Imgenberg-Kreuz et al. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus

Supplementary text

Methods

Subjects and samples: In the discovery phase SLE patients visiting the rheumatology clinics at the university hospitals in Uppsala, Sweden, and Linköping,1 Sweden, (n=400) and control individuals from the Uppsala BioResource (n=400) of healthy blood donors visiting the Department of Transfusion Medicine, Uppsala University Hospital, Sweden, were included. As replication SLE patients (n=201) and controls (n=187) from the Stockholm Karolinska Institute, Sweden, were included. All subjects provided informed consent to participate in the study, and the study was approved by the regional ethics board (Uppsala EPN dnr 00-227, 09-013, 16-155). Genomic DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Clinical and medication data were collected from patient charts. Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) ≥ 5 in the discovery cohort and Systemic Lupus Activity Measure (SLAM) ≥10 in the replication cohort were used to define more active SLE.2,3 548 patients fulfilling at least four of the eleven American College for Rheumatology criteria4 and who were not genetic population outliers were included in association analyses.

DNA methylation analysis: DNA methylation levels at 485,577 CpG sites were determined using the HM450k BeadChip (Illumina Inc., San Diego, CA, USA).5 Bisulfite treatment was performed on 500 ng of DNA using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA). Samples were randomized on the BeadChips to avoid batch effects. The BeadChips were scanned on an iScan SQ instrument (Illumina), and signal intensities were exported as intensity data files and parsed into the Minfi R package for quality control (QC) and Subset-quantile Within Array Normalization (SWAN).6,8 Methylation beta-values were calculated as the fraction of the signal intensity from the methylated CpG sites over the total intensity (range 0-1, corresponding to 0-100% methylation). Only CpG sites with complete data across all samples, signal detection p >0.01, unambiguously mapping to autosomes, containing no SNPs in the 10 bp at the 3’-end of the probe (1000 genomes CEU populations, release 3) and not overlapping any known copy number variant were kept for further analysis.9 Samples identified as outliers in PCA, with >5% missing data, and not falling within the major European cluster in PCA using the Infinium ImmunoChip (Illumina) were excluded. The post-QC dataset comprised 385,851 CpG sites, 347 SLE patients and 400 controls for the discovery phase and 201 patients and 188 controls for the replication phase.

Genotyping: Genotype data generated on the Infinium ImmunoChip (Illumina) containing 196,524 probes were available for 527 SLE patients and 567 of the healthy control individuals with HM450k data in our study. Genotype QC at the sample level was performed as previously described10 and 133,838 SNPs fulfilling ≥98% call rate, a Hardy-Weinberg equilibrium test p >1x10^-4 and a minor allele frequency (MAF) of ≥1% were included.

SLE case-control association analysis: The genetic case-control association analysis for SLE included a larger set of 1,135 Swedish SLE patients and 2,931 Swedish control individuals from the university hospital rheumatology clinics in Uppsala, Stockholm Karolinska Solna, Linköping, Lund and the four
northernmost counties of Sweden. SNPs were tested for association with SLE using logistic regression in PLINK.\textsuperscript{11}

**Epigenome-wide association analysis:** Relative blood cell-type composition of the samples was determined using the method by Houseman \textit{et al.} \textsuperscript{12} as implemented in the Minfi R package\textsuperscript{2} and using publicly available reference DNA methylation signatures of flow sorted blood cell types\textsuperscript{13} (Supplementary Figure 6). To determine differential methylation between SLE patients and controls a linear regression model containing cell count estimates, age and sex as covariates was fitted, predicting methylation at each CpG site as a function of disease. Differentially methylated CpG sites (DMCs) were called in the discovery phase if they had a $p < 1.3 \times 10^{-7}$ for association based on Bonferroni correction (385,851 tested CpG sites, $\alpha=0.05$) and an average difference in methylation beta values between cases and controls of $>0.05$ regardless of the direction of the effect ($|\Delta \beta| >0.05$). Association results from the discovery phase were plotted using the software Circos.\textsuperscript{14}

Similarly, the role of methylation in different disease manifestations was investigated in a case-case analysis focusing on the eleven criteria defined by the American College for Rheumatology (ACR)\textsuperscript{2} as was the association between different medications and methylation. In the medication case-case analysis disease activity status (active/inactive SLE) was included as a covariate. CpG sites on the sex chromosomes were analysed separately in females and males. Significance in the replication phase was determined as $p < 0.05$ divided by the number of tested CpG sites. Functional gene-set enrichment analysis was performed using the ToppGene Suite database\textsuperscript{15} testing the 500 most significantly associated DMCs (based on association $p$-value in the discovery cohort) with a gene name annotation. For classification of IFN-regulated genes the Interferome v2.01 database was used.\textsuperscript{16}

**Functional genomic annotation:** Probe mapping and annotation in relation to gene regions were performed as previously described\textsuperscript{9}. The regional distribution of probes that passed the QC filtering was compared to the distribution of associated CpGs using the $\chi^2$-test, where $p<0.005$ was regarded as significant (Bonferroni correction). Publicly available chromatin modification datasets from primary CD3+ T cells and CD19+ B cells were obtained from the NIH Roadmaps Epigenomics Project for the histone modification marks H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9me3, H3K27me3 and DNase I hypersensitive sites (DHS).\textsuperscript{17} Chromatin mark peaks from the reference cells were analysed for overlap with the genomic coordinates of SLE associated CpG sites applying a Bonferroni corrected $\chi^2$-test.

**Methylation quantitative trait loci (meQTL) analysis:** 527 SLE patients and 567 healthy controls from the discovery and replication cohorts with both HM450k array and ImmunoChip data passing QC were included in the meQTL analysis. Methylation levels were tested in PLINK for genotype association separately in patients and controls assuming an additive model and including chip, chip position, age, sex and estimated cell-type distribution as covariates.\textsuperscript{11} For each CpG site all polymorphisms on the ImmunoChip within a ±100 kb flanking region were tested in a $\textit{cis}$-meQTL analysis (in total 7,703,665 association tests, 385,851 CpG sites and 130,115 SNPs) (Supplementary Figure 3). A Bonferroni corrected $\alpha<0.05$ was considered significant ($p<6.5 \times 10^{-9}$). Methylation variance was calculated as the difference between a subject's methylation value and the genotype-specific mean,\textsuperscript{18} and var-meQTLs were tested using and additive model and the same significance threshold as for meQTLs ($p<6.5 \times 10^{-9}$).
References