Molecular Pathways in Patients with Systemic Lupus Erythematosus Revealed by Gene Centred DNA Sequencing
Sandling et al.

Contents:

Supplementary Figures S1-S8
**Filter schematic for SNPs and samples**

- **Alignment**
- **Genotyping**
- **VQSR PASS (only SNPs)**
- **Depth > 8**
- **Genotype Quality > 20**
- **Polymorphic positions**
  - $0.2 < \text{Allelic Balance} < 0.8$
  - $\text{HWE} p > 1 \times 10^{-6}$ (controls)
- **Variant call rate > 90%**
- **Polymorphic positions**
- **Differential missingness**
  - `plink --test-missing Bonferroni alfa 0.05`
- **Sample filter I**
  - Average coverage > 10x
- **Sample filter II**
  - QC outliers (pseq i-stats)
  - Total variants
  - Missingness
  - Singletons
  - Het/Hom ratio
  - Ti/Tv
  - KING
  - Concordance
  - LASER
- **Sample filter III**
  - Call rate > 80%

**Supplementary Figure S1.** Targeted sequencing variant calling and quality control pipeline.
**Supplementary Figure S2.** Overview of analyses.

- **1832 gene regions**
  - 287,354 SNVs
  - 958 SLE patients, 1026 HC

- **Aggregate analyses**
  - Case-control association
    - SKAT-O
      - 35 pathways
        - KEGG
      - 5 gene-sets
        - literature
      - 1832 gene regions
    - Case-control association, with annotation
      - GenePy
      - 1832 gene regions
  - Polygenic risk scores
    - KEGG
    - Cluster analysis
    - 1832 gene regions

- **Single-variant analyses**
  - Case-control association
    - PLINK
    - Replication
      - 835 SLE, 1225 HC
  - Case-only SNVs
    - SweGen, GnomAD
Supplementary Figure S3. Pathway polygenic risk score density plots for SLE patients and controls for each tested KEGG pathway. P-values represent differences in PRS between SLE patients (SLE) and control individuals (HC), uncorrected P-values are presented (Bonferroni corrected threshold P=0.00143). The dashed line indicates the PRS 97.5 percentile in control individuals.
Supplementary Figure S4. Pathway SLE polygenic risk score density plots for the four groups of SLE patients identified for each tested KEGG pathway. P-values represent differences in PRS between clusters of SLE patients, uncorrected P-values are presented (Bonferroni corrected threshold P=0.00143).
Supplementary Figure S5. Results of SLE case-control association analyses with P-values for association plotted against chromosomal location: a) Gene-based aggregate association testing (SKAT-O) where each point represents a gene region. Gene names are indicated for the top gene regions. The red line represents a Bonferroni corrected significance threshold and the black line FDR 0.05. Novel loci are indicated in bold. b) Single variant association testing where each point represents a SNV. The red line represents a Bonferroni corrected significance threshold and the black line the suggestive significance threshold (P<1×10^-4). Novel loci are indicated in bold. c-e) Single variant association result regional association plots for the CAPN13, MOB3B/IFNK and HAL regions respectively. The colour scale indicates linkage disequilibrium (r²) between SNVs.
Supplementary Figure S6. Distribution of different classes of SNVs in SLE patients and control individuals: a) rare non-synonymous variants, b) rare synonymous variants, c) rare constrained variants (GERP RS score >2), d) rare non-coding variants (any of the following snpeff annotations: sequence feature, upstream, downstream, intergenic, TF binding site variant). P-values represent difference between SLE patients and control individuals, uncorrected P-values are presented (Bonferroni corrected threshold P=0.0125).
Supplementary Figure S7. Genetic population structure of study individuals. a) Study samples mapped on population reference samples. b) Principal components for population stratification within study, PC1 vs PC2 c) PC2 vs PC3 d) PC3 vs PC4.
Supplementary Figure S8. SNV genotype average concordance between targeted sequencing and genotyping by a beadchip array (Illumina ImmunoChip). Concordance for two types of SNVs, common SNVs (MAF≥0.05) and low frequency SNVs (MAF<0.05), are displayed.