

# The Lupus Epigenome Relates to Genetics, Transcription and Serological Profiles with Dependency on Molecular Subtypes and Informs Drug Discovery

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

\* Authors share contribution

§ Author correspondence: Elena Carnero-Montoro & Marta E. Alarcón-Riquelme

## Affiliations

1. GENYO. Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, 18016 Granada, Spain
2. Usher Institute of Population Health Sciences and Informatics. University of Edinburgh Medical School, EH8 9YL, Edinburgh, UK.
3. IBPLN-CSIC, Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Granada 18016, Spain
4. Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), 08916, Badalona, Barcelona, Spain.
5. Università degli Studi di Milano and Istituto Auxologico Italiano, Milan, Italy
6. Sanofi, Early Development and Research, Cambridge, MA, USA
7. Sanofi, Precision Medicine & Computational Biology (PMCB), R&D, Cambridge, MA, USA
8. Institute for Environmental Medicine, Karolinska Institutet, 171 67 Solna, Sweden.

## ABSTRACT

**Objective.** 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.

**Methods.** 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.

**Results.** 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 potential drug targets for SLE.

**Conclusion.** This study expands the number of genes associated with SLE and reveals novel pathways of disease. The findings reveal possible genetic drivers and consequences of epigenetic variability on SLE heterogeneity and disentangles the DNAm mediation role on SLE genetic risk and the genetic architecture of DNAm in different molecular contexts. Finally, novel targets for drug development were discovered.

## INTRODUCTION

Systemic lupus erythematosus (SLE) is an 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 (1). Treating and diagnosing SLE is challenged by the patients' heterogeneity in terms of diversity of symptoms, manifestations (2), 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 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 (3,4). A strong *HLA* genetic association with the presence of autoantibodies such as anti-SSB, anti-SSA, anti-RNP, anti-SM (5) and anti-dsDNA production has been recognized since long (5,6). 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(7) as well as in other SADS(8,9) . A strong interaction between HLA genetic variation, the production of anti-SSA AAb and the epigenetic IFN signature has been reported in primary Sjogren's syndrome (9), however the genetic and autoantibody determinants of the IFN signature in SLE are still not clear. Recently, the study by *Barturen et al.* (8) 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 taking into account to overcome SLE heterogeneity.

Heritability studies showed that genetics explains only a small fraction, less than 10% of SLE susceptibility (10–12), suggesting the important contribution of environmental or non-genetic factors (13) 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 (14). Dysregulation at the epigenetic level has been identified in association with SLE and some SLE manifestations (15–17). 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 epigenetic signature but, intriguingly, not all of them. Functional genomics deciphers the regulatory role of many disease-associated non-coding genetic variants (18), but studies are scarce in SLE (17). Approaches such as methylation quantitative trait loci (meQTL) (18) allows identifying genetic variants influencing disease by modifying DNA methylation (DNAm) (17). 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 (9,18,19)

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 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.

## METHODS

An overview of the study design is depicted in **Figure 1**.

**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 450K array) from the PRECISESADS project (8). **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 (20) 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.* (8).

**DNA methylation (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 (21). 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(22) and Haplotype Reference Consortium as reference panel(23). Filtering of genetic variants and quality controls were performed by PLINK (24). 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 (EWAS).** 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.* (8) and by autoantibody (AABs) 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 (**Figure 1**) and 11, 7 and 19 SLE patients in the replication cohort as previously defined by Barturen et al. (8). 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 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.

**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 (25) implemented by *enrichPathway function* in R library ReactomePA (26). As background distribution we used all genes interrogated in the EPIC array. We separately analyzed 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 (27), 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 SNPs located no farther than 1Mb 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 effects in different groups, with the ensuing conditions i) FDR  $< 0.05$  in one group AND  $P < 0.05$  in another group, or ii)  $P_{\text{interaction}} < 0.05$  AND  $P < 0.05$  in both groups AND  $\text{FDR}_{\text{interaction}} < 0.005$ .

**Genetic association and mediation analyses.** We ran a genome-wide association study (GWAS) in an European sample of 4,212 SLE patients and 4,065 controls previously described (28), 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.* (29) 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:  $\text{Prop. Mediated} = (a*b)/(a*b + c')$ , 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 (30) as previously described (31). 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  $P < 0.05/1,198 = 4E-5$  was used to claim statistical significance.

**Identification of drug targets within epigenetic signals.** We used different informatics platforms and data sources (**Supplementary File 1**) 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 **Figure 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

## RESULTS

### 1. Genome-wide DNAm patterns associated with different molecular SLE subtypes

We identified 262 SLE-DMPs (97% replication rate in 450K independent sample) (**Figure 2A**, **Supplementary Table 2**), 64% of them exhibiting hypomethylation and 36% hypermethylation effects. Not unexpectedly, the top SLE-DMPs implicated large reductions of DNAm at IRGs and genes enriched in IFN pathways (**Supplementary Table 3**), as for example *IFI44L*, *MX1*, *NLRC5*, *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 (32).

Stratifying SLE patients into the molecular subtypes allowed us to raise the number of associations despite reducing the sample size (**Figure 2A**). We identified 314 inflammatory-DMPs (96.18% hypomethylation, **Supplementary Table 4**), 41 lymphoid-DMPs (90.24% hypomethylation,



**Supplementary Table 5**) and 659 IFN-DMPs (54.48% hypomethylation, **Supplementary Table 6**), among which 214, 3, and 420, respectively, had not been previously detected when all SLE patients were pooled together (**Figure 2E**). The IFN subtype exhibited the largest effects at genes involved in IFN pathways, but a persistent epigenIFNsig and similar enrichment were also observed in patients within the lymphoid and the inflammatory subtypes (**Figure 2B-D**, **Supplementary Table 3**). Inflammatory-DMG were also enriched in synaptic functions. (**Supplementary Table 7**).

We identified 71 inflammatory-dependent DMPs, 1 lymphoid-dependent DMPs and 232 IFN-dependent DMPs (**Supplementary Figure 1A-C**, **Supplementary Table 8**). 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 ( $P < 0.05$ ) (**Figure 2F-G**). For the IFN subtype, we found large DNAm differences at *IRF7* and *TRIM22* genes, (**Figure 2H-I**) both of which are related to type I IFN pathways (33) .

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

We stratified patients based on their positivity for the most prevalent AAbs. We identified 388 anti-SSA-DMPs, 223 anti-chromatin-DMPs, 256 anti-dsDNA-DMPs and 164 anti-U1-RNP-DMPs (**Figure 3A**, **Supplementary Tables 9-12**) 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 (**Figure 3B-F**). However, the epigenetic signature was still persistent and dominant in SLE patients positive for other autoantibodies (**Figure 3B-E**). A small proportion of 27 anti-chromatin-DMPs and 38 anti-DNA-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. (**Figure 3G**, **Supplementary Table 14**). Functional enrichment analyses on AAb-DMPs identified IFN pathways as well as antigen processing and other pathways regulatory of immune responses (**Supplementary Table 13**). 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.

## 3. 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,  $FDR < 0.05$ ) (**Figure 4A**, **Supplementary Table 15**). Up to 31 loci involved in *cis*-meQTL-DMPs

also associate with SLE (SLE-*cis*-meQTLs-DMPs, Bonferroni significance  $P < 8 \times 10^{-05}$ ) (**Figure 4A**, **Supplementary Table 15**). Mediation analyses on SLE-*cis*-meQTLs-DMPs (**Supplementary Table 16**) revealed that SLE genetic risk is significantly reduced when DNAm is incorporated in the model (**Figure 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 (**Figure 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 (**Figure 4D**).

#### 4. 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 (**Supplementary Table 17-19**). We identified SLE-dependent meQTLs for 394 DMPs (**Supplementary Table 20**), among which significant disease-dependent effects were observed in *CDHR5* and *MAML2* (**Figure 4E-F**). Likewise, we discovered inflammatory- and IFN- dependent meQTLs for 283 and 316 DMPs, respectively (**Supplementary Table 21 and 22**). 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 (**Figure 4G**). IFN-dependent meQTL effects were observed for example for *HLA-F* (**Figures 4H**). Intriguingly, we also identified meQTLs with strong opposite effects (op-meQTLs) between SLE and CTRLs (119 DMPs) (**Supplementary Table 23**), or between the IFN and inflammatory subtypes (35 DMPs) among which *IFI44* gene shows the greatest opposite effect (**Figure 4I**) (**Supplementary Table 24**). Some genetic variants involved in context-dependent meQTLs were associated with SLE at a Bonferroni-corrected threshold ( $P < 1.2 \times 10^{-04}$ ) (**Supplementary Table 17**). 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 ( $P_{int} = 6.5 \times 10^{-05}$ ) in the regulation of DNAm within the *CDHR5* gene at chromosome 11, that involve SNPs strongly associated with SLE ( $P = 1.9 \times 10^{-07}$ ) (**Figure 4E**).

#### 5. 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* (34). Here, we identified up to 61 different TFs whose activity correlated with DNAm at several DMPs ( $P < 0.05$ ) (**Supplementary Table 25**). Interestingly, the large interaction between DNAm and TFactivity for *IRF9*, *IRF1*, *STAT2*, *STAT1*, *STAT3*, *TFDP1*, *FOXMI*, *E2F3*, *E2F2*, *GLI2* and *RUNX3* was restricted to SLE patients (**Figure 5A-B**, **Supplementary Table 26**). 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 *FOXMI* correlated also with DNAm at IRGs such as *IFI44L*, *MXI*, *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 (**Supplementary Table 27**). For example, *NR2F2* TFact is strongly negatively associated with DNAm at the *LETMI* gene, but such effect is not observed in the IFN subtype (**Figure 5C**). Likewise, *ZEB1* TF-act is associated with DNAm at *LGALS9*, *RSAD2*, *TMEM123*, *HECA* and *IFI44L* genes only in IFN patients (**Figure 5D**).

We explored whether DMPs could be associated with cytokine production in each molecular group (**Supplementary Tables 28-30**). 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*, *CMPK2*) and IP10 (*NLRC5*, *B2M*) which were not observed in CTRLs (**Supplementary Table 29**). We also observed inflammatory-dependent association for TNFRI (*RAPGEF1*) and CRP (*OAS3*) (**Supplementary Table 30**). **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 *MXI* in IFN subtype but it is positively associated in the inflammatory subtype (**Figure 5F**), similarly MCP2 showed opposite effects at the DNAm level for the *LGALS3BP* gene for the inflammatory and IFN subtypes (**Figure 5G**). Altogether, our results indicate that the relationship between DNAm, TFact and cytokine production is determined in a subtype-dependent manner.

## 6. 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 31-32, Figure 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” (**Figure 6**, see **Methods** for scoring details). We showed strong enrichment of validated targets within the groups of genes with scores equal or higher than one (**Supplementary Figure 2**). 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*, *TAPI1*, *TYMP*, *QPCT*, *GPD2*, *PSMB9* and *LAMBI*) overlapped with interferon specific epigenetic signatures (**Supplementary Table 32**).

## 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(7). Here we show that the epigenIFN sig 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 *IRF7* and *TRIM22* genes. *IRF7* has been established as an SLE genetic risk locus that alters IFN type I expression (33) 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(35) 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 (31), 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)(36).

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 IFNepig signature in Sjogren's syndrome is restricted to SSA positive patients and driven by HLA genetic variation(9). 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 (7). 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. 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* (37,38) and associated with systemic sclerosis (SS) (38). The candidate gene *CDHR5* is a member of the cadherin family which interacts with the  $\beta$ -catenin pathway (39). 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.

TF activity (TFact) has been linked to SLE recently (31). 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 *FOXMI* whose activity associates with DNAm at IRGs. On the contrary, we identified that *RUNX3* activity, a susceptibility gene in SLE and systemic sclerosis (SSc) (40,41), 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 (42,43), 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 epigIFNsign. 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 (44–49). Another interesting target candidate is the transcription Factor *ETS1*, which also had a significant strong SLE genetic association (50) 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 (51,52). 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 (31,53), as *STAT4* is for the IFN subtype and *VDR* for the inflammatory.

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 analyses we show that the strong association between epigenetics and genetics, TF activity and cytokine production is highly dependent on disease and molecular context.

## REFERENCES

1. Rider V, Abdou NI, Kimler BF, Lu N, Brown S, Fridley BL. Gender Bias in Human Systemic Lupus Erythematosus: A Problem of Steroid Receptor Action? *Front Immunol.* 2018 Mar 28;9:611.
2. Nehar-Belaid D, Hong S, Marches R, Chen G, Bolisetty M, Baisch J, et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat Immunol.* 2020 Sep;21(9):1094–106.
3. Bentham J, Morris DL, Cunninghame Graham DS, Pinder CL, Tomblason P, Behrens TW, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet.* 2015 Dec;47(12):1457–64.
4. Wang YF, Zhang Y, Lin Z, Zhang H, Wang TY, Cao Y, et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat Commun.* 2021 Feb 3;12(1):772.
5. Kariuki SN, Franek BS, Kumar AA, Arrington J, Mikolaitis RA, Utset TO, et al. Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. *Arthritis Research & Therapy.* 2010 Jul 26;12(4):R151.
6. Chung SA, Taylor KE, Graham RR, Nititham J, Lee AT, Ortmann WA, et al. Differential Genetic Associations for Systemic Lupus Erythematosus Based on Anti-dsDNA Autoantibody Production. *PLoS Genet.* 2011 Mar 3;7(3):e1001323.
7. Teruel M, Sawalha AH. Epigenetic Variability in Systemic Lupus Erythematosus: What We Learned from Genome-Wide DNA Methylation Studies. *Curr Rheumatol Rep.* 2017 Jun;19(6):32.
8. Barturen G, Babaei S, Català-Moll F, Martínez-Bueno M, Makowska Z, Martorell-Marugán J, et al. Integrative Analysis Reveals a Molecular Stratification of Systemic Autoimmune Diseases. *Arthritis Rheumatol.* 2020 Dec 8;
9. Teruel M, Barturen G, Martínez-Bueno M, Castellini-Pérez O, Barroso-Gil M, Povedano E, 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.* 2021 Dec 2;11(1):23292.
10. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet.* 2009 Nov;41(11):1228–33.
11. Guerra SG, Vyse TJ, Cunninghame Graham DS. The genetics of lupus: a functional perspective. *Arthritis Research & Therapy.* 2012 May 29;14(3):211.
12. Teruel M, Alarcón-Riquelme ME. The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *Journal of Autoimmunity.* 2016 Nov 1;74:161–75.
13. Ulf-Møller CJ, Svendsen AJ, Viemose LN, Jacobsen S. Concordance of autoimmune disease in a nationwide Danish systemic lupus erythematosus twin cohort. *Seminars in Arthritis and Rheumatism.* 2018 Feb 1;47(4):538–44.
14. Dupont C, Armant DR, Brenner CA. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Semin Reprod Med.* 2009 Sep;27(5):351–7.
15. Joseph S, George NI, Green-Knox B, Treadwell EL, Word B, Yim 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.* 2019 Jan;96:147–57.

16. Rask-Andersen M, Martinsson D, Ahsan M, Enroth S, Ek WE, Gyllensten U, et al. Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum Mol Genet.* 2016 Nov 1;25(21):4739–48.
17. Imgenberg-Kreuz J, Carlsson Almlöf J, Leonard D, Alexsson A, Nordmark G, Eloranta ML, et al. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann Rheum Dis.* 2018 May;77(5):736–43.
18. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet.* 2021 Sep;53(9):1311–21.
19. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017 Jan;49(1):131–8.
20. L LL, Pe J, M AR, C J, Jo P, undefined. Standardization procedure for flow cytometry data harmonization in prospective multicenter studies. *Sci Rep.* 2020 Jul 14;10(1):11567–11567.
21. Teruel M, Barturen G, Martínez-Bueno M, Barroso M, Castelli O, Povedano E, et al. An integrative multi-omics approach in Sjögren's Syndrome identifies novel genetic drivers with regulatory function and disease-specificity. *medRxiv.* 2020 Oct 1;2020.09.14.20192211.
22. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016 Oct;48(10):1284–7.
23. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 2016 Nov;48(11):1443–8.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007 Sep;81(3):559–75.
25. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol Cell Proteomics.* 2020 Dec;19(12):2115–24.
26. Ulgen E, Ozisik O, Sezerman OU. pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. *Frontiers in Genetics* [Internet]. 2019 [cited 2022 Jan 20];10. Available from: <https://www.frontiersin.org/article/10.3389/fgene.2019.00858>
27. Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012 May 15;28(10):1353–8.
28. Martínez-Bueno M, Alarcón-Riquelme ME. Exploring Impact of Rare Variation in Systemic Lupus Erythematosus by a Genome Wide Imputation Approach. *Front Immunol.* 2019 Feb 26;10:258.
29. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science.* 2010 Feb;25(1):51–71.
30. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 2019 Aug;29(8):1363–75.
31. Lopez-Dominguez R, Toro-Dominguez D, Martorell-Marugan J, Garcia-Moreno A, Holland CH, Saez-Rodriguez J, et al. Transcription Factor Activity Inference in Systemic Lupus Erythematosus. *Life.* 2021 Apr;11(4):299.
32. Rönnblom L, Leonard D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. *Lupus Science & Medicine.* 2019 Aug 1;6(1):e000270.
33. Liu W, Li M, Wang Z, Wang J. IFN- $\gamma$  Mediates the Development of Systemic Lupus Erythematosus. *Biomed Res Int.* 2020 Oct 17;2020:7176515.

34. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet.* 2016 Aug 1;17(9):551–65.
35. Jones EL, Laidlaw SM, Dustin LB. TRIM21/Ro52 - Roles in Innate Immunity and Autoimmune Disease. *Front Immunol.* 2021 Sep 6;12:738473.
36. Kerick M, González-Serna D, Carnero-Montoro E, Teruel M, Acosta-Herrera M, Makowska Z, et al. Expression Quantitative Trait Locus Analysis in Systemic Sclerosis Identifies New Candidate Genes Associated With Multiple Aspects of Disease Pathology. *Arthritis & Rheumatology.* 2021;73(7):1288–300.
37. Carmona FD, Gutala R, Simeón CP, Carreira P, Ortego-Centeno N, Vicente-Rabaneda E, et al. Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. *Ann Rheum Dis.* 2012 Jan;71(1):114–9.
38. European Scleroderma Group†, Australian Scleroderma Interest Group (ASIG), López-Isaac E, Acosta-Herrera M, Kerick M, Assassi S, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun.* 2019 Dec;10(1):4955.
39. Rinne SJ, Sipilä LJ, Sulo P, Jouanguy E, Béziat V, Abel L, et al. Candidate Predisposition Variants in Kaposi Sarcoma as Detected by Whole-Genome Sequencing. *Open Forum Infect Dis.* 2019 Jul 17;6(10):ofz337.
40. Wang W, Ma Y, Xu Y, Sheng Y, Gao J, Zuo X, et al. RUNX3 gene polymorphisms are associated with clinical features of systemic lupus erythematosus in Chinese Han population. *Journal of Dermatological Science.* 2015 Oct 1;80(1):69–71.
41. O'Reilly S. Epigenetic regulation of RUNX3 in systemic sclerosis pathogenesis: time to target? *Annals of the Rheumatic Diseases.* 2019 Sep 1;78(9):1149–50.
42. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr).* 2014 Jun;36(3):9648.
43. Verschuur CP, McEwen LM, Kohli V, Wolfson C, Bowdish DME, Raina P, et al. The relation between DNA methylation patterns and serum cytokine levels in community-dwelling adults: a preliminary study. *BMC Genet.* 2017 Jun 21;18:57.
44. Hagberg N, Joelsson M, Leonard D, Reid S, Eloranta ML, Mo J, 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.* 2018 Jul;77(7):1070–7.
45. Liang Y, Pan HF, Ye DQ. Therapeutic potential of STAT4 in autoimmunity. *Expert Opin Ther Targets.* 2014 Aug;18(8):945–60.
46. Menke J, Bork T, Kutska B, Byrne KT, Blanford M, Relle M, et al. Targeting transcription factor Stat4 uncovers a role for interleukin-18 in the pathogenesis of severe lupus nephritis in mice. *Kidney Int.* 2011 Feb;79(4):452–63.
47. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med.* 2007 Sep 6;357(10):977–86.
48. Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, Nilsson O, 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.* 2008 Sep 15;17(18):2868–76.
49. Taylor KE, Remmers EF, Lee AT, Ortmann WA, Plenge RM, Tian C, et al. Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS Genet.* 2008 May 30;4(5):e1000084.
50. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-Wide Association Study in Asian Populations Identifies Variants in ETS1 and WDFY4 Associated with Systemic Lupus Erythematosus. *PLoS Genet.* 2010 Feb 12;6(2):e1000841.
51. Garrett-Sinha LA, Kearly A, Satterthwaite AB. The role of the transcription factor Ets1 in

lupus and other autoimmune diseases. *Crit Rev Immunol*. 2016;36(6):485–510.

52. Kim CJ, Lee CG, Jung JY, Ghosh A, Hasan SN, Hwang SM, et al. The Transcription Factor Ets1 Suppresses T Follicular Helper Type 2 Cell Differentiation to Halt the Onset of Systemic Lupus Erythematosus. *Immunity*. 2018 Dec 18;49(6):1034-1048.e8.

53. Shao WH, Cohen PL. The role of tyrosine kinases in systemic lupus erythematosus and their potential as therapeutic targets. *Expert Rev Clin Immunol*. 2014 May;10(5):573–82.

## **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 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.

Funding for this work has been provided by the Lupus Research Alliance/BMS Award to M.E.A.R. and the Diversity Supplement to O.C.P.

This project has received funding 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 programme and EFPIA.

## **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.

## **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.

### Data availability

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

### Competing interests

Weiliang Qiu, Cheng Zhu and Srinivas Shankara are employees of SANOFI. All other authors have no competing interests.

### PRECISESADS Clinical Consortium

**Lorenzo Beretta<sup>9</sup>, Barbara Vigone<sup>9</sup>, Jacques-Olivier Pers<sup>10</sup>, Alain Saraux<sup>10</sup>, Valérie Devauchelle-Pensec<sup>10</sup>, Divi Cornec<sup>10</sup>, Sandrine Jousse-Joulin<sup>10</sup>, Bernard Lauwerys<sup>11</sup>, Julie Ducreux<sup>11</sup>, Anne-Lise Maudoux<sup>11</sup>, Carlos Vasconcelos<sup>12</sup>, Ana Tavares<sup>12</sup>, Esmeralda Neves<sup>12</sup>, Raquel Faria<sup>12</sup>, Mariana Brandão<sup>12</sup>, Ana Campar<sup>12</sup>, António Marinho<sup>12</sup>, Fátima Farinha<sup>12</sup>, Isabel Almeida<sup>12</sup>, Miguel Angel Gonzalez-Gay Mantecón<sup>13</sup>, Ricardo Blanco Alonso<sup>13</sup>, Alfonso Corrales Martínez<sup>13</sup>, Ricard Cervera<sup>14</sup>, Ignasi Rodríguez-Pintó<sup>14</sup>, Gerard Espinosa<sup>14</sup>, Rik Lories<sup>15</sup>, Ellen De Langhe<sup>15</sup>, Nicolas Hunzelmann<sup>16</sup>, Doreen Belz<sup>16</sup>, Torsten Witte<sup>17</sup>, Niklas Baerlecken<sup>17</sup>, Georg Stummvoll<sup>18</sup>, Michael Zauner<sup>18</sup>, Michaela Lehner<sup>18</sup>, Eduardo Collantes<sup>19</sup>, Rafaela Ortega Castro<sup>19</sup>, Ma Angeles Aguirre-Zamorano<sup>19</sup>, Alejandro Escudero-Contreras<sup>19</sup>, Ma Carmen Castro-Villegas<sup>20</sup>, Norberto Ortego<sup>20</sup>, María Concepción Fernández Roldán<sup>20</sup>, Enrique Raya<sup>21</sup>, Inmaculada Jiménez Moleón<sup>21</sup>, Enrique de Ramon<sup>22</sup>, Isabel Díaz Quintero<sup>22</sup>, Pier Luigi Meroni<sup>23</sup>, Maria Gerosa<sup>23</sup>, Tommaso Schioppo<sup>23</sup>, Carolina Artusi<sup>23</sup>, Carlo Chizzolini<sup>24</sup>, Aleksandra Zuber<sup>24</sup>, Donatienne Wynar<sup>24</sup>, Laszlo Kovács<sup>25</sup>, Attila Balog<sup>25</sup>, Magdolna Deák<sup>25</sup>, Márta Bocskai<sup>25</sup>, Sonja Dulic<sup>25</sup>, Gabriella Kádár<sup>25</sup>, Falk Hiepe<sup>26</sup>, Velia Gerl<sup>26</sup>, Silvia Thiel<sup>26</sup>, Manuel Rodriguez Maresca<sup>27</sup>, Antonio López-Berrio<sup>27</sup>, Rocío Aguilar-Quesada<sup>27</sup> & Héctor Navarro-Linares<sup>27</sup>**

<sup>9</sup>Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milano, Italy. <sup>10</sup>Centre Hospitalier Universitaire de Brest, Hospital de la Cavale Blanche, Brest, France. <sup>11</sup>Pôle de pathologies rhumatismales systémiques et inflammatoires, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. <sup>12</sup>Centro Hospitalar do Porto, Porto, Portugal. <sup>13</sup>Servicio Cantabro de Salud, Hospital Universitario Marqués de Valdecilla, Santander, Spain. <sup>14</sup>Hospital Clinic I Provincia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. <sup>15</sup>Katholieke Universiteit Leuven, Leuven, Belgium. <sup>16</sup>Klinikum der Universitaet zu Koeln, Cologne, Germany. <sup>17</sup>Medizinische Hochschule Hannover, Hannover, Germany. <sup>18</sup>Medical University Vienna, Vienna, Austria. <sup>19</sup>Servicio Andaluz de Salud, Hospital Universitario Reina Sofía Córdoba, Córdoba, Spain. <sup>20</sup>Servicio Andaluz de Salud, Complejo hospitalario

Universitario de Granada (Hospital Universitario San Cecilio), Granada, Spain. <sup>21</sup>Servicio Andaluz de Salud, Complejo hospitalario Universitario de Granada (Hospital Virgen de las Nieves), Granada, Spain. <sup>22</sup>Servicio Andaluz de Salud, Hospital Regional Universitario de Málaga, Málaga, Spain. <sup>23</sup>Università degli studi di Milano, Milan, Italy. <sup>24</sup>Hospitaux Universitaires de Genève, Genève, Switzerland. <sup>25</sup>University of Szeged, Szeged, Hungary. <sup>26</sup>Charite, Berlin, Germany. <sup>27</sup>Andalusian Public Health System Biobank, Granada, Spain.

### **PRECISESADS Flow Cytometry Study Group**

**Montserrat Alvarez<sup>28</sup>, Damiana Alvarez-Errico<sup>29</sup>, Nancy Azevedo<sup>30</sup>, Nuria Barbarroja<sup>31,32</sup>, Anne Buttgereit<sup>33</sup>, Qingyu Cheng<sup>34</sup>, Carlo Chizzolini<sup>28</sup>, Jonathan Cremer<sup>35</sup>, Aurélie De Groof<sup>36</sup>, Ellen De Langhe<sup>37</sup>, Julie Ducreux<sup>36</sup>, Aleksandra Dufour<sup>28</sup>, Velia Gerl<sup>34</sup>, Maria Hernandez-Fuentes<sup>38</sup>, Laleh Khodadadi<sup>34</sup>, Katja Kniesch<sup>39</sup>, Tianlu Li<sup>34</sup>, Chary Lopez-Pedrerá<sup>36</sup>, Zuzanna Makowska<sup>33</sup>, Concepción Marañón<sup>41</sup>, Brian Muchmore<sup>41</sup>, Esmeralda Neves<sup>30</sup>, Bénédicte Rouvière<sup>42</sup>, Quentin Simon<sup>42</sup>, Elena Trombetta<sup>40</sup>, Nieves Varela<sup>39</sup> & Torsten Witte<sup>39</sup>**

<sup>28</sup>Immunology and Allergy, University Hospital and School of Medicine, Geneva, Switzerland.

<sup>29</sup>Chromatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. <sup>30</sup>Serviço de Imunologia EX-CICAP, Centro Hospitalar e Universitário do Porto, Porto, Portugal. <sup>31</sup>IMIBIC, Reina Sofía Hospital, University of Cordoba, Córdoba, Spain. <sup>32</sup>Bayer AG, Berlin, Germany. <sup>33</sup>Pharmaceuticals Division, Bayer Pharma, Berlin, Germany. <sup>34</sup>Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany.

<sup>35</sup>Department of Microbiology and Immunology, Laboratory of Clinical Immunology, KU Leuven, Leuven, Belgium. <sup>36</sup>Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, Brussels, Belgium. <sup>37</sup>University Hospitals Leuven and Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium. <sup>38</sup>UCB, Slough, UK. <sup>39</sup>Klinik für Immunologie Und Rheumatologie, Medical University Hannover, Hannover, Germany. <sup>40</sup>Laboratorio di Analisi Chimico Cliniche e Microbiologia - Servizio di Citofluorimetria, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy. <sup>41</sup>GENYO, Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain. <sup>42</sup>NSERM, UMR1227, CHRU Morvan, Lymphocytes B et Autoimmunité, University of Brest, BP 824, Brest, France.

### **FIGURE LEGENDS**

**Figure 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.



**Figure 2. Epigenetic signatures of SLE molecular subtypes.** **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 stars dots represents subtypes-dependent DMPs.

**Figure 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  $\log_{10}(P)$  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.

**Figure 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= 4,212) and a group of CTRL (N= 4,065). 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  $P < 1 \times 10^{-6}$  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-J** 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.

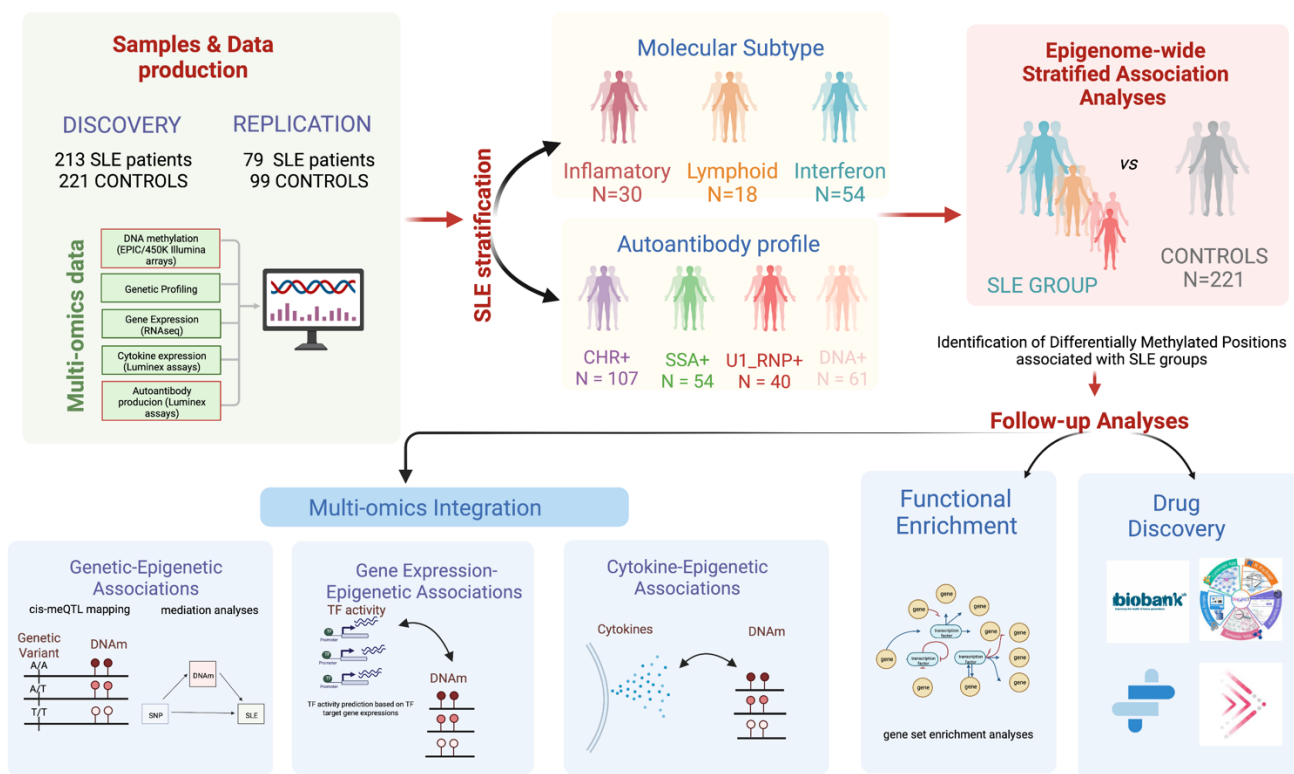
**Figure 5. Relationship between SLE-epigenetic signatures, transcription factor activity, and cytokine production.** **A.** Heatmap representing SLE-dependent associations between DNAm at DMPs and TFact inferred from RNAseq data. **B.** SLE dependent example showing the effect between TFact STAT2 and 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.

**Figure 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 Table 32**.

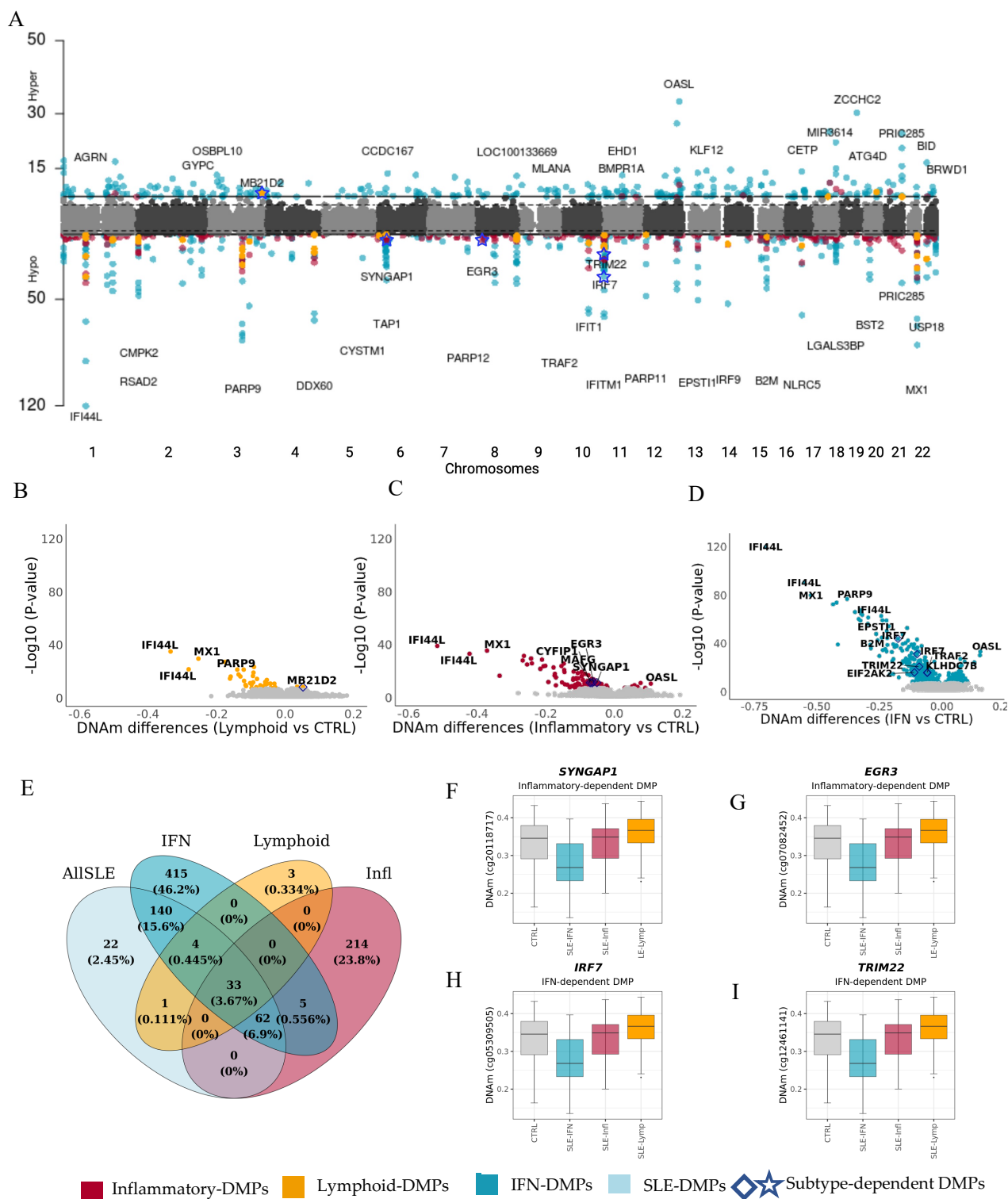
**Supplementary Figure 1. Effect Size Comparison comparing results from different EWAS.** **A.** Effect size comparison between inflammatory-dependent DMPs and lymphoid-dependent DMPs. **B.** Effect size comparison between inflammatory-dependent DMPs and IFN-dependent DMPs. **C.** Effect size comparison between IFN-dependent DMPs and lymphoid-dependent DMPs.

**Supplementary Figure 2. Relationship between SLE Epigenomic signature genes and known drug targets.** **A.** Bar plot with number of methylated genes and known drug targets that are in these methylated genes under different score thresholds. **B.** Bar plot with number of methylated genes and the rate of known drug targets in these methylated genes under different score thresholds.

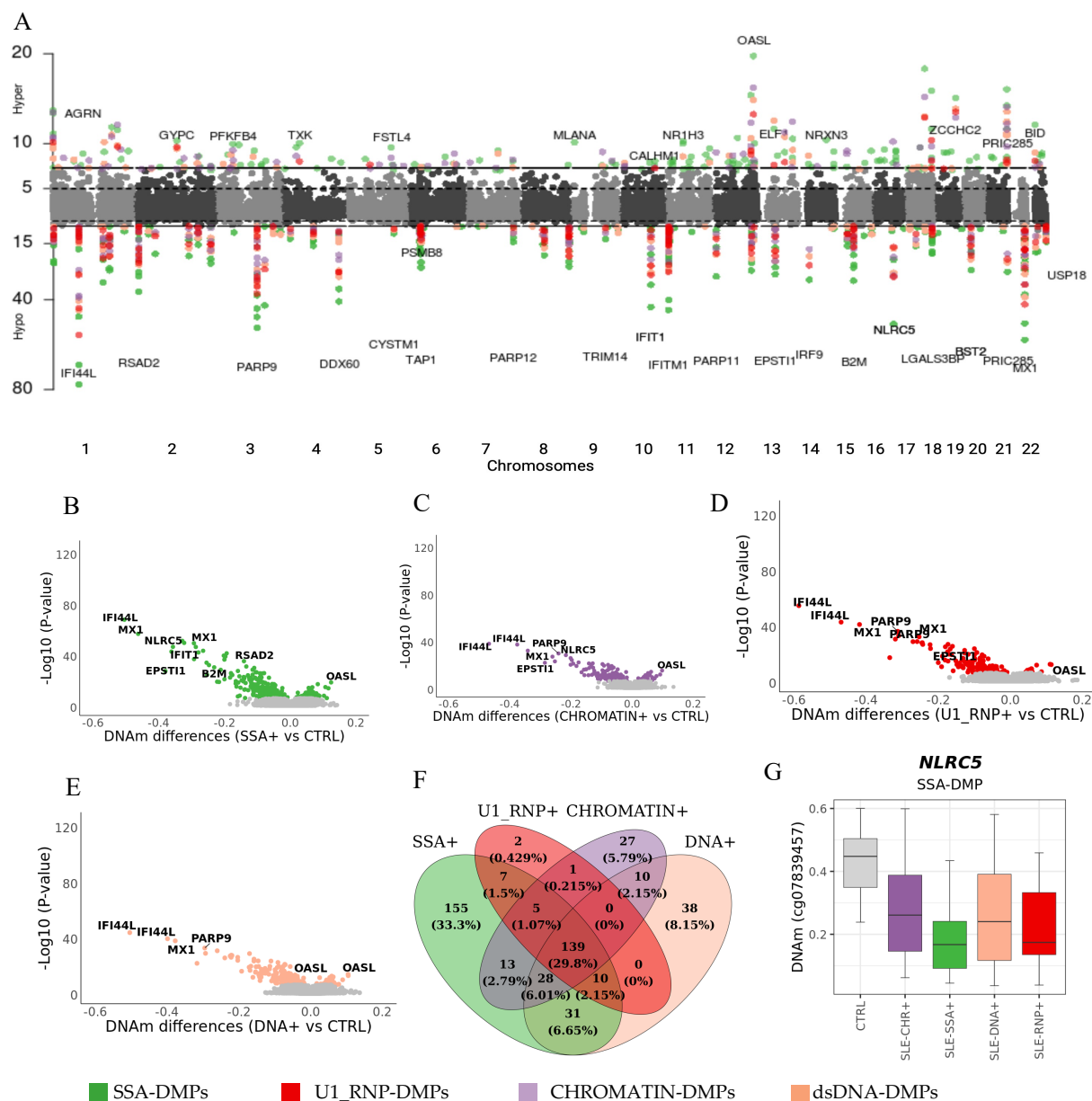
**Figure 1. Overview of Study Design**



**Figure 2.**

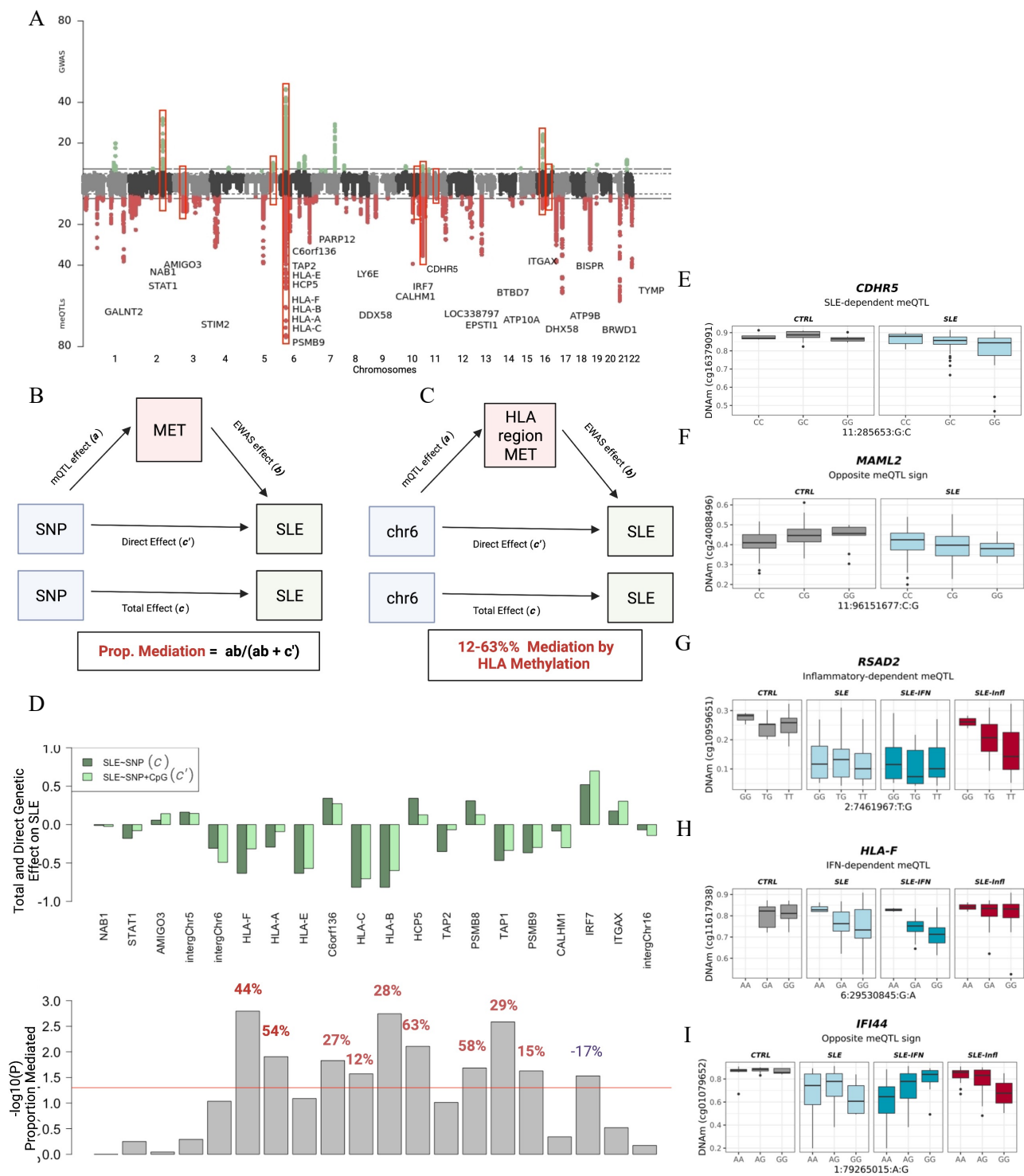


**Figure 3**

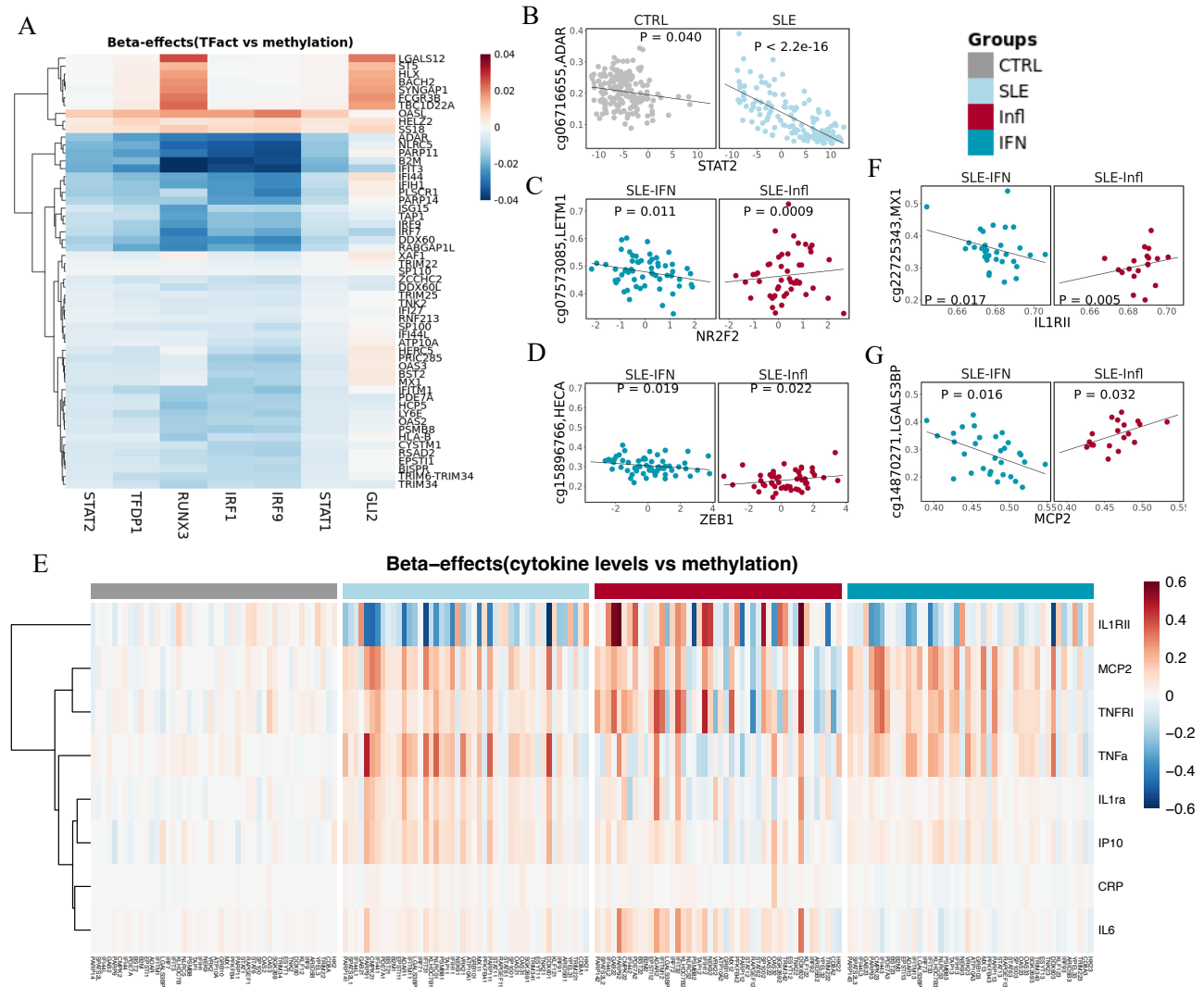




**Figure 4**



**Figure 5**



**Figure 6**

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			
<i>SYK</i>	cg13782919(chr9:93588842)									3	1	SLE
<i>JAK3</i>	cg20959703(chr19:17959783)									3	1	SLE
<i>BCL2</i>	cg01863731(chr18:60810665)									3	1	SLE
<i>PIK3CD</i>	cg04287259(chr1:9730401)									3	1	SLE
<i>VDR</i>	cg06195428(chr12:48295148)									3	1	Infl
<i>STAT4</i>	cg15636519(chr2:191894418)									3	2	IFN
<i>BTLA</i>	cg19281794(chr3:112218761)									3	1	SLE
<i>GPD2</i>	cg02039987(chr2:157291262)									3	1	SLE
<i>NDUFS8</i>	cg12070987(chr11:67804055)									3	1	IFN
<i>FGR</i>	cg09370867(chr1:27961680)									2	1	SLE
<i>ETS1</i>	cg03702171(chr11:128333168)									2	2	SLE
<i>CD8A</i>	cg26730653(chr2:87022189)									2	1	SLE
<i>FCGR3B</i>	cg00974864(chr1:161601053)									2	1	SLE
<i>STAT1</i>	cg00676801(chr2:191876673)									2	1	SLE
<i>CBLB</i>	cg12614898(chr3:105581164)									2	1	SLE
<i>INPP5D</i>	cg01218053(chr2:233940417)									2	1	SLE
<i>INSR</i>	cg19961945(chr19:7192407)									2	1	SLE
<i>RARA</i>	cg16680942(chr17:38500161)									2	1	IFN
<i>HIST1H2BL</i>	cg19832381(chr6:27776524)									2	2	SLE
<i>TXK</i>	cg11132334(chr4:48136298)									2	1	Infl