (SS)<sup>32</sup>. The candidate gene *CDHR5* is a member of the cadherin family which interacts with the  $\beta$ -catenin pathway<sup>33</sup>. These observations are in agreement with a growing body of evidence that highlights that genetic effects are largely context and time-specific, and implicates that future research and data collection of pathological and molecular status within longitudinal larger autoimmune populations will be able to decipher many more genetic variants with important regulatory functions in autoimmunity<sup>34</sup>.

Transcription factor activity has been linked to SLE recently<sup>24</sup>. In this work, we identified a group of TFs whose activity was associated with SLE-associated DNAm changes. The activity of STAT and IRFs regulators show the strongest association with DMPs at IRGs, and here we show that this relationship is specific of SLE population and not observed in healthy individuals. We also reveal other TFs not previously implicated in SLE, such as TFDPI, E2F3, and FOXM1 whose activity associates with DNAm at IRGs. On the contrary, we identified that *RUNX3* activity, a susceptibility gene in SLE and systemic sclerosis (SSc)<sup>35,36</sup>, correlates with DNAm at non-IFN-related genes. Our results suggest that TFact could play a relevant role in autoimmunity by altering epigenetic programs.

In SLE pathogenesis, an inflammatory cascade is mediated by altered cytokine production. Despite the growing recognition of the high potential of DNAm changes as surrogate and biomarkers of pro-inflammatory proteins, especially in ageing phenotypes<sup>37,38</sup>, no study, to the best of our knowledge, had reported epigenetic-cytokine correlations within autoimmune diseases. Here we show a strong association in the SLE population between inflammatory cytokines and methylation changes, not observed within healthy individuals, being *IL2RA* and *IP10* the cytokines showing the strongest association with the epigenIFNsig. Our results also show that many epigenetic-cytokine associations are only observed in certain molecular contexts.

Finally, we discovered several interesting potential new drug targets based on the SLE epigenetic profiles. Among the novel SLE candidate genes with higher score is *STAT4*, which is in direct PPI connections with SLE drug targets and exhibit strong genetic associations<sup>39-44</sup>. Another interesting target candidate is the transcription factor *ETS1*, which also had a significant strong SLE genetic association<sup>45</sup> and which is in direct PP interaction with SLE drug targets such as *JAK3*. This protein has been observed to suppress T follicular helper type 2 cell differentiation and halt SLE onset<sup>46,47</sup>. Of interest, *FGR*, a member of the Src family of protein tyrosine kinases (PTKs) was scoring high for being a drug target for non-SLE conditions, but in close PPI connections with SLE validated targets, making it a potential candidate for drug repositioning. It functions as a positive regulator of cell migration and regulates cytoskeleton reorganization via *RAC1* activation. It also phosphorylates *SYK* and promotes *SYK*-dependent activation of *AKT1* and MAP kinase signaling. *SYK* is one of the top genes in the candidate drug target gene list for SLE<sup>24,48</sup>, as *STAT4* is for the IFN subtype and *VDR* for the inflammatory.

This study has some limitations. First, our study focused on discerning molecular alterations within samples derived from whole blood. This composite includes a variety of immune cell types, many of which play direct roles in the autoimmune processes associated with SLE. The statistical methodology employed in this work allowed us to report epigenetic changes present with and without dependency on cell type abundance. However, it is crucial to note that our investigation did not unveil specific molecular changes exclusive to particular cell types and the identification of specific pathways influenced by these cell type-dependent epigenetic changes remains elusive due to the limited number of significant DMPs found in this analysis. Despite this limitation, exploring these potential pathways could be a compelling avenue for future research endeavors, particularly with larger sample sizes and DNAm data derived from sorted cells or single cell analyses. We were also unable to detect unique epigenetic alterations linked to

specific AAbs positivity. For most SLE patients, AAbs are not mutually exclusive, implying that individuals positive for one AAb are likely also exhibiting other AAbs. Future investigations focusing on AAbs combinations in autoimmune diseases will enhance our comprehension of the epigenetic variations associated with AAb profiles and SLE.

To wrap up, our study disentangles epigenetic signatures in SLE with regards to different heterogeneity aspects and identifies potential novel drug targets. By means of integrative multi-omics stratifying analyses we show that the strong association between epigenetics, genetics, TFact and cytokine production is highly dependent on disease and molecular context. Our results serve to motivate future epigenome-wide and genomic studies to molecularly stratified autoimmune patients in order to gain biological and statistical power and unravel novel genes and molecular mechanisms involved in disease pathogenesis.

## Methods

### Samples

We included 292 SLE patients and 320 healthy controls (CTRL). The sample was divided into a discovery set (213 SLE patients and 221 healthy controls with DNAm data based on the EPIC array), and a replication set (79 SLE patients and 99 with DNAm data based on the 450 K array) from the PRECISESADS project<sup>8</sup>. Supplementary Table 1 describes the main characteristics of the study sample, together with the groups and traits analyzed. Blood cell proportions were obtained using flow cytometry<sup>49</sup> while autoantibodies, cytokines and other inflammatory mediators were analyzed from serum. Autoantibodies (anti-chromatin, anti-dsDNA, anti-U1RNP, anti-SSA/Ro, anti-SSB/La, anti-SM, anti- $\beta$ 2 glycoprotein 1, anti- $\beta$ 2 microglobulin, IgG anti-cardiolipin, IgM anti-cardiolipin, rheumatoid factor, and anti-ENA), cytokines and inflammatory mediators (BAFF, BLC, CRP, FASL, GDF15, IL1RII, IL1RA, IL6, IP10, MCP2, MCP4, MIP1B, MMP2, MMP8, TARC, TGF $\beta$ 1, TNFRI, and TNF $\alpha$ ) were measured in serum samples as described by Barturen et al.<sup>8</sup>.

### Ethical declaration

An ethical protocol was prepared, achieving consensus among all partners, both academic and industrial, and was subsequently translated into the languages of all participants. It received approval from each local ethical committee at the clinical recruitment centers, as did all experimental protocols. For a comprehensive list of local committees and centers involved in PRECISESADS, please refer to Supplementary Note 1. All recruited patients were 18 years of age or older and provided signed informed consent forms. All methodologies were conducted in strict accordance with relevant guidelines and regulations, adhering to the standards set forth by the International Conference on Harmonization and Good Clinical Practice, as well as ethical principles stemming from the Declaration of Helsinki (2013). Confidentiality of records containing identifiable information was upheld in accordance with EU Directive 2001/20/EC and the applicable national and international data protection requirements in each participating country.

### DNAm profiling

DNAm data was produced using the Illumina Infinium HumanMethylationEPIC BeadChip array and the Illumina Infinium HumanMethylation450 BeadChip array, which covers up to 850,000 and 485,512 CpG sites respectively. DNA was extracted from peripheral blood samples from which the genome was amplified and hybridized to the Illumina arrays. Standard methodological procedures for quality control and probe filtering were performed as described previously<sup>50</sup>. 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. After QC, the raw

methylation beta values were background corrected and normalized using the functional normalization within the meffil R-package. A beta value ranging from 0 to 1 was used to measure DNAm in whole blood, being 1 the methylated status with 100% of cells being methylated at a given CpG and 0 the unmethylated status with 0% of cells being methylated.

### Genetic profiling

Genotyping was performed using InfiniumCore from Illumina. Imputation was performed using the Michigan Imputation Server<sup>51</sup> and Haplotype Reference Consortium as reference panel<sup>52</sup>. Filtering of genetic variants and quality controls were performed by PLINK<sup>53</sup>. A total of 4,553,097 variants with a minor allele frequency (MAF) higher than 0.05 were used for subsequent analyses.

### Epigenome-wide association analyses

We performed a series of EWAS interrogating DNAm differences between groups at each autosomic CpG site. We used linear regression models adjusting by sex, age and blood cell composition (B cells, CD4 T cells, CD8 T cells, monocytes, neutrophils and natural killer cells) as well as technical confounder effects (Sample\_Plate and Sample\_Position). SLE-associated differentially methylated CpG positions (SLE-DMPs) were identified comparing DNAm between SLE patients and CTRLs. We stratified SLE patients into the molecular subtypes described in Barturen et al.<sup>8</sup> and by autoantibody (AAb) profile and compared their DNAm separately with CTRL. In total, 30 SLE patients were classified into the inflammatory subtype, 18 into the lymphoid subtype and 54 to the IFN subtype in the discovery cohort (Fig. 1) and 11, 7, and 19 SLE patients in the replication cohort as previously defined by Barturen et al.<sup>8</sup>. We only performed EWAS on those AABs that exhibited positivity in more than 20 SLE patients and corrected models for the presence of all other autoantibodies. We applied a Bonferroni significance threshold ( $P < 5 \times 10^{-08}$ ) to claim genome-wide significance. We assessed replication ( $P < 0.05$ ) in an independent cohort from based on 450K-methylation data and determined replication rate for those CpGs present in both arrays. To determine group-dependent differentially methylated positions (DMPs) (ex. SLE-dependent DMPs), we used strict significant threshold for the specific DMP passing a Bonferroni threshold in one group while exhibiting a  $P > 0.001$  in the others. We separately identify cell-type-interacting DMPs through linear regressions models with an interaction term between cell proportion and disease or molecular status, while adjusting by the same covariates used in the EWAS model.

### Functional enrichment analyses

Gene-set enrichment function analyses were performed to reveal if genes annotated to DMP were enriched in particular functional pathways. We used the Reactome database<sup>54</sup> implemented by *enrichPathway* function in R library ReactomePA<sup>55</sup>. As background distribution, we used all genes interrogated in the EPIC array. We separately analyzed genes showing DMP with hypomethylation and hypermethylation events.

### Methylation quantitative trait loci analysis

We performed methylation quantitative trait locus (meQTL) mapping using the Matrix eQTL R package<sup>56</sup>, by means of linear regressions in which the minor allele dosage effect on DNAm was tested while adjusting for age, sex, batch effects, cells proportions, the first genetic principal component, and disease status. We defined *cis*-meQTLs as single nucleotide polymorphisms (SNPs) located no farther than 1 Mb from the interrogated DMP. *Cis*-meQTLs were separately discovered in different groups: i) whole sample, ii) SLE patients, iii) healthy controls, iv) inflammatory SLE patients, and vi) IFN SLE patients. Lymphoid subtype was discarded from this analysis due to the low sample size. We used a permutation-based False Discovery Rate (FDR)  $< 0.05$  to claim significance. We investigated the interaction effect between SNP and group for the following groups: SLE vs CTRL and inflammatory vs IFN. To determine group-dependent meQTLs

we used the following statistical conditions:  $FDR < 0.05$  in one group and  $P > 0.05$  in the other group and  $P < 0.05$  significant interaction effect. We also looked for meQTLs with an opposite and significant genetic effect in different groups, with the ensuing conditions i)  $FDR < 0.05$  in one group AND  $P < 0.05$  in another group, or ii)  $P_{interaction} < 0.05$  AND  $P < 0.05$  in both groups AND  $FDR_{interaction} < 0.005$ .

### Genetic association and mediation analyses

We ran a genome-wide association study (GWAS) in an European sample of 4212 SLE patients and 4065 controls previously described<sup>57</sup>, and extracted the results for those SNPs involved in meQTLs regulating 148 DMPs. We established a Bonferroni significance threshold of  $P < 0.05/148 = 0.00033$  to claim statistical significance. We performed mediation analyses with the R package mediation using the method development by Imai et al.<sup>58</sup> to quantify mediation effect of DNAm on SLE genetic risk. We report the Total Genetic Effect of the SNP on SLE (c), the Direct Genetic Effect of the SNP on SLE from a model adjusted by DNAm (c'), and estimated the proportion of the Total Genetic Effect on SLE explained by DNAm Mediation by the formula:  $Prop. Mediated = (a*b)/(a*b + c')$ , where *a* is the effect of the SNP on DNAm and *b* is the effect of DNAm in SLE. Test statistics for these measurements were estimated by 10,000 Monte Carlo simulations.

### Epigenetic associations with cytokine levels and transcription factor activities (TFact)

We inferred TFact for 119 transcription factors in the PRECISESADS data using whole blood RNAseq data and the R package DoRothEA<sup>59</sup> as previously described<sup>24</sup>. Linear regression models were performed between DNAm at DMP and log transformed cytokine levels for 18 different inflammatory cytokines (see supplementary Table 1) or with TFact in different groups (SLE, CTRL, inflammatory, and IFN) and corrected for the same covariates used for the EWAS/meQTLs. A Bonferroni threshold of  $P < 0.05/1198 = 4E-5$  was used to claim statistical significance.

### Identification of drug targets within epigenetic signals

We used different informatics platforms and data sources (Supplementary Methods) to identify drug targets within epigenetic signals. Specifically, for each gene in the list of the 549 differentially methylated genes, a "Total Score" was calculated as the sum of individual sub-scores, as described below (see also Fig. 6):

- (I) "known drug target" in OpenTargets OR Informa (clinical phase 1 or above), sub-score = 1;
- (II) "known drug target for SLE" in OpenTargets OR Informa (clinical phase 1 or above), sub-score = 1
- (III) having direct interactions with an SLE drug target (from OpenTargets OR Informa) in the PPI network, sub-score = 1
- (IV) reported in DisGeNet to have GDA score of association with SLE  $> 0.3$ , sub-score = 1
- (V) reported from UK BioBank with significant genetic variant associations ( $P < 5 \times 10^{-08}$ ) to SLE, sub-score = 1

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

ELIXIR Luxembourg (<https://elixir-luxembourg.org>) serves as the host for the data. Due to restrictions mandated by the ethics committee and/or informed consent agreements with participant, we are unable to deposit the data into a secure access-controlled repository. Upon request, the data is accessible, and the access procedure is outlined on the data landing page with the following <https://doi.org/10.17881/th9v-xt85>. Online availability of DNA methylation data can be found at <https://bioinfo.genyo.es/precisesadsdata/>.

## Code availability

All code utilized for data analysis and figure generation will be made accessible upon request. The computational tools are built within the R framework and detailed in the Methods section. Interested parties can contact the corresponding authors for access to the code. We used the latest R (R version 4.3.1) and associated package versions: *ReactomePA* (version 1.44.0), *MatrixEQTL* (version 2.3), *DoRothEA* (version 1.14.1) and *ggplot* (version 3.4.3) and *karyoploteR* (version 1.28.0) for figures.

Received: 3 April 2023; Accepted: 17 May 2024;

Published online: 16 July 2024

## References

- Rider, V. et al. Gender bias in human systemic lupus erythematosus: a problem of steroid receptor action? *Front. Immunol.* **9**, 611 (2018).
- Nehar-Belaid, D. et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020).
- Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
- Wang, Y.-F. et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* **12**, 772 (2021).
- Kariuki, S. N. et al. Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. *Arthritis Res. Ther.* **12**, R151 (2010).
- Chung, S. A. et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet.* **7**, e1001323 (2011).
- Teruel, M. & Sawalha, A. H. Epigenetic variability in systemic lupus erythematosus: what we learned from genome-wide DNA methylation studies. *Curr. Rheumatol. Rep.* **19**, 32 (2017).
- Barturen, G. et al. Integrative analysis reveals a molecular stratification of systemic autoimmune diseases. *Arthritis Rheumatol.* <https://doi.org/10.1002/art.41610> (2020).
- Teruel, M. et al. Integrative epigenomics in Sjögren's syndrome reveals novel pathways and a strong interaction between the HLA, autoantibodies and the interferon signature. *Sci. Rep.* **11**, 23292 (2021).
- Gateva, V. et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1228–1233 (2009).
- Guerra, S. G., Vyse, T. J. & Cunninghame Graham, D. S. The genetics of lupus: a functional perspective. *Arthritis Res. Ther.* **14**, 211 (2012).
- Teruel, M. & Alarcón-Riquelme, M. E. The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *J. Autoimmun.* **74**, 161–175 (2016).
- Ulf-Møller, C. J., Svendsen, A. J., Viemose, L. N. & Jacobsen, S. Concordance of autoimmune disease in a nationwide Danish systemic lupus erythematosus twin cohort. *Semin. Arthritis Rheum.* **47**, 538–544 (2018).
- Dupont, C., Armant, D. R. & Brenner, C. A. Epigenetics: definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* **27**, 351–357 (2009).
- Joseph, S. et al. Epigenome-wide association study of peripheral blood mononuclear cells in systemic lupus erythematosus: Identifying DNA methylation signatures associated with interferon-related genes based on ethnicity and SLEDAI. *J. Autoimmun.* **96**, 147–157 (2019).
- Rask-Andersen, M. et al. Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum. Mol. Genet.* **25**, 4739–4748 (2016).
- Imgenberg-Kreuz, J. et al. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann. Rheum. Dis.* **77**, 736–743 (2018).
- Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**, 1311–1321 (2021).
- Bonder, M. J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49**, 131–138 (2017).
- Rönnblom, L. & Leonard, D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. *Lupus Sci. Med.* **6**, e000270 (2019).
- Liu, W., Li, M., Wang, Z. & Wang, J. IFN- $\gamma$  Mediates the development of systemic lupus erythematosus. *Biomed. Res. Int.* **2020**, 7176515 (2020).
- Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).
- Jones, E. L., Laidlaw, S. M. & Dustin, L. B. TRIM21/Ro52 - roles in innate immunity and autoimmune disease. *Front. Immunol.* **12**, 738473 (2021).
- Lopez-Dominguez, R. et al. Transcription factor activity inference in systemic lupus erythematosus. *Life* **11**, 299 (2021).
- 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.* **73**, 1288–1300 (2021).
- Castell, S. D., Harman, M. F., Morón, G., Maletto, B. A. & Pistoresi-Palencia, M. C. Neutrophils which migrate to lymph nodes modulate CD4 + T cell response by a PD-L1 dependent mechanism. *Front. Immunol.* **10**, 105 (2019).
- Peng, D. et al. Common variants in or near ZNRF1, COLEC12, SCYL1BP1 and API5 are associated with diabetic retinopathy in Chinese patients with type 2 diabetes. *Diabetologia* **58**, 1231–1238 (2015).
- Bethunaickan, R., Berthier, C. C., Zhang, W., Kretzler, M. & Davidson, A. Comparative transcriptional profiling of 3 murine models of SLE nephritis reveals both unique and shared regulatory networks. *PLoS ONE* **8**, e77489 (2013).
- Osthoff, A.-K. R. et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. *Ann. Rheum. Dis.* **77**, 1251–1260 (2018).
- Gutierrez-Arcelus, M. et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat. Genet.* **52**, 247–253 (2020).
- Carmona, F. D. et al. Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. *Ann. Rheum. Dis.* **71**, 114–119 (2012).
- European Scleroderma Group†. et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.* **10**, 4955 (2019).
- Rinne, S. J. et al. Candidate predisposition variants in kaposi sarcoma as detected by whole-genome sequencing. *Open Forum Infect. Dis.* **6**, ofz337 (2019).
- Gutierrez-Arcelus, M., Rich, S. S. & Raychaudhuri, S. Autoimmune diseases — connecting risk alleles with molecular traits of the immune system. *Nat. Rev. Genet.* **17**, 160–174 (2016).
- Wang, W. et al. RUNX3 gene polymorphisms are associated with clinical features of systemic lupus erythematosus in Chinese Han population. *J. Dermatol. Sci.* **80**, 69–71 (2015).
- O'Reilly, S. Epigenetic regulation of RUNX3 in systemic sclerosis pathogenesis: time to target? *Ann. Rheum. Dis.* **78**, 1149–1150 (2019).
- Steegenga, W. T. et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr.)* **36**, 9648 (2014).
- Verschoor, C. P. et al. The relation between DNA methylation patterns and serum cytokine levels in community-dwelling adults: a preliminary study. *BMC Genet.* **18**, 57 (2017).

39. Hagberg, N. et al. The STAT4 SLE risk allele rs7574865[T] is associated with increased IL-12-induced IFN- $\gamma$  production in T cells from patients with SLE. *Ann. Rheum. Dis.* **77**, 1070–1077 (2018).
40. Liang, Y., Pan, H.-F. & Ye, D.-Q. Therapeutic potential of STAT4 in autoimmunity. *Expert Opin. Ther. Targets* **18**, 945–960 (2014).
41. Menke, J. et al. Targeting transcription factor Stat4 uncovers a role for interleukin-18 in the pathogenesis of severe lupus nephritis in mice. *Kidney Int* **79**, 452–463 (2011).
42. Remmers, E. F. et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
43. Sigurdsson, S. et al. A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. *Hum. Mol. Genet.* **17**, 2868–2876 (2008).
44. Taylor, K. E. et al. Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS Genet.* **4**, e1000084 (2008).
45. Yang, W. et al. Genome-wide association study in asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.* **6**, e1000841 (2010).
46. Garrett-Sinha, L. A., Kearly, A. & Satterthwaite, A. B. The role of the transcription factor Ets1 in lupus and other autoimmune diseases. *Crit. Rev. Immunol.* **36**, 485–510 (2016).
47. Kim, C. J. et al. The transcription factor Ets1 suppresses T follicular helper type 2 cell differentiation to halt the onset of systemic lupus erythematosus. *Immunity* **49**, 1034–1048.e8 (2018).
48. Shao, W.-H. & Cohen, P. L. The role of tyrosine kinases in systemic lupus erythematosus and their potential as therapeutic targets. *Expert Rev. Clin. Immunol.* **10**, 573–582 (2014).
49. L, L. L. et al. Standardization procedure for flow cytometry data harmonization in prospective multicenter studies. *Sci. Rep.* **10**, 11567–11567 (2020).
50. Teruel, M. et al. An integrative multi-omics approach in Sjögren's Syndrome identifies novel genetic drivers with regulatory function and disease-specificity. *medRxiv* <https://doi.org/10.1101/2020.09.14.20192211> (2020).
51. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
52. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
53. 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).
54. Griss, J. et al. ReactomeGSA - efficient multi-omics comparative pathway analysis. *Mol. Cell Proteom.* **19**, 2115–2124 (2020).
55. Ulgen, E., Ozisik, O. & Sezerman, O. U. pathfindR: an R package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front. Genet.* **10**, 858 (2019).
56. Shabalín, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
57. Martínez-Bueno, M. & Alarcón-Riquelme, M. E. Exploring impact of rare variation in systemic lupus erythematosus by a genome wide imputation approach. *Front. Immunol.* **10**, 258 (2019).
58. Imai, K., Keele, L. & Yamamoto, T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* **25**, 51–71 (2010).
59. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

## Acknowledgements

E.C.M and O.C.P have been funded by Grants for the financing of Research, Development, and Innovation (R + D + I) in Biomedicine and Health

Sciences in Andalusia for the year 2021, project co-financed by 80% by funds of the FEDER Operational Program of Andalusia 2014-2020 (File Number: PECOVID-0072-2020). Funding for this work has been provided by the Plan Estatal, Ministerio de Ciencia e Innovación a MEAR No. PID2020-113776GB-I00, as well as to the Lupus Research Alliance/BMS Award to M.E.A.R. and the Diversity Supplement to O.C.P. Funding for the preparation of this manuscript has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no 115,565, resources composed of the financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and the EFPIA companies' in kind contribution and from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 831434 (3TR). The JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. The contributions of the consortia to this study are gratefully acknowledged.

## Author contributions

E.C.-M., M.E.A.-R., E.dR., and O.C.P. 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. O.C.P., E.C.-M., G.B., M.K., R.L.T., D.R., A.I., E.dR., W.Q., C.Z., S.S., A.S., and M.M.-B. contributed to pre-processing and/or analyzing the data. M.E.A.-R., J.M., C.M., E.dR., 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.

## Funding

Open access funding provided by Karolinska Institute.

## Competing interests

W.Q., C.Z., S.S., and E.dR. are employees of SANOFI. All other authors have no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41525-024-00420-0>.

**Correspondence** and requests for materials should be addressed to Elena Camero-Montoro or Marta E. Alarcón-Riquelme.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

**PRECISEADS Clinical Consortium**

Lorenzo Beretta<sup>13</sup>, Barbara Vigone<sup>13</sup>, Jacques-Olivier Pers<sup>14</sup>, Alain Saraux<sup>14</sup>, Valérie Devauchelle-Pensec<sup>14</sup>, Divi Cornec<sup>14</sup>, Sandrine Jousse-Joulin<sup>14</sup>, Bernard Lauwerys<sup>15</sup>, Julie Ducreux<sup>15</sup>, Anne-Lise Maudoux<sup>15</sup>, Carlos Vasconcelos<sup>16</sup>, Ana Tavares<sup>16</sup>, Esmeralda Neves<sup>16</sup>, Raquel Faria<sup>16</sup>, Mariana Brandão<sup>16</sup>, Ana Campar<sup>16</sup>, António Marinho<sup>16</sup>, Fátima Farinha<sup>16</sup>, Isabel Almeida<sup>16</sup>, Miguel Angel Gonzalez-Gay Mantecón<sup>17</sup>, Ricardo Blanco Alonso<sup>17</sup>, Alfonso Corrales Martínez<sup>17</sup>, Ricard Cervera<sup>18</sup>, Ignasi Rodríguez-Pintó<sup>18</sup>, Gerard Espinosa<sup>18</sup>, Rik Lories<sup>19</sup>, Ellen De Langhe<sup>19</sup>, Nicolas Hunzelmann<sup>20</sup>, Doreen Belz<sup>20</sup>, Torsten Witte<sup>21</sup>, Niklas Baerlecken<sup>21</sup>, Georg Stummvoll<sup>22</sup>, Michael Zauner<sup>22</sup>, Michaela Lehner<sup>22</sup>, Eduardo Collantes<sup>23</sup>, Rafaela Ortega Castro<sup>23</sup>, Ma Angeles Aguirre-Zamorano<sup>23</sup>, Alejandro Escudero-Contreras<sup>23</sup>, Ma Carmen Castro-Villegas<sup>24</sup>, Norberto Ortego<sup>24</sup>, María Concepción Fernández Roldán<sup>24</sup>, Enrique Raya<sup>25</sup>, Inmaculada Jiménez Moleón<sup>25</sup>, Enrique de Ramon<sup>26</sup>, Isabel Díaz Quintero<sup>26</sup>, Pier Luigi Meroni<sup>27</sup>, Maria Gerosa<sup>27</sup>, Tommaso Schioppo<sup>27</sup>, Carolina Artusi<sup>27</sup>, Carlo Chizzolini<sup>28</sup>, Aleksandra Zuber<sup>28</sup>, Donatienne Wynar<sup>28</sup>, Laszlo Kovács<sup>29</sup>, Attila Balog<sup>29</sup>, Magdolna Deák<sup>29</sup>, Márta Bocskai<sup>29</sup>, Sonja Dulic<sup>29</sup>, Gabriella Kádár<sup>29</sup>, Falk Hiepe<sup>30</sup>, Velia Gerl<sup>30</sup>, Silvia Thiel<sup>30</sup>, Manuel Rodriguez Maresca<sup>31</sup>, Antonio López-Berrio<sup>31</sup>, Rocío Aguilar-Quesada<sup>31</sup> & Héctor Navarro-Linares<sup>31</sup>

<sup>13</sup>Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milano, Italy. <sup>14</sup>Centre Hospitalier Universitaire de Brest, Hospital de la Cavale Blanche, Brest, France. <sup>15</sup>Pôle de pathologies rhumatismales systémiques et inflammatoires, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. <sup>16</sup>Centro Hospitalar do Porto, Porto, Portugal. <sup>17</sup>Servicio Cantabro de Salud, Hospital Universitario Marqués de Valdecilla, S, Santander, Spain. <sup>18</sup>Hospital Clinic I Provincia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. <sup>19</sup>Katholieke Universiteit Leuven, Leuven, Belgium. <sup>20</sup>Klinikum der Universitaet zu Koeln, Cologne, Germany. <sup>21</sup>Medizinische Hochschule Hannover, Hannover, Germany. <sup>22</sup>Medical University Vienna, Vienna, Austria. <sup>23</sup>Servicio Andaluz de Salud, Hospital Universitario Reina Sofía Córdoba, Córdoba, Spain. <sup>24</sup>Servicio Andaluz de Salud, Complejo hospitalario Universitario de Granada (Hospital Universitario San Cecilio), Granada, Spain. <sup>25</sup>Servicio Andaluz de Salud, Complejo hospitalario Universitario de Granada (Hospital Virgen de las Nieves), Granada, Spain. <sup>26</sup>Servicio Andaluz de Salud, Hospital Regional Universitario de Málaga, Málaga, Spain. <sup>27</sup>Università degli studi di Milano, Milan, Italy. <sup>28</sup>Hospitiaux Universitaires de Genève, Genève, Switzerland. <sup>29</sup>University of Szeged, Szeged, Hungary. <sup>30</sup>Charité, Berlin, Germany. <sup>31</sup>Andalusian Public Health System Biobank, Granada, Spain.

**PRECISEADS Flow Cytometry Study Group**

Montserrat Alvarez<sup>32</sup>, Damiana Alvarez-Errico<sup>33</sup>, Nancy Azevedo<sup>34</sup>, Nuria Barbarroja<sup>35,36</sup>, Anne Buttgerit<sup>37</sup>, Qingyu Cheng<sup>38</sup>, Carlo Chizzolini<sup>32</sup>, Jonathan Cremer<sup>39</sup>, Aurélie De Groof<sup>40</sup>, Ellen De Langhe<sup>41</sup>, Julie Ducreux<sup>40</sup>, Aleksandra Dufour<sup>32</sup>, Velia Gerl<sup>38</sup>, Maria Hernandez-Fuentes<sup>42</sup>, Laleh Khodadadi<sup>38</sup>, Katja Kniesch<sup>43</sup>, Tianlu Li<sup>38</sup>, Chary Lopez-Pedraza<sup>40</sup>, Zuzanna Makowska<sup>37</sup>, Concepción Marañón<sup>1</sup>, Brian Muchmore<sup>1</sup>, Esmeralda Neves<sup>34</sup>, Bénédicte Rouvière<sup>44</sup>, Quentin Simon<sup>44</sup>, Elena Trombetta<sup>45</sup>, Nieves Varela<sup>43</sup> & Torsten Witte<sup>43</sup>

<sup>32</sup>Immunology and Allergy, University Hospital and School of Medicine, Geneva, Switzerland. <sup>33</sup>Chromatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. <sup>34</sup>Serviço de Imunologia EX-CICAP, Centro Hospitalar e Universitário do Porto, Porto, Portugal. <sup>35</sup>IMIBIC, Reina Sofia Hospital, University of Cordoba, Córdoba, Spain. <sup>36</sup>Bayer AG, Berlin, Germany. <sup>37</sup>Pharmaceuticals Division, Bayer Pharma, Berlin, Germany. <sup>38</sup>Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany. <sup>39</sup>Department of Microbiology and Immunology, Laboratory of Clinical Immunology, KU Leuven, Leuven, Belgium. <sup>40</sup>Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, Brussels, Belgium. <sup>41</sup>University Hospitals Leuven and Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium. <sup>42</sup>UCB, Slough, UK. <sup>43</sup>Klinik für Immunologie Und Rheumatologie, Medical University Hannover, Hannover, Germany. <sup>44</sup>NSERM, UMR1227, CHRU Morvan, Lymphocytes B et Auto-immunité, University of Brest, BP 824, Brest, France. <sup>45</sup>Laboratorio di Analisi Chimico Cliniche e Microbiologia - Servizio di Citofluorimetria, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy.